The genetic, molecular, and neuroendocrine basis of behavioral evolution in deer mice

Natalie Niepoth

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

under the Executive Committee

of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2024

Abstract

The genetic, molecular, and endocrine basis of behavioral evolution in deer mice

Natalie Niepoth

Despite the extraordinary diversity of behavior across the animal kingdom, the genes and molecules that contribute to such natural diversity are largely unknown. In this thesis, I leverage the dramatic divergence in behavior between two closely related species of deer mice (genus *Peromyscus*) to investigate the genetic, cellular, and neuroendocrine basis of behavior. In chapter 2, I show that the monogamous oldfield mouse (*Peromyscus polionotus subgriseus*) has evolved a novel cell type in the adrenal gland that expresses the enzyme AKR1C18, which converts progesterone into 20α-hydroxyprogesterone (20α-OHP). I then demonstrate that 20α-OHP is more abundant in oldfield mice than in the closely-related promiscuous prairie deer mouse (*P. maniculatus bairdii*) and that it increases monogamous-typical parental behaviors when administered to both monogamous fathers. Using quantitative trait locus mapping in a cross between these species, I discover interspecific genetic variation that drives expression of the glycoprotein tenascin N and ultimately contributes to gain of adrenal AKR1C18 expression in oldfield mice. In chapter 3, I investigate the genetic architecture underlying the striking difference in exploratory behavior between prairie deer mice and oldfield mice. Through congenic fine-mapping, I identify a 15-Mb locus that strongly contributes to species differences in exploratory behavior. I then investigate the potential contributions of one of the 18 genes in the locus, Olfm4, which harbors *cis*-regulatory variants that drives its expression in the oldfield hypothalamus. Taken together, my research advances our understanding of the genetic and molecular causes that drive rapid behavioral divergence between species.

Table of Contents

List of Tables and Figures

Acknowledgments

First and foremost, I am immensely grateful to my advisor, Dr. Andrés Bendesky, for inspiring me, for believing in me, for challenging me, and for always bringing out the best in my work. Your mentorship has profoundly influenced my development as a scientist.

To my lab mates in the Bendesky lab, your friendship and support have made my time here equal parts scientifically stimulating and fun. What a pleasure to work alongside such a talented group of scientists. I am especially grateful to my fellow mouseketeers Jenny Merritt, Michelle Uminski, and Emily Lei, with whom I worked closely over the past six years and whose friendships I will always cherish. Claire and Madison, thank you for all of the laughs and shenanigans. Kerel and Sophie, thank you both for your friendship and for keeping the lab running smooth as butter.

I am grateful to the members of my committee: Dustin, Molly, Stavros, and René—your insight over the last six years has been invaluable. I have also had the great pleasure of working with brilliant collaborators who have made important contributions to the work in this dissertation, including Kiran Soma, Asmita Poudel, Stephanie Rudolph, Tim Yeh, Victoria Esquibies, Christoph Gebhardt, and Sarah Wacker. I am extremely grateful to each of you for sharing your time and expertise with me.

Finally, to my husband Mick, de liefde van mijn leven—dankjewel voor alles.

Dedication

To my parents, whose quality of biparental care is unrivaled even by oldfield mice

Chapter 1: Introduction

Published as:

Niepoth, N., & Bendesky, A. (2020). How Natural Genetic Variation Shapes Behavior. *Annual review of genomics and human genetics, 21*, 437–463. https://doi.org/10.1146/annurev-genom-111219-080427

Animal behavior is characterized by its complexity. It is generated by the integration of sensory cues with internal states to direct motor output via precise signaling in sophisticated neuronal circuits. These circuits are remarkably malleable and are constantly remodeled developmentally and by experience and learning, allowing animals to adapt to both recurring patterns and changes in their environment. Behavior is further influenced by innate variation in neuronal anatomy and function. Thus, behaviors are plastic within individuals throughout their lives as well as variable among individuals.

Behavior can be difficult to measure, particularly in natural settings, where the conditions animals experience over their lifetimes are difficult to control. Even under controlled laboratory environments, behavior is notoriously susceptible to subtle environmental perturbations[1]. These challenges make it difficult to measure the environmental and genetic variables that influence behavior. Therefore, our knowledge of the genetic underpinnings of behavior lags behind what we currently know about morphology, physiology, and disease risk. However, technological and statistical methods for studying genetic contributions to behavior are advancing quickly, unlocking new opportunities. Though much is still unknown, patterns in the field have begun to emerge. We have reached an opportune moment to study these patterns and make inferences about the larger processes that govern the evolution of behavior.

In this review, we survey examples of natural genetic variants that modulate behavior within and among populations of a species and that contribute to differences in behavior among

species. We focus on three common targets of genetic influences on behavior: sensation of environmental cues, higher-order processing in the central nervous system, and interactions with environmental molecules outside of the nervous system. We then discuss the molecular types of variants observed and how these variants arise and are maintained in populations, and we conclude with a summary of emerging patterns in the field and outstanding questions.

Measuring the genetic basis of behavior in natural populations

Analyses of the physiological roles of genes on behavior started with the pioneering forward genetics screens of Seymour Benzer in *Drosophila melanogaster* in the 1970s. More recently, powerful reverse genetics tools have been applied to laboratory animals—mainly nematodes, flies, and mice—to study specific genes. However, these forward and reverse genetics approaches reveal little about the genetic bases of behavioral variation in nature. For example, the Mouse Genome Informatics database contains more than 10,000 examples of artificial mutations that affect behavior in laboratory mice[2]. If these mutations were to arise spontaneously in nature, many would be too detrimental for survival and thus quickly removed from populations. Therefore, these mutations likely do not represent the types of variation that occur and segregate in nature.

What is clear, however, is that variation in most natural behaviors has a substantial genetic component. A sweeping meta-analysis of 17,804 traits from 2,748 twin study publications showed that the heritability of behavioral traits in humans is comparable to that of nonbehavioral traits[3]. Most human behavioral traits studied are 30–60% heritable, though this estimate may be skewed by ascertainment and publication biases; for example, geneticists may choose to measure the heritability of traits they think are likely to have a genetic component. In

nonhuman animals, heritability is also similar among behavioral, life history, and morphological traits[4], although these estimates may suffer from similar biases.

Pinpointing the specific genes and variants that contribute to trait heritability is a central goal of behavioral genetics. Different approaches have been classically used to parse this genetic component of behavior into contributory variants and the genes they affect. In recent years, quantitative genetics research has encouragingly shifted from candidate gene studies to analyses of variants throughout the whole genome. Unbiased approaches such as genome-wide association studies (GWASs), quantification of population differentiation, and quantitative trait locus (QTL) mapping are powerful methods for discovering genetic loci associated with variation in behavior (Table 1.1).

Table 1.1: Methods for identifying genetic loci linked to behavior.

Method	Scale of comparison	Mapping resolution	Sample size required	Advantages	Disadvantages
GWAS	Within populations	High (to scale of LD)	Large	Can study in nature	Has biases from population stratification, cannot capture associations with rare variants, also measures indirect effects
QTL mapping	Within pedigrees, between populations or sister species	Low or moderate (depending on the number of recombination events)	Moderate	Can use controlled laboratory conditions; family structure avoids stratification	Potentially overestimates effect sizes, can generate false negatives from closely linked variants with opposite effects, captures only effects of variants present in founders of cross
Population differentiation scan	Between populations or closely related species	High (to scale of LD)	Small	Can study in nature, provides evidence of selection	Is agnostic to phenotypic traits
Comparative gene expression	Within and between populations or species	Transcript level	Small	Offers temporal and tissue-specific insights into cellular function	Is agnostic to genetic variation; many correlated genes will not

					contribute directly to behavior

Abbreviations: GWAS, genome-wide association study; LD, linkage disequilibrium; QTL, quantitative trait locus.

Genome-wide association studies. GWASs test for association between a trait (such as a specific behavior) and genetic variants that are common in populations (frequency > ~5%). However, most common variants have exceedingly small effects, so large samples—on the order of tens to hundreds of thousands of subjects—are usually necessary to gain sufficient power[5,6]. In recent years, public consortia and private companies have compiled genetic samples and phenotypic information from more than 4 million individuals (e.g. [7–9]). This data accessibility has enabled the analysis of dozens of human behavioral traits, and in the past 10 years, more than 700 publications have reported the results of GWASs for specific human behaviors[10]. These publications demonstrate that nearly all behaviors are highly polygenic: An individual behavior within a population is influenced by many genetic loci, each of mostly small effect.

GWASs are not designed to detect associations with rare variants, yet nearly all human genetic variants are rare: More than 90% of variants segregating in human populations have minor allele frequencies below 1%[11]. Because rare variants tend to have larger effects than common variants[12,13], they may be important contributors to the behaviors of specific individuals or families where these variants segregate. The extent of the contribution of rare variants to the heritability of behavior within populations is still an open question.

Genetic associations in traditional GWAS designs measure not only direct genetic effects of variants on people in the study but also the indirect contributions of these variants through their effects on relatives with whom the subjects interact[14,15]. For example, variants may affect the behavior of subjects indirectly by modulating parental behaviors. Future studies that jointly analyze the genotypes of test subjects and their parents will help to alleviate these problems and

provide better estimates of direct and indirect genetic effects—both of which are necessary for a comprehensive understanding of behavioral variation[16]. GWASs are also subject to false-positive associations resulting from population stratification and assortative mating, which are difficult to fully control for using statistical methods[16]. For example, studies of the use of chopsticks in a city would likely capture genetic variants common in Asians even if none of these variants directly affect chopstick use[17].

Population differentiation. Populations often differ in their behavior, yet it is very difficult to estimate the contribution of genetic differences to such behavioral variation. People with mixed ancestry from these populations can be used to find associations between local genomic ancestry and behavior in an approach called admixture mapping. Because genomic segments from different populations are present in admixed individuals, the argument is that this curtails problems of environmental and genetic interaction confounders that occur when estimating the effect of genes in two separate populations. However, ancestry proportions in admixed individuals have been found to correlate with socioeconomic status[18], which can result from familial and cultural contingencies as well as discrimination and can severely confound genetic analyses. The proportion of ancestry can also be correlated with many other environmental variables, such as diet, religion, and education, further confounding behavioral inferences from genetics.

Another powerful population-genetics approach to study differences among populations involves scanning the genomes of two or more populations that differ in a behavior of interest. Regions that are particularly differentiated between populations (i.e., outliers in F_{ST}-based statistics) may contain variants that explain trait variation. For example, a genome-wide F_{ST} scan between two populations of warblers identified a candidate gene contributing to the choice in

winter migration to either Central or South America[19]. Because populations often differ in more than one trait, differentiated regions may affect not the behavior of interest but rather a correlated trait. Nevertheless, these F_{ST}-based methods can be useful in identifying candidate genomic regions for local adaptation that can be further studied using other approaches.

Quantitative trait locus mapping. An alternative method to probe the genetic basis of behavioral variation is QTL mapping. The goal of QTL mapping is to identify loci that co-segregate with a trait in families or in experimental crosses. The family structure of these mapping populations generally avoids issues with population stratification. Laboratory crosses allow for careful behavioral measurements while affording control of environmental conditions and biological variables such as age. However, QTL mapping crosses usually originate from small numbers of founding animals, limiting the number of haplotypes that can be analyzed and failing to fully capture natural allele frequency distributions. Moreover, because many experimental crosses comprise only a few generations where meiotic recombination can take place, mapping resolution is usually low. Advanced intercross schemes greatly improve mapping resolution[20], and QTLs can also be fine-mapped using follow-up targeted crosses. Multiple genes and specific genetic variants affecting behavior have been found using QTL mapping and follow-up fine-mapping, particularly in *Caenorhabditis elegans* and *Drosophila melanogaster*[21–23]. Even when QTL mapping does not lead to the identification of causative variants, important inferences can be made about the genetic architecture of behavior, including estimation of pleiotropy and sex-specific effects as well as quantification of effect sizes[24–26].

Comparative gene expression. Unbiased whole-transcriptome analysis technologies, such as RNA sequencing (RNA-seq), allow for correlating gene expression with variation in a behavior within populations or between populations or species. While this is a common

approach, it is not always a successful method for identifying causal genes. For instance, comparing expression in relevant tissues between two species with different innate behaviors can identify thousands of differentially expressed genes. Thus, complementary methods such as QTL mapping or experimental manipulations are usually required to narrow down the list of correlated genes to the most likely candidates (e.g., [24]). For example, an *Aedes aegypti* mosquito olfactory receptor (OR) was first identified through RNA-seq comparisons between mosquitoes attracted to humans and those attracted to other animals; the expression of this gene was then found to correlate with behavior in a cross between these two types of mosquitos[27].

RNA-seq can be a powerful method in carefully designed comparisons of specific organs or tissues, brain regions, or cell types. For example, RNA-seq was used to identify the molecular bases of the evolution of infrared sensors in snakes and vampire bats. TRPA1 channels of pit vipers and some species of boas and pythons harbor mutations that make them very heat sensitive[28]. Moreover, these channels are expressed at much higher levels in the trigeminal ganglia—which innervate the heat-sensitive pit—than in the dorsal root ganglia, which transmit other somatosensory information[28]. TRPA1 is not upregulated in the trigeminal ganglia in snakes without infrared-detecting pits[28]. Vampire bats have evolved high trigeminal ganglion expression of an isoform of the heat-sensing TRPV1 that is particularly heat sensitive. This high TRPV1 isoform expression was also identified by measuring alternative splicing using RNA-seq in vampire bats and fruit bats[29].

Artificial selection on behavior. The methods described above are typically applied to study behavioral variation within or among natural populations. Behavioral differences can also be exaggerated through artificial selection over many generations to create strains with extreme behavior. The genetic differences among selected lines can then be probed by genome-

wide F_{ST} scans, through QTL mapping, or by comparing gene expression in relevant tissues or brain areas. Molecular signatures of selection can also be searched for within a selected line. These types of artificial selection studies have pointed to genetic regions and specific genes implicated in aggression in various species, including flies[30], rats[31,32], and foxes[14].

Identifying genes that modulate behavior

There are no genes that specify behavior[33]; rather, genetic variation modulates biochemical and cellular pathways and shapes neuronal circuits that ultimately give rise to behavior. Genetic variation can therefore affect behavior by acting at different levels: by altering sensory perception, by modulating higher-order circuits of the central nervous system, or by affecting metabolic processes outside of the nervous system.

As more genetic variants are identified, important evolutionary questions can begin to be answered. For example, are particular classes of genes or biological processes more often implicated in behavioral variation and evolution? And how often does independent evolution of similar behaviors converge on the same molecular pathways? Though we still know very little about the genetic mechanisms underlying most behaviors, the case studies highlighted in the following sections provide some of the first clues to the answers.

Genetic variation in sensory systems alters behavioral responses to external cues. Before environmental information is processed by the nervous system, signals must be detected by sensory receptors that are often housed in specialized sensory organs. Genetic variation that alters sensory detection—for example, by affecting the function of these receptors—can cause a direct behavioral shift by disrupting signal input. Alternatively, variation downstream from receptors can affect how a stimulus is processed after it is detected.

Mutations in sensory receptors are characteristic of behavioral shifts in natural populations. Cockroaches taste sugar using hair-like sensory structures protruding from their mouthparts that house neurons expressing taste receptors. Most cockroaches, like other insects, are attracted to sugars, including glucose; however, several populations of cockroaches have recently evolved glucose aversion. Glucose is a component of many commercial cockroach baits designed to poison the animals; therefore, behavioral attraction to glucose has negative fitness consequences in populations of cockroaches under this selection regime. The taste neurons that sense sugars—and mediate attraction—in wild cockroaches have become less sensitive to glucose in cockroaches from populations that avoid glucose, while the neurons that detect bitter compounds—and mediate aversion—have become more sensitive to glucose[34]. Thus, mutations affecting sensory neurons have changed the valence of glucose from attractive to aversive.

In humans, food and drink preference is modulated by smell. Some people are genetically predisposed to detecting the taste of cilantro as unpleasantly soapy, which affects diet choice and cilantro preference. A GWAS identified a genetic region significantly associated with this soapy-taste detection contained within a cluster of OR genes on chromosome 11. One such OR, OR6A2, has a high binding specificity for several aldehydes that give cilantro its characteristic odor[35].

Variation in food preference among species can arise from species-specific adaptations in sensory receptors. Most birds, including chickens, turkeys, and finches, have lost the ability to sense sugars, as they lack the sweet taste receptor gene *TAS1R2*[36]. Hummingbirds, however, are specialists that feed exclusively on nectar and have regained sugar sensation by repurposing the umami receptor (a dimer encoded by the genes *TAS1R1* and *TAS1R3*). Mutations in these genes

transform the receptor from one that detects savory amino acids into one that detects sugars, thereby permitting the characteristic specialization of nectar-feeding behavior in these birds[36].

Some animals communicate using chemicals called pheromones when they signal within species and kairomones when they signal between species. In moths and other insects, females produce and secrete sex pheromones that attract males. Female Asian corn borer moths (*Ostrinia furnacalis*) produce the sex pheromones (*E*)-12- and (*Z*)-12-tetradecenyl acetate, whereas female European corn borer moths (*Ostrinia nubilalis*) produce slightly different isomers, (*E*)-11- and (*Z*)-11-tetradecenyl acetate[37]. *O. furnacalis* and *O. nubilalis* males are attracted to the distinct pheromone blend from conspecific females due to a nonsynonymous mutation in the *O. furnacalis* pheromone receptor gene *Or3* that reduces *O. furnacalis* male response to the *O. nubilalis* pheromone 14-fold[38]. Therefore, genetic variation affecting pheromone receptors can mediate interspecies specificity in mate attraction.

Genetic variation modulating chemical communication also affects behavioral interactions between distantly related species. Some species of nematodes commensally infest live insects; the insects provide the nematode with dispersal opportunities, a food source, and, after the insect dies, a substrate for the nematode to continue its life cycle[39]. Some natural populations of the nematode *Pristionchus pacificus* associate with the oriental scarab beetle (*Anomala orientalis*)[40]. These beetles produce the chemical (*E*)-11-tetradecenyl acetate, and certain strains of *P. pacificus* nematodes are highly attracted to this kairomone[41]. Differences in (*E*)-11-tetradecenyl acetate attraction among *P. pacificus* strains map to variation not in a sensory receptor but rather in the protein kinase EGL-4[41], a component of the cGMP signaling pathway that regulates olfaction in *C. elegans*[42].

In *C. elegans*, pheromones that accumulate in high local population densities bind to the pheromone receptors SRG-36 and SRG-37 to stimulate progression into an alternative diapause-like state called dauer[23]. Under high-density selection regimes in the laboratory, two *C. elegans* strains have independently acquired resistance to dauer progression from nearly identical deletions affecting both *srg-36* and *srg-37*[23]. While no wild strain has been identified that harbors deletions affecting both genes, 18% of wild *C. elegans* strains from around the globe harbor a putative loss-of-function deletion within *srg-37* that is identical by descent[43]. Interestingly, there is enrichment of this allele in *C. elegans* populations that have colonized a rotting fruit niche (which provides bacteria that worms eat), suggesting that there may be particularly strong selection against dauer formation in worms exploiting resources that support reproductive growth[43].

Loss of sensitivity to environmental molecules may therefore be adaptive in certain habitats and can underlie the evolution of behavior. In highveld mole rats, multiple genetic changes affect both the protein sequence of the TRPA1 channel and the expression of a second channel protein, NALCN, in sensory neurons. These mutations confer insensitivity to the painful substance allyl isothiocyanate, a defensive compound produced by some insects and plants, and thus shapes mole rat behavior by permitting both feeding on pungent food sources and coexistence with aggressive stinging ants[44].

Decision-making must integrate both environmental signals and internal states such as hunger; therefore, sensation of internal cues can be as important as that of external cues. In *C. elegans*, the decision to abandon an area with food is modulated by noncoding variation affecting the G protein–coupled catecholamine receptor gene *tyra-3*. The receptor encoded by this gene is

expressed in sensory neurons yet binds internal biogenic amines (tyramine), suggesting that the gene modulates responses to the environment by integrating internal information[21].

Genetic variation alters behavior by modulating central nervous system circuitry. Behavior can also be modulated by variation affecting higher-order nervous system processing rather than sensory perception. Pioneering work on the genetic modulation of social behavior implicated variation in neuropeptide receptors in the brain, but more recent examples show that variation in other classes of neuronal molecules is also important for generating the diversity of social behavior observed both within and between species.

The G protein–coupled receptors of the neuropeptides arginine vasopressin (AVP) and oxytocin (OT) are classic examples of genes underlying natural variation in mating systems among species of voles[45]. Molecular approaches comparing the monogamous prairie vole (*Microtus ochrogaster*) and the promiscuous montane vole (*Microtus montanus*) implicated species-specific distribution patterns of AVP and OT receptors in the brain in many behavioral aspects of monogamy, including pair bonding, paternal care, and mate guarding[46]. Interestingly, QTL mapping in a different rodent clade—comparing the monogamous oldfield mouse (*Peromyscus polionotus*) and the promiscuous prairie deer mouse (*Peromyscus maniculatus*)—also implicated the AVP system in variation in parental care[24]. However, in *Peromyscus* mice, *cis*-regulatory variation affecting the expression of the AVP ligand in the hypothalamus, rather than the AVP receptor, is linked to the elaborate nest building characteristic of monogamous parents.

Insects also have diverse social structures; for example, some populations of the sweat bee *Lasioglossum albipes* produce solitary nests, whereas others are eusocial. A GWAS approach linked noncoding variation in syntaxin 1a (syx1a), a protein that mediates synaptic

vesicle release, to intraspecific variation in eusociality. A single intronic polymorphism of *syx1a* altered its expression in in vitro assays, consistent with in vivo expression differences between social and solitary bees[47].

C. elegans nematodes also vary in their social behaviors. For example, they differ in their propensity to aggregate with each other, a behavior that is influenced by pheromonal communication as well as by environmental variables such as food availability and oxygen levels[48–50]. Differences in this behavior were initially found to be strongly affected by a single amino-acid difference in the receptor *npr-1*, a gene homologous to the mammalian neuropeptide-Y receptor[51]. A later study, based on analyses of more than 200 wild *C. elegans* strains, strongly suggested that this mutation arose during the domestication of *C. elegans* in the laboratory[52]. Further quantitative genetics approaches implicated naturally occurring polymorphisms affecting the expression of EXP-1, a receptor for the neurotransmitter GABA, in the propensity of *C. elegans* to aggregate with each other[53].

As described in the previous section, females of some species attract conspecific males by emitting sex pheromones; however, changes in the valence of sex pheromones—whether they are perceived as attractive or aversive—can arise from genetic differences affecting central neural circuitry rather than peripheral sensory perception. Two species of *Drosophila*, *D. melanogaster* and *D. simulans*, are closely related yet are largely reproductively isolated due to differences in pheromone signaling between the species. *D. melanogaster* females, but not *D. simulans* females, produce the sex pheromone 7,11-heptacosadiene, which is highly attractive to *D. melanogaster* males but aversive to *D. simulans* males[54]. Interestingly, sensory neurons respond similarly to 7,11-heptacosadiene in *D. melanogaster* and *D. simulans*, but differences in

how the signal is propagated in downstream circuits in the fly brain explain this variation in behavioral response[54].

D. melanogaster and *D. simulans* diverge not only in their pheromone signaling but also in the male courtship songs that attract conspecific females. Differences in aspects of the courtship song between laboratory strains of these species mapped to a retroelement insertion into the intron of the *slowpoke* (*slo*) gene, which encodes a calcium-activated potassium channel expressed throughout the central nervous system[22]. The existence or prevalence of this mutation in natural populations is unknown.

Differences in social and mating behaviors may involve the coordination of many genetic variants that are selected upon through subtle changes in allele frequencies at many loci simultaneously (known as polygenic selection). For example, male cichlid fishes of Lake Malawi build bowers that attract females: Certain species dig pit bowers, whereas others build castles. Genetic variants across each of the 22 linkage groups in 20 diverse species of pit diggers and castle builders have elevated F_{ST} values, suggesting that the divergence in genetic architecture between the species is highly complex yet consistent across species, perhaps due to introgression[55]. F_1 hybrids between a pit-digging species and a castle-building species display both bower-building behaviors sequentially: During the pit-digging epoch, a suite of alleles inherited from the pit-digging parent become upregulated in the F_1 brain, while during the castle-building epoch, alleles inherited from the castle-building parent become upregulated. The temporal specificity of this allele-specific expression indicates that modular synchronization of transcriptomic responses can underlie the display of highly complex behaviors[55].

Variation in genes outside the nervous system affects behavior. Variation affecting genes that function outside of the nervous system can also modulate behavior (**Figure 1.1**). For

example, variation in metabolism can affect what and how much an animal chooses to eat or drink. Among mammals, variation in the copy number of the gene for amylase, a digestive enzyme that breaks down starch, correlates with starch preference. The more copies of the amylase gene a species has, the more starch it tends to eat as part of its diet. It is unclear, however, whether species-specific preference for starch drove the evolution of amylase copy number variation or whether copy number variation preceded starch preference[56]. In humans, individual differences in salivary amylase level and copy number affect the perception of texture and perhaps even flavor and likely affect starchy food preference[57,58].

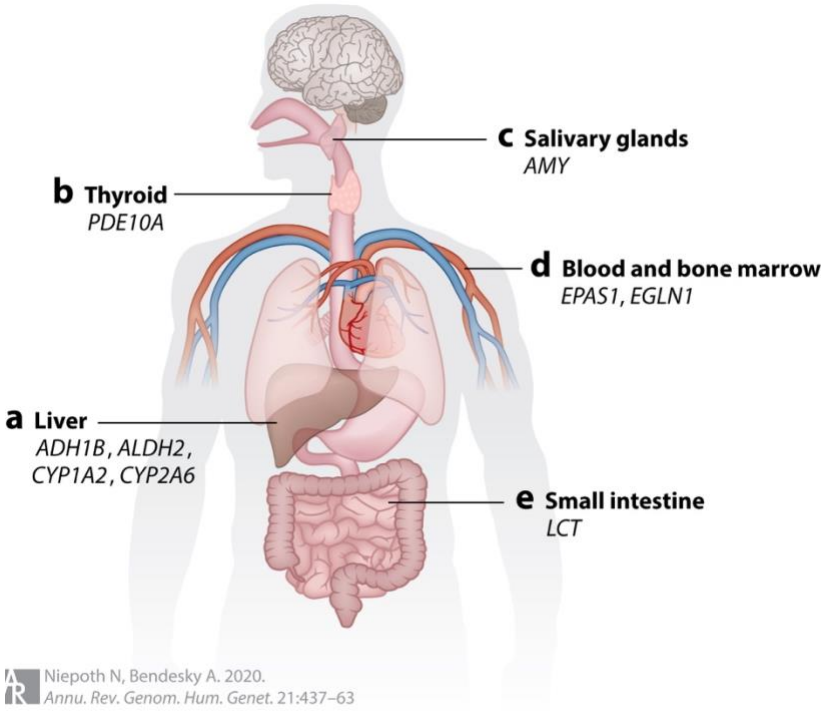

Figure 1.1: Non-neuronal genes that affect the behavior of humans and other vertebrates. (a) Polymorphisms in ADH1B and ALDH2 affect the rate at which alcohol is metabolized in the liver, affecting alcohol dependence[59,60]; noncoding single-nucleotide polymorphisms (SNPs) near the caffeine-metabolizing-enzyme gene CYP1A2 are associated with increased coffee drinking and caffeine consumption[61]; and coding SNPs in the nicotine-metabolizing-enzyme gene CYP2A6 are associated with cigarette-smoking behavior[62]. (b) Polymorphisms in the PDE10A gene of Bajau people cause increased spleen size, likely by modulating hormones released by the thyroid, thereby allowing specialized diving behaviors[63].

(c) Copy number variation of the AMY gene, whose product metabolizes starch, is correlated with the amount of starch mammals eat[56,58]. (d) Variants affecting EPAS1 and EGLN1 permit high-altitude adaptation and perhaps habitat preference in humans, Peromyscus mice, ducks, and other vertebrates (e.g., [64–67]). (e) Noncoding variants cause lactase persistence by prolonging expression of the LCT gene in the small intestine into adulthood[68].

A similar example in humans is the repeated evolution of the ability to digest lactose into adulthood, which affects how much dairy people eat. Most mammals cannot easily digest lactose after weaning; however, variants in and around the lactase gene drive its continued expression in the small intestine into adulthood, facilitating lactose digestion[69]. This phenotype arose independently in different pastoral human populations that came to rely on dairy as an important source of nutrition. The haplotype containing variants that contribute to lactose tolerance is identical by descent in Europeans and Indians[70] but different in Africans[71]. Both lactose-tolerance haplotypes have experienced a selective sweep over the past 7,000 years[71].

The amount of alcohol and coffee that people drink is strongly modulated by genetic variation. Alcohol is broken down into acetaldehyde and from there into acetic acid by the liver enzymes alcohol dehydrogenase (ADH) and aldehyde dehydrogenase (ALDH), respectively. Variants in paralogs of each enzyme, specifically ADH1B and ALDH2, are associated with alcohol consumption and are the common genetic variants with the strongest known effects on human behavior[60]. Because people with low-activity *ADH1B* alleles accumulate toxic acetaldehyde more slowly than people with high-activity alleles, they are less prone to the acetaldehyde symptoms characteristic of alcohol excess, including nausea and headache, and have a threefold-higher risk of developing alcoholism[72]. Low-activity *ALDH2* variants, which lead to an accumulation of acetaldehyde, have an even stronger effect on alcohol consumption and dependence. These variants are common only in some East Asian populations,

whereas *ADH1B* variants are also common in Europeans and Africans[73]. Similarly, coffee drinking is shaped by variation in genes involved in the metabolism of caffeine[61].

More than 1 billion people worldwide smoke cigarettes, and this behavior is largely mediated by dependence on nicotine, a highly addictive component of tobacco. Variation in the cytochrome P450 2A6 (CYP2A6) liver enzyme, which is essential for nicotine metabolism, is associated with the number of cigarettes smoked per day[74]. People who carry versions of CYP2A6 with reduced activity (and therefore slower metabolism of nicotine) smoke fewer cigarettes per day and usually find it easier to quit smoking[75]. However, variation in genes expressed in the central nervous system also modulates smoking: Nearly all nicotinic acetylcholine receptor (nAChR) genes expressed in the brain exhibit variation associated with smoking behaviors[74,76].

Adaptation to extreme environments can also arise from selection on variation affecting genes outside of the nervous system. Different human populations, including Tibetans and Andeans, have independently colonized extremely high-altitude environments and are genetically adapted to low-oxygen (hypoxic) conditions. High-altitude-adapted Tibetans carry a variant of *EPAS1*, which encodes a transcription factor regulating the production of hemoglobin and the development of new blood vessels, that helps them use oxygen more efficiently at high altitudes[67]. Both Tibetans and Andeans have signatures of positive selection on the *EGLN1* gene, whose product interacts with EPAS1[64]. Ducks adapted to high altitudes also carry variants of *EPAS1* and *EGLN1* at higher frequencies than lowland ducks[65], and *Peromyscus* mice harbor molecular signatures of selection at *Epas1*[66], suggesting that high-altitude habitat choice may converge on similar genetic mechanisms in distant species. Mutations in the *C. elegans* homolog of *EGLN1*, *egl-9*, strongly affect preference for high or low oxygen (aerotaxis)[77], suggesting that

variation in hypoxia-related genes could also affect vertebrate preference for different oxygen concentrations.

The choice of plants that herbivores eat is influenced by attraction, preference, and resistance to plant defensive compounds. Cardiac glycoside compounds produced by the milkweed plant are toxic for many herbivorous species. However, parallel evolution of cardiac glycoside resistance has permitted feeding on milkweed across many orders of insects[78]. For example, the monarch butterfly (*Danaus plexippus*) feeds on milkweed in its larval stage and sequesters the toxic chemical to deter predators as a butterfly. Cardiac glycoside resistance in many species is conferred by three mutations that alter three amino acids in the protein pump Na^+,K^+-ATPase, the molecular target of cardiac glycoside. Two studies have recently used phylogenetic comparative approaches and genetic engineering to prove that the order in which these three mutations evolved matters[79,80], highlighting how genetic interactions (epistasis) can constrain the paths through which behaviors evolve.

Two *Drosophila* species have independently evolved specialization to the toxic noni fruit *Morinda citrifolia*: a population of *D. yakuba* from the island of Mayotte and the noni specialist *D. sechellia* from the nearby Seychelles archipelago[81,82]. Variation in genomic regions linked to noni fruit adaptation overlaps more often than expected by chance between the Mayotte and Seychelles noni specialists, suggesting a parallel molecular basis to this specialization[82]. Mayotte *D. yakuba* showed strong signatures of selection in several detoxification genes compared with mainland *D. yakuba* generalists, including a major toxin tolerance locus previously identified in *D. sechellia*[82]. Species-specific attraction to noni fruit in *D. sechellia* is influenced not only by detoxification genes but also by variants affecting olfactory receptor tuning to noni fruit volatile chemicals[83,84]. Transgenic experiments demonstrate that OR22a,

which mediates long-range attraction to these volatiles, contains three naturally occurring amino-acid substitutions that each increase sensitivity to noni volatiles[85].

Social behavior is modulated by visual, auditory, mechanical, and chemical signaling between partners. Genetic variation that alters social signals can therefore strongly affect the behavior of animals receiving the signal. Divergence in female sex pheromone synthesis can acutely alter male attraction and promote divergence of male preference over longer timescales. In the European corn borer moth, two populations have begun to diverge in their pheromone signaling, leading to reproductive isolation. Female moths from the *E* population produce a blend of pheromone containing 98% (*E*)-11-tetradecenyl acetate and 2% (*Z*)-11-tetradecenyl acetate, whereas the *Z* population produces 3% (*E*)-11-tetradecenyl acetate and 97% (*Z*)-11-tetradecenyl acetate[86]. This divergence is caused by multiple nonsynonymous substitutions in a single fatty-acyl reductase gene involved in the synthesis of precursors to (*E*)-11- and (*Z*)-11-tetradecenyl acetate[86].

In nature, most *C. elegans* individuals are hermaphrodites with the ability to self-fertilize, while males occur at a frequency of less than 1%. An Australian strain of *C. elegans* exhibits male–male mating behavior caused by a natural loss-of-function mutation in a single gene (*plep-1*) expressed in the excretory pore; males homozygous for the *plep-1* mutation attract copulations from other males[87]. This result shows that even behaviors that appear complex can sometimes arise from mutations in single genes.

Lessons from Genetic Mapping of Behavioral Diversity. In the introduction to this section, I posed two major questions: What types of genes and biological processes does variation most often impact, and do similar behaviors evolve through similar or distinct molecular mechanisms? We do not yet have enough information to fully answer these questions,

but we can identify two major patterns. First, genetic variants affecting sensory receptors are very common. Additionally, genetic variation affecting the expression or protein sequence of other classes of genes in the brain, such as neuropeptide and neurotransmitter receptors, also characterizes behavioral divergence. In most cases, however, the specific mechanisms underlying these genetic effects are not well understood, even if a general biological pathway can be implicated in the behavior. Second, despite an expectation that parallel evolution of behavior—the independent evolution of a behavior based on changes in the same genes or pathways—might be rare due to the complexity of the genetic and neuronal bases of behavior, there are many examples of parallelism both within and among species. For example, the same hypoxia-inducible factor pathway is involved in adaptation to high altitude in humans, deer mice, and ducks, and variation in the vasopressin system affects monogamous behaviors in both voles and deer mice. Additionally, the courtship song of Hawaiian crickets has evolved in three independent pairs of species through changes at overlapping QTLs, suggesting parallel evolution[88].

Importantly, demonstrating that particular genetic variants influence behavior through their effects on specific genes (quantitative trait genes) is very challenging. The gold standard is the reciprocal hemizygosity test (and related tests)[89,90], which is rarely performed outside the powerful genetic model organisms *Drosophila* and *C. elegans*. Thus, many studies implicate genes based on protein-coding changes, allele-specific expression differences, proximity to mapped variants, and experimental manipulations. Each of these approaches has limitations, but together they can provide more convincing evidence for the effects of particular genes on behavior.

Types of molecular variation that modulate behavior

Genetic variation ranges from that affecting single nucleotides, to additions or deletions of thousands to millions of bases, to large-scale chromosomal rearrangements. Mutations affecting a small genetic region, such as single-nucleotide polymorphisms (SNPs) and short insertions or deletions (indels), are a major source of variation that can affect protein sequence or gene regulation. Other, larger-scale mutations, such as supergenes and gene expansions, are also prominent contributors to behavioral evolution.

Regulatory versus coding mutations. Mutations can alter the temporal and spatial regulation of gene expression or modify protein-coding sequences themselves. Both types of molecular changes have been shown to contribute to behavioral diversity. Protein-sequence changes may be particularly important in the evolution of sensory receptor tuning to various environmental cues. For example, coding variation in taste receptors permits hummingbird attraction to sugar[36], while sweet taste receptor genes in many carnivorous mammals have been pseudogenized[91,92] (**Figure 1.2**). However, protein-coding changes in genes that are widely expressed in the central nervous system can detrimentally disrupt essential networks, while regulatory variants alter gene expression more modularly[93]. Therefore, regulatory mutations rather than coding mutations are likely the primary type of variants affecting genes that are broadly expressed in the brain or are essential for neural development[93]. For example, regulatory variation affecting *slo*, a gene expressed ubiquitously in the fly brain, contributes to the evolution of the *Drosophila* courtship song[22].

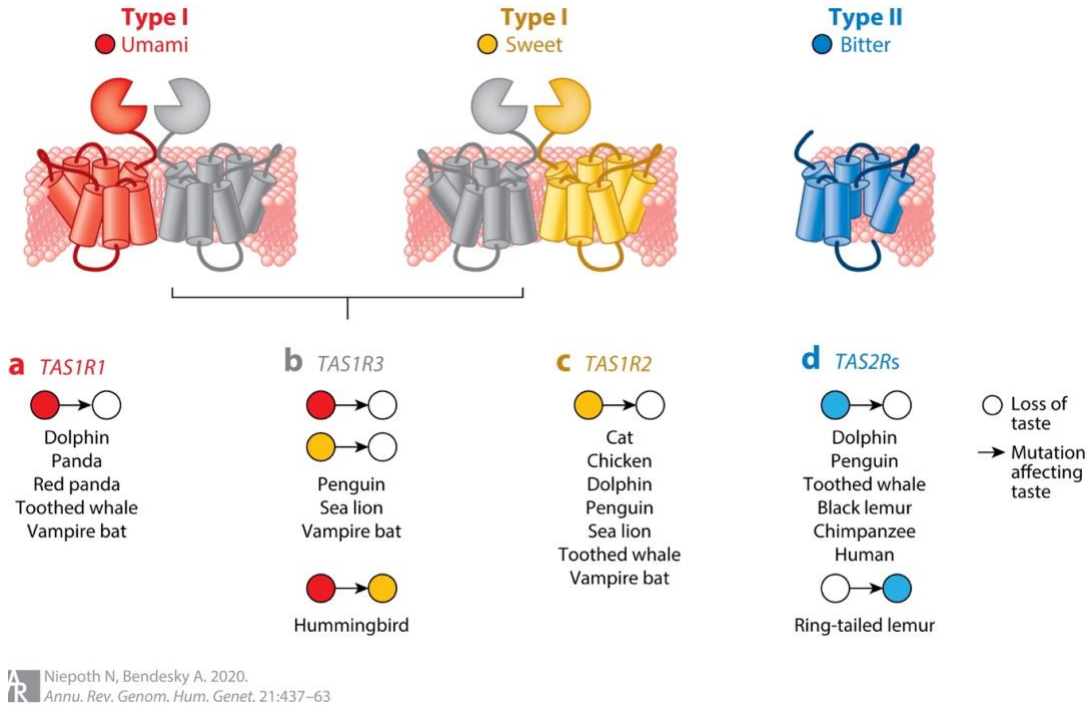

Figure 1.2. Molecular evolution of vertebrate taste receptor genes. In most vertebrates, the proteins TAS1R1 and TAS1R3 dimerize to form the umami taste receptor, while TAS1R2 and TAS1R3 dimerize to form the sweet taste receptor. Type II taste receptors (TAS2Rs) primarily detect bitter taste. Genetic variation across the animal kingdom has altered the protein structure and function of these genes, causing loss of function (pseudogenization or other coding changes) or gain of novel functionality, which alters taste perception. (a,c) Independent pseudogenization of the TAS1R1 and TAS1R2 genes caused loss of umami and sweet taste perception, respectively, in many taxa[91,92,94–100]. (b) Pseudogenization of the TAS1R3 receptor conferred loss of both umami and sweet taste perception in penguins, sea lions, and vampire bats[36,91,98,99], while coding mutations in TAS1R3 in hummingbirds transformed the TAS1R1/TAS1R3 heterodimer from detecting umami to detecting sweet tastes. (d) Pseudogenization of TAS2R receptors in dolphins, penguins, and toothed whales caused loss of bitterness perception[91,94,98]. Lemur-specific amino-acid substitutions in TAS2R16 changed the receptor response to arbutin from agonism to inverse agonism, thereby reducing sensitivity to salicin bitterness. However, ring-tailed lemurs regained the ability to recognize arbutin as a TAS2R16 agonist via coding mutations affecting the TAS2R16 sequence[101]. Independent mutations to TAS2R38 eliminate sensitivity to the bitter compound phenylthiocarbamide in some chimpanzees and humans[102].

Supergenes. Variation affecting behavior tends to be spread across the genome, where chromosomal segregation and recombination unlink variants that have beneficial effects on traits.

Genomic rearrangements that prevent recombination ensure that a block of the genome is inherited together and can therefore spread in a population. These supergenes can accumulate further genes and variants. Supergenes have strong effects on multiple behaviors across species (**Table 2**). In fire ants, a 13-Mb supergene contributes to variation in social organization[103]. Contained within each of the two nonrecombining supergene alleles, social B (*sB*) and social b (*sb*), are specific variants of the gene *Gp-9*, which encodes an odorant-binding protein that dictates whether colonies will accept multiple queens[104]. The supergene alleles also confer a difference in colony-level aggression[104]. Honeybees from highland and lowland populations in East Africa have rampant gene flow between them, with the exception of two haplotype blocks on two chromosomes that result from inversions[105]. Many genes within these supergene-like haplotypes influence honeybee behavior that may be adaptive in these divergent environments; for example, one haplotype contains nearly all of the octopamine receptor genes in the honeybee genome, and these genes play essential roles in learning and foraging behavior[105].

Table 2: Supergenes that affect animal behavior

Animal	Locus	Haplotype length	Behaviors affected	Nonbehavioral traits affected	References
White-throated sparrow (*Zonotrichia albicollis*)	*ZAL2*	98 Mb	Parental care, singing, mate preference, aggression, courtship	Plumage color	[106,107]
Ruff (*Philomachus pugnax*)	*Faeder*/*Satellite*	4.5 Mb	Mating strategy, territoriality	Plumage color, body size	[108,109]
Fire ant (*Solenopsis invicta*)	*SB*/*Sb*	13 Mb	Tolerance of multiple queens, aggression	Body size, queen fecundity	[103,110]
House mouse (*Mus musculus*)	*t* haplotype	40 Mb	Dispersal, migration	Spermatogenesis manipulation (meiotic drive)	[111–113]

Variation in mating behavior in the ruff, a wading bird, is also caused by linked variation within a 4.5-Mb inversion. The ruff has three male morphs (independents, faeders, and satellites) that differ in behavior, color, and size, representing three lekking strategies during which males aggregate and compete for access to females. Independents have retained the ancestral genotype (no inversion), while the faeder allele arose from an initial inversion and the satellite allele likely originated from an unlikely recombination event between faeder and independent alleles[109]. Another avian species, the white-throated sparrow, contains a chromosomal rearrangement at the *ZAL2* locus. Genetic variants within the *ZAL2* inversion (*ZAL2*^{m}) increase the expression of estrogen receptor α (*ESR1*) in specific brain regions, which causes heightened aggression[114]. Sparrows containing the *ZAL2*^{m} allele also have alternate territorial song, nestling-provisioning, and mate-guarding behaviors compared with sparrows that do not contain the inversion[115,116].

Gene expansions. While behavioral diversity due to supergenes acts within species, a major source of evolutionary divergence among species is large-scale gene expansion. Gene duplication can relax selective constraint on one of the copies and allow the gain of novel functionality (neofunctionalization) from new mutations in paralogs. Across species, large expansions or contractions of gene repertoires can shape species-specific behavior. There are many well-documented expansions of sensory gene repertoires, suggesting that sensory gene evolution has been a steadfast process powering behavioral evolution. Sensory gene radiation across mammals has occurred to the greatest extent in olfactory and vomeronasal receptors[117,118]. Because of combinatorial olfactory perception for most odorants—a regime under which individual receptors participate in the detection of specific odors but are neither necessary nor sufficient—olfactory genes may be particularly mutable across deep evolutionary timescales and can be prime sources of genetic variation affecting behavior.

Hundreds of gains and losses of OR genes have occurred across different lineages of reptiles and mammals[117]. For example, primates have fewer than 400 functional OR genes, while dogs and rodents have two and three times as many OR genes, respectively. Variation in the number of functional OR genes among different lineages appears to be driven by ecological adaptation. In birds and reptiles, for example, patterns of OR expansion correlate with the ecological requirements of the lineage. In diverse bird species, specific OR family expansions coincide with aquatic adaptations (water birds), vocal learning, and land specialization[119]. The expansion of *OR5*, *OR8*, and *OR9* occurred in both predatory birds and alligators, suggesting an adaptive role for those genes in carnivory[119]. Surprisingly, large expansions of these genes are actually linked to herbivory in mammals[120]. Though neutral evolutionary processes likely contribute somewhat to the rapid duplication and pseudogenization of ORs[121], it has been shown that ORs in great apes are under selective constraints[122]. Furthermore, the correlation between OR evolution and ecological requirements suggests that at least some families of OR genes are likely under positive selection in diverse animal taxa and that neofunctionalization of ORs may play an important role in behavioral adaptation.

Aquatic mammals, such as whales, are characterized by a reduction in ORs relative to their land ancestors, concordant with the evolution of other sensory modalities, such as echolocation[121,123]. Similarly, a reduction of functional OR genes in primates may be related to their acquisition of three-dimensional color vision (trichromacy) due to adaptive variation in pigments (called opsins) that allow vision in vertebrates[124]. In primates, color vision likely has important consequences for behaviors such as foraging, mate choice, predator avoidance, and navigation[124]. The most light-sensitive type of opsin—responsible for vision in dimly lit conditions—is the rod opsin gene rhodopsin (*RH1*), and most vertebrate taxa possess just a

single *RH1* gene. However, three deep-sea teleost lineages have independently gained additional copies of *RH1*[125,126], suggesting that these expansions have permitted these lineages to live in the deep sea. The deep-sea silver spinyfin in particular has expanded its RH1 repertoire to 38 rod opsins, the largest number known in any vertebrate[126]. Protein regeneration and simulation have shown that these spinyfin RH1s are tuned to a wide range of light wavelengths that encompass the bioluminescence spectrum of the deep sea, suggesting this expansion may allow spinyfins to better perceive bioluminescent signals important for adaptation.

Vomeronasal receptors that bind pheromones are also among the fastest-evolving genes in mammals and have gone through huge expansions in some species of rodents and loss of all functional genes in catarrhine primates and dogs[127,128]. Mice have not only more than 1,000 ORs but also more than 350 vomeronasal receptors (V1Rs and V2Rs) that allow specialized olfaction of pheromones essential for regulating social behaviors such as mating, parenting, and aggression[129,130]. The Lake Victoria cichlid fish *Haplochromis chilotes* also has an expanded repertoire of vomeronasal type II receptor-like genes (*OlfC* genes), which has been suggested to contribute to its extraordinary feeding behavior diversification by allowing for the detection of a wide range of amino acids[131].

Sources of genetic variation contributing to behavior

Genetic variation fundamentally arises through mutation and spreads within and between populations through migration and mating. New mutations, standing genetic variation, and gene flow between populations and species are important sources of variation that contribute to behavioral evolution.

New mutations. The ultimate source of genetic variation is new mutations. While most mutations are deleterious and disappear quickly, some are maintained in the population at low frequencies. A third, rare outcome is the selective sweep, whereby a beneficial mutation spreads rapidly due to positive selection. A classic example of a selective sweep of a behavior-modulating variant is the spread of lactase persistence alleles in Europeans (Section 3.3). These alleles were not detected in ancient DNA samples from early Neolithic Europeans, suggesting that they arose recently[132]. In horses, a mutation in the gene *DMRT3*, encoding a transcription factor that affects the differentiation of spinal cord interneurons, likely arose within the last 10,000 years[133]. While most horses with the ancestral *DMRT3* allele have a limited locomotive repertoire (walk, trot, or gallop), horses containing this recent variant of *DMRT3* exhibit unusual gaited locomotive patterns[134]. This variant was artificially selected for by humans, presumably based on its interesting effect on horse locomotion, producing tens of gaited horse breeds that exist today.

Standing genetic variation. Selection can alter the allele frequencies of either new mutations or preexisting genetic variation in the population. Standing genetic variants may persist at low frequencies in the population in the absence of selection and then segregate at intermediate frequencies in response to soft sweeps, genetic drift, or balancing selection. Due to cryptic genetic variation, variants that confer small or no phenotypic effects in particular environments can allow for behavioral adaptation when environments change[135].

Selection on standing genetic variation underlies variation in schooling behavior between marine and freshwater stickleback fishes. Sticklebacks from marine populations overwhelmingly carry the ancestral allele of the gene *Eda*, but the alternate allele persists in the population at low frequencies and has repeatedly become fixed in many independent populations that have

colonized freshwater habitats[136]. Marine and freshwater sticklebacks differ in various aspects of their schooling behavior, including the angles of their bodies during schooling. Differences in this body position map to variation at the *Eda* locus[137], and follow-up transgenic experiments confirmed the functional effect of *Eda* expression on schooling behavior variation[138].

It has been argued that soft sweeps on standing genetic variation are more common in human adaptation (including behavioral adaptation) than hard sweeps following new beneficial mutations[139]. For example, the *PDE10A* allele that increases spleen size and helps breath-holding diving in the Bajau is present in 37% of Bajau people but less than 7% of people in closely related populations[63]. In humans, genes involved in central nervous system development appear to be particularly enriched for adaptation from standing genetic variation[139].

***Gene flow and adaptive introgression*.** Hybridization with other populations or other species can also introduce genetic variation that affects behavior. Neanderthals and Denisovans colonized Europe and Asia hundreds of thousands of years before modern humans left Africa[140]. When modern humans expanded out of Africa, they mixed with Neanderthals and Denisovans, and gene flow from those archaic humans provided modern humans with genetic variants that facilitated their adaptation to their new environments. For example, the *EPAS1* gene in Tibetans, which now permits high-altitude living (Section 3.3), was introgressed from Denisovans[141]. Present-day Europeans also bear genomic signatures of gene flow with Neanderthals, and introgressed Neanderthal DNA affects many behavioral traits, including sleeping patterns, mood, and smoking[142].

The evolvability of behavior

Is behavior more evolutionarily labile than other traits? Phylogenetic patterns across species indeed suggest that behavior may be particularly evolvable[143]. For example, in primates, the phylogenetic signal—the conservation of a trait among lineages across evolutionary time—is typically lower for behavioral traits such as diet choice, sociability, and foraging patterns than for morphological and life-history traits[144]. A phylogeny of *Polyrhachis* ants contains many evolutionary transitions of highly intricate social nest-weaving behavior[145], suggesting that even complex behaviors can readily evolve in different species.

Sensory receptors are encoded by some of the most evolutionarily labile genes in the animal kingdom, perhaps allowing for rapid evolution of signal perception while bypassing potential negative pleiotropy of genetic change to higher-order circuits. The types of natural genetic variation that affect behavior are nonetheless incredibly diverse: An individual behavior may be modulated by many types of genes either inside or outside of the nervous system (Section 3). However, certain systems may be more adaptable than others, promoting evolutionary parallelism (Section 3.4). Population-level mechanisms that maintain genetic diversity can provide the variation necessary for rapid evolution. In particular, balancing selection likely plays an essential role in maintaining behavioral variation by preserving multiple alleles in a species (Section 4.3.1). Furthermore, standing and cryptic genetic diversity provides an adaptive substrate for selection when environmental pressures change (Section 4.2.2).

Emerging Patterns and Outstanding Questions

There is an extraordinary diversity of behavior across the animal kingdom, and we still have much to learn about the genetic contributions to such diversity. However, a few general

patterns are beginning to emerge. In general, many genes and many genetic variants contribute to specific behaviors, and these variants can affect gene regulation or protein sequence. Tentatively, protein-coding changes appear to be enriched in genes that interact with environmental molecules to modulate behavior, such as those encoding sensory receptors and enzymes. Furthermore, there are many examples of genetic variation affecting sensory systems, but it is not yet clear whether this represents a primary source of adaptation or is merely a system where genetic effects can be more easily detected or dissected.

An important remaining question is to what extent the genetic architecture of behavior differs from those of nonbehavioral traits. Unlike other quantitative traits, such as metabolite concentrations or gene expression levels, behavioral traits are not discrete molecules that can be measured, but are rather more arbitrary constructs whose magnitude and scale depend on how they are defined and measured. Thus, it is difficult to quantitatively compare the number and effect size of loci associated with behavior with those of other traits. Qualitatively, however, the genetic architecture of behavior appears to be similar to those of other traits: Multiple loci of small effects usually contribute to variation in behavior within species and among closely related species.

On the other hand, emerging evidence suggests that balancing selection is a particularly important evolutionary force shaping the function of the brain and behavioral patterns compared with other traits[146,147]. In addition, certain molecular events, such as large-scale changes in particular classes of genes and the contribution of supergenes, appear to be particularly prominent in behavioral evolution. How these forces and molecular mechanisms constrain or facilitate behavioral evolution remains an open question.

The Future of Behavioral Genetics

Whereas genetic mapping approaches have yielded many loci linked to behavioral traits, a pressing issue in behavioral genetics is how to identify the genes affected by the variants linked to behavioral variation. A common approach is to assume that the gene closest to the peak of linkage or association is the causative (quantitative trait) gene. Benchmarking using well-curated molecular traits indicates that 70% of causative genes are closest to peaks of association in GWASs[148], but this proximity might be lower for behavioral traits because neuronal genes tend to have highly elaborate regulatory mechanisms[149,150]. Thus, other lines of evidence are necessary to implicate specific genes in trait variation. The gold standard is the reciprocal hemizygosity test[90], but this test is difficult to perform in animals that lack powerful genetic tools.

With the increasing number of behavioral GWASs in humans and the development of polygenic scores to predict traits, some might be tempted to use such scores to study the genetic bases of behavioral differences among populations. However, because of gene–gene and gene–environment/culture interactions, population stratification, and lack of knowledge about causal variants (in most cases we know only of associated haplotypes in specific populations), translating polygenic scores estimated in one population to another is highly problematic[151].

Convincing cases of genetic contributions to differences in behavior between human populations have identified peaks of genetic differentiation (e.g., [63,67]). These studies also find evidence that variation in the trait within a population is associated with polymorphisms within these peaks of genetic differentiation that fall near genes implicating specific biological functions[63]. These three pillars—loci strongly differentiated among populations of interest,

association of loci with behavior within populations, and functional evidence supporting links between variants or genes and behavior—are good guideposts for future population-genetics-based studies in humans and other species.

Our knowledge of behavioral variation has been historically limited to select groups (i.e., to laboratory model species or, in human genetics, to European populations). However, novel and low-cost technologies now allow geneticists to study essentially any species, which can help to answer questions about preferred targets of behavioral diversity (parallelism) and to discover new genes that affect behavior. New methods for gene editing in nonmodel organisms may also advance our understanding of the biological mechanisms underlying variation in behavior. Expanding genetic analyses of behavior to other human populations will have substantial impacts on psychiatric genetics and on public health throughout the world.

Chapter 2: Genetic causes and behavioral consequences of a newly evolved adrenal cell type in monogamous mice

Adapted from manuscript in review:

Niepoth, N., Merritt, J.R., Uminski, M., Lei, E., Gebhardt, C., Lutzu, S., Wacker, S.A., Rudolph, S., Bendesky, A. Genetic causes and phenotypic consequences of a newly evolved adrenal cell type in monogamous mice.

Introduction

Recent advances in molecular profiling have revealed that distinct cell types modulate specific social behaviors. For example, scientists have used single-cell RNA-sequencing of neuronal populations in the rodent hypothalamus to identify cell populations that are essential for mating, maternal aggression, and parental care[152–154]. Using single-cell sequencing, scientists have also discovered how assemblages of cell types differ between taxa and how novel cell types arose over evolutionary time[155–162]. However, with some exceptions (e.g. [163]), the function of evolutionarily novel cell types has not been firmly established, and the genetic basis underlying the evolution of new cell types is largely unknown[164].

Here, I leverage the genetic similarity between two sister species of *Peromyscus* mice, the prairie deer mouse *P. maniculatus bairdii* (henceforth "deer mouse") and the oldfield mouse *P. polionotus subgriseus*, to probe the genetic causes and to quantify the biochemical and behavioral consequences of a recently-evolved cell type in the oldfield mouse adrenal cortex. Due to their recent evolutionary divergence (~1.8 million years[165,166]) these two species can hybridize, permitting forward genetic analysis of behavioral, anatomical, and molecular traits that have evolved between the two species. One such trait is parental behavior: deer mice are promiscuous and fathers do not exhibit care towards pups, whereas oldfield mice evolved

monogamy and biparental care[24,167–170]. Additionally, as I show in this chapter, deer- and oldfield mice differ dramatically in their adrenal gland morphology and function. Using histology and transcriptomics, I characterize the vast divergence of the adrenal gland at single-cell resolution, and I identify a novel cell type in oldfield mouse adrenals. I quantify the phenotypic effect of this adrenal cell type via pharmacology and electrophysiology, as well as use quantitative genetic mapping methods to identify the genetic loci contributing to the evolution of this cell type.

The adrenals of *Peromyscus* sister species have diverged in size and morphology

The adrenals are bipartite endocrine glands that secrete catecholamines from the medulla and steroid hormones from the cortex. Catecholamines, mainly in the form of adrenaline, acutely modulate physiology, whereas steroid hormones have both fast and slow effects on neuronal activity and behavior by acting on both membrane receptors and nuclear receptors[171]. I discovered that the adrenal glands of adult deer mice and oldfield mice differ significantly in size and weight. Whereas the adrenal glands of deer mice and house mice (*Mus musculus*) are similar in size, the adrenals of oldfield mice are 4-fold heavier than those of deer mice, and 6-fold heavier after adjusting for body weight (**Figure 2.1a,c**). The oldfield mouse adrenal is already two-fold larger than the deer mouse adrenal at birth, indicating that the size difference has an embryonic origin (**Figure A1.1a**). The oldfield adrenal, but not the deer mouse adrenal, then continues to grow throughout adulthood (**Figure A1.1b**). This extraordinary divergence in internal organ size between closely related species is unprecedented.

The adrenal medulla and the adrenal cortex have different embryonic origins: the medulla is derived from the neural crest, whereas the cortex is derived from the mesoderm[172]. Because the overgrowth could be limited to one of these two tissues, I next measured them individually. I

discovered that the oldfield cortex is 8.9 times larger by volume than the deer mouse cortex, controlling for body weight, whereas the adrenal medulla is only 2.6 times larger (**Figure 2.1b,d; Figure A1.2a**). The enlargement of the oldfield cortex is partly due to bigger cells (**Figure A1.2b**), and partly due to a 4.8-fold increase in the number of cells (**Figure 2.1e**). Thus, the oldfield mouse adrenal cortex is much larger than deer mouse adrenal cortex due to a combination of having more cells and larger cells.

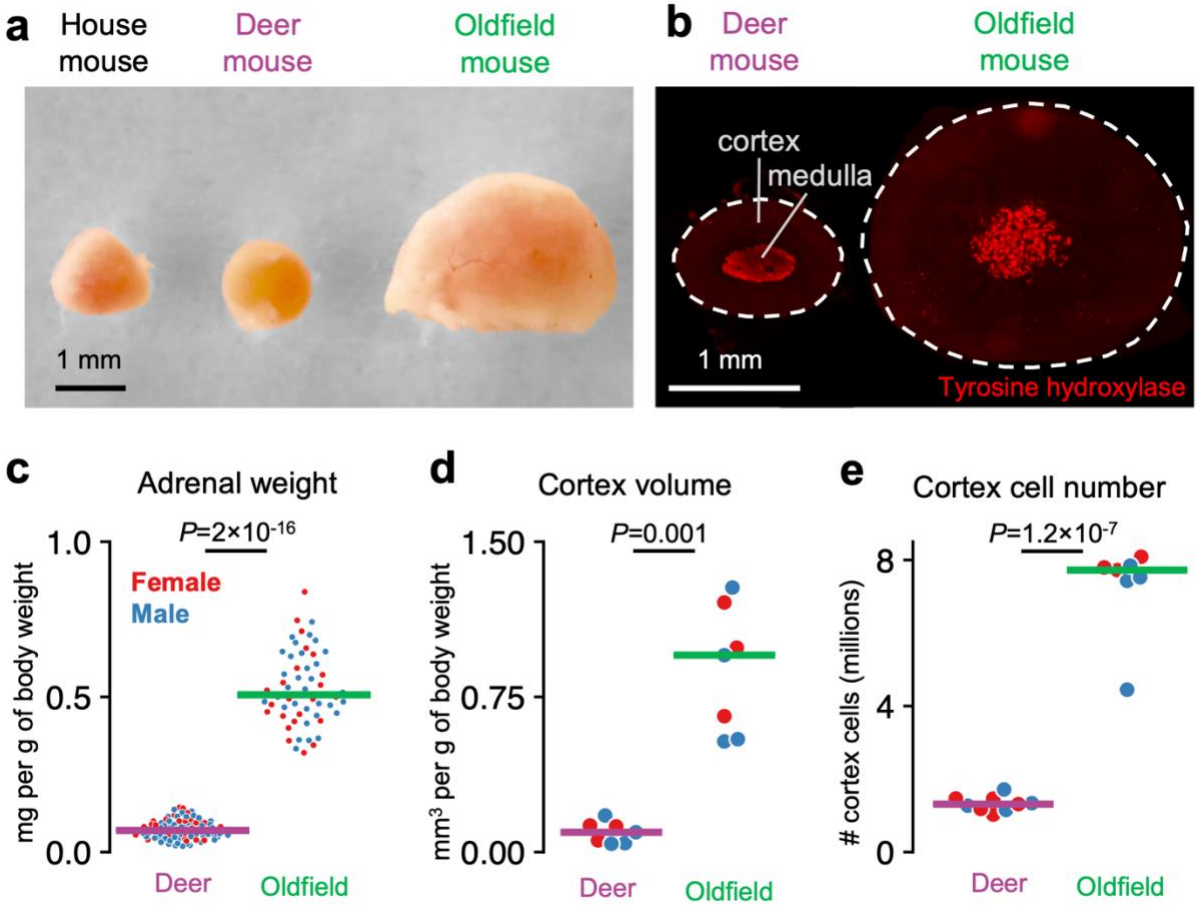

Figure 2.1. Oldfield mice have enlarged adrenal glands. a, Photo of adrenal glands from house mouse *Mus musculus*, deer mouse *Peromyscus maniculatus bairdii* and oldfield mouse *Peromyscus polionotus subgriseus*. **b**, Representative mid-adrenal sections from deer- and oldfield mice labeled with antibody against tyrosine hydroxylase, a marker of the adrenal medulla. **c**, Adrenal weight (species $P=2\times10^{-16}$, sex *n.s.*, species × sex *n.s.*, generalized linear model; deer: N=126, oldfield: N=72). **d**, Cortex volume (species $P=0.001$, sex *n.s.*, species × sex *n.s.*, generalized

linear model; N=7 per species). **e**, Cortex cell number (species P=1.2×10^{-7}, sex *n.s.*, species × sex *n.s.*, generalized linear model; deer: N=9, oldfield: N=7). All lines at median.

Oldfield mouse adrenals have a novel cell type

To begin to characterize the molecular differences between deer- and oldfield mouse adrenals, I performed RNA sequencing (RNA-seq) of the whole gland. Paralleling the anatomical divergence between species, I found more than 50% of genes expressed in the gland were differentially expressed between the species (**Figure A1.3**). To determine if the two species differ in the cell types that compose their adrenal glands, I conducted adrenal single nucleus RNA-sequencing (snRNA-seq) in juvenile and adult mice of both species. To analyze both species in a single framework, I integrated the transcriptomic data using reciprocal principal component analysis[173] and then used a graph-based clustering algorithm to identify cell types[174]. This analysis identified multiple cell clusters, including endothelial, stromal, glial, medullar, and steroid-producing cortex cells, consistent with the known histology of the mammalian adrenal gland (**Figure 2.2, Figure A1.4**).

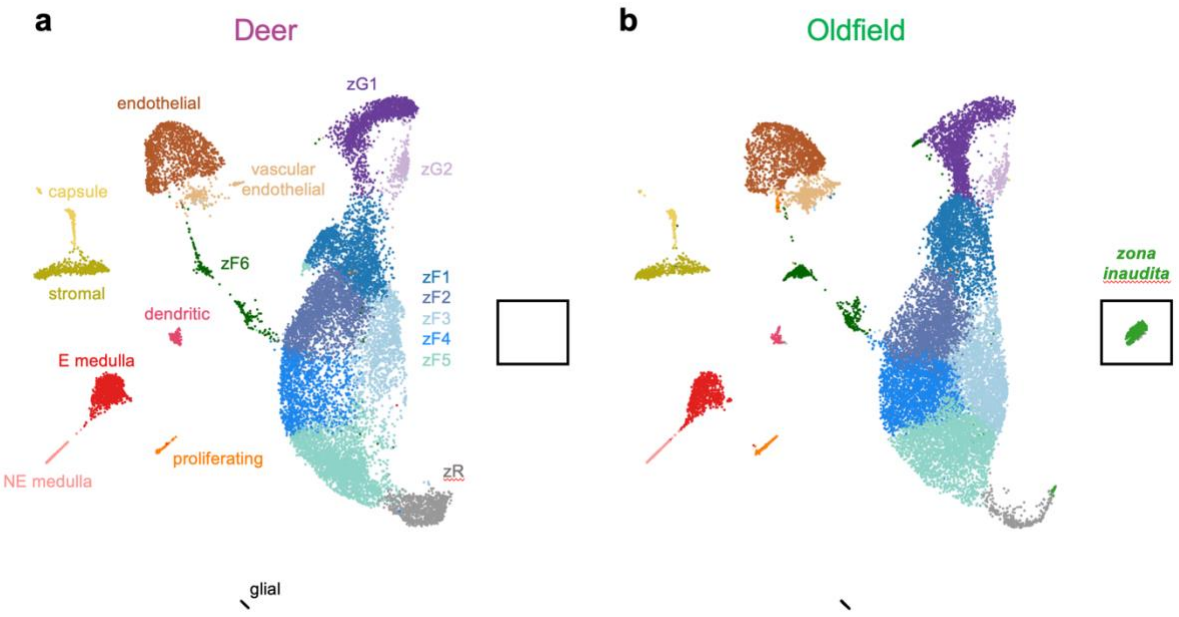

Figure 2.2. Cell types of the adult deer- and oldfield mouse adrenal gland. UMAP of adrenal snRNA-seq after reciprocal PCA integration of deer mouse and oldfield mouse cells.

The mammalian adrenal cortex is organized into concentric layers with specialized steroidogenic functions[171]. An outer *zona glomerulosa* ("zG") produces the mineralocorticoid aldosterone; a middle *zona fasciculata* ("zF") produces glucocorticoids such as corticosterone in most rodents; and an innermost *zona reticularis* ("zR") produces androgens in some species including humans but not in many rodents. Based on house mouse, brown rat, and human markers for cells in these zones[175–179], coupled with histological analyses and integrations with snRNA-seq data of house mouse adrenal glands, I identified clusters in the single-cell data that define all three layers in both *Peromyscus* species (**Figure 2.3, 2.4**; **A1.6**).

Consistent with their steroidogenic function, cells in all three layers express steroidogenic acute regulatory protein (Star), cholesterol side-chain cleavage enzyme (Cyp11a1), and 3β-hydroxysteroid dehydrogenase (3β-Hsd), which form the first three steps in the steroidogenesis pathway from cholesterol to progesterone (**Figure 2.4**). As in house mice and

rats[180], I found no expression of the androgen-synthesizing enzyme cytochrome p450 17a1 (Cyp17a1) in any cell of the adrenal gland, suggesting that deer- and oldfield mice do not produce androgens in these glands. Cells in the zF in both species were marked by steroid 11β-hydroxylase (Cyp11b1, which synthesizes corticosterone (**Figure 2.3, 2.4**). Altogether, I identify cell types that correspond to the three typical adrenal cortex zones in both species.

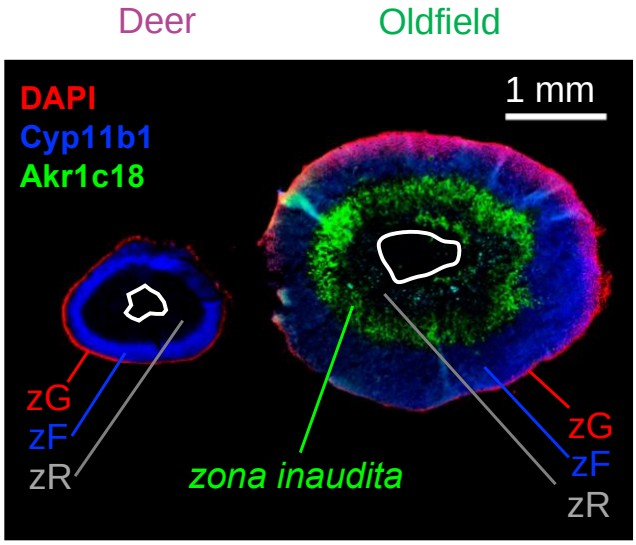

Figure 2.3. Histological characterization of the *zona inaudita*. I*n situ* hybridization of Cyp11b1 (*zF*) and Akr1c18 (*zI*) for visualization of adrenocortical zonation. The *zG* is marked by high cell density; see Figure A1.5.

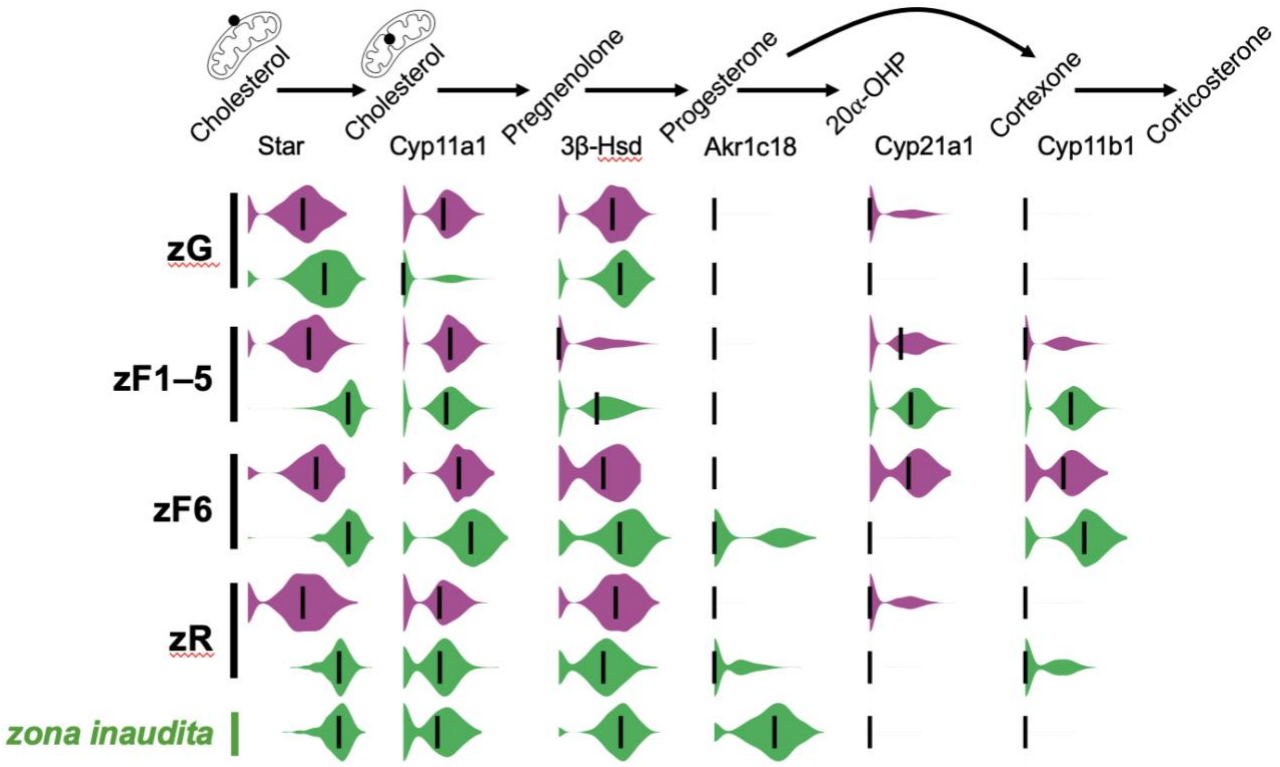

Figure 2.4: Expression of steroidogenic enzymes across cell types of the deer- and oldfield adrenal cortex Violin plots denoting the distribution of expression of steroidogenic enzymes in the corticosterone- and 20α-OHP synthesis pathways across cortex cell types (purple: deer mouse, green: oldfield mouse)

Notably, in addition to the three classic steroidogenic cell types, I identified a fourth cellular cluster that also expressed genes necessary for steroidogenesis. This cluster was present only in the oldfield mouse adrenal and not in the deer mouse adrenal nor in the house mouse adrenal gland (**Figure 2.5**; **Figure A1.4**, **A1.7**). Through histological staining of aldo-keto reductase 1c18 (Akr1c18) and tenascin N (Tnn), which mark this cell cluster, I find that it forms a prominent layer between the zF and the zR (**Figure 2.3; Figure A1.5**). Because a layer in this

location has not been described in mammalian adrenal glands before, I called this layer the *zona inaudita*, or "previously unheard-of zone".

In house mice, Akr1c18 is a marker of a transient layer abutting the medulla called the "X zone"[181,182], which disappears during puberty in males and during pregnancy or around 3–7 months in the absence of pregnancy in females[183,184]. However, integration of snRNA-seq data between oldfield mice and the C57BL/6 strain of house mice demonstrates that the *zona inaudita* and the X zone are not homologous cell types (**Figure 2.5**; **Figure A1.7**). Furthermore, the *zona inaudita* follows a different developmental trajectory than cells of the X zone: after *zona inaudita* cells arise on postnatal day 21–25, they persist in the oldfield adrenal gland into adulthood in both sexes and after parenthood (**Figure A1.8**). These results indicate that oldfield mice evolved a novel cell type forming a *zona inaudita* layer without precedent in other rodents.

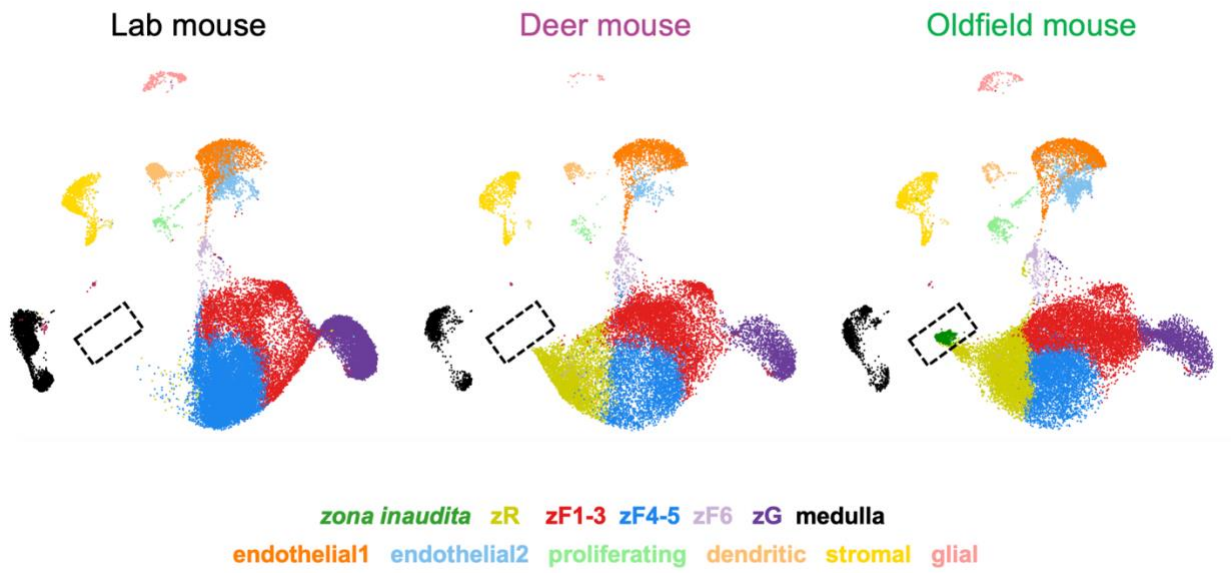

Figure 2.5: *Zona inaudita* is not present in house mouse adrenals. UMAP visualization of cells from C57BL/6 house mouse, deer mouse, and oldfield mouse after RPCA integration of cell types. Dotted box: the oldfield-specific *zona inaudita* cell type cluster.

To characterize how the molecular function of the *zona inaudita* cell type has diverged from the other cortex zones, I next identified genes that are more highly expressed in *zona inaudita* cells compared to other cortical cell types. I found 188 genes that distinguish *zona inaudita* cells from other adrenocortical cells (**Table A1.1**). Notably, the *zona inaudita* has very high expression of Akr1c18, whose murine and human homologs encode 20α-hydroxysteroid dehydrogenase (20α-HSD) enzymes that catalyze the conversion of progesterone to 20α-hydroxyprogesterone (20α-OHP; **Figure 2.4**)[185–187]. While Akr1c18 is virtually absent from all deer mouse cells, it is also expressed in a subset of zR and zF6 cells in the oldfield mouse, suggesting a convergent gain of expression of this enzyme in multiple oldfield cell types. In addition to gain of Akr1c18 expression, I found that the *zona inaudita* contains an enrichment of markers whose protein product localizes to the extracellular matrix (ECM; enrichment $P_{adjusted}=10^{-12}$), including core ECM proteins such as Tnn and podocan-like 1 (Podnl1), as well as ECM-associated genes such as plasminogen activator inhibitor 1 (Serpine1; **Figure 2.6**). Akr1c18 and many ECM *zona inaudita* markers are among the most differentially expressed genes between the deer- and the oldfield mouse adrenals (**Figure A1.3**). Additionally, I found that the *zona inaudita* has higher expression of several transcription factors compared to other cortical cell types (**Figure 2.7**), including Batf and Ezr, which were not expressed in any cortex cell in deer mice. Altogether, I found that the *zona inaudita* cell type is characterized not only by high levels of Akr1c18 expression and the upregulation of several transcription factors, but also by its production of a unique ECM that may contribute to the differentiation and/or maintenance of this adrenal layer.

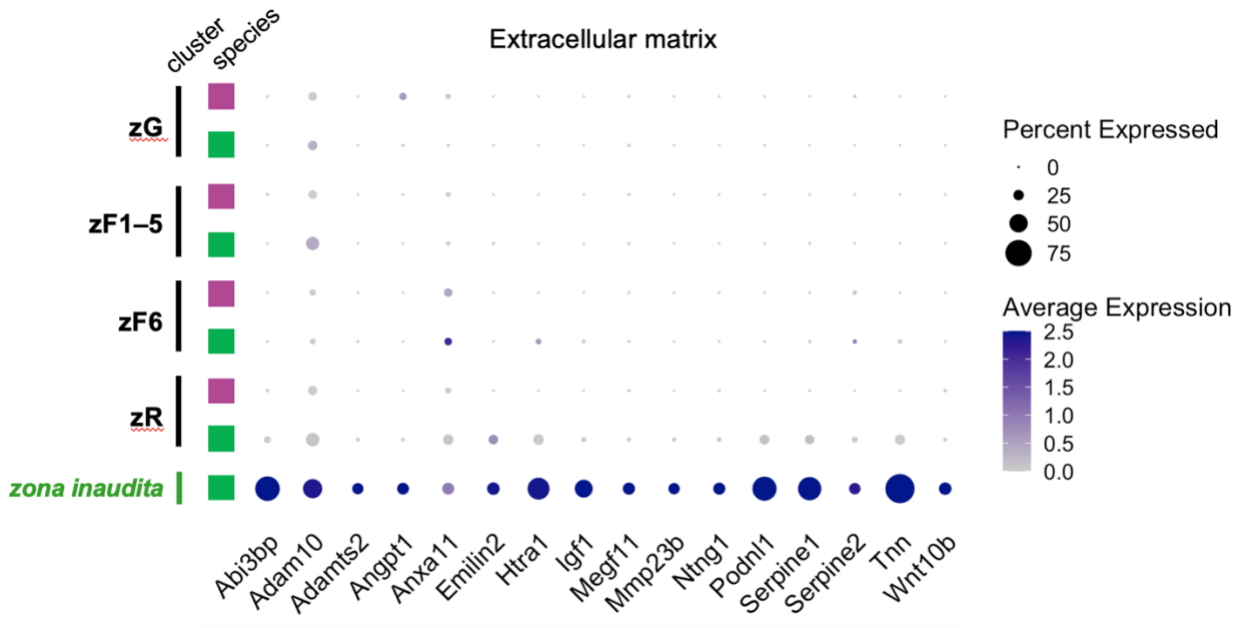

Figure 2.6. Expression of extracellular matrix genes upregulated in the *zona inaudita*. zI ECM genes. Expression of markers localized to the extracellular matrix across cortical cell types of both species

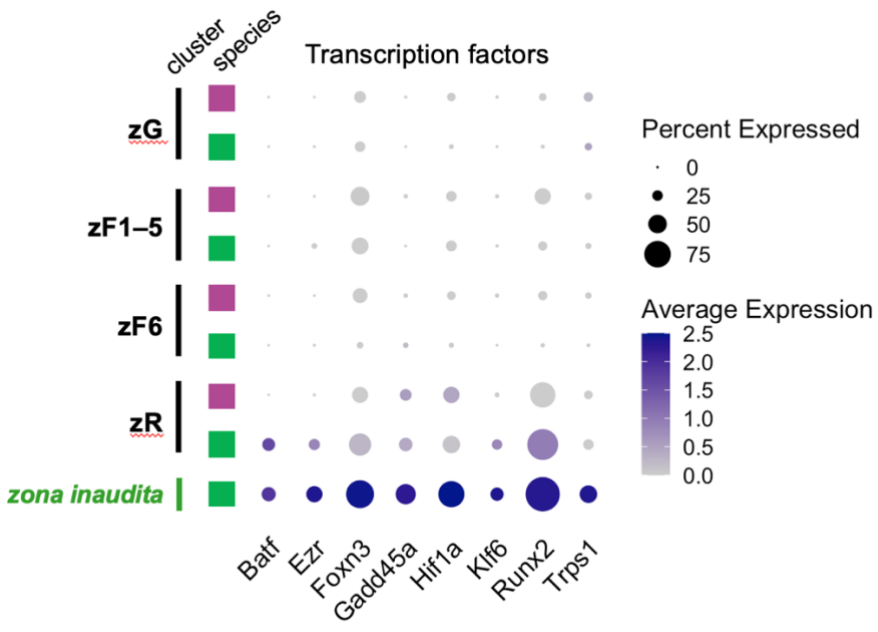

Figure 2.7. Expression of transcription factors upregulated in the *zona inaudita*. Expression of transcription factor markers of the cortical cell types of both species

Oldfield *zona inaudita* cells produce 20α-OHP

Akr1c18 is not only a highly significant marker of oldfield *zona inaudita* cells, but it is actually one of the 0.16% most highly expressed of all oldfield adrenal transcripts, constituting 1 of every 500 polyadenylated RNA molecules in the gland (**Figure A1.3**). Akr1c18 encodes an enzyme whose homolog in house mice and brown rats converts progesterone to 20α-OHP, a poorly studied steroid hormone produced in the ovaries of those species[185,186,188]. Whereas deer- and oldfield mouse AKR1C18 do not differ in their amino acid sequence, they are only 85% identical to house mouse AKR1C18 (**Figure 2.8a**), raising the possibility that its biochemical activity is not conserved. To test this possibility, we incubated progesterone with recombinant deer/oldfield mouse AKR1C18 *in vitro* and observed that AKR1C18 indeed converts progesterone into 20α-OHP (**Figure 2.8b**). In addition to Akr1c18, which uses progesterone as a substrate, oldfield *zona inaudita* cells express all other enzymes needed for progesterone production (Star, Cyp11a1, and 3β-Hsd; **Figure 2.4**), indicating these cells are capable of turning cholesterol into 20α-OHP.

Given the capacity of oldfield *zona inaudita* cells to produce 20α-OHP, and the lack of this cell type in deer mice, we hypothesized that the concentration of 20α-OHP would be higher in oldfield mouse adrenals and plasma than in deer mice. Indeed, levels of 20α-OHP present in both oldfield mouse adrenal gland tissue as well as levels circulating in oldfield plasma was higher than in deer mouse (**Figure 2.9**). Oldfield females had >10 times higher levels than deer mouse females, and levels in deer mouse males were below the limit of detection. Female oldfield and deer mice have higher levels of circulating 20α-OHP than males of their own species likely as a result of Akr1c18 expression in the ovaries (**Figure A1.8**). Thus, the exclusive

presence of *zona inaudita* in oldfield mice is consistent with their elevated 20α-OHP levels compared to deer mice.

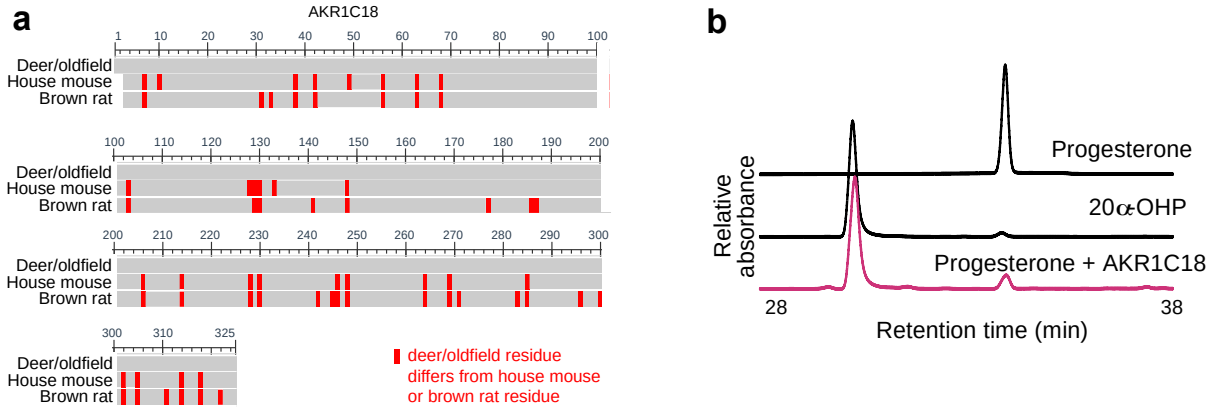

Fig. 2.8: Despite molecular evolution of AKR1C18 across muroid rodents, AKR1C18 of deer- and oldfield mice also catalyzes the conversion from progesterone to 20α-OHP. a, Alignment of AKR1C18 amino acid sequences between deer- and oldfield mouse (identical sequence between species), house mouse, and brown rat. Red bars indicate positions where house mouse or brown rat residues differ from the deer/oldfield residue. **b**, High performance liquid chromatography (HPLC) trace of progesterone standard, 20α-OHP standard, and the reduction of progesterone to 20α-OHP by oldfield mouse AKR1C18 (in collaboration with Sarah Wacker).

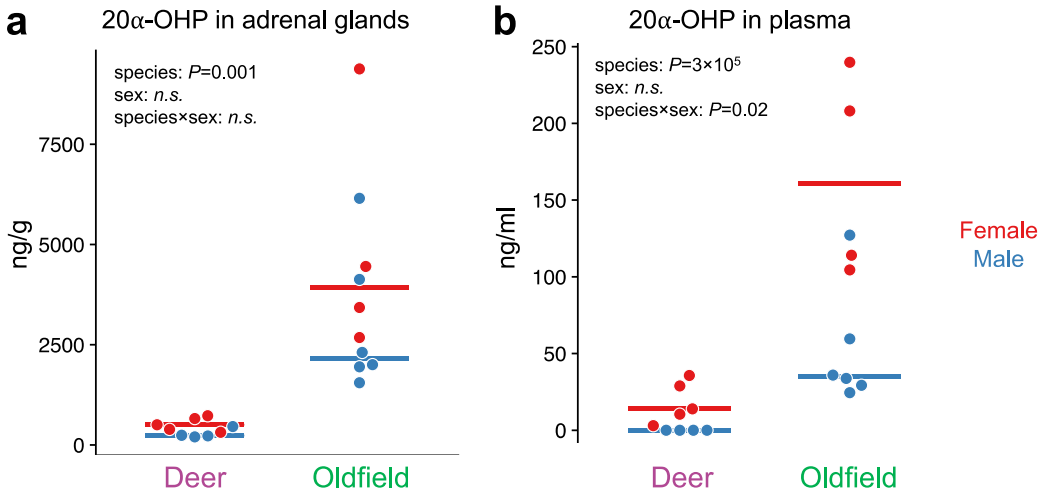

Figure 2.9: Oldfield mice have high levels of 20α-OHP in adrenals and plasma. Concentration of 20α-OHP in adrenals plasma by liquid chromatography tandem mass

spectrometry. *P* values by generalized linear model (deer: N=9, oldfield: N=10). Measured in collaboration with Asmita Poudel and Kiran Soma.

20α-OHP increases parental behavior

Though evolutionarily closely related, deer- and oldfield mice have vastly different parental care strategies. Specifically, deer mouse fathers exhibit very low levels of paternal care towards pups, while oldfield mouse fathers exhibit high levels of paternal care[24]. Given the prominent role of steroid hormones in regulating parental behaviors[189] and the higher levels of 20α-OHP and paternal behavior in oldfield mice, we hypothesized that 20α-OHP promotes parental care. To test this hypothesis, we delivered a single intraperitoneal injection of 20α-OHP into oldfield fathers, and then measured parenting behavior 20 hours later. Remarkably, this single injection of 20α-OHP had a large effect on many aspects of paternal care in oldfield mice (**Figure 2.10**), including increasing the number of pups retrieved as well as the quality of their nests. 20α-OHP-injected fathers also spent 4 times longer grooming their pups compared to vehicle-injected controls (**Figure 2.10**). This dramatic behavioral effect suggests that the evolution of the 20α-OHP-producing *zona inaudita* has likely contributed to the evolution of monogamous-typical parental behaviors in oldfield mice.

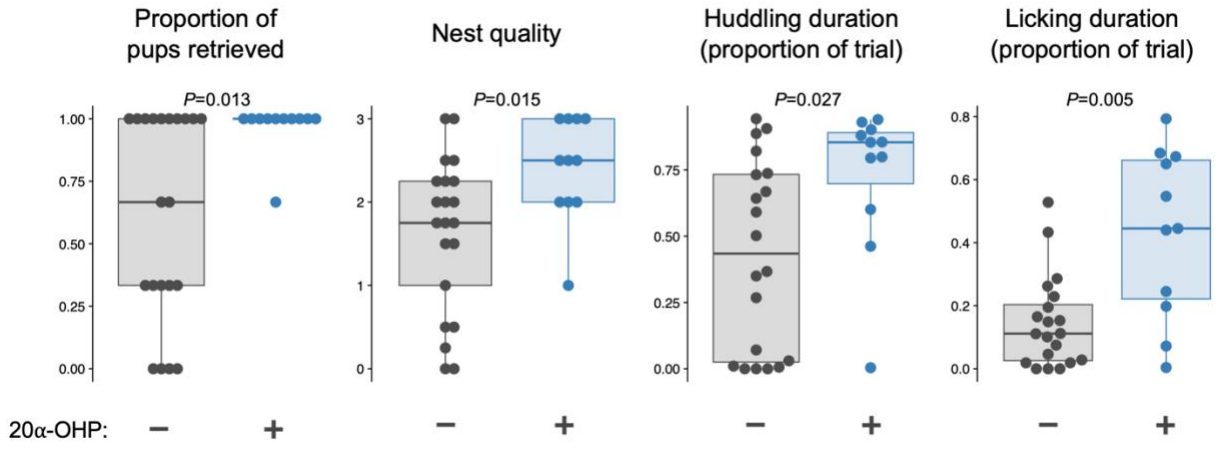

Figure 2.10: 20α-OHP injection alters parental behavior in oldfield fathers. Four metrics of parental behavior (the proportion of pups retrieved, nest quality, time spent huddling pup, and time spent licking pup) scored 20 hours after intraperitoneal injection of saline vehicle or 5 mg/kg 20α-OHP in experienced oldfield fathers. *P* values by Mann-Whitney U test (vehicle: N=21, 20α-OHP: N=11) Behavior experiments conducted in collaboration with Jennifer Merritt and Victoria Esquibies.

20α-OHP is metabolized into allo-diol in the deer- and oldfield brain

20α-OHP has weak affinity for the nuclear progesterone receptor[190], suggesting that its behavioral activity could be mediated by one of its derivatives. Indeed, 20α-OHP can be further metabolized into 5α-pregnane-3α,20α-diol (allopregnanediol or allo-diol) through the actions of 5α-reductase (5α-R) and 3α-hydroxysteroid dehydrogenase (3α-HSD)[191,192] (**Figure 2.11a**). These enzymes are broadly expressed in brains of rodents and humans[193–196], where they also metabolize progesterone into 5α-pregnane-3α-ol-20-one (allopregnanolone) (**Figure 2.11a**), a neurosteroid with rapid non-genomic effects on neuronal excitability[197]. Thus, in brain regions where neurosteroid synthesis is high, elevated 20α-OHP could lead to an accumulation of allo-diol instead of other neurosteroids like allopregnanolone. Indeed, we found that 20α-OHP was converted to allo-diol *in vitro* by the hypothalamus and cerebellum of both deer- and oldfield mice, whereas there was no change in levels of allopregnanolone (**Figure 2.11b,c**). In oldfield mice but not in deer mice, 20α-OHP incubation also led to a small increase in progesterone, but the absolute levels of allo-diol produced from 20α-OHP was over 20 times higher than progesterone, suggesting that while small amounts of 20α-OHP could be back-converted to progesterone, a much higher quantity of 20α-OHP is metabolized into allo-diol. Thus, high levels of circulating 20α-OHP in oldfield mice may lead to increased paternal care via its conversion to allo-diol, which then acts on neuronal circuits important for parenting.

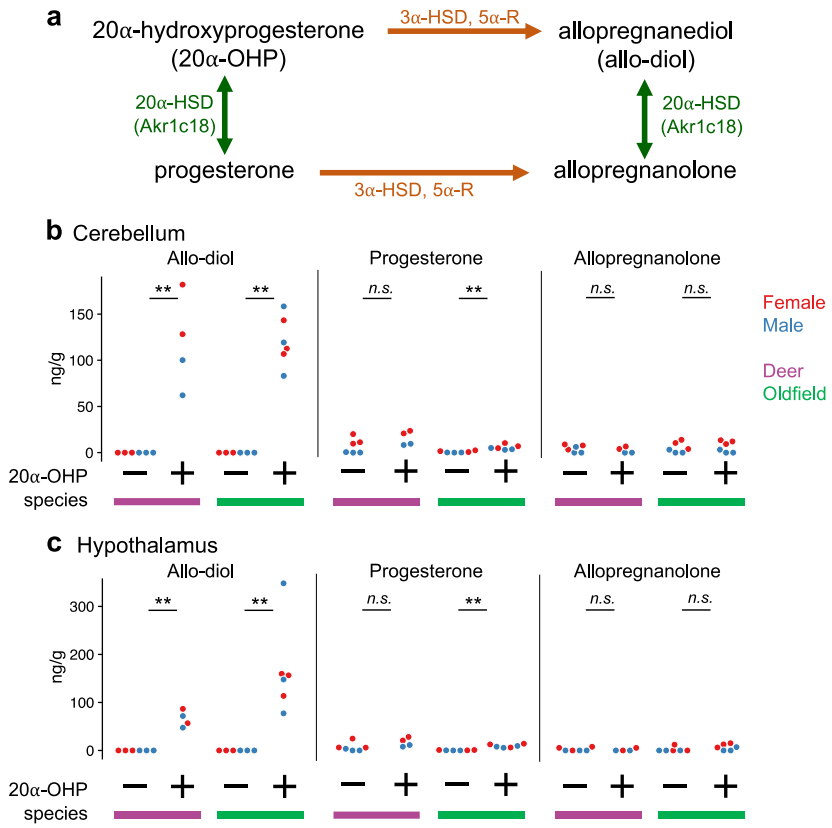

Fig. 2.11: Allo-diol is a primary metabolite of 20α-OHP in cerebellum and hypothalamus. a, Synthesis pathway between 20α-OHP, progesterone, allopregnanolone, and allo-diol. **b** and **c**, Concentration of allo-diol, progesterone, and allopregnanolone extracted from minced deer- or oldfield cerebellum (**b**) or hypothalamus (**c**) after a 2-hour incubation in medium containing 1 uM 20α-OHP, measured by liquid chromatography tandem mass spectrometry in collaboration with Asmita Poudel and Kiran Soma. ** $P<0.01$, *n.s.* not significant; by Wilcoxon test.

Allo-diol inhibits tonic GABAergic currents

Allo-diol is a poorly characterized steroid that is structurally similar (α-hydroxylated carbon 3 and α-reduced carbon 5 of the steroid nucleus) to neurosteroids that alter neuronal excitability through allosteric modulation of $GABA_A$ receptors containing the δ subunit ($\delta GABA_AR$)[198–200]. Furthermore, $\delta GABA_AR$ regulate parental behavior[198,201], raising the possibility that allo-diol modulates parental care at least partially through $\delta GABA_AR$. Previous

research has shown that allo-diol has neurosteroid activity through its positive allosteric modulation of $GABA_AR$[202,203]; however, these measurements were made in synaptoneurosomes[203], which largely exclude the δ subunit of $GABA_AR$[204], and in oocytes overexpressing $GABA_AR$ that do not contain the δ subunit[202]. Therefore, we revisited this question by performing whole-cell patch clamp experiments in cerebellar granule cells, neurons that express high levels of $\delta GABA_AR$ and display large tonic currents with known neurosteroid sensitivity[198,205]. We found that allo-diol reduces GABAergic tonic currents (mediated by $\delta GABA_AR$) and associated channel noise, a relative measure of channel opening (**Figure 2.12**). By contrast, allo-diol did not affect phasic currents mediated by $GABA_AR$ lacking the δ subunit in molecular layer interneurons, suggesting that the effect of allo-diol was specific to tonic GABAergic currents (**Figure A1.10**). In contrast to allo-diol, 20α-OHP had no effect on either phasic or tonic current amplitude and noise, consistent with its known activity as a weak progestin rather than a $GABA_AR$ modulator[190] (**Figure 2.12, A1.10**). Unlike other neurosteroids derived from progesterone, including allopregnanolone and 5α-pregnane-3α,21-diol-20-one (THDOC), which are positive allosteric modulators of $GABA_AR$[197,198], we discovered that allo-diol can act as a negative modulator of tonic GABAergic currents. Altogether, our behavioral and electrophysiological results are consistent with a model in which adrenal 20α-OHP increases parental care at least in part through its conversion into allo-diol, which acts on $\delta GABA_AR$ to alter neuronal excitability.

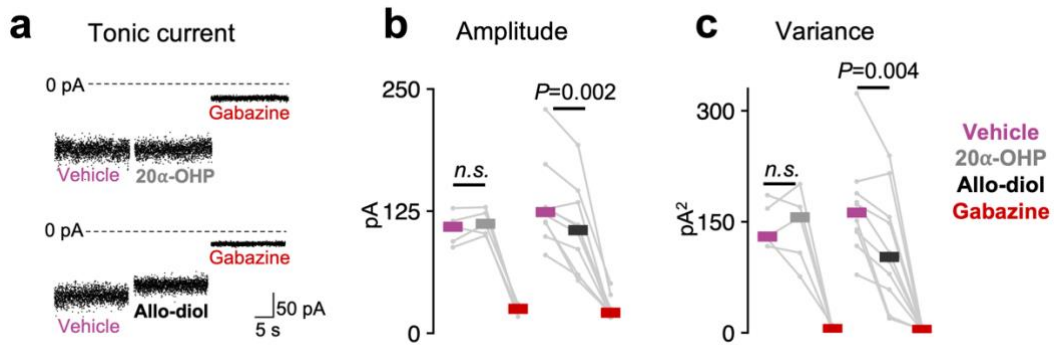

Figure 2.12: Allo-diol but not 20α-OHP is a negative allosteric modulator of the GABA$_A$ receptor. Representative tonic GABAergic current trace recorded from house mouse cerebellar granule cells at baseline (vehicle), after addition of 1 uM 20α-OHP or 1 uM allo-diol, and after addition of 50 uM Gabazine, a GABA$_A$ receptor antagonist. *P* values by Holm-Šídák's multiple comparisons test. Experiments conducted by Stephano Lutzu and Stephanie Rudolph.

Interspecific divergence of tenascin N underlies the evolution of *zona inaudita* cells

Having identified the *zona inaudita* cell of the oldfield mouse as a newly-evolved cell type, and its production of 20α-OHP as a mediator of the higher parental behavior of oldfield mouse fathers, I next wanted to understand how this cell type evolved. To that end, I took a quantitative genetics approach to uncover the evolutionary genetic bases of the emergence of the *zona inaudita* cell. I took advantage of the ability of deer- and oldfield mice to interbreed and used a cohort of second-generation (F_2) intercross hybrids (**Figure 2.13a**). Each F_2 individual contains a unique mixture of the genomes of the two species, allowing us to genetically map the interspecific variants that regulate the expression of *zona inaudita* marker genes.

I began by measuring the expression of each adrenal gene by performing 3'-biased bulk RNA-seq (TM3'seq)[206] of the adrenals of 705 F2 hybrids. Next, I calculated the correlation between the expression of each gene and that of the *zona inaudita* marker Akr1c18, reasoning that genes whose expression is correlated are regulated by an overlapping set of interspecific genetic variants. Notably, markers of the *zona inaudita* were significantly ($P=10^{-16}$, Kruskal-

Wallis rank sum test) more correlated with Akr1c18 (median r=0.24) than genes that are not enriched in that cell type (median r=0.06, **Figure 2.13b**, **Figure A1.11**). Furthermore, several marker genes, including Tnn, cyclin-dependent kinase inhibitor 2A (Cdkn2a), Podnl1, and Serpine1 were in the top 99.9th percentile of correlations with Akr1c18. Therefore, the expression of some *zona inaudita* genes is likely modulated by genetic variants that also affect the expression of Akr1c18.

Next, to map the genetic basis of the expression of *zona inaudita* marker genes, I performed quantitative trait locus mapping of expression (eQTL mapping) of each *zona inaudita* marker. Using genome-wide genotypes of 369 male and 325 female F_2 hybrids, I identified the loci at which the genotype (homozygous deer mouse, homozygous oldfield mouse, or heterozygous) affects expression of markers of the *zona inaudita* that were highly correlated with Akr1c18 expression in F_2 hybrids. I focused the quantitative genetic analyses to males to avoid confounding from estrous cycle and reproductive state variation effects on the adrenal in females[207–210]; however, many of the male findings are similar in females (**Figure A1.12**). Although we identified many eQTL across the genome, some seem to be acting as *trans* eQTL hotspots, as they modulate the expression of multiple genes. Therefore, the regulation of *zona inaudita* genes has a complex genetic architecture, but some genetic loci appear to have a particularly prominent role by regulating multiple genes.

One such prominent *trans* eQTL hotspot, located on chromosome 11, is particularly noteworthy because genetic variation at this locus drives the expression of Akr1c18 itself (**Figure 2.13c**). Akr1c18 lacks a *cis* eQTL, suggesting that local genetic variants do not contribute substantially to expression of Akr1c18 in the adrenal gland; instead, Akr1c18 expression is modulated by genetic variation at this *trans* eQTL hotspot as well as by additional

eQTLs on chromosomes 15 and 21. The oldfield allele at the chromosome 11 hotspot not only drives expression of Akr1c18 in F_2 males but also increases expression in *trans* of five other *zona inaudita* genes: cyclin-dependent kinase inhibitor 1A (Cdkn1a), Podnl1, Cdkn2a, Serpine1, and TIMP metallopeptidase inhibitor 1 (Timp1), and Tensin-4 (Tns4, **Figure 2.13c**). The peak of this eQTL lies 5.3 Mb from the ECM glycoprotein gene tenascin N (Tnn), also known as tenascin W^{211} (**Figure 2.13c**). Tnn is a highly significant marker of the oldfield *zona inaudita* (**Table A1.1**), is essentially undetectable in deer mouse adrenals (**Figure 2.6**), and is one of the most highly correlated genes with Akr1c18 in F_2 adrenals (**Figure 2.13b**, **Figure A1.11**). These results suggest that interspecific variation in or near the Tnn gene contributes significantly to the expression of multiple genes that define the *zona inaudita* cell

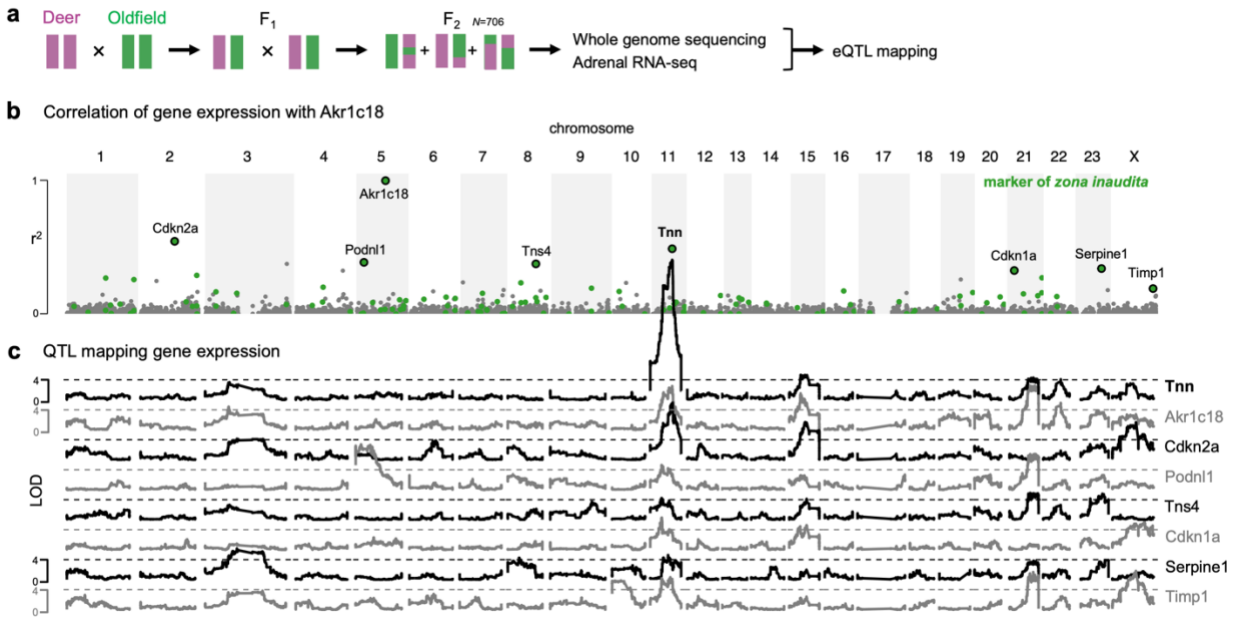

Fig. 2.13: Mapping the genetic basis of the oldfield *zona inaudita* cell. a, Schematic of experimental design for expression quantitative trait locus (eQTL) mapping of adrenal *zona inaudita* genes. **b**, Correlation of expression of each gene with Akr1c18 expression by gene position in F_2 hybrids. **c**, Logarithm of the odds (LOD) of adrenal expression of labeled genes by genotype across the genome in male F_2 hybrids. Dashed lines denote genome- wide threshold of significance (α=0.05).

The eQTL at Tnn that modulates *zona inaudita* gene expression could be caused by coding or by noncoding variation at the Tnn locus. There were no differences in the Tnn amino acid sequence between deer- and oldfield mice, suggesting that interspecific variation at the eQTL might instead affect the expression of Tnn in *cis*. We first tested whether *cis* variation regulated Tnn levels by measuring the expression of each of the two alleles in the adrenals of deer- and oldfield F_1 hybrids, in which both alleles are present in a common "*trans*" environment, allowing for the measurement of the *cis* effects. Strikingly, in these F_1 hybrids, nearly all Tnn expression originated from the oldfield allele (**Figure A1.13**). Consistent with a strong *cis* effect, when we performed QTL mapping of the expression of Tnn, a highly significant peak on chromosome 11 encompassing Tnn explained 30% of the variance in expression among the F_2 hybrid males used for QTL mapping (**Figure 2.13c**). Finally, we performed a mediation analysis to test whether the QTL at the Tnn locus that regulates *zona inaudita* gene expression does so through the levels of Tnn expression. Indeed, when we control for Tnn expression in the eQTL model of Akr1c18 expression, the QTL at Tnn disappears (**Figure A1.14**). Altogether, the results indicate that oldfield genetic variants at the Tnn locus promote Tnn expression and that elevated levels of Tnn then drive the expression of other *zona inaudita* genes.

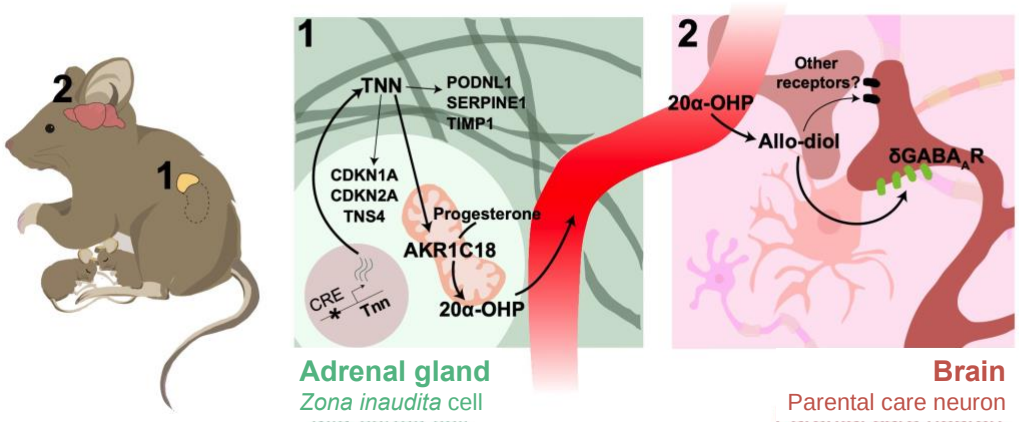

Fig. 2.14: Model of the genetic causes and phenotypic consequences of the newly evolved *zona inaudita* cell type of the oldfield mouse adrenal gland. In the *zona inaudita*, a *cis*-regulatory element drives the expression of Tnn, which in turn upregulates other markers of this cell type, including Akr1c18. In the brain, 20α-OHP is converted to allo-diol, which inhibits tonic GABAergic signaling and may bind to other receptors, to increase parental care in oldfield mice. Illustration by Claire Everett.

Discussion

By combining histology with single-nucleus RNA sequencing, I discovered an adrenal cell type not previously described and without apparent homology to cells in other rodents—the *zona inaudita* cell. Using biochemistry, pharmacology, and electrophysiology, we then found that the *zona inaudita* cell contributed to the evolution of parental care in monogamous oldfield mice by producing a derivative of progesterone. Finally, using transcriptomics and quantitative genetics I discovered the genetic bases of the evolution of this cell type (**Figure 2.14**).

Experimental genetic manipulation in multiple species has shown that transcription factors are essential for the development of particular cell types[212–214]. Given this essential role in development, genetic variation affecting transcription factors has long been hypothesized to drive the evolution of novel cell types[164], but empirical evidence to support this hypothesis is scant[215,216]. Here, I used unbiased genome-wide genetic mapping, rather than a targeted analysis of candidate genes, to discover what causes the presence of the *zona inaudita* cell in one species but not in another. I found that genetic variation affecting tenascin N—an extracellular matrix protein—is a likely cause of the gain of novel enzymatic function of this cell type. Transcription factors are likely involved in the development of the *zona inaudita* cell, as several transcription factors including Runx2, Batf, and Ezr are also markers of this cell type. However, my genetic mapping experiments indicate that interspecific genetic variation at or near these transcription

factor genes is not a major force driving the presence of Akr1c18, an enzyme essential for *zona inaudita* function.

Tnn is one of four members of the tenascin gene family of glycoproteins, which are abundant in the ECM during mammalian embryonic development and are often upregulated in tumors and in adult stem cells[217]. Tenascins have been shown to regulate cell migration, survival, proliferation and differentiation[218–220] due to their interactions with many ECM proteins including cell surface receptors, and due to their capacity for autocrine signaling[218,221]. Tnn is expressed in the *zona inaudita* cell of oldfield mice, but very sparsely in other cells of the oldfield adrenal cortex and is essentially absent from the adrenals of the deer mouse, the house mouse, and humans. I found that this novel *zona inaudita*-specific expression was caused by local non-coding genetic variation. Thus, a novel *cis*-regulatory element in oldfield mice led to an evolutionarily unique expression pattern for this gene, which in turn contributes to the molecular functionality of a novel cell type.

Variation in steroid hormone receptors is known to contribute to differences in social behaviors across animals[114]. By contrast, I discovered that high-parenting oldfield mice have higher levels of 20α-OHP than low-parenting deer mice as a consequence of a *zona inaudita* exclusive to oldfield mice that converts progesterone into 20α-OHP. 20α-OHP is a poorly studied steroid hormone that is present throughout mammals, including humans[222,223]. I found that 20α-OHP administration increases paternal behavior in oldfield mice, likely through the actions of its metabolite allo-diol, another poorly studied steroid hormone also present in humans[224]. Allo-diol, but not 20α-OHP, negatively modulates tonic GABAergic currents mediated by extrasynaptic δGABA$_A$Rs, yet it is possible that their behavioral effects *in vivo* are mediated by additional receptors.

Taken together, our results establish how *cis*-regulatory variation between two sister species drives the adrenal expression of Tnn exclusively in oldfield mice. In turn, Tnn facilitates the expression of multiple genes that define the *zona inaudita*, including Akr1c18. The *zona inaudita* forms a prominent layer of the oldfield adrenal cortex and converts progesterone into 20α-OHP. I found that 20α-OHP promotes high parental care—a distinguishing feature of the monogamous oldfield mouse—likely through the actions of its metabolite allo-diol. By examining both the genetic causes and phenotypic consequences of this adrenal cell population, my work provides insight into the processes by which novel cell types can arise and their role in the evolution of animal behavior, even when these cells evolve outside the brain.

Methods

Animal husbandry. Deer mice *Peromyscus maniculatus bairdii* (strain BW) and oldfield mice *Peromyscus polionotus subgriseus* (strain PO) were originally obtained from the *Peromyscus* Stock Center at the University of South Carolina and colonies were established at Columbia University. Mice were housed in 19.4 cm x 18.1 cm x 39.8 cm (500 cm^2 floor space) ventilated cages (NexGen Mouse 500, Allentown) under barrier conditions with a 16:8 light:dark cycle at 22 ºC. We provided each cage with Enviro-dri (Shepherd Specialty Papers) and cotton nestlets as nest building material. Mice had access to food (PicoLab Rodent Diet 5053 for nonbreeders, 5058 for breeders) and water *ad libitum*. All procedures were carried out in accordance with the guidelines established by the NIH Guide for the Care and Use of Laboratory Animals and the Animal Experimentation Guidelines from the Columbia University. All animal protocols were approved by Columbia University's Institutional Animal Care and Use Committee.

Adrenal weight. Adrenals of deer mice and oldfield mice were dissected and the periadrenal fat was removed using forceps and fine surgical scissors. To measure adrenal weight, one gland was drop fixed in 4% paraformaldehyde (PFA) for 30 minutes and then washed three times in PBS. Fixed glands were blotted dry before being weighed on a Mettler Toledo ME103TE Precision Balance.

Adrenal sectioning and marker gene staining. For histology, adrenal glands were dissected, flash frozen in dry ice, and then embedded in cryomolds in optimal cutting temperature (OCT) compound. Glands were sectioned on the horizontal plane using a Leica CM3050S cryostat and mounted on Superfrost microscope slides. Adrenals were sectioned at 10 µm for visualizing cell and nucleus size, and 25-30 µm for visualizing adrenal zonation and the spatial distribution of cell-type markers.

The adrenal medulla was stained using a rabbit anti-tyrosine hydroxylase primary antibody (EMD Millipore AB152, 1:1000), and the *zona glomerulosa* was stained using mouse anti-Rbfox1 (EMD Millipore MABE159, 1:1000). All other genes were stained by *in situ* hybridization using hybridization chain reaction (HCR) v3.0[225]. Split-initiator probes were designed against the deer mouse sequences of aldo keto reductase family 1 member C18 (Akr1c18, annotated as LOC102910062 in the HU_Pman2.1.3 genome), cytochrome P450 11B1 (Cyp11b1, annotated as LOC102923554), tenascin N (Tnn), and cholesterol side-chain cleavage enzyme (Cyp11b1, annotated as LOC102905324) using the HCR 3.0 Probe Maker tool[226] (up to 33 probe pairs per gene) and ordered as an oligo pool (oPool) from Integrated DNA Technologies. HCR reagents and hairpin amplifiers were purchased from Molecular Instruments (Los Angeles, CA). HCR was performed using the protocol published in ref. [225], and slides were

counterstained with DAPI and imaged on a Nikon AZ100 Multizoom microscope or a Nikon Eclipse upright microscope.

Adrenal volume and hypertrophy. To measure adrenal cortex and medulla volume, adrenal glands were fixed in 4% PFA and then cleared and whole-mount immunostained using the iDISCO+ protocol[227] from http://www.idisco.info. Tyrosine hydroxylase was labeled with rabbit anti-Tyrosine Hydroxylase (Millipore AB152, 1:1000) and secondary donkey anti-rabbit Alexa Fluor 647 (Invitrogen A-31573) antibodies to mark the adrenal medulla. Cleared glands were imaged using a LaVision BioTec UltraMicroscope II light sheet microscope. For each gland, the cortex and medulla were delineated in 20 equidistant z-stack images based on tyrosine hydroxylase expression, and the cortex and medulla areas were calculated for each image. Volumes of the cortex and medulla, respectively, were estimated by multiplying each area × the number of z-stack images per gland × the 3-micron z-stack step size.

The number of cells per gland was estimated by counting the number of DAPI-positive cortical nuclei in a 100×100 pixel region of a 10-μm thick cryosectioned adrenal gland. For each individual, cortical nuclei were manually counted in ImageJ in six 100×100 pixel regions of a single adrenal section (three regions from the *zona fasciculata*, three from the *zona reticularis*) and the median nuclei count was calculated. Total cells per gland was estimated as median number of nuclei per 100×100×100 pixel volume multiplied by the median species volume of the cortex as calculated above (deer: 1.87 mm^3, oldfield: 12.28 mm^3).

Bulk RNA-seq of deer- and oldfield mice adrenal glands and ovaries. Adrenal glands were collected from male and female deer mice (N=8, half of each sex), oldfield mice (N=8, half of each sex), and F_1 hybrids (N=8, half of each sex) immediately following decapitation, flash frozen in dry ice, and stored at -70 ºC until use. Frozen adrenals were transferred to Trizol and

lysed using a motorized homogenizer (PRO250, Pro Scientific). Total RNA was extracted from the lysate using the Direct-zol RNA mini-prep kit (Zymo Research), and mRNA was isolated using the NEBNext Poly(A) mRNA Magnetic Isolation Kit (New England Biolabs). Adrenal RNA-seq libraries were prepared using the NEBNext® Ultra II Directional RNA Library Prep Kit for Illumina (New England Biolabs). Ovaries (N=4 per species) were homogenized in lysis buffer and mRNA was extracted using the Dynabeads™ mRNA DIRECT™ Purification Kit. Following mRNA extraction, ovary libraries were created using Tagmentation-Mediated 3' Sequencing (TM3'-seq)[206] using homebrew Tn5 transposase enzyme[228]. All libraries were sequenced on an Illumina NextSeq550 with paired-end reads (2x75bp) and adapters were trimmed using Trimmomatic v0.36[229].

To quantify differential expression of all genes between deer- and oldfield mouse adrenals, and of Akr1c18 in ovaries, reads were aligned to the deer mouse genome assembly HU_Pman_2.1.3 using STAR v2.6.0a[230] in two-pass mode. Transcript levels were quantified with RSEM v1.3.3[231] and differentially expressed genes (Benjamini-Hochberg $P_{adjusted}<0.05$) were identified using DESeq2 v1.36.0[232] after filtering lowly expressed genes (transcripts less than 2 counts per million in ≥80% of individuals of either species). To quantify allele-specific expression of Tnn, F_1 hybrid adrenals reads were pseudoaligned to a hybrid diploid genome of deer- and oldfield mouse using kallisto v0.46.0[233] and allelic expression was quantified using mmseq v1.0.11[234].

Single-nucleus RNA-seq. Adrenal glands were collected following decapitation and flash frozen in dry ice for the following 6 treatments: 3 species (deer, oldfield, or C57BL/6J *Mus musculus* lab mice, ~60 days old) × 2 sexes, with adrenals from 3 mice pooled per treatment. Nuclei were extracted using the Chromium Nuclei Isolation Kit with RNase Inhibitor (10X

Genomics) and nuclei concentration was determined with trypan blue staining on a hemocytometer. Nuclei suspensions were loaded onto a 10X Chromium Chip G and GEMs were generated using the 10X Chromium Controller (10X Genomics). Libraries were creating using the Chromium Next GEM Single Cell 3' Kit v3.1 and then sequenced on an Illumina NextSeq 2000 with an insert size of 90 bp.

Peromyscus reads were aligned to the *P. maniculatus* genome HU_Pman_2.1.3 and C57BL/6 reads were aligned to the mm39 genome assembly. Feature-barcode matrices generated using 10X Genomics CellRanger v7.1. Ambient RNA was removed *in silico* by CellBender v0.3.0, and gene expression matrices were then analyzed in Seurat v4.3.0. First, low-quality cells and likely multiplets were discarded, and then integration of the datasets from deer mice and oldfield mice was performed. Briefly, anchors for dataset integration were identified using a reciprocal principal components analysis (rPCA), in which the deer mouse dataset is projected onto the reduced PCA space of the oldfield mouse data and vice versa, with 1800 anchor features and 20 PCA dimensions. Clusters were identified in Seurat using the FindClusters() function with a cluster resolution parameter of 0.3. Upregulated differentially expressed genes (i.e. markers) of the *zona inaudita* cell type were determined using the FindMarkers() Seurat function comparing all cells of the *zona inaudita* cluster to all other cell types in deer- and oldfield mice. *Zona inaudita* markers had an adjusted *P* value <0.05 of differential expression against all other cell clusters and surpassed a $\log_2$ fold change threshold of 1. Gene ontology (GO) term enrichment analysis identified that this set of markers is enriched for genes encoding proteins localized to the extracellular matrix ($P_{\text{adjusted}}=10^{-12}$, GO:0031012).

To compare cell type homology between *Peromyscus* and *Mus musculus*, deer- and oldfield datasets were integrated with C57BL/6 data after filtering the included features to the set

of 16,410 one-to-one orthologs shared between the deer mouse and C57BL/6 genome annotations. Orthologous genes between *Peromyscus* and *Mus* were inferred using TOGA[235]. Integration and clustering was performed as described above.

Purification of deer- and oldfield mouse AKR1C18. The AKR1C18 coding sequence of deer- and oldfield mice (which are identical) was synthesized, codon optimized for *Escherichia coli*, and cloned into the pET16b plasmid with an N-terminal His-tag followed by a Factor Xa Protease cleavage site by Genscript (Piscataway, NJ). The protein was overexpressed in *E. coli* BL21(DE3) at 30 °C for 13 hours in the presence of 100 µg/mL ampicillin and 1 mM isopropylthio-β-galactoside. The harvested cells were resuspended in lysis buffer (50 mM potassium phosphate (pH 7.8), 100 mM NaCl, 10% (v/v) glycerol, 10 mM imidazole, and 2 mM 2-mercaptoethanol) and stored at -80 °C. Recombinant protein was purified using Ni-affinity chromatography. Briefly, the resuspended cells were supplemented with 0.1 mg/mL DNase and protease inhibitors (Pierce) and kept on ice for twenty minutes. Cells were lysed by sonication and the lysate was clarified by centrifugation at 47,800 rcf in a Sorvall RC5C centrifuge for one hour. The clarified lysate was combined with 1 mL of Ni-NTA agarose (Invitrogen) and allowed to bind for 2 hours. The resin-lysate mixture was poured into a poly-prep column (Bio-Rad), which was washed with 150 mL lysis buffer. AKR1C18 protein was eluted using buffer containing 37.5 mM potassium phosphate pH 7.8, 75 mM NaCl, 7.5% glycerol, 1.5 mM 2-mercaptoethanol, and 250 mM imidazole. The AKR1C18 protein was dialyzed against buffer containing 50 mM potassium phosphate pH 7.8, 100 mM NaCl, 10% glycerol, 2.5 mM EDTA, and 2 mM 2-mercaptoethanol. Protein was analyzed using SDS-PAGE and was stored at 4°C, before using it within two weeks.

Progesterone reduction by AKR1C18. The reduction of progesterone was performed as described[236]. In summary, the assay was run in a 5 mL volume containing 30 μM progesterone (Thermo Scientific 225651000), 180 μM NADH, 4% acetonitrile, 0.1 M potassium phosphate (pH 6.0) along with AKR1C18 enzyme. Assays were run for 1 h at 37 °C. All samples were extracted twice with 2 mL ethyl acetate and dried by vacuum centrifugation before reconstitution in 400 μL of 60% acetonitrile in water. Product formation was detected via reversed-phase high performance liquid chromatography linked to UV detection (RP-UV/HPLC) using a Shimadzu LC-2030 Plus liquid chromatograph. Extracted samples were run on a Thermo Scientific ODS Hypersil 4.6 × 250 mm column with 5 μm particles using a constant flow rate of 0.75 mL/min with the solvents water (A) and acetonitrile (B). The method started at 15% B, which it held for 1 min and then increased to 75% B over 34 min. Then, it increased to 95% B over 5 min, decreased to 5% B over 1 min, held for 14 min. Elution of progesterone and 20α-hydroxyprogesterone (20α-OHP) was monitored at 242 nm. Peaks were integrated and experimental assays were compared to standards of progesterone (Sigma-Aldrich, P0130) and 20α-OHP (Steraloids, Q3600).

Adrenal and plasma concentrations of 20α-OHP. Adrenal glands and 1mL intracardial blood was obtained from virgin deer- and oldfield mice immediately following death by carbon dioxide inhalation between 1–3 pm (Zeitgeber time 10–12). Adrenals were flash frozen in dry ice, and blood was immediately transferred to a tube containing 10 μL 250 mM EDTA, and then centrifuged at 1,500 rcf for 15 minutes at 4 ^{0}C and the plasma supernatant collected. Adrenal glands and plasma were stored at -70 ºC until sample processing by the Soma lab. Steroids were extracted from 5mg of adrenal tissue and 5mL of plasma as described in [237]. Absolute quantification of 20α-OHP, 5α-pregnan-3α, 20α-diol ("allo-diol"), progesterone, and

allopregnanolone was performed on a Sciex QTRAP 6500 UHPLC-MS/MS system as previously described[237].

Effect of steroid manipulation on parental care. To quantify parental behavior in deer- and oldfield mice after pharmacological manipulation, 5 mg/kg 20α-OHP (Steraloids, Q3600) or saline vehicle was delivered to experienced oldfield fatheres (80-120 days old) 16–20 hours before behavior testing. This timing allowed for detection of both acute and genomic effects of 20α-OHP delivery and allowed for ample recovery from the stress of handling and injection.

Animals were transferred to the cage rack adjacent to the testing room 20–24 hours before behavioral testing. Each injected test animal was placed individually in a new, clean cage with 0.625 g of compacted nestlet and left undisturbed to habituate for 1 hour prior to behavior testing. Following habituation, the trial begins. The experimenter places one unfamiliar, conspecific pup inside the cage away from the nesting material. After 15 minutes and again two minutes after that, we added another pup from the same litter to the cage, at least 5 cm from the pup(s) already in the cage. The trial ended after 20 minutes, after all three pups had been added to the cage. Habituation and behavior testing occurred in the dark under 870 nm infrared LED lighting between Zeitgeber time 8–17.

Pups were used for behavior testing between postnatal days 2–6, an age in which they cannot return to the nest without assistance. Between behavior trials, pups were kept in an incubator at 37 °C; at the end of testing, each litter was returned to their parental cage and observed to ensure that they were licked and/or retrieved by a parent.

Behavior trials were video recorded using 2 Raspberry Pi 3 with NOIR Camera Boards to record the animals in their cage from two different angles (top and side). Video footage from the top and side views of the cage were analyzed blind to the treatment each animal received. For the

first 15 minutes of trial, the duration of time the test animal spent huddling over the pup and time spent grooming or licking the pup were calculated. The proportion of pups retrieved (out of three pups added to the cage over the 20-minute trial) was also scored. Nest quality at the end of the 20-minute trial was scored on a 4-point scale described in [24].

Recording tonic and phasic GABAergic currents. Acute sagittal slices were prepared from young C57BL/6J adult mice (6–8 weeks) of males and females. Mice were anesthetized with isoflurane and intracardially perfused with ice cold cutting solution containing 110 mM CholineCl, 7 mM $MgCl_2$, 2.5 mM KCl, 1.25 mM NaH_2PO_4, 0.5 mM $CaCl_2$, 25 mM Glucose, 11.5 mM Na-ascorbate, 3 mM Na-pyruvate, 25 mM $NaHCO_3$, equilibrated with 95% O_2 and 5% CO_2. The brain was rapidly dissected, and the cerebellum was cut into 250 µm thick parasagittal slices in the same solution on a vibratome (VT1200S, Leica). Slices were then transferred to artificial cerebrospinal fluid (ACSF) at 34°C containing 125 mM NaCl, 26 mM $NaHCO_3$, 1.25 mM NaH_2PO_4, 2.5 mM KCl, 1 mM $MgCl_2$, 1.5 mM $CaCl_2$, and 25 mM glucose, equilibrated with 95% O_2 and 5% CO_2 and incubated for 30 min. Slices were then stored at room temperature until recording for up to 6 hours.

Tonic currents were recorded from cerebellar granule cells and phasic currents were recorded from parvalbumin-expressing interneurons of the molecular layer at ~32 °C with an internal solution containing 110 mM CsCl, 10 mM CsGluconate, 10 mM HEPES, 10 mM MgATP, 0.5 mM NaGTP, 5 mM phosphocreatine-$tris_2$, and 5 mM phosphocreatine-Na_2, and 0.1 mM Alexa 594 (pH adjusted to 7.2 with CsOH, osmolality adjusted to 310 mOsm/kg). The reversal potential for chloride was adjusted to ~0 mV to maximize GABAergic current amplitude at -65 mV holding potential. Visually-guided whole-cell recordings were obtained with patch pipettes of ~5 MΩ resistance pulled from borosilicate capillary glass (BF150-86-10, Sutter

Instrument, Novato, CA). Electrophysiology data was acquired using a Sutter dPatch amplifier (Sutter Instruments), digitized at 10 kHz and filtered at 5 kHz. For isolating inhibitory currents in voltage clamp the following receptor antagonists were added to the bath solution: 2 µM R-CPP, 5 NBQX, 1 µM strychnine, 1.5 µM CGP to block NMDA, AMPA, glycine and GABAB receptors. 1 µM of GABA was included in the bath for tonic current recordings. GABAergic currents were blocked with 50 µM SR95531 ("gabazine"). Receptor antagonists were purchased from Abcam (Cambridge, MA) and Tocris (Bristol, UK). 20α-OHP and allo-diol were purchased from Steraloids (Q3600, P1950) and used at a concentration of 1 µM during bath application over the time course of 15 mins to assure full equilibration. After each experiment, the rig was washed with ethanol and distilled water for 10 min to assure removal of drugs from the tubing and recording chamber. Recordings were performed blind to the recording condition and unblinded after conclusion of the analysis.

Electrophysiology data were analyzed using AxographX and IgorPro (Wavemetrics). Tonic current amplitude was determined by generating all-points histograms of continuously recorded current under control conditions, in the presence of steroid and in gabazine. The histograms were fit with a Gaussian, yielding the mean current[238]. The current noise as a relative measure of channel opening was determined by calculating the variance of the current. Phasic spontaneous inhibitory postsynaptic currents (IPSCs) were detected with a sliding template function. To determine the average amplitude under control conditions and in the presence of steroids, IPSCs detected in a 3-minute time window were averaged and amplitudes compared. Statistical analysis was performed using repeated measures ANOVA with Geisser-Greenhouse correction and multiple comparisons Holm-Šídák's test (tonic and phasic current).

F_2 adrenal RNA-seq. F_2 hybrids of deer mice and oldfield mice were generated from a founding population of four deer mouse mothers and four oldfield mouse fathers[24]. The adrenals of 705 adult F_2 hybrids were dissected and stored at -70 ºC. Adrenals were lysed using a motorized homogenizer in lysis buffer and mRNA was extracted using the Dynabeads™ mRNA DIRECT™ Purification Kit. Following mRNA extraction, I created Tagmentation-Mediated 3' Sequencing (TM3'-seq)[206] libraries using homebrew Tn5 transposase enzyme[228] which allowed for high-throughput sample processing. Adrenal transcriptomes were sequenced using 1x76 bp reads on an Illumina NextSeq 550 to an average depth of 3.3 million reads (minimum 2 million reads per sample). Reads were adaptor trimmed, aligned, and quantified as described for *Bulk RNA-seq in deer mice and oldfield mice adrenals*. The correlation of $\log_{10}$-transformed F_2 hybrid gene expression to $\log_{10}$-transformed Akr1c18 expression was calculated using Pearson correlation.

eQTL mapping in F_2 hybrids. Determination of genome-wide ancestry for each F_2 animal was described in Bendesky et al. (2017)[24]. Briefly, the genomic positions of fixed SNPs between the species were determined from ddRAD-sequencing of genomic DNA from the eight founders of the F_2 cross. Libraries for ddRAD-seq were then created for each of 694 F_2 hybrid mice (369 male, 325 female) and a hidden Markov model was used to calculate genotype probabilities along each chromosome from the number of reads mapping to the deer mouse versus the oldfield mouse SNP at each fixed variant position.

Genotype probabilities were imported into R/qtl[239] using read.cross.msg.1.5.R (https://github.com/dstern/ read_cross_msg/) and quantitative trait locus (QTL) mapping of *zona inaudita* gene expression (transcripts per million of each marker gene) was conducted for males and females separately using scanone under a nonparametric model. The scanone function was

also used to compute the genome-wide LOD significance threshold at $\alpha<0.05$ using 1000 permutations.

Chapter 3: Fine-mapping the genetic architecture of exploratory behavior

Manuscript in preparation

Niepoth, N. & Bendesky, A. Genetic dissection of exploratory behavior in deer mice.

Introduction

Exploration is an essential animal behavior: animals must explore their surroundings to find food, mates, and resources. However, exploring can be dangerous, particularly in the presence of predators and other threats. Natural selection can act on exploratory behavior in environments where the benefits or risks of exploring are predictable[240]; thus, different exploratory behavioral strategies can evolve in species adapted to different habitats. However, the genetic mechanisms that modulate exploratory behavior within natural populations, and that cause behavior to diverge between species, are largely unknown.

As the most abundant mammal in North America, *Peromyscus* mice have been studied extensively in the field by naturalists and biologists for over 100 years[241]. The genus contains around 60 species of mice that have expanded into diverse habitats across the continent and therefore exhibit a broad range of behaviors and adaptations[241]. Furthermore, many *Peromyscus* sister species can interbreed in the lab, allowing for forward genetic analyses of behaviors that vary dramatically and have evolved in nature.

The prairie deer mouse *Peromyscus maniculatus bairdii* and the oldfield mouse *Peromyscus polionotus subgriseus* inhabit very different habitat types and are thus exposed to different selective pressures that may favor evolution of divergent exploratory behavioral strategies. Prairie deer mice (henceforth “deer mice”) inhabit prairies with dense, sheltering

vegetation throughout North America, while oldfield mice live in open sandy fields in the southeastern United States. These divergent ecologies have likely reinforced divergent exploratory strategies that influence how each species moves in space and explores novel stimuli.

Animals increase or decrease their exploratory behavior based on internal states including hunger, thirst, arousal, and motivation. The internal state resembling what humans define as anxiety has a particularly strong effect on exploratory behavior in rodents[242,243]. A classic behavioral paradigm that measures rodent exploratory behavior as a readout of an anxiety-related internal state is known as the elevated plus maze (EPM). The EPM apparatus contains two walled arms and two open arms that extend perpendicularly several feet off the floor. Increased avoidance of the open arms of the EPM indicates heightened anxiety-related behavior, as drugs that reduce anxiety in humans increase time spent exploring the open arms[244], and open arm avoidance is mediated by circuitry homologous to the circuits underlying human anxiety[242]. Natural genetic variation that increases exploration in oldfield mice compared to deer mice may therefore do so by attenuating an anxiety-like internal state in oldfield mice.

Here, I discover that two sister species of *Peromyscus* mice adapted to exploring very different environments display stark differences in exploratory behavior, and I leverage their genetic similarity to discover novel regulators of this ecologically-important behavior. I accomplish this by fine-mapping two genomic regions that most strongly mediate differences in exploration between the species through six generations of targeted introgression. Ultimately, I identify a 15-Mb locus on chromosome 9 that contributes to the species difference in exploratory behavior, and I investigate the genes within this locus that may regulate behavior.

Deer and oldfield mice differ in their propensity to explore

Deer- and oldfield mice differ in behavioral response to threats[245], suggesting that their propensity to explore potentially threatening novel spaces, conspecifics, and objects may also differ between the species. To quantify species differences in exploratory behavior, I first tested whether deer- and oldfield mice differed in how they explored space. I discovered that oldfield mice spent more time exploring both the open arms of the EPM as well as the open center of a large open field arena compared to deer mice (**Figure 3.1a**,**b**). Additionally, I found species differences in other axes of exploratory behavior: specifically, oldfield mice more readily investigated and interacted with novel conspecific mice in a three-chamber apparatus and novel objects placed in their home cage (**Figure 3.1c**,**d**). Taken together, these results suggest that oldfield mice are less hesitant to explore potentially threatening stimuli than deer mice, potentially reflecting a low-anxiety internal state.

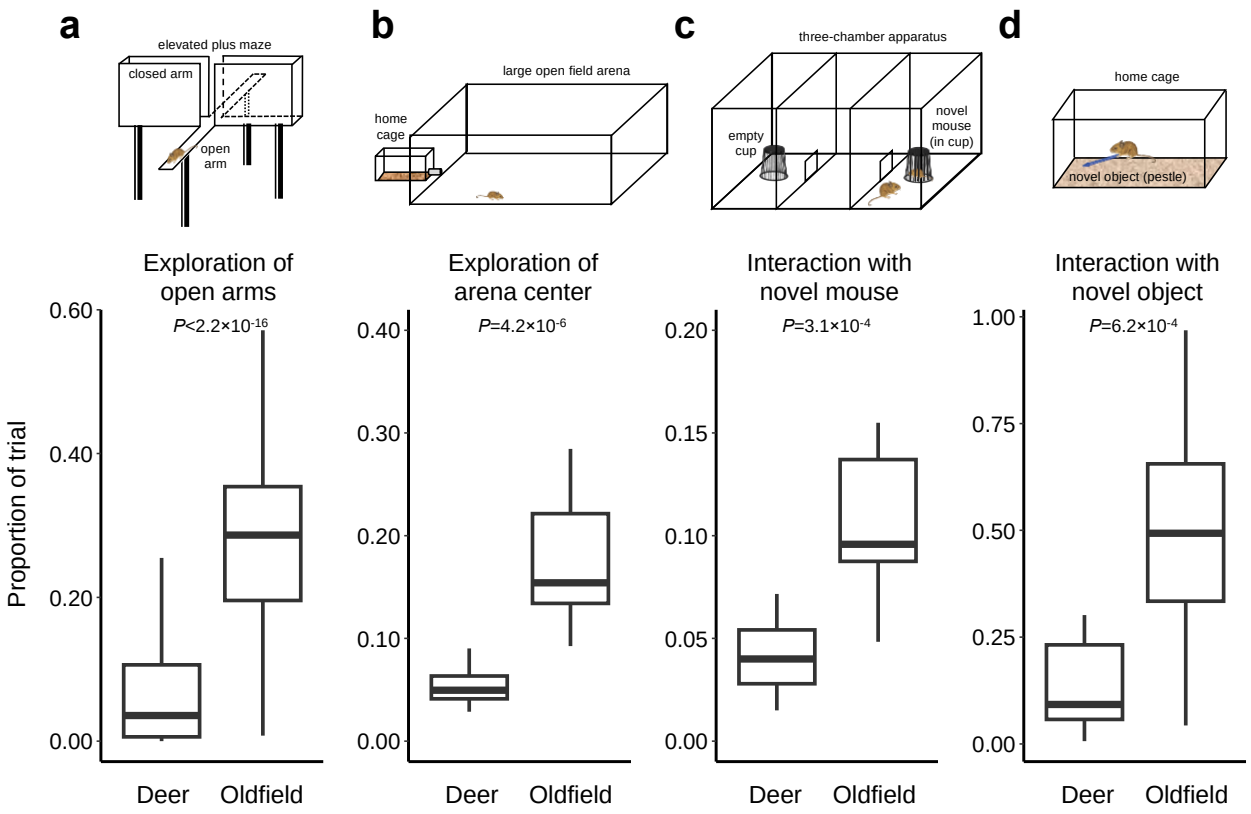

Figure 3.1. Behavior of deer- and oldfield mice differs across four laboratory tests of exploration. a, Proportion of time spent exploring the open arms of an elevated plus maze; Kruskal-Wallis test, $P< 2.2\times10^{-16}$ **b**, Proportion of time spent exploring the center of a 4x4-ft. arena connected to the subject's home cage during an 8-hour trial; *t*-test, $P=4.2\times10^{-6}$ **c**, Proportion of time spent exploring a novel conspecific mouse in a three-chamber apparatus, *t*-test, $P=3.1\times10^{-4}$ **d**, Proportion of time spent exploring a novel object (a plastic pestle) introduced into the subject's home cage, *t*-test, $P=6.2\times10^{-4}$

Mapping the genetic architecture of exploratory behavior

Interspecific variation in exploratory behavior between deer- and oldfield mice may represent divergent adaptive strategies that benefit each species in their natural habitats. Previous unpublished work from the Bendesky lab identified the genomic regions linked to this interspecific variation in exploratory behavior on the elevated plus maze by quantitative trait locus (QTL) mapping in more than 1,500 F_2 hybrids of deer- and oldfield mice.

QTL mapping is a technique used to locate genomic regions associated with variation in quantitative traits like behavior. Here, the exploratory behavior of each F_2 hybrid was tested. Then, because hybrid genomes are a recombinant mix of deer and oldfield ancestry, each hybrid was genotyped at many markers (i.e. single-nucleotide variants fixed in each parental species) spread throughout their genome. Then, the association between genotype at each marker and EPM behavior was quantified: genomic locations linked to exploratory behavior were identified as QTL—loci that contribute to trait variation. The unbiased genome-wide approach of QTL mapping allows for robust detection of genetic markers linked to a behavior.

QTL mapping identified genetic loci on eleven chromosomes that contributed to variation in open arm avoidance on the elevated plus maze; however, the genetic regions with the largest effect included a 25.5-Mb region on chromosome 4 and a 23-Mb region on chromosome 9 (**Figure 3.2**). These two QTL each explained ~2.5% of the total variation in open arm avoidance behavior among the F_2 hybrids, and in each case, the oldfield mouse allele conferred increased propensity to explore the open arms (**Figure 3.3**). Interestingly, the QTL on chromosomes 4 and 9 did not overlap with previously-identified QTL of exploratory behavior or anxiety between inbred mice strains[246–248], including QTL that first identified the contribution of *Rgs2* and *Cofilin-1* variation to anxiety-related behavior on the EPM in lab mice.

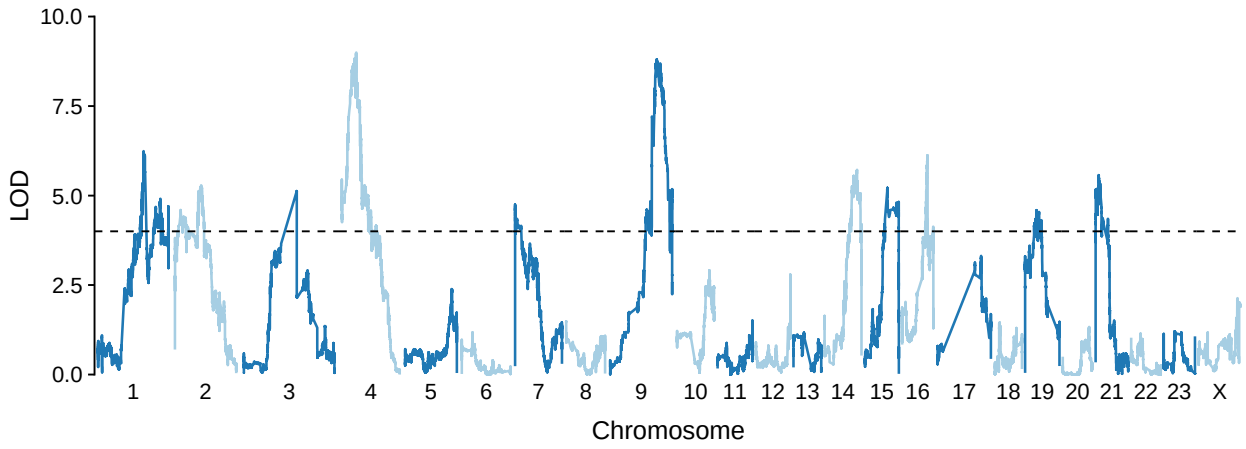

Fig. 3.2: QTL mapping open arm avoidance behavior on the EPM. The linkage (LOD score) across the genome to the proportion of time in the closed arms of the EPM. Dashed line denotes genome-wide threshold of significance. Dashed line denotes genome-wide threshold of significance (α=0.05). Data collected by Andres Bendesky.

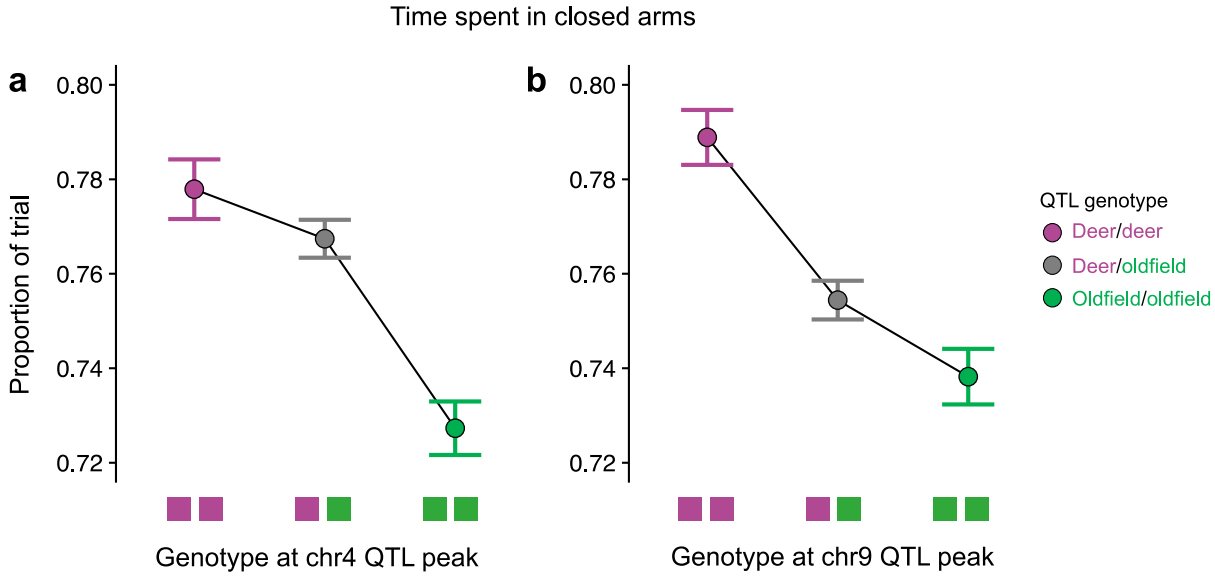

Fig. 3.3: Effect of genotype on EPM behavior at the two most significant QTL. Mean time spent in the closed arms (±s.e.m.) by F_2 genotype at the peak of the QTL on chr4 (**a**) and on chr9 (**b**).

Fine-mapping the two most significant QTL of exploratory behavior

F_2 QTL mapping is a powerful method for identifying loci that explain variation in behavioral traits. However, given the low number of recombination events per chromosome in F_2 hybrids—typically less than two—mapping resolution is fairly low compared to that of advanced intercrosses. For example, a marker associated with a phenotype in F_2 hybrids may be millions of bases from the genetic variation that modulates behavior due to genetic linkage. The top QTLs on chromosomes 4 and 9, for example, contain 152 and 35 protein-coding genes, respectively, including many plausible candidate genes such as transcription factors, ion channels, and cell-adhesion molecules involved in the assembly of neuronal circuits.

Additionally, the F_2 genome contains, on average, half of its ancestry from each founder species, but the ancestral segments differ across the genome between individuals due to meiotic recombination. The same ancestral allele at a QTL might interact differently with background loci depending on their genotype, dampening the strength of linkage of the QTL to the trait. By isolating an oldfield mouse allele at the QTL in a homogenous deer mouse background, the QTL allele is released from these epistatic constraints due to interaction with other oldfield alleles[249,250], and it is possible that the QTL genotype could have a stronger effect on behavior.

To overcome these two limitations of F_2 QTL mapping (low mapping resolution and epistasis with a heterogeneous genome), I created three congenic strains to "fine-map" the two strongest QTL, with the goal of isolating the minimal genetic region that affects exploratory behavior. Each congenic strain contains nearly >99% deer mouse ancestry except for a small donor segment of oldfield ancestry within a subset of the ~25Mb QTL on either chromosome 4 or chromosome 9.

Each congenic strain was created using a targeted backcrossing approach (**Figure 3.4**). Each backcross generation, I used ultra-low coverage (0.02–0.05X) sequencing to impute ancestry of each individual to select breeders for the next generation[251]. Chosen breeders had the highest proportion of deer mouse ancestry while maintaining a portion of oldfield mouse ancestry at the QTL of interest, and mice containing recombinant breakpoints within the QTL were preferentially selected. After six generations of backcrossing, I maintained three congenic lines: strains I – II, which fine-mapped the chr4 QTL into three segments (**Figure 3.5**), and strain III, which fine-mapped the chr9 QTL into two segments (**Figure 3.7**).

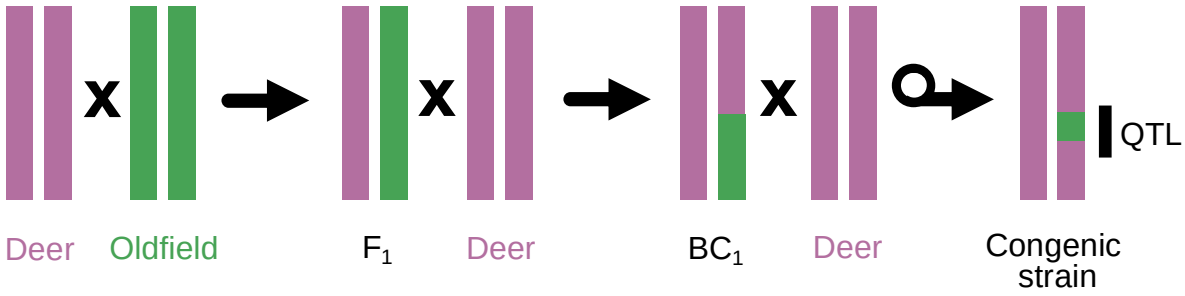

Fig. 3.4: Introgression mapping strategy. A genomic interval (QTL) that affects EPM behavior is introduced from exploratory oldfield mice (green) to the genome of the less exploratory deer mouse (purple) by targeted backcrossing for six generations. Shown is chromosome 9 from a representative mouse at each generation. All additional chromosomes of the final congenic strain are of deer mouse ancestry.

Testing the contribution of congenic genotype to exploratory behavior

After producing these three congenic strains, each with a distinct oldfield mouse donor segment within the QTL of chr4 or chr9, I crossed congenic heterozygous females to heterozygous males within each strain. The result was a cohort of offspring that possessed one of three genotypes at their respective congenic locus: (1) two copies of the deer mouse allele (i.e.

genetically identical to a wildtype deer mouse), (2) one deer mouse allele and one oldfield allele, or (3) two copies of the oldfield mouse allele. Then, I quantified the effect of genotype at the congenic locus on EPM behavior for each congenic strain.

Congenic strains I and II allowed me to fine-map the 25.5Mb chr4 QTL into three consecutive segments spanning 7, 8.5, and 10Mb, respectively (**Figure 3.5**). A difference in EPM behavior between genotypes within each strain could then be used to localize the genomic position of exploration-modulating variants. For example, if genotype were to correlate with open arm avoidance in strain I but not strain II, we could reason that the causal variants must lie in segment 1. If genotype was linked to behavior in both strains, then the causal variants would lie in segment 2, which is shared between strain I and strain II. Causal variants in segment 3 would explain linkage of genotype to behavior observed in strain II but not strain I.

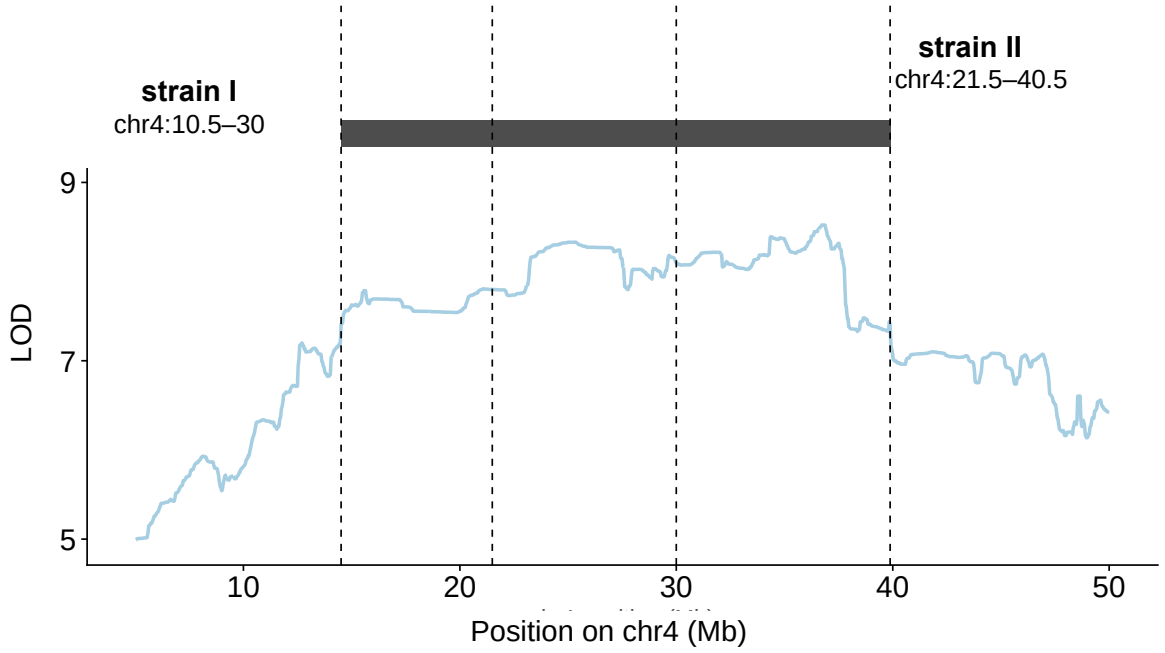

Fig. 3.5: Fine-mapping chr4 QTL. Green bars indicate the chromosomal position of the oldfield mouse allele introgressed in congenic strains I and II. The dark grey bar spans the 95% Bayes

credible interval of the QTL. Blue trace is the linkage (LOD score) of normalized time in the closed arms of the EPM to each genetic marker. Dotted lines represent the three segments of overlap between the congenic alleles and the QTL.

Unfortunately, genotype did not contribute to behavior in either strain I or strain II (**Figure 3.6**). It is possible that the QTL on chromosome 4 actually harbors several smaller QTLs that are broken up by recombination. Another potential explanation is that the exploration-increasing allele in the chr4 QTL interacts with other oldfield genetic variants to exert its behavioral effect and is unable to do so in this congenic genetic background. Curiously, all genotypes of strain I were very exploratory, which may be due to maternal or paternal effects, or the incidental accumulation of exploration-increasing alleles outside of the congenic region in this particular cohort of deer mice. Together, the lack of genotype effect in either congenic strain, combined with the interesting genotype-independent increase in exploratory behavior in strain I, highlights how exploratory behavior, like other behaviors, has complex inheritance patterns that can be difficult to untangle using standard mapping approaches. Thus, as I was unable to recapitulate the F_2 QTL results in the chr4 congenic strains, I did not further investigate the potential contribution of this locus to behavior.

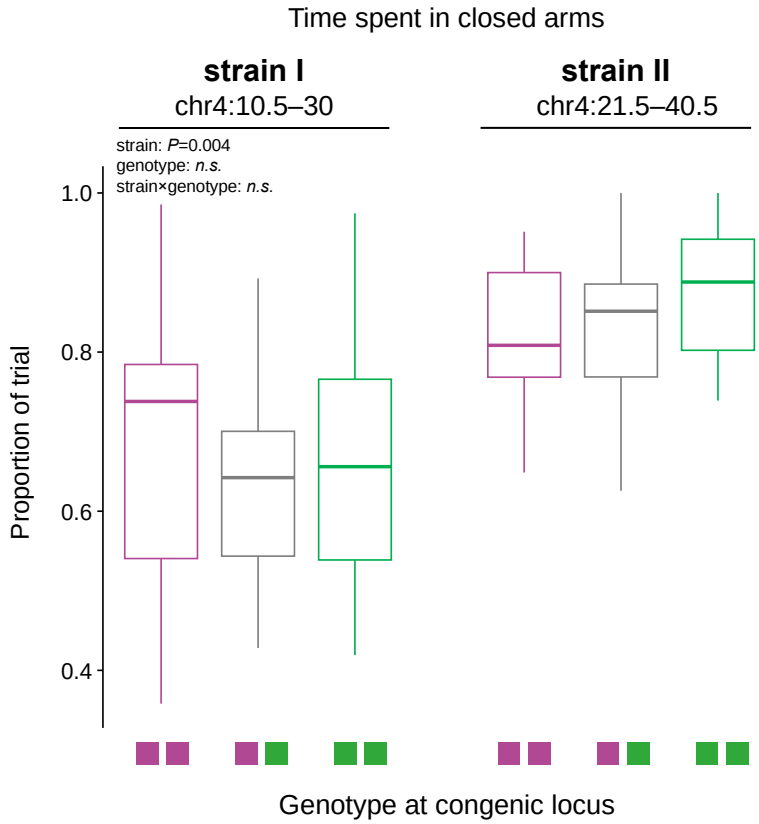

Fig. 3.6: Genotype at congenic locus in strains I and II does not contribute to EPM behavior. Time spent in the closed arms of the elevated plus maze by genotype in congenic mice, strains I and II.

Excitingly, however, there was a strong effect of congenic genotype on open arm avoidance behavior in strain III, which dissected the 23Mb-long QTL on chromosome 9 into segments of 15Mb and 8Mb containing 18 and 17 protein-coding genes, respectively (**Figure 3.7**). As observed in F_2 hybrids, congenic mice that inherited two copies of the oldfield allele were more exploratory than mice homozygous for the deer mouse allele, and heterozygotes displayed an intermediate phenotype (**Figure 3.8**). However, while F_2 hybrids homozygous for the oldfield allele spent just 6.5% more time in the open arms of the EPM compared to F_2 hybrids homozygous for the deer mouse allele, in the congenic strain, oldfield homozygotes

spent >11% more time in the open, suggesting that the contribution of genotype to exploratory behavior in the congenic strain is stronger than in F_2 hybrids. Indeed, the strength of association is also stronger in the congenic strain: while the oldfield allele explained 2.5% of phenotypic variance in F_2 mice, this allele explained 9.9% of the variance in the congenic mice (Spearman's ρ^2). This increase in phenotypic effect is consistent with the idea that this locus may interact with other oldfield alleles that dampen the strength of linkage in F_2 QTL mapping.

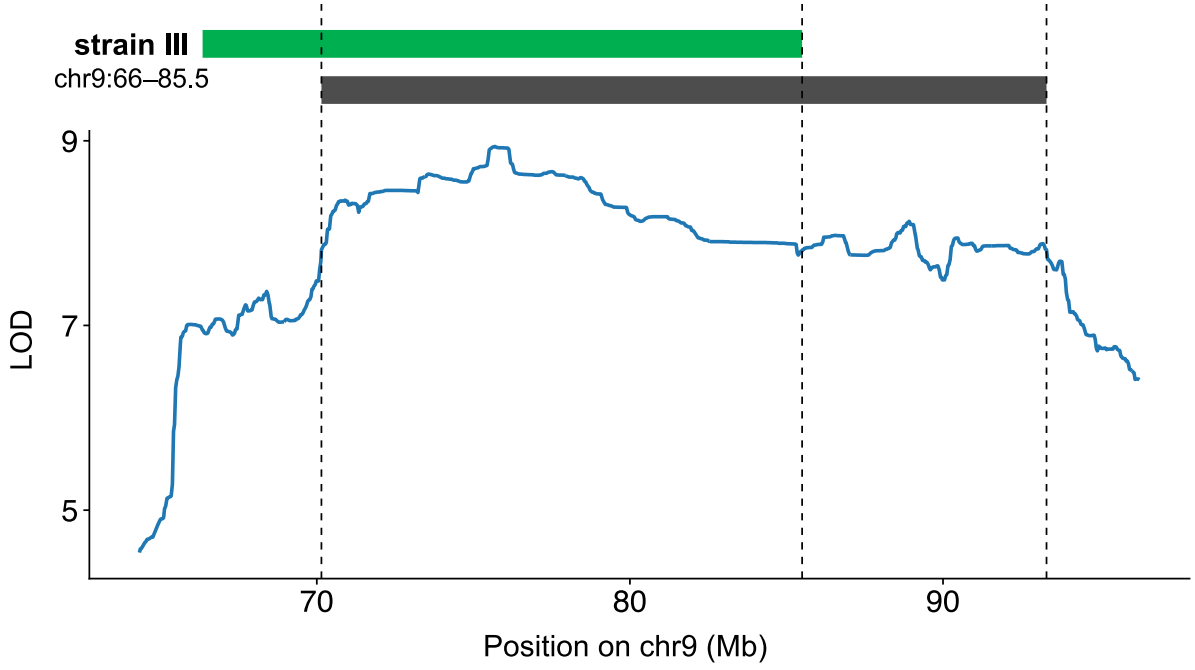

Fig. 3.7: Fine-mapping the chr9 QTL. Green bar indicates the chromosomal position of the oldfield allele introgressed in congenic strain III. The dark grey bar spans the 95% Bayes credible interval of the QTL. Blue trace is the linkage (LOD score) of time in the closed arms of the EPM to each genetic marker. Dotted lines denote the two segments of overlap between the congenic allele and the QTL.

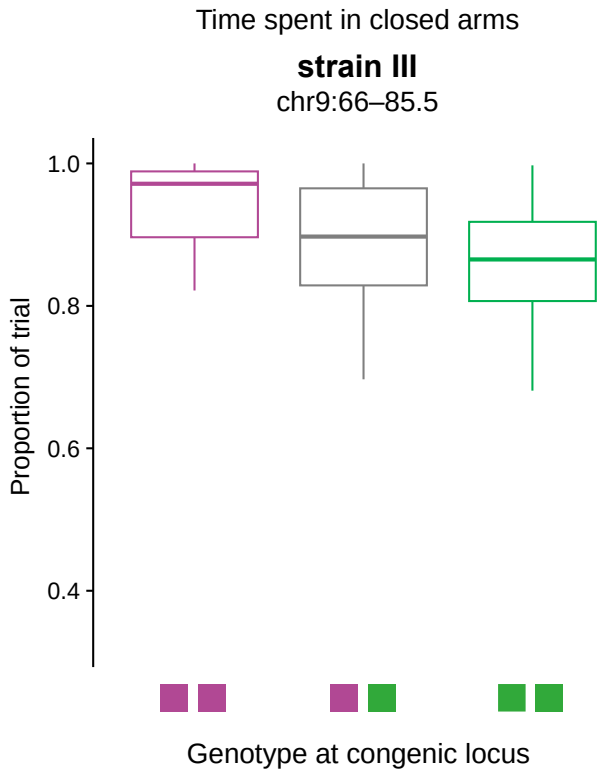

Fig. 3.8: Genotype at strain III congenic locus contributes to EPM behavior. Time spent in the closed arms of the elevated plus maze by genotype in strain III congenic mice. Kruskal-Wallis χ^2 = 12.102, P=0.0024. N=122.

Given the effect of strain III genotype on behavior, the causal variation contributing to open arm avoidance behavior on the EPM likely lies within the first 15Mb of the chr9 QTL, which is the shared region between the F_2 QTL and the congenic allele. This fine-mapped segment, from 70.1 - 85.5Mb on chr9, contains 18 protein-coding genes, listed in **Table 3.1**. It is likely that a species difference affecting one or more of these 18 genes contributes to the species difference in exploratory behavior.

Table 3.1: Protein-coding genes in the finemapped chr9 locus (chr9:70.1–85.5Mb)

Gene symbol	Gene name	Distance of TSS to QTL peak (Mb)
LOC102925380	40S ribosomal protein S4, X isoform-like	0.11
Pcdh17	protocadherin 17	0.53
Diaph3	diaphanous related formin 3	1.96
Tdrd3	tudor domain containing 3	2.23
Pcdh20	protocadherin 20	3.15
Olfm4	olfactomedin 4	4.26
Pcdh8	protocadherin 8	4.47
Cnmd	chondromodulin	4.58
Sugt1	SGT1 homolog, MIS12 kinetochore complex assembly cochaperone	4.66
Elf1	E74 like ETS transcription factor 1	4.72
Wbp4	WW domain binding protein 4	4.76
Kbtbd7	kelch repeat and BTB domain-containing protein 7	4.83
LOC121832382	40S ribosomal protein S27-like	4.84
Mtrf1	mitochondrial translation release factor 1	4.86
Naa16	N-alpha-acetyltransferase 16, NatA auxiliary subunit	4.87
Rgcc	regulator of cell cycle	4.97
Vwa8	von Willebrand factor A domain containing 8	5.40
Pcdh9	protocadherin 9	8.72

Identifying causal genes in the fine-mapped locus

Next, to identify which of these 18 genes most likely mediates exploratory behavior, I pinpointed genes that differed in amino acid sequence between deer- and oldfield mice, which could indicate a change in protein function. Only one gene, Sugt1, contained fixed missense

variants between the two species, at positions 60 and 62 (**Figure 3.9**). These sites were variable across a multiple species alignment of Sugt1, and both mutations (K60R in deer mice and R62K in oldfield mice) were conservative substitutions. Thus, it is not likely that these two variants have a meaningful effect on protein function.

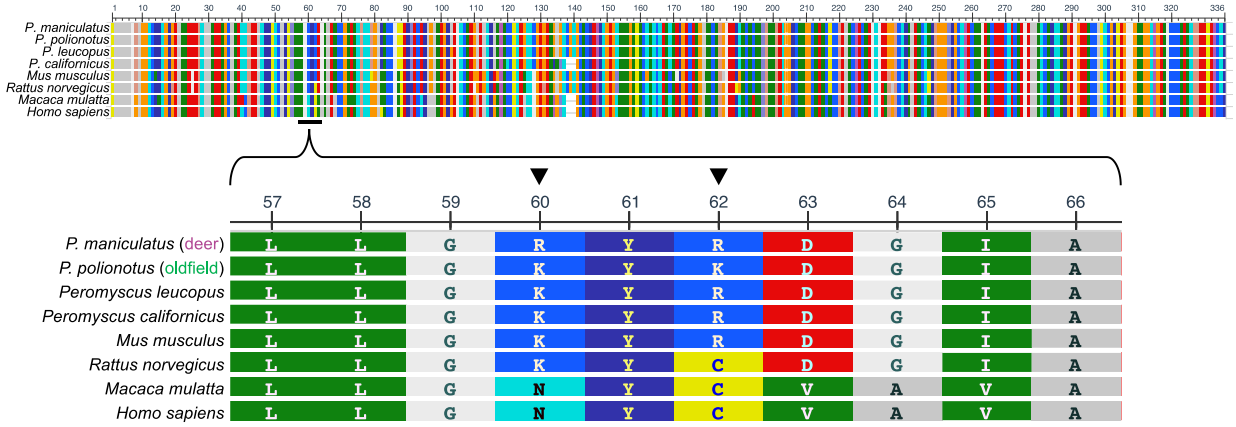

Fig. 3.8: Multiple species alignment of Sugt1. Protein alignment of the Sugt1 protein across deer mice (*Peromyscus maniculatus*), oldfield mice (*Peromyscus polionotus*), four additional rodent species (*P. leucopus, P. californicus*, *M. musculus*, and *R. norvegicus*), and two primates (*M. mulatta* and *H. sapiens*). Residues are colored according to the RasMol color scheme, denoting similarities in polarity and charge.

Alternatively, causal genetic variation may not affect protein sequence but rather influence the level of expression of nearby genes. I would expect such *cis*-regulatory variation to drive expression differences of genes in brain regions known to mediate exploratory behavior and anxiety, including the amygdala, hippocampus, hypothalamus, or medial prefrontal cortex (mPFC). Therefore, to search for *cis*-regulatory elements that affect EPM behavior, I next looked for genes that (1) had differential gene expression between deer- and oldfield mice and (2) had differential allelic expression in F_1 hybrids in these key brain regions. Because F_1 hybrids are heterozygous at all genomic positions, genes with biased allelic expression (i.e. in which the

oldfield allele is more highly expressed than the deer mouse allele or vice versa) indicates local *cis*-regulation rather than *trans*-acting regulatory variation that acts upon both alleles.

Of the 18 genes in **Table 3.1**, four had both differential expression in F_0 mice and allele-specific expression in F_1 hybrids in one or more of the aforementioned brain regions implicated in anxiety: Pcdh17, Olfm4, Sugt1, and Naa16. *Cis*-regulatory variation drove species differences in expression across all four brain regions in Naa16 and Sugt1, while *cis*-regulation specifically impacted Pcdh17 and Olfm4 expression in particular regions (Pcdh17 in the hippocampus and mPFC; Olfm4 in the hypothalamus) (**Figure 3.9**). Taken together, after analyzing protein sequence conservation and allele-specific expression, I narrowed down the likely list of causal genes from 18 to 4.

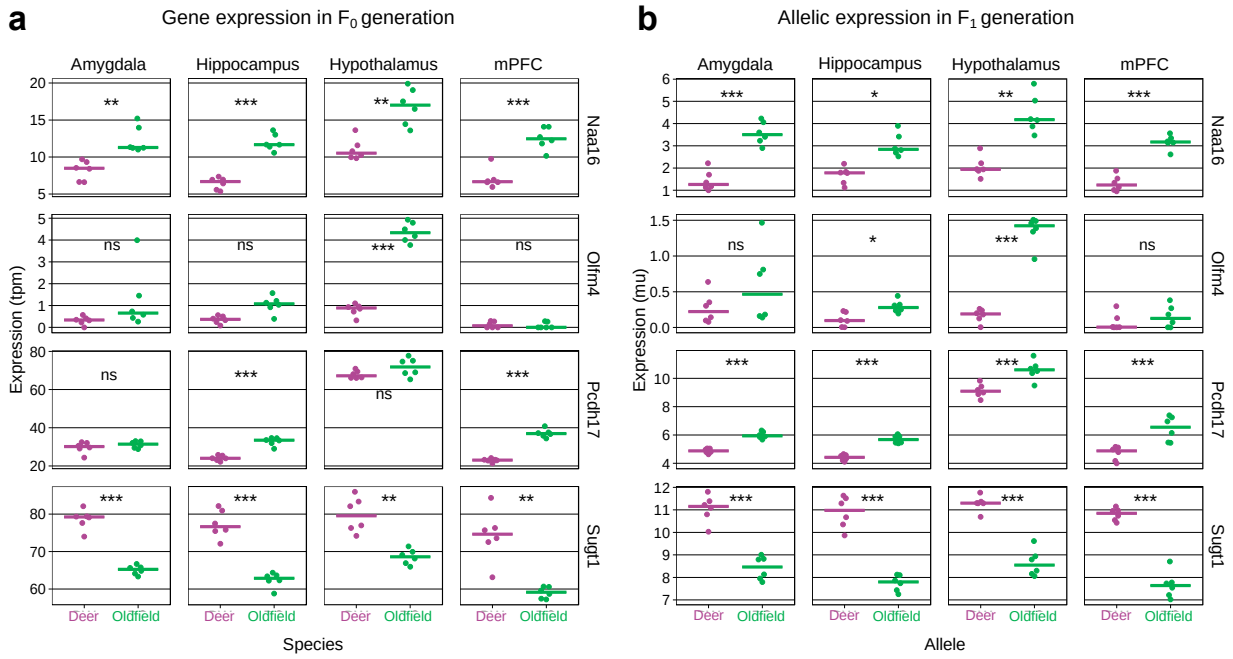

Fig. 3.9: Expression of strain III fine-mapping candidate genes in select brain regions that contribute to anxiety. a, Bulk RNA-seq expression of genes in deer- and oldfield mice from four brain regions. *P* values by unpaired *t*-test. **b**, Allele-specific expression (tpm: transcripts per million; mu: mmdiff expression parameter; mPFC: medial prefrontal cortex). *P* values by paired *t*-test; n.s. not significant; * *P*=<0.05; ** *P*=<0.01; *** *P*=<0.001. Gene expression data plotted from dataset published in [252].

Next, I reasoned that genes previously found to modulate behavior in lab mice or impact the risk of psychiatric disease in humans would be the most likely candidates to mediate exploratory behavior. Three of the four genes—Pcdh17, Olfm4, and Naa16—have been linked to behavior in rodents or humans (**Table 3.2**). In lab mice, knockout of Pcdh17 produces an anti-depressive-like phenotype but has no effect on anxiety-related behavior in the EPM[253]. In humans, genetic variation at the Pcdh17 locus is linked to altered amygdala structure and function as well as an increased risk of mood disorders[254]. The contribution of Olfm4 to rodent behavior is not well characterized; however, a genetic variant in the intron of the Olfm4 gene in humans is strongly linked to the risk of major depression—it is the most significant risk variant of depression across 44 independent and significant loci[255,256]—as well as the incidence of insomnia[257,258]. In lab mice, Naa16 knockout causes hyperactivity[259]; however, there is no clear link to human behaviors or psychiatric traits.

Table 3.2: Congenic locus genes with protein sequence mutation or differential gene regulation

Gene symbol	Coding change?	Differential gene expression (in F_0) and allelic expression (F_1) in relevant brain region	Linked to behavior in lab mice?	Linked to psychiatric risk in humans?
Pcdh17		✓	✓[253]	✓[254]
Olfm4		✓		✓[255–258]
Sugt1	✓	✓		
Naa16		✓	✓[259]	

Pcdh17, Olfm4, and Naa16 are each interesting candidates that could plausibly contribute to differences in exploratory behavior between deer- and oldfield mice. However, I decided to further interrogate the role of Olfm4 on EPM behavior—instead of Pcdh17 or Naa16—based on the following observations. First, the fold change in gene expression of hypothalamic Olfm4 between deer- and oldfield mice was the highest across all genes and brain regions depicted in **Figure 3.9**, and I reasoned that highly differential genes were poised to have the most significant effects on behavior. Second, that a common human genetic variant located in the intron of Olfm4 has an extremely strong association to major depression indicates that there may be a functional link between Olfm4 and the neurobiological endophenotypes that underlie anxiety and depression. Given the dearth of research characterizing the role of Olfm4 in the central nervous systems of rodents and humans alike, any level of functional characterization of Olfm4 as a causal mediator of behavior would not only advance our understanding of *Peromyscus* evolution but also of a devastating human psychiatric disease.

Therefore, I decided to further examine the role of Olfm4 on EPM behavior using two strategies. First, I quantified the expression of Olfm4 in the hypothalamus of F_2 hybrids using bulk RNA-seq to determine if higher levels of Olfm4 expression conferred higher exploratory behavior in the EPM, controlling for the chr9 QTL genotype. While the chr9 QTL genotype had a very strong effect on Olfm4 hypothalamic expression, indicating strongly biased allelic expression as observed in F_1 hybrids, the expression of Olfm4 did not contribute to time spent in the closed arms of the EPM (**Figure 3.10**). Thus, although the oldfield mouse allele at the Olfm4 locus contributes to both exploratory behavior and hypothalamic Olfm4 expression in F_2 hybrids, I did not identify a within-genotype effect of Olfm4 expression on behavior.

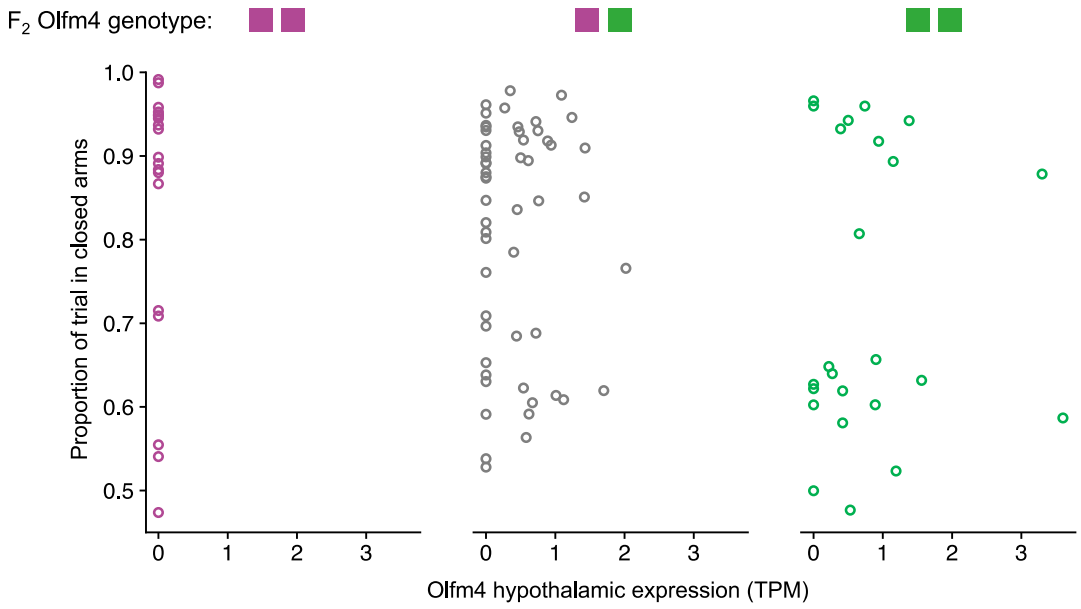

Fig. 3.10: Hypothalamic expression of Olfm4 in F_2 adult mice but does not contribute to variation in time spent in the closed arms of the EPM after controlling for Olfm4 genotype. Time spent in the closed arms of the elevated plus maze by Olfm4 genotype in F2 mice. Generalized linear model: genotype: $P < 2.2 \times 10^{-16}$, Olfm4 expression: P=0.99, interaction: P=0.99. N=97.

A more direct test of a potential Olfm4 contribution to behavior is the targeted manipulation of Olfm4 expression. Thus, I next tested the effect of Olfm4 knockout on exploratory behavior. Due to the technical limitations of genetic manipulation in *Peromyscus* mice, I tested EPM behavior in an existing Olfm4-null strain of C57BL/6 lab mice[260]. I crossed the Olfm4-null strain to wildtype C57BL/6 to produce heterozygous Olfm4+/- mice, and then crossed these heterozygotes to produce a cohort of offspring containing either zero, one, or two copies of the knockout locus. Ultimately, I found that Olfm4 knockout had no effect on exploratory behavior (**Figure 3.11**), suggesting that loss of Olfm4 expression does not contribute to anxiety in C57BL/6 lab mice.

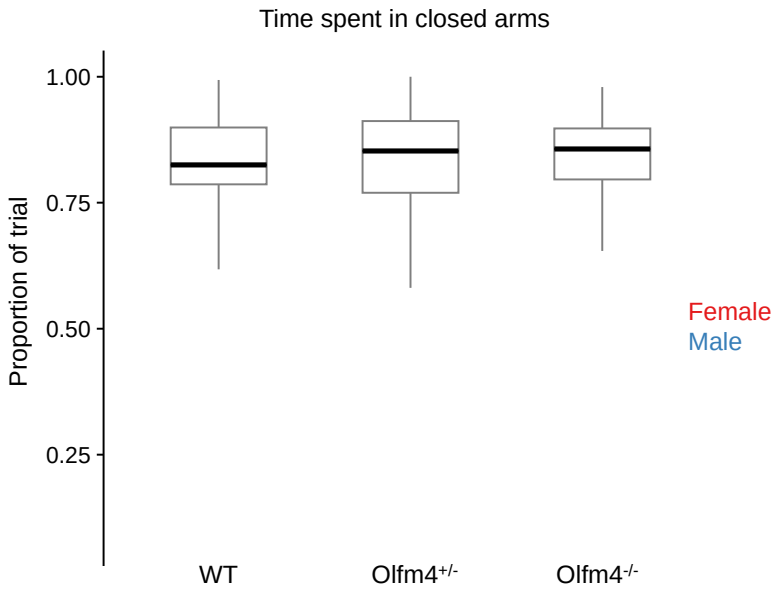

Fig. 3.11: Olfm4 knockout in C57BL/6 mice does contribute to EPM behavior. Time spent in the closed arms of the elevated plus maze in C57BL/6 mice by Olfm4 allele. WT: wildtype. Generalized linear model: genotype: P=0.47, sex: P=0.80, genotype×sex: P=0.74. N=150. Olfm4-null strain provided by Matthew N. Alder.

Discussion

In this chapter, I investigated the genetics underlying the striking difference in exploratory behavior between sister species of *Peromyscus*. Through a combination of QTL mapping and fine-mapping strategies, I identified a 15-Mb genetic locus on chromosome 9 that contributes to variation in behavior. This locus encompasses 18 protein-coding genes, including Pcdh17, Olfm4, and Naa16, which have been previously linked to behavioral traits in other mammals and harbor *cis*-regulatory variants that drive patterns of differential expression in key brain regions underlying anxiety-related behavior. While initial experiments did not discover a functional connection between Olfm4 expression and exploration, future research will further interrogate the molecular divergence of this fine-mapped region.

Olfactomedin-4 (Olfm4) is a secreted glycoprotein that is expressed strongly in the intestines and has anti-apoptotic and cell adhesive properties[261–263]. Despite harboring noncoding variants that strongly predict depression in humans, its mechanistic link to behavior remains unknown. I found that while Olfm4 displayed strong genotype-dependent effects on expression levels in the F_2 hypothalamus, its expression did not predict exploratory behavior in *Peromyscus* hybrids. Furthermore, there was no effect of an Olfm4 null mutation in a laboratory strain of *Mus musculus* mice (C57BL/6). Despite these initial negative results, further investigation into the role of Olfm4 in *Peromyscus* exploratory behavior is still warranted, given the limitations of my initial experiments combined with the promising contribution of other members of the olfactomedin gene family to neural circuits and anxiety-related behavior.

Hypothalamic Olfm4 expression in F_2 adults did not correlate with exploration after controlling for Olfm4 genotype; however, this result does not discount the potential contribution of Olfm4 during specific developmental time periods. Indeed, the expression of certain olfactomedin proteins during development is essential for proper circuit wiring: the olfactomedin domain of the protein gliomedin mediates Schwann cell-axon interactions and Node of Ranvier formation in peripheral nerves[264], and olfactomedin-2 (Olfm2) expression is required in zebrafish for maturation of the anterior central nervous system[265]. In lab mice, olfactomedin-1 (Olfm1) is required for proper brain growth and synaptic protein interaction, and knockout of both Olfm1 and Olfm2 results in perturbations to anxiety-related behavior[266,267]. While the developmental time course of Olfm4 expression in the *Peromyscus* central nervous system is not known, it is possible that its expression early in development impacts the formation of anxiety-related neural circuitry that modulates behavior in adulthood.

Another possibility is that adult Olfm4 expression might contribute to exploratory behavior when expressed in specific hypothalamic nuclei or perhaps in other parts of the brain. I conducted bulk RNA-seq of the entire F_2 hypothalamus, which is comprised of many transcriptionally-distinct cell types and neuronal populations that contribute not only to anxiety but to sleep, hunger, body temperature, blood pressure, and arousal. The low level of Olfm4 expression in my bulk RNA-seq data (<4 transcripts per million from all F_2 hybrids characterized) suggests that Olfm4 may be expressed in a small subset of hypothalamic cells; therefore, attempts to correlate exploration with Olfm4 expression in F_2 hypothalami using single cell methods may be more fruitful as it would confer higher resolution by restricting analysis to only the cell types in which Olfm4 is expressed. Alternatively, it is possible that Olfm4 may influence behavior through its expression in regions outside of the hypothalamus. In rats, Olfm4 is very lowly expressed in the hypothalamus yet has high expression in the frontal cortex[268], and in humans, Olfm4 expression in the dorsolateral prefrontal cortex is associated with major depression[269]. Thus, the expression of Olfm4 in more frontal brain regions, rather than the hypothalamus, may be important for modulating behavior in *Peromyscus*.

While it was low-hanging fruit to test the contribution of Olfm4 to EPM behavior in Olfm4-null C57BL/6 lab mice, they are not an ideal model for understanding the role of Olfm4 in the central nervous system. Unlike in rats, Olfm4 is expressed at extremely low levels across all brain regions in adult lab mice[270]. For example, across nearly 350,000 hypothalamic cells collected in C57BL/6 embryos, juveniles, and adults, over 99.7% of cells had zero Olfm4 transcription[271]. Since hypothalamic transcript levels of Olfm4 are so low across all developmental stages, knockout of this gene would not be expected to have a meaningful effect on hypothalamic function and therefore behavior. While Olfm4 may not play a significant role in

the nervous system of lab mice, it is possible that it plays a more prominent role in the neural circuits underlying behavior in humans and oldfield mice.

Of course, it may be the case that there is no connection of Olfm4 to behavior in *Peromyscus*. It is possible that another gene in the fine-mapped region—perhaps protocadherin-17 (Pcdh17)—underlies a portion of the species difference in exploratory behavior. Pcdh17 is a member of the non-clustered δ2-protocadherin family and, like Olfm4, has been identified as a likely susceptibility gene for major mood disorders[254]. High resolution QTL mapping in highly recombinant outbred populations of lab mice identified a QTL for total distance travelled in the EPM that contains just one gene: Pcdh17[272]. However, it is not clear that variation in the total distance traveled in the EPM reflects differences in anxiety-related exploratory behavior; rather, it is likely a more general readout of locomotor activity. Between deer- and oldfield mice, for instance, I do not observe a difference in total distance traveled in the EPM despite the large difference in open arm avoidance. In lab mice, knockout of Pcdh17 did not affect the time spent in the closed arms of the EPM nor did it influence locomotor activity (total distance traveled in the open field test). However, Pcdh17-null mice do exhibit heightened "anti-depressive-like" behavior including reduced immobility in both the tail suspension test and the forced swim test, indicating that altering its expression can produce behavioral effects in rodents. Intriguingly, in our genetic mapping of deer- and oldfield F_2 hybrids, the Pcdh17 promoter lies just 500 kb from the QTL peak. Targeted manipulation of Pcdh17 expression in *Peromyscus* would also be an informative next step to undercover the role of this fine-mapped locus in modulating exploratory behavior.

Previous introgression fine-mapping efforts in nematodes, flies, and lab mice have successfully identified single genes—and even specific genetic variants—that modulate behavior

(e.g. [21,52,53,273–275]). While I did not convincingly demonstrate the contribution of a particular gene to exploratory behavior, I was able to strongly implicate a 15-Mb locus containing 18 genes, and narrowed down the likely genes to a shortlist of three: Olfm4, Pcdh17, and Naa16. Ultimately, my results advance our understanding of the genetic architecture of exploratory behavior, a hallmark of animal behavior that is commonly dysregulated in psychiatric disease.

Methods

Elevated plus maze (EPM) assay. The EPM apparatus contained two closed arms and two open arms (12" long and 1.5" wide; wall height: 17.5") made from acrylic. The open arms were lined with a ½" acrylic ledge to discourage mice from falling or jumping off the apparatus. The platform was 24" from the ground and brightly illuminated with LED strip lights affixed to the ceiling.

Before each trial, the subject was transferred from its home cage to a 5"x3.5"x1.5" acrylic box, and the box was placed in the EPM for a two-minute acclimation period. After two minutes, the box was opened at the center of the maze without disturbing the mouse. After release from the acclimation box, the trial begins, and the subject is recorded exploring the EPM using a raspberry pi camera. The centroid position of the mouse was tracked using a custom Python script, and duration spent in the closed and open arms was calculated.

In the event that a mouse climbed or jumped off of the EPM, the mouse was retrieved and returned to the acclimation box for two minutes, and the trial began again. Mice that jumped off the apparatus more than once were discarded from further analysis. All surfaces of the EPM were cleaned with 70% ethanol between trials.

Free exploration open field assay. A subject's home cage was connected by tubing to a novel 4'x4' arena that the mouse can choose to enter and explore at will over the course of a 10-hour overnight trial. The trial begins one hour before lights-off, when the barrier between the home cage and arena is removed. For the next 10 hours (1 hour lights-on, 8 hours lights-off, 1 hour lights-on), the position of the mouse is recorded from above using an infrared raspberry pi camera as it moves freely between the home cage and the arena. At the end of the trial, the centroid position of the mouse was tracked using a custom Python script, and duration spent in center of the arena (at least 6" from the walls) was calculated. Food and water was provided *ad libitum* in the home cage.

Three-chamber sociability test. The three-chamber apparatus consisted of a clear acrylic box with two inner walls that delimited three 60 cm x 30 cm x 30 cm chambers, with openings allowing for free movement between chambers. The test is conducted in dim lighting. Before each trial, the test mouse was placed in the middle chamber for habituation for 10 minutes; the openings to the left and right outer chambers were blocked with removable clear plexiglass. After acclimation, an inverted metal pencil cup containing an unfamiliar mouse of the same sex was placed in the right chamber and an empty inverted pencil cup was placed in the left chamber. The pencil cup contained slits which allowed for nose protrusion from the unfamiliar mouse and thus the potential for physical contact with the test mouse. At the start of the trial, the chamber entrance barriers were removed, and the test mouse was allowed to explore all three chambers freely for 10 minutes. Movement of the test mouse was recorded above using a raspberry pi camera. The time spent interacting with the novel mouse was then manually scored from the video footage. A mouse was considered to be interacting with the novel conspecific if its head was oriented towards and nose within one inch from the cup.

Novel object introduction paradigm. A novel object (a sterile plastic pestle) is introduced into the subject's home cage, and the subject's behavior is recorded using a raspberry pi camera over a 10-minute trial. The time spent interacting with the object was then manually scored from the video footage. Interaction was scored as any physically handling of the object including biting, licking, huddling, or physically carrying it around the cage.

QTL mapping EPM behavior. QTL mapping of EPM behavior was conducted in 1571 F_2 hybrid mice (808 males, 763 females). Details of EPM protocol and determination of marker genotypes is described in [24]. Time spent in the closed arms of the EPM ("open arm avoidance") was transformed using the following formula to normalize the data due to negative skew: $y = 1/(\max(x) + 1 - x)$.

Normalized open arm avoidance was then QTL mapped using R/qtl[239] by Haley-Knott regression with sex added as covariate. The scanone function was used to compute the genome-wide LOD significance threshold at $\alpha<0.05$ using 1000 permutations.

Congenic fine-mapping. Creation of congenic strains began from three breeding pairs of founders (three cages whereby a deer mouse females mated with an oldfield mouse male). From this initial cross, 10 F_1 hybrid males were then backcrossed to deer mouse females to produce the first backcrossed (BC_1) generation. For this and each successive BC generation, I collected ear clip tissue at weaning age from which to extract genomic DNA using the Omega Mag-Bind® Blood & Tissue DNA HDQ 96 Kit. DNA libraries were created via tagmentation using homebrew Tn5 transposase and barcoding by PCR. Libraries were sequenced on a NextSeq550 at 0.2–0.5x coverage (1x76bp reads).

The set of fixed SNP locations was determined from variant calling of high-coverage sequencing of deer- and oldfield samples by the Hoekstra lab, now published in [276,277] (raw read

data available on the NCBI short-read archive). SNPs were considered to be fixed if they were homozygous for the reference allele in all deer mouse samples and homozygous for the alternate allele in all oldfield samples and passed the following variant calling thresholds: MQ>=59, AC>=3, QD>=35, GQ>=45. I then called variants using deer- and oldfield mice from our own colony and used this VCF to prune the set of fixed SNPs to a total of 1,116,155 variants.

Each backcrossed library sample was aligned to the deer mouse genome HU_Pman2.1.3 using bwa-mem and samtools mpileup was used to generate variant information at all 1,116,155 variant sites cross the genome. The hidden markov model AncestryHMM[251] was then used to predict ancestry along each chromosome from each genotype panel. Based on each mouse's genome-wide ancestry, a subset of mice were chosen as breeders for the successive backcross generation. After six generations of backcrossing, three congenic strains were created, each with deer mouse ancestry across all genomic locations except for the following locations containing oldfield ancestry: chr4:10.5–30Mb in Strain I, chr4:21.5–40.5Mb in Strain II, and chr9:66–85.5Mb in Strain III.

Identifying congenic genes harboring missense mutations. RNA-seq reads from various brain regions from deer- and oldfield mouse were obtained from [252] and variants were called using bcftools. The VCF was then filtered for sites where all deer mouse samples were heterozygous for one version of the allele and all oldfield samples for the other allele, and then filtered to contain only genotype information from the congeni region (chr9:70.1–85.5Mb). SNPeff version 4.3t was used to identify which of these variants resulted in an animo acid difference between species. Multiple sequence alignment of Sugt1 was conducted using the COBALT multiple sequence alignment tool with protein sequences for each species obtained from NCBI.

Identifying congenic genes harboring *cis*-regulatory variation. Gene expression data from in deer mouse, oldfield mouse, and F_1 hybrid brain regions was obtained from [252]. Refer to [252] for details of tissues dissection, sequencing, and expression analysis. Differential gene expression between deer and oldfield mice in the amygdala, hippocampus, hypothalamus, and medial prefrontal cortex was tested for all congenic genes (see Table 3.1) by unpaired *t*-test. Allele-specific expression between F_1 alleles was tested via paired *t*-test.

Correlation of Olfm4 expression in F_2 hypothalami and EPM behavior. The brains from 96 F_2 hybrids of deer- and oldfield mice (a subset of mice from the original F_2 QTL mapping population described in *QTL mapping EPM behavior*) were dissected and stored at -70 ºC in cryomolds in OCT. Each OCT-embedded mold was sectioned using a Leica CM3050S cryostat on a horizontal plane in the ventral-to-dorsal direction until the hypothalamus was visible through the OCT. A precision tissue punch tool was then used to punch out the whole hypothalamus (a 3mm-diameter, 2-mm deep tissue section from the ventral side of the brain). Hypothalamic punches were transferred to an Eppendorf tube and stored at -70 ºC until RNA-seq library prep. RNA-seq libraries were made, sequenced, and analyzed using the protocol described in *F_2 adrenal RNA-seq*. The contribution of Olfm4 genotype and expression to EPM behavior was determined using a generalized linear model testing for the effect of Olfm4 genotype, Olfm4 expression, and their interaction.

Contribution of Olfm4-null allele to EPM behavior in C57BL/6 mice. An Olfm4-null strain of C57BL/6 *Mus musculus* mice was produced by and obtained from the Alder lab[260]. Olfm4-null homozygotes were crossed with wildtype C57BL/6 mice to produce a cohort of F_1 hybrid heterozygotes. F_1 hybrids were then crossed to produce F_2 hybrids containing either 0,1, or 2 copies of the Olfm4-null locus. Their EPM behavior was tested and the contribution of

genotype to behavior was calculated using a generalized linear model testing for effect of genotype, sex, and their interaction.

Chapter 4: Discussion and Future Directions

After 800 million years of animal evolution, the striking diversity of behavioral adaptations across the animal kingdom is nothing short of astonishing. However, we still understand little about the genes and molecules that underlie behavioral evolution. My dissertation promised to explore the genetic, molecular, and neuroendocrine basis of behavior; therefore, in this final chapter, I would like to examine each of these categories—genes, molecules, and neuroendocrinology—one by one, to summarize what I have discovered during my PhD research and to discuss the many open questions left to be answered.

The genetic basis of behavior

As I wrote in the introduction, there are no genes for behavior[33]. Instead, genes interact with cellular pathways and shape neuronal circuitry that ultimately generates behavior. In chapter 2, I discovered that the gene Akr1c18 is a potent modulator of the molecular and neuroendocrine pathways that give rise to behavior, and the species difference in adrenal expression of Akr1c18 may be a key contributor underlying the divergence in parenting strategy between promiscuous deer mice and monogamous oldfield mice. I found that biparental oldfield mice have evolved a new cell type in their adrenal glands that highly expresses Akr1c18, a gene not expressed in deer mouse adrenals, and that the enzymatic product of Akr1c18—the steroid 20α-OHP—increases parenting behavior. Through quantitative genetic dissection of the adrenal transcriptome, I discovered that this species difference in Akr1c18 expression is due to *cis*-regulatory variation that drives the expression of the gene Tnn in oldfield adrenals.

The mechanism underlying how Tnn, a core member of the extracellular matrix (ECM), could drive the emergence of a new cell type is yet to be discovered. My genetic results suggest that Tnn expression drives the expression of genes whose protein products localize to the mitrochondria (Akr1c18), cytoplasm (Tsf4), nucleus (Cdkn1a and Ckn2a), and ECM (Podnl1, Serpine1, and Timp1). These genes are all markers of the *zona inaudita* cell type, suggesting that signaling cascades mediated by extracellular TNN may coordinate the expression of these genes via autocrine signaling within individual *zona inaudita* cells and/or via close-range paracrine signaling between these cells. In the lab mouse, Tnn has been shown to increase both canonical Wnt signaling [220,278] and hedgehog signaling[279], and Tnn is likely also a ligand for integrins[280], transmembrane receptors that activate transduction pathways impacting cell growth and differentiation[281]. Future research should interrogate how Tnn coordinates the emergence of the *zona inaudita* cell by determining which signaling pathways are impacted by variation in Tnn expression in oldfield adrenals.

An informative next experiment would be to knock down Tnn expression in the oldfield adrenal gland just before *zona inaudita* cells begin to develop. This could be accomplished by injecting a recombinant adeno-associated virus (AAV)—preferably of serotype AAV9, which has high tropism for the adrenal cortex[282]—either systemically through the bloodstream or intraadrenally via microinjection. The AAV would express either short hairpin RNAs or CRISPR reagents to silence Tnn transcripts. Following adrenal Tnn knockdown, one could then characterize the histological and transcriptomic effect in these glands and monitor perturbations in *zona inaudita* development. Interestingly, Tnn knockout in lab mice causes an enlargement of the spleen and reduction in size of the liver[259]. While the tissue distribution of Tnn expression in the oldfield mouse is not yet known, systemic delivery of a Tnn-knockdown AAV that targets

multiple organs might demonstrate that Tnn not only impacts the size of the adrenal gland (through the differentiation and accumulation of a new cell type) but also the size of other organs.

In chapter 3, I identified a 15-Mb genetic locus that contributes to increased exploratory behavior in oldfield mice compared to deer mice, bringing one step closer to uncovering genes that underlie this species difference in behavior. Future attempts to pinpoint the causal variation in the chr9 congenic locus could benefit from following additional experiments. First, it would be beneficial to characterize the molecular and functional impact of the congenic locus on the central nervous system. I might start by visualizing the distribution of expression of Olfm4, Pcdh17, and Naa16—genes with *cis*-regulatory variants in the congenic locus—using *in situ* hybridization in serial brain sections between congenic mice with and without oldfield ancestry at the congenic locus. Brain regions where congenic genotype correlates with gene expression could be further characterized through single-cell RNA-seq to identify all genes that differ in expression in particular cell types due to ancestry at the congenic locus.

Future experiments should also continue to investigate the connection between Olfm4 expression and EPM behavior through functional perturbations of gene expression. One strategy would be to deliver an AAV to drive Olfm4 expression under a ubiquitous or neuron-specific promoter via intracerebroventricular injection or stereotaxic injection into the hypothalamus of newborn deer mice. The impact of Olfm4 overexpression on exploratory behavior could then be tested 6–8 weeks later. Using an opposite approach, one could deliver an AAV to drive the expression of molecules to knock down Olfm4 expression in oldfield mice, using a similar viral construct proposed earlier to knockdown Tnn. Both tactics would be useful experiments to functionally tie Olfm4 expression to exploratory behavioral differences between deer- and

oldfield mice. If modifying levels of Olfm4 expression does not impact exploratory behavior, one could use a similar strategy to target other genes in the congenic locus including Pcdh17.

Are there overarching lessons about the genetic basis of behavior that we can glean from chapters 2 and 3? One similarity is in the nature of mutations, as noncoding variation rather than protein sequence evolution seems to underlie the behavioral contributions of Tnn, Akr1c18, and the chr9 QTL. The role of *cis*-regulatory variation can be particularly important in behavioral evolution, as it allows for modular changes to gene expression patterns that are restricted to particular tissues or developmental stages[89]. Another interesting connection between chapters 2 and 3 is the potential contribution of genes that are expressed outside of the central nervous system to behavior. Most notably, the gain of Akr1c18 expression in adrenal gland, rather than in the brain, contributes to differences in parenting behavior between deer- and oldfield mice. Furthermore, while I focused on characterizing Olfm4 expression in the hypothalamus, it should be noted that Olfm4 is highly expressed outside of the central nervous system, namely in the gut. The mechanism underlying the Olfm4 contribution to depression in humans is not understood; however, it is intriguing that the Olfm4 intronic variant associated with depression is also highly associated with obesity, indicating that Olfm4 might contribute to both diseases through modulation of the gut-brain axis[283]. Ultimately, my research highlights that the evolution of behavior is not necessarily caused by genetic changes restricted to the central nervous system, and that certain tissues outside of the brain could serve as potential targets for novel psychiatric and behavioral therapies.

The molecular basis of behavior

Steroid hormones have long been recognized as important signaling molecules with powerful effects on animal behavior. However, nearly all of what we understand about the contribution of steroids to variation in behavior stems from just a handful of steroid hormones, namely cortisol/corticosterone, progesterone, estradiol, and testosterone. The wealth of research on the impacts of these classical steroids far outweighs what is known about the contributions of their steroid derivatives. In chapter 2, I showed that a scarcely-studied derivate of progesterone, 20α-OHP, powerfully increases parental care behaviors when administered to deer- and oldfield mice. This result suggests that many steroid derivatives, long characterized as "biologically inactive" (e.g.[284–286]), could potentially also influence on behavioral evolution.

While I focused on the emergence of the *zona inaudita* in my dissertation, my single cell transcriptomic comparison between deer- and oldfield adrenals reveals a high level of molecular divergence between these species. For example, the enzyme that synthesizes the glucocorticoid corticosterone, Cyp11b1, is more highly expressed in oldfield adrenals compared to deer mouse adrenals. Glucocorticoids not only cause physiological changes in immune function, cardiovascular function, and cognition[287], but also behavioral changes including heightened anxiety or depressive-like symptoms[288]. These behavioral outcomes can arise acutely in response to a short stressor[289] or can persist during chronic glucocorticoid elevation, as in patients with Cushing's syndrome[290]. Unpublished data from the lab suggests that oldfield mice have >20-fold higher levels of circulating corticosterone than deer mice, and that the oldfield glucocorticoid receptor harbors mutations that reduce its sensitivity to corticosterone binding. The impact of these evolved differences in glucocorticoid signaling on behavioral differences between deer-

and oldfield mice—as well as the contribution of high corticosterone to oldfield adrenal enlargement—are interesting questions worth further investigation.

20α-OHP and corticosterone are just two of the many hormones secreted by the oldfield adrenal gland that may impact behavior. To further understand the adrenal contribution to behavioral evolution in oldfield mice, I propose using untargeted metabolomics to characterize the full suite of adrenal-secreted hormones in order to quantify which hormones differ between deer- and oldfield mice. Hormones whose levels differ between the species could indicate molecules with important effects on behavior. As geneticists have moved away from a candidate gene approach toward more unbiased methods, so too could endocrinologists move toward an unbiased consideration all steroids, including lesser studied metabolites, in characterizing the hormonal contribution to behavior. Indeed, my results suggest that some of the behavioral effects thought to be mediated by certain steroids may actually be caused by their metabolism into downstream molecules with different functional properties[291]. Ultimately, considering how steroid metabolism influences behavior is a fruitful research direction that would undoubtedly advance our understanding of animal behavior.

The neuroendocrine basis of behavior

The influence of adrenal-derived 20α-OHP on behavior is a fascinating and novel finding—to my knowledge, we are the first to characterize the contribution of 20α-OHP to parenting behavior. Yet, the mode of action by which 20α-OHP affects parenting is yet to be firmly established. As I have shown in chapter 2, it is possible that these effects are largely mediated through the actions of allo-diol via the inhibition of tonic GABAergic currents. However, it is also possible that 20α-OHP or its metabolites bind to as-yet-undiscovered receptors in the central

nervous system to shape parental behaviors. Interestingly, previous studies have characterized allo-diol as a positive rather than negative allosteric modulator of $GABA_AR$[202,203]. This raises the possibility that allo-diol could have differential effects on $GABA_AR$ signaling depending on its local concentration or other factors, and future research should characterize how the balance between positive and negative $GABA_AR$ modulation might contribute parental behavior. Characterizing the mechanism by which 20α-OHP and allo-diol modulate behavior is a pivotal next step in understanding the ultimate contributions of these molecules to evolved species differences in behavior.

A widely held notion in the field of behavioral genetics is that behavioral variation arises from the contributions and interactions of many genes[292]. It may be helpful to consider the neuroendocrine contribution to behavior in a similar framework: that the integration of signaling from many hormones ultimately influences animal behavior. For example, competition between steroids that bind the same enzymes can have cascading effects on local levels of metabolites, each with unique biochemical properties. As such, high concentrations of 20α-OHP in the brains of oldfield mice suggests increased competition between 20α-OHP and progesterone as substrates for the enzymes 5α-R and 3α-HSD, which convert 20α-OHP to allo-diol and progesterone to allopregnanolone. This substrate competition could underlie a variety of functional effects on parental behavior circuitry. First, high rates of 20α-OHP metabolism leads to the accumulation of 20α-hydroxylated steroids like allo-diol, which we showed are negative modulators of GABA signaling. Second, higher competition with 20α-OHP leads to lower rates of progesterone metabolism, which not only reduces rates of allopregnanolone synthesis but also increases levels of progesterone signaling. Thus, it is possible that the effect of 20α-OHP on parenting lies in the delicate balance between the genomic actions of progesterone and non-

genomic actions of neurosteroidal derivatives of progesterone and 20α-OHP. Fully understanding how steroids compete and interact to shape neuronal circuits is a difficult task; yet, the more we investigate these complex interactions, the greater understanding we will have of the neuroendocrine contribution to behavioral evolution.

Future experiments should additionally attempt to identify the neuronal populations in which 20α-OHP acts to ultimately alter parental behavior. Local infusions of 20α-OHP or allo-diol to particular brain regions, followed by parental behavior testing, could help us pinpoint which aspects of the parental care circuitry respond to 20α-OHP and its metabolites. Understanding where and how 20α-OHP shapes behavior in *Peromyscus* could also advance our understanding of the neuroendocrine basis of parenting in humans. While the contribution of 20α-OHP to parenting in humans is not known, my research raises the tantalizing possibility that we might one day be able to pharmacologically treat parental neglect of children through delivery of a 20α-OHP-like drug. While we are still in the early stages of understanding the influence of 20α-OHP on parental behavior, this novel biochemical link opens a new research avenue into the biological underpinnings of parenting in humans and rodents alike.

Summary

Throughout this dissertation, I have combined forward genetics and unbiased genome-wide sequencing techniques to search for the genetic and molecular changes that drive behavioral evolution. I have shown that divergent parental behaviors in deer- and oldfield mice have arisen in part from genetic variation that drove the evolution of a new cell type, which conferred novel biochemical functionality to the oldfield adrenal gland. Furthermore, I have identified a genetic locus that directly contributes to species differences in exploratory behavior,

an essential animal behavior with implications for understanding anxiety in humans. Altogether, my work has advanced the field of behavioral genetics by deepening our understanding of the genetic, molecular, and neuroendocrine mechanisms that underlie the evolution of behavior.

References

1. Crabbe, J. C., Wahlsten, D. & Dudek, B. C. Genetics of mouse behavior: Interactions with laboratory environment. *Science* **284**, 1670–1672 (1999).
2. Bult, C. J., Blake, J. A., Smith, C. L., Kadin, J. A. & Richardson, J. E. Mouse Genome Database (MGD) 2019. *Nucleic Acids Res.* **47**, D801–D806 (2019).
3. Polderman, T. J. C. *et al.* Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat. Genet.* **47**, 702–709 (2015).
4. Stirling, D. G., Réale, D. & Roff, D. A. Selection, structure and the heritability of behaviour. *J. Evol. Biol.* **15**, 277–289 (2002).
5. Karlsson Linnér, R. *et al.* Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. *Nat. Genet.* **51**, 245–257 (2019).
6. Ripke, S. *et al.* Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).
7. Baselmans, B. M. L. *et al.* Multivariate genome-wide analyses of the well-being spectrum. *Nat. Genet.* **51**, 445–451 (2019).
8. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
9. Nakka, P. *et al.* Characterization of Prevalence and Health Consequences of Uniparental Disomy in Four Million Individuals from the General Population. *Am. J. Hum. Genet.* **105**, 921–932 (2019).

10. Buniello, A. *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **47**, D1005–D1012 (2019).

11. Lappalainen, T., Scott, A. J., Brandt, M. & Hall, I. M. Genomic analysis in the age of human genome sequencing. *Cell* **177**, 70–84 (2019).

12. Hernandez, R. D. *et al.* Ultrarare variants drive substantial cis heritability of human gene expression. *Nat. Genet.* **51**, 1349–1355 (2019).

13. Long, T. *et al.* Whole-genome sequencing identifies common-to-rare variants associated with human blood metabolites. *Nat. Genet.* **49**, 568–578 (2017).

14. Kukekova, A. V. *et al.* Red fox genome assembly identifies genomic regions associated with tame and aggressive behaviours. *Nat. Ecol. Evol.* **2**, 1479–1491 (2018).

15. Young, A. I. *et al.* Relatedness disequilibrium regression estimates heritability without environmental bias. *Nat. Genet.* **50**, 1304–1310 (2018).

16. Young, A. I., Benonisdottir, S., Przeworski, M. & Kong, A. Deconstructing the sources of genotype-phenotype associations in humans. *Science* **365**, 1396–1400 (2019).

17. Lander, E. S. & Schork, N. J. Genetic dissection of complex traits. *Science* **265**, 2037–2048 (1994).

18. Campbell, D. D. *et al.* Amerind ancestry, socioeconomic status and the genetics of type 2 diabetes in a Colombian population. *PLoS ONE* **7**, e33570 (2012).

19. Toews, D. P. L., Taylor, S. A., Streby, H. M., Kramer, G. R. & Lovette, I. J. Selection on *VPS13A* linked to migration in a songbird. *Proc. Natl. Acad. Sci.* **116**, 18272–18274 (2019).

20. Rockman, M. V. & Kruglyak, L. Recombinational landscape and population genomics of *Caenorhabditis elegans*. *PLoS Genet.* **5**, e1000419 (2009).

21. Bendesky, A., Tsunozaki, M., Rockman, M. V., Kruglyak, L. & Bargmann, C. I. Catecholamine receptor polymorphisms affect decision-making in *C. elegans*. *Nature* **472**, 313–318 (2011).

22. Ding, Y., Berrocal, A., Morita, T., Longden, K. D. & Stern, D. L. Natural courtship song variation caused by an intronic retroelement in an ion channel gene. *Nature* **536**, 329–332 (2016).

23. McGrath, P. T. *et al.* Parallel evolution of domesticated *Caenorhabditis* species targets pheromone receptor genes. *Nature* **477**, 321–325 (2011).

24. Bendesky, A. *et al.* The genetic basis of parental care evolution in monogamous mice. *Nature* **544**, 434–439 (2017).

25. Flint, J. Analysis of quantitative trait loci that influence animal behavior. *J. Neurobiol.* **54**, 46–77 (2003).

26. Weber, J. N., Peterson, B. K. & Hoekstra, H. E. Discrete genetic modules are responsible for complex burrow evolution in *Peromyscus* mice. *Nature* **493**, 402–405 (2013).

27. McBride, C. S. *et al.* Evolution of mosquito preference for humans linked to an odorant receptor. *Nature* **515**, 222–227 (2014).

28. Gracheva, E. O. *et al.* Molecular basis of infrared detection by snakes. *Nature* **464**, 1006–1011 (2010).

29. Gracheva, E. O. *et al.* Ganglion-specific splicing of TRPV1 underlies infrared sensation in vampire bats. *Nature* **476**, 88–92 (2011).

30. Dierick, H. A. & Greenspan, R. J. Molecular analysis of flies selected for aggressive behavior. *Nat. Genet.* **38**, 1023–1031 (2006).

31. Albert, F. W. *et al.* Genetic architecture of tameness in a rat model of animal domestication. *Genetics* **182**, 541–554 (2009).

32. Heyne, H. O. *et al.* Genetic influences on brain gene expression in rats selected for tameness and aggression. *Genetics* **198**, 1277–1290 (2014).

33. Robinson, G. E., Fernald, R. D. & Clayton, D. F. Genes and social behavior. *Science* **322**, 896–900 (2008).

34. Wada-Katsumata, A., Silverman, J. & Schal, C. Changes in taste neurons support the emergence of an adaptive behavior in cockroaches. *Science* **340**, 972–975 (2013).

35. Eriksson, N. *et al.* A genetic variant near olfactory receptor genes influences cilantro preference. *Flavour* **52**, 1–7 (2012).

36. Baldwin, M. W. *et al.* Evolution of sweet taste perception in hummingbirds by transformation of the ancestral umami receptor. *Science* **345**, 929–933 (2014).

37. Ishikawa, Y. *et al. Ostrinia* spp. in Japan: their host plants and sex pheromones. *Entomol. Exp. Appl.* **91**, 237–244 (1999).

38. Leary, G. P. *et al.* Single mutation to a sex pheromone receptor provides adaptive specificity between closely related moth species. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 14081–14086 (2012).

39. Kiontke, K. & Fitch, D. H. A. The phylogenetic relationships of *Caenorhabditis* and other rhabditids. *WormBook* (2005) doi:10.1895/wormbook.1.11.1.

40. Herrmann, M., Mayer, W. E. & Sommer, R. J. Nematodes of the genus *Pristionchus* are closely associated with scarab beetles and the Colorado potato beetle in Western Europe. *Zoology* **109**, 96–108 (2006).

41. Hong, R. L., Witte, H. & Sommer, R. J. Natural variation in *Pristionchus pacificus* insect pheromone attraction involves the protein kinase EGL-4. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 7779–7784 (2008).

42. L'Etoile, N. D. *et al.* The cyclic GMP-dependent protein kinase EGL-4 regulates olfactory adaptation in *C. elegans*. *Neuron* **36**, 1079–1089 (2002).

43. Lee, D. *et al.* Selection and gene flow shape niche-associated variation in pheromone response. *Nat. Ecol. Evol.* **3**, 1455–1463 (2019).

44. Eigenbrod, O. *et al.* Rapid molecular evolution of pain insensitivity in multiple African rodents. *Science* **364**, 852–859 (2019).

45. Young, L. J., Wang, Z. & Insel, T. R. Neuroendocrine bases of monogamy. *Trends Neurosci.* **21**, 71–75 (1998).

46. Insel, T. R. The Challenge of Translation in Social Neuroscience: A Review of Oxytocin, Vasopressin, and Affiliative Behavior. *Neuron* **65**, 768–779 (2010).

47. Kocher, S. D. *et al.* The genetic basis of a social polymorphism in halictid bees. *Nat. Commun.* **9**, (2018).

48. Gray, J. M. *et al.* Oxygen sensation and social feeding mediated by a *C. elegans* guanylate cyclase homologue. *Nature* **430**, 317–322 (2004).

49. Hodgkin, J. & Doniach, T. Natural variation and copulatory plug formation in *Caenorhabditis elegans*. *Genetics* **146**, 149–164 (1997).

50. Macosko, E. Z. *et al.* A hub-and-spoke circuit drives pheromone attraction and social behaviour in *C. elegans*. *Nature* **458**, 1171–1175 (2009).

51. De Bono, M. & Bargmann, C. I. Natural variation in a neuropeptide Y receptor homolog modifies social behavior and food response in *C. elegans*. *Cell* **94**, 679–689 (1998).

52. McGrath, P. T. *et al.* Quantitative mapping of a digenic behavioral trait implicates globin variation in *C. elegans* sensory behaviors. *Neuron* **61**, 692–699 (2009).

53. Bendesky, A. *et al.* Long-range regulatory polymorphisms affecting a GABA receptor constitute a quantitative trait locus (QTL) for social behavior in *Caenorhabditis elegans*. *PLoS Genet.* **8**, e1003157 (2012).

54. Seeholzer, L. F., Seppo, M., Stern, D. L. & Ruta, V. Evolution of a central neural circuit underlies *Drosophila* mate preferences. *Nature* **559**, 564–569 (2018).

55. York, R. A. *et al.* Behavior-dependent cis regulation reveals genes and pathways associated with bower building in cichlid fishes. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E11081–E11090 (2018).

56. Pajic, P. *et al.* Independent amylase gene copy number bursts correlate with dietary preferences in mammals. *eLife* **8**, e44628 (2019).

57. De Wijk, R. A., Prinz, J. F., Engelen, L. & Weenen, H. The role of α-amylase in the perception of oral texture and flavour in custards. *Physiol. Behav.* **83**, 81–91 (2004).

58. Mandel, A. L., Des Gachons, C. P., Plank, K. L., Alarcon, S. & Breslin, P. A. S. Individual differences in *AMY1* gene copy number, salivary α-amylase levels, and the perception of oral starch. *PLoS ONE* **5**, e13352 (2010).

59. Clarke, T. K. *et al.* Genome-wide association study of alcohol consumption and genetic overlap with other health-related traits in UK biobank (N=112117). *Mol. Psychiatry* **22**, 1376–1384 (2017).

60. Luczak, S. E., Glatt, S. J. & Wall, T. J. Meta-analyses of *ALDH2* and *ADH1B* with alcohol dependence in asians. *Psychol. Bull.* **132**, 607–621 (2006).

61. Cornelis, M. C. *et al.* Genome-wide meta-analysis identifies regions on 7p21 (AHR) and 15q24 (CYP1A2) as determinants of habitual caffeine consumption. *PLoS Genet.* **7**, e1002033 (2011).

62. Thorgeirsson, T. E. *et al.* Sequence variants at *CHRNB3-CHRNA6* and *CYP2A6* affect smoking behavior. *Nat. Genet.* **42**, 448–453 (2010).

63. Ilardo, M. A. *et al.* Physiological and genetic adaptations to diving in Sea Nomads. *Cell* **173**, 569–580 (2018).

64. Bigham, A. *et al.* Identifying signatures of natural selection in Tibetan and Andean populations using dense genome scan data. *PLoS Genet.* **6**, e1001116 (2010).

65. Graham, A. M. & McCracken, K. G. Convergent evolution on the hypoxia-inducible factor (HIF) pathway genes *EGLN1* and *EPAS1* in high-altitude ducks. *Heredity* **122**, 819–832 (2019).

66. Schweizer, R. M. *et al.* Physiological and genomic evidence that selection on the transcription factor *Epas1* has altered cardiovascular function in high-altitude deer mice. *PLoS Genet.* **15**, e1008420 (2019).

67. Yi, X. *et al.* Sequencing of 50 Human Exomes Reveals Adaptation to High Altitude. *Science* **329**, 75–78 (2010).

68. Bersaglieri, T. *et al.* Genetic signatures of strong recent positive selection at the lactase gene. *Am. J. Hum. Genet.* **74**, 1111–1120 (2004).

69. Swallow, D. M. Genetics of Lactase Persistence and Lactose Intolerance. *Annu. Rev. Genet.* **37**, 197–219 (2003).

70. Gallego Romero, I. *et al.* Herders of Indian and European cattle share their predominant allele for lactase persistence. *Mol. Biol. Evol.* **29**, 249–260 (2011).

71. Tishkoff, S. A. *et al.* Convergent adaptation of human lactase persistence in Africa and Europe. *Nat. Genet.* **39**, 31–40 (2007).

72. Bierut, L. J. *et al. ADH1B* is associated with alcohol dependence and alcohol consumption in populations of European and African ancestry. *Mol. Psychiatry* **17**, 445–450 (2012).

73. Walters, R. K. *et al.* Transancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders. *Nat. Neurosci.* **21**, 1656–1669 (2018).

74. Liu, M. *et al.* Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nat. Genet.* **51**, 237–244 (2019).

75. Ray, R., Tyndale, R. F. & Lerman, C. Nicotine dependence pharmacogenetics: Role of genetic variation in nicotine-metabolizing enzymes. *J. Neurogenet.* **23**, 252–261 (2009).

76. Berrettini, W. *et al.* Alpha-5/alpha-3 nicotinic receptor subunit alleles increase risk for heavy smoking. *Mol. Psychiatry* **13**, 368–373 (2008).

77. Chang, A. J. & Bargmann, C. I. Hypoxia and the HIF-1 transcriptional pathway reorganize a neuronal circuit for oxygen-dependent behavior in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 7321–7326 (2008).

78. Zhen, Y., Aardema, M. L., Medina, E. M., Schumer, M. & Andolfatto, P. Parallel molecular evolution in an herbivore community. *Science* **337**, 1634–1637 (2012).

79. Karageorgi, M. *et al.* Genome editing retraces the evolution of toxin resistance in the monarch butterfly. *Nature* **574**, 409–412 (2019).

80. Taverner, A. M. *et al.* Adaptive substitutions underlying cardiac glycoside insensitivity in insects exhibit epistasis in vivo. *eLife* **8**, e48224 (2019).

81. Louis, J. & David, J. R. Ecological specialization in the *Drosophila melanogaster* species subgroup: A case study of *D. sechellia*. *Acta Oecol* **7**, 215–229 (1985).

82. Yassin, A. *et al.* Recurrent specialization on a toxic fruit in an island *Drosophila* population. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 4771–4776 (2016).

83. Prieto-Godino, L. L. *et al.* Olfactory receptor pseudo-pseudogenes. *Nature* **539**, 93–97 (2016).

84. Prieto-Godino, L. L. *et al.* Evolution of acid-sensing olfactory circuits in Drosophilids. *Neuron* **93**, 661–676 (2017).

85. Auer, T. O. *et al.* Olfactory receptor and circuit evolution promote host specialization. *Nature* **579**, 402–408 (2020).

86. Lassance, J. M. *et al.* Allelic variation in a fatty-acyl reductase gene causes divergence in moth sex pheromones. *Nature* **466**, 486–489 (2010).

87. Noble, L. M. *et al.* Natural variation in *plep-1* causes male-male copulatory behavior in C. elegans. *Curr. Biol.* **25**, 2730–2737 (2015).

88. Blankers, T., Oh, K. P. & Shaw, K. L. Parallel genomic architecture underlies repeated sexual signal divergence in Hawaiian Laupala crickets. *Proc. Biol. Sci.* **286**, 20191479 (2019).

89. Bendesky, A. & Bargmann, C. I. Genetic contributions to behavioural diversity at the gene–environment interface. *Nat. Rev. Genet.* **12**, 809–820 (2011).

90. Stern, D. L. Identification of loci that cause phenotypic variation in diverse species with the reciprocal hemizygosity test. *Trends Genet.* **30**, 547–554 (2014).

91. Jiang, P. *et al.* Major taste loss in carnivorous mammals. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 4956–4961 (2012).

92. Li, X. *et al.* Pseudogenization of a sweet-receptor gene accounts for cats' indifference toward sugar. *PLoS Genet.* **1**, e3 (2005).

93. Torgerson, D. G. *et al.* Evolutionary processes acting on candidate cis-regulatory regions in humans inferred from patterns of polymorphism and divergence. *PLoS Genet.* **5**, e1000592 (2009).

94. Feng, P., Zheng, J., Rossiter, S. J., Wang, D. & Zhao, H. Massive losses of taste receptor genes in toothed and baleen whales. *Genome Biol. Evol.* **6**, 1254–1265 (2014).

95. Hu, Y. *et al.* Comparative genomics reveals convergent evolution between the bamboo-eating giant and red pandas. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 1081–1086 (2017).

96. Li, R. *et al.* The sequence and de novo assembly of the giant panda genome. *Nature* **463**, 311–317 (2010).

97. Shi, P. & Zhang, J. Contrasting modes of evolution between vertebrate sweet/umami receptor genes and bitter receptor genes. *Mol. Biol. Evol.* **23**, 292–300 (2006).

98. Zhao, H., Li, J. & Zhang, J. Molecular evidence for the loss of three basic tastes in penguins. *Curr. Biol.* **25**, R141–R142 (2015).

99. Zhao, H., Xu, D., Zhang, S. & Zhang, J. Genomic and genetic evidence for the loss of umami taste in bats. *Genome Biol. Evol.* **4**, 73–79 (2012).

100. Zhao, H. *et al.* Evolution of the sweet taste receptor gene Tas1r2 in bats. *Mol. Biol. Evol.* **27**, 2642–2650 (2010).

101. Itoigawa, A., Hayakawa, T., Suzuki-Hashido, N. & Imai, H. A natural point mutation in the bitter taste receptor TAS2R16 causes inverse agonism of arbutin in lemur gustation. *Proc. R. Soc. B Biol. Sci.* **286**, (2019).

102. Wooding, S. *et al.* Independent evolution of bitter-taste sensitivity in humans and chimpanzees. *Nature* **440**, 930–934 (2006).

103. Wang, J. *et al.* A Y-like social chromosome causes alternative colony organization in fire ants. *Nature* **493**, 664–668 (2013).

104. Ross, K. G. & Keller, L. Genetic control of social organization in an ant. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 14232–14237 (1998).

105. Wallberg, A., Schoning, C., Webster, M. T. & Hasselmann, M. Two extended haplotype blocks are associated with adaptation to high altitude habitats in East African honey bees. *PloS Genet.* **13**, e1006792 (2017).

106. Thomas, J. W. *et al.* The chromosomal polymorphism linked to variation in social behavior in the white-throated sparrow (*Zonotrichia albicollis*) is a complex rearrangement and suppressor of recombination. *Genetics* **179**, 1455–1468 (2008).

107. Thorneycroft, H. B. A cytogenetic study of the white-throated sparrow, *Zonotrichia albicollis* (Gmelin). *Int. J. Org. Evol.* **29**, 611–621 (1975).

108. Küpper, C. *et al.* A supergene determines highly divergent male reproductive morphs in the ruff. *Nat. Genet.* **48**, 79–83 (2015).

109. Lamichhaney, S. *et al.* Structural genomic changes underlie alternative reproductive strategies in the ruff (*Philomachus pugnax*). *Nat. Genet.* **48**, 84–88 (2015).

110. Keller, L. & Parker, J. D. Behavioral genetics: A gene for supersociality. *Curr. Biol.* **12**, 180–181 (2002).

111. Lyon, M. F. Transmission Ratio Distortion in Mice. *Annu. Rev. Genet.* **37**, 393–408 (2003).

112. Runge, J. N. & Lindholm, A. K. Carrying a selfish genetic element predicts increased migration propensity in free-living wild house mice. *Proc. R. Soc. B Biol. Sci.* **285**, 20181333 (2018).

113. Silver, L. Mouse *t* Haplotypes. *Annu. Rev. Genet.* **19**, 179–208 (1985).

114. Merritt, J. R. *et al.* A supergene-linked estrogen receptor drives alternative phenotypes in a polymorphic songbird. *Proc. Natl. Acad. Sci.* **117**, 21673–21680 (2020).

115. Horton, B. M. *et al.* Estrogen receptor a polymorphism in a species with alternative behavioral phenotypes. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 1443–1448 (2014).

116. Zinzow-Kramer, W. M. *et al.* Genes located in a chromosomal inversion are correlated with territorial song in white-throated sparrows. *Genes Brain Behav.* **14**, 641–654 (2015).

117. Niimura, Y. & Nei, M. Extensive gains and losses of olfactory receptor genes in mammalian evolution. *PLoS ONE* **2**, e708 (2007).

118. Young, J. M., Massa, H. F., Hsu, L. & Trask, B. J. Extreme variability among mammalian V1R gene families. *Genome Res.* **20**, 10–18 (2010).

119. Khan, I. *et al.* Olfactory receptor subgenomes linked with broad ecological adaptations in Sauropsida. *Mol. Biol. Evol.* **32**, 2832–2843 (2015).

120. Hughes, G. M. *et al.* The birth and death of olfactory receptor gene families in mammalian niche adaptation. *Mol. Biol. Evol.* **35**, 1390–1406 (2018).

121. Nei, M., Niimura, Y. & Nozawa, M. The evolution of animal chemosensory receptor gene repertoires: Roles of chance and necessity. *Nat. Rev. Genet.* **9**, 951–963 (2008).

122. Gilad, Y., Bustamante, C. D., Lancet, D. & Pääbo, S. Natural selection on the olfactory receptor gene family in humans and chimpanzees. *Am. J. Hum. Genet.* **73**, 489–501 (2003).

123. Kishida, T., Kubota, S., Shirayama, Y. & Fukami, H. The olfactory receptor gene repertoires in secondary-adapted marine vertebrates: Evidence for reduction of the functional proportions in cetaceans. *Biol. Lett.* **3**, 428–430 (2007).

124. Kawamura, S. Color vision diversity and significance in primates inferred from genetic and field studies. *Genes Genomics* **38**, 779–791 (2016).

125. Lin, J. J., Wang, F. Y., Li, W. H. & Wang, T. Y. The rises and falls of opsin genes in 59 ray-finned fish genomes and their implications for environmental adaptation. *Sci. Rep.* **7**, 15568 (2017).

126. Musilova, Z. *et al.* Vision using multiple distinct rod opsins in deep-sea fishes. *Science* **364**, 588–592 (2019).

127. Young, J. M. & Trask, B. J. V2R gene families degenerated in primates, dog and cow, but expanded in opossum. *Trends Genet.* **23**, 212–215 (2007).

128. Zhang, J. & Webb, D. M. Evolutionary deterioration of the vomeronasal pheromone transduction pathway in catarrhine primates. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 8337–8341 (2003).

129. Kimchi, T., Xu, J. & Dulac, C. A functional circuit underlying male sexual behaviour in the female mouse brain. *Nature* **448**, 1009–1014 (2007).

130. Wu, Z., Autry, A. E., Bergan, J. F., Watabe-Uchida, M. & Dulac, C. G. Galanin neurons in the medial preoptic area govern parental behaviour. *Nature* **509**, 325–330 (2014).

131. Nikaido, M. *et al.* Lineage-specific expansion of vomeronasal type 2 receptor-like (OlfC) genes in cichlids may contribute to diversification of amino acid detection systems. *Genome Biol. Evol.* **5**, 711–722 (2013).

132. Burger, J., Kirchner, M., Bramanti, B., Haak, W. & Thomas, M. G. Absence of the lactase-persistence-associated allele in early Neolithic Europeans. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 3736–3741 (2007).

133. Staiger, E. A. *et al.* The evolutionary history of the *DMRT3* 'Gait keeper' haplotype. *Anim. Genet.* **48**, 551–559 (2017).

134. Andersson, L. S. *et al.* Mutations in *DMRT3* affect locomotion in horses and spinal circuit function in mice. *Nature* **488**, 642–646 (2012).

135. Paaby, A. B. & Rockman, M. V. Cryptic genetic variation: Evolution's hidden substrate. *Nat. Rev. Genet.* **15**, 247–258 (2014).

136. Colosimo, P. F. *et al.* Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *Science* **307**, 1928–1933 (2005).

137. Greenwood, A. K., Wark, A. R., Yoshida, K. & Peichel, C. L. Genetic and neural modularity underlie the evolution of schooling behavior in threespine sticklebacks. *Curr. Biol.* **23**, 1884–1888 (2013).

138. Greenwood, A. K., Mills, M. G., Wark, A. R., Archambeault, S. L. & Peichel, C. L. Evolution of schooling behavior in threespine sticklebacks is shaped by the *Eda* gene. *Genetics* **203**, 677–681 (2016).

139. Schrider, D. R. & Kern, A. D. Soft sweeps are the dominant mode of adaptation in the human genome. *Mol. Biol. Evol.* **34**, 1863–1877 (2017).

140. Higham, T. *et al.* The timing and spatiotemporal patterning of Neanderthal disappearance. *Nature* **512**, 306 (2014).

141. Huerta-Sánchez, E. *et al.* Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature* **512**, 194–197 (2014).

142. Dannemann, M. & Kelso, J. The contribution of Neanderthals to phenotypic variation in modern humans. *Am. J. Hum. Genet.* **101**, 578–589 (2017).

143. Blomberg, S. P., Garland, T. & Ives, A. R. Testing for phylogenetic signal in comparative data: Behavioral traits are more labile. *Evolution* **57**, 717–745 (2003).

144. Kamilar, J. M. & Cooper, N. Phylogenetic signal in primate behaviour, ecology and life history. *Philos. Trans. R. Soc. B Biol. Sci.* **368**, 20120341 (2013).

145. Robson, S. K. A., Kohout, R. J., Beckenbach, A. T. & Moreau, C. S. Evolutionary transitions of complex labile traits: Silk weaving and arboreal nesting in *Polyrhachis* ants. *Behav. Ecol. Sociobiol.* **69**, 449–458 (2015).

146. Comeron, J. M. Background selection as baseline for nucleotide variation across the *Drosophila* genome. *PLoS Genet.* **10**, e1004434 (2014).

147. Croze, M. *et al.* A genome-wide scan for genes under balancing selection in *Drosophila melanogaster*. *BMC Evol. Biol.* **17**, 1–12 (2017).

148. Stacey, D. *et al.* ProGeM: A framework for the prioritization of candidate causal genes at molecular quantitative trait loci. *Nucleic Acids Res.* **47**, 1–13 (2019).

149. de la Torre-Ubieta, L. *et al.* The Dynamic Landscape of Open Chromatin during Human Cortical Neurogenesis. *Cell* **172**, 289–304 (2018).

150. Won, H. *et al.* Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature* **538**, 523–527 (2016).

151. Martin, A. R. *et al.* Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* **51**, 584–591 (2019).

152. Knoedler, J. R. *et al.* A functional cellular framework for sex and estrous cycle-dependent gene expression and behavior. *Cell* **185**, 654-671.e22 (2022).

153. Kim, D.-W. *et al.* Multimodal analysis of cell types in a hypothalamic node controlling social behavior. *Cell* **179**, 713-728.e17 (2019).

154. Moffitt, J. R. *et al.* Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. *Science* **362**, (2018).

155. Tosches, M. A. *et al.* Evolution of pallium, hippocampus, and cortical cell types revealed by single-cell transcriptomics in reptiles. *Science* **360**, 881–888 (2018).

156. Bakken, T. E. *et al.* Single-cell and single-nucleus RNA-seq uncovers shared and distinct axes of variation in dorsal LGN neurons in mice, non-human primates, and humans. *eLife* **10**, e64875 (2021).

157. Krienen, F. M. *et al.* Innovations present in the primate interneuron repertoire. *Nature* **586**, 262–269 (2020).

158. Zeng, H. *et al.* Large-scale cellular-resolution gene profiling in human neocortex reveals species-specific molecular signatures. *Cell* **149**, 483–496 (2012).

159. Hain, D. *et al.* Molecular diversity and evolution of neuron types in the amniote brain. *Science* **377**, eabp8202 (2022).

160. Woych, J. *et al.* Cell-type profiling in salamanders identifies innovations in vertebrate forebrain evolution. *Science* **377**, eabp9186 (2022).

161. Bakken, T. E. *et al.* Comparative cellular analysis of motor cortex in human, marmoset and mouse. *Nature* **598**, 111–119 (2021).

162. Callaway, E. M. *et al.* A multimodal cell census and atlas of the mammalian primary motor cortex. *Nature* **598**, 86–102 (2021).

163. Brückner, A. *et al.* Evolutionary assembly of cooperating cell types in an animal chemical defense system. *Cell* **184**, 6138-6156.e28 (2021).

164. Arendt, D. *et al.* The origin and evolution of cell types. *Nat. Rev. Genet.* **17**, 744–757 (2016).

165. Upham, N. S., Esselstyn, J. A. & Jetz, W. Inferring the mammal tree: Species-level sets of phylogenies for questions in ecology, evolution, and conservation. *PLOS Biol.* **17**, e3000494 (2019).

166. Khadraoui, M., Merritt, J. R., Hoekstra, H. E. & Bendesky, A. Post-mating parental behavior trajectories differ across four species of deer mice. *PloS One* **17**, e0276052 (2022).

167. Dewsbury, D. A. Aggression, copulation, and differential reproduction of deer mice (*Peromyscus maniculatus*) in a semi-natural enclosure. *Behaviour* **91**, 1–23 (1984).

168. Dewsbury, D. A. & Lovecky, D. V. Copulatory behavior of old-field mice (*Peromyscus polionotus*) from different natural populations. *Behav. Genet.* **4**, 347–55 (1974).

169. Birdsall, D. A. & Nash, D. Occurrence of successful multiple insemination of females in natural populations of deer mice (*Peromyscus maniculatus*). *Evolution* **27**, 106–110 (1973).

170. Foltz, D. W. Genetic evidence for long-term monogamy in a small rodent, *Peromyscus polionotus*. *Am. Nat.* **117**, 665–675 (1981).

171. Melmed, S., Koenig, R., Rosen, C., Auchus, R. & Goldfine, A. *Williams Textbook of Endocrinology, 14th edition*. (Elsevier Health Sciences, 2019).

172. Barresi, M. J. F. & Gilbert, S. F. *Developmental Biology, 13th edition*. (Oxford University Press, 2023).

173. Andreatta, M. & Carmona, S. J. STACAS: Sub-type anchor correction for alignment in Seurat to integrate single-cell RNA-seq data. *Bioinformatics* **37**, 882–884 (2021).

174. Stuart, T. & Satija, R. Integrative single-cell analysis. *Nat. Rev. Genet.* **20**, 257–272 (2019).

175. Dolfi, B. *et al.* Unravelling the sex-specific diversity and functions of adrenal gland macrophages. *Cell Rep.* **39**, 110949 (2022).

176. Hanemaaijer, E. S. *et al.* Single-cell atlas of developing murine adrenal gland reveals relation of Schwann cell precursor signature to neuroblastoma phenotype. *Proc. Natl. Acad. Sci. U. S. A.* **118**, (2021).

177. Huang, L. *et al.* Single-cell transcriptomes reveal characteristic features of cell types within the human adrenal microenvironment. *J. Cell. Physiol.* **236**, 7308–7321 (2021).

178. Lai, S. *et al.* Mapping a mammalian adult adrenal gland hierarchy across species by microwell-seq. *Cell Regen.* **9**, 1–12 (2020).

179. Mitani, F. Functional zonation of the rat adrenal cortex: the development and maintenance. *Proc. Jpn. Acad. Ser. B Phys. Biol. Sci.* **90**, 163–83 (2014).

180. Pihlajoki, M., Dörner, J., Cochran, R. S., Heikinheimo, M. & Wilson, D. B. Adrenocortical zonation, renewal, and remodeling. *Front. Endocrinol.* **6**, 1–14 (2015).

181. Jeff Huang, C. C. & Kang, Y. The transient cortical zone in the adrenal gland: The mystery of the adrenal X-zone. *J. Endocrinol.* **241**, R51–R63 (2019).

182. Hershkovitz, L., Beuschlein, F., Klammer, S., Krup, M. & Weinstein, Y. Adrenal 20α-hydroxysteroid dehydrogenase in the mouse catabolizes progesterone and 11-deoxycorticosterone and is restricted to the X-zone. *Endocrinology* **148**, 976–988 (2007).

183. Howard-Miller, E. A transitory zone in the adrenal cortex which shows age and sex relationships. *Am. J. Anat.* **40**, 251–293 (1927).

184. Tamura, Y. Structural changes in the suprarenal gland of the mouse during pregnancy. *J. Exp. Biol.* **4**, 81–92 (1926).

185. Mao, J., Duan, R. W., Zhong, L., Gibori, G. & Azhar, S. Expression, purification and characterization of the rat luteal 20 alpha-hydroxysteroid dehydrogenase. *Endocrinology* **138**, 182–190 (1997).

186. Pratt-Hyatt, M., Lickteig, A. J. & Klaassen, C. D. Tissue distribution, ontogeny, and chemical induction of aldo-keto reductases in mices. *Drug Metab. Dispos.* **41**, 1480–1487 (2013).

187. Penning, T. M. The aldo-keto reductases (AKRs): Overview. *Chem. Biol. Interact.* **234**, 236–246 (2015).

188. Frye, C. A. & Duncan, J. E. Progesterone metabolites, effective at the $GABA_A$ receptor complex, attenuate pain sensitivity in rats. *Brain Res.* **643**, 194–203 (1994).

189. Numan, M. & Insel, T. R. *The Neurobiology of Parental Behavior*. (Springer New York, 2006).

190. Ogle, T. F. & Beyer, B. K. Steroid-binding specificity of the progesterone receptor from rat placenta. *J. Steroid Biochem.* **16**, 147–150 (1982).

191. Nowak, F. V. Distribution and metabolism of 20α-hydroxylated progestins in the female rat. *J. Steroid Biochem. Mol. Biol.* **80**, 469–479 (2002).

192. Nowak, F. V., Nuti, K. M. & Karavolas, H. J. Quantitative changes in the metabolism of 20α-hydroxy-4-pregnen-3-one by rat hypothalamus and pituitary during proestrus. *Steroids* **28**, 509–520 (1976).

193. Khanna, M., Qin, K.-N. & Cheng, K.-C. Distribution of 3α-hydroxysteroid dehydrogenase in rat brain and molecular cloning of multiple cDNAs encoding structurally related proteins in humans. *J. Steroid Biochem. Mol. Biol.* **53**, 41–46 (1995).

194. Penning, T. M. *et al.* Human 3alpha-hydroxysteroid dehydrogenase isoforms (AKR1C1-AKR1C4) of the aldo-keto reductase superfamily: functional plasticity and tissue distribution reveals roles in the inactivation and formation of male and female sex hormones. *Biochem. J.* **351**, 67–77 (2000).

195. Agís-Balboa, R. C. *et al.* Characterization of brain neurons that express enzymes mediating neurosteroid biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 14602–14607 (2006).

196. Russell, D. W. & Wilson, J. D. Steroid 5 alpha-reductase: two genes/two enzymes. *Annu. Rev. Biochem.* **63**, 25–61 (1994).

197. MacKenzie, G. & Maguire, J. Neurosteroids and GABAergic signaling in health and disease. *Biomol. Concepts* **4**, 29–42 (2013).

198. Stell, B. M., Brickley, S. G., Tang, C. Y., Farrant, M. & Mody*, I. Neuroactive steroids reduce neuronal excitability by selectively enhancing tonic inhibition mediated by δ subunit-containing GABAA receptors. *Proc. Natl. Acad. Sci.* **100**, 14439–14444 (2003).

199. Maguire, J. & Mody, I. $GABA_AR$ plasticity during pregnancy: relevance to postpartum depression. *Neuron* **59**, 207–213 (2008).

200. Maguire, J., Ferando, I., Simonsen, C. & Mody, I. Excitability Changes Related to $GABA_A$ Receptor Plasticity during Pregnancy. *J. Neurosci.* **29**, 9592–9601 (2009).

201. Rudolph, S. *et al.* Cerebellum-Specific Deletion of the GABAA Receptor δ Subunit Leads to Sex-Specific Disruption of Behavior. *Cell Rep.* **33**, 108338 (2020).

202. Priestley, C. M., Williamson, E. M., Wafford, K. A. & Sattelle, D. B. Thymol, a constituent of thyme essential oil, is a positive allosteric modulator of human $GABA_A$ receptors and a homo-oligomeric GABA receptor from *Drosophila melanogaster*. *Br. J. Pharmacol.* **140**, 1363–1372 (2003).

203. Belelli, D. & Gee, K. W. 5α-pregnan-3α,20α-diol behaves like a partial agonist in the modulation of GABA-stimulated chloride ion uptake by synaptoneurosomes. *Eur. J. Pharmacol.* **167**, 173–176 (1989).

204. MacKenzie, G. & Maguire, J. Neurosteroids and GABAergic signaling in health and disease. *Biomol. Concepts* **4**, 29–42 (2013).

205. Farrant, M. & Nusser, Z. Variations on an inhibitory theme: phasic and tonic activation of GABA(A) receptors. *Nat. Rev. Neurosci.* **6**, 215–229 (2005).

206. Pallares, L. F., Picard, S. & Ayroles, J. F. TM3'seq: a tagmentation-mediated 3' sequencing approach for improving scalability of RNA-seq experiments. *G3 Genes Genomes Genet.* **10**, 143–150 (2020).

207. Brunton, P. J., Russell, J. A. & Douglas, A. J. Adaptive responses of the maternal hypothalamic-pituitary-adrenal axis during pregnancy and lactation. *J. Neuroendocrinol.* **20**, 764–776 (2008).

208. Duthie, L. & Reynolds, R. M. Changes in the maternal hypothalamic-pituitary-adrenal axis in pregnancy and postpartum: influences on maternal and fetal outcomes. *Neuroendocrinology* **98**, 106–115 (2013).

209. Atkinson, H. C. & Waddell, B. J. Circadian variation in basal plasma corticosterone and adrenocorticotropin in the rat: sexual dimorphism and changes across the estrous cycle. *Endocrinology* **138**, 3842–3848 (1997).

210. Bailey, K. J. Diurnal progesterone rhythms in the female mouse. *J. Endocrinol.* **112**, 15–21 (1987).

211. Degen, M., Scherberich, A. & Tucker, R. P. Tenascin-W: discovery, evolution, and future prospects. *Front. Immunol.* **11**, 1–5 (2021).

212. Sengupta, P. & Bargmann, C. I. Cell fate specification and differentiation in the nervous system of *Caenorhabditis elegans*. *Dev. Genet.* **18**, 73–80 (1996).

213. Hobert, O. & Kratsios, P. Neuronal identity control by terminal selectors in worms, flies, and chordates. *Curr. Opin. Neurobiol.* **56**, 97–105 (2019).

214. Dasen, J. S. & Jessell, T. M. Hox networks and the origins of motor neuron diversity. *Curr. Top. Dev. Biol.* **88**, 169–200 (2009).

215. Lynch, V. J., May, G. & Wagner, G. P. Regulatory evolution through divergence of a phosphoswitch in the transcription factor CEBPB. *Nature* **480**, 383–386 (2011).

216. Lynch, V. J. *et al.* Adaptive changes in the transcription factor HoxA-11 are essential for the evolution of pregnancy in mammals. *Proc. Natl. Acad. Sci.* **105**, 14928–14933 (2008).

217. Chiquet-Ehrismann, R., Orend, G., Chiquet, M., Tucker, R. P. & Midwood, K. S. Tenascins in stem cell niches. *Matrix Biol.* **37**, 112–123 (2014).

218. Pesheva, P., Gloor, S., Schachner, M. & Probstmeier, R. Tenascin-R Is an intrinsic autocrine factor for oligodendrocyte differentiation and promotes cell adhesion by a sulfatide-mediated mechanism. *J. Neurosci.* **17**, 4642–4651 (1997).

219. Kimura, H., Akiyama, H., Nakamura, T. & Crombrugghe, B. de. Tenascin-W inhibits proliferation and differentiation of preosteoblasts during endochondral bone formation. *Biochem. Biophys. Res. Commun.* **356**, 935–941 (2007).

220. Tucker, R. P. & Degen, M. The expression and possible functions of tenascin-W during development and disease. *Front. Cell Dev. Biol.* **7**, 1–10 (2019).

221. Czopka, T., Von Holst, A., Schmidt, G., Ffrench-Constant, C. & Faissner, A. Tenascin C and tenascin R similarly prevent the formation of myelin membranes in a RhoA-dependent

manner, but antagonistically regulate the expression of myelin basic protein via a separate pathway. *Glia* **57**, 1790–1801 (2009).

222. Abdel-Khalik, J., Björklund, E. & Hansen, M. Simultaneous determination of endogenous steroid hormones in human and animal plasma and serum by liquid or gas chromatography coupled to tandem mass spectrometry. *J. Chromatogr. B Analyt. Technol. Biomed. Life. Sci.* **928**, 58–77 (2013).

223. Florensa, E., Harrison, R., Johnson, M. & Youssefnejadian, E. Plasma 20 alpha-dihydroprogesterone, progesterone and 17-hydroxyprogesterone in normal human pregnancy. *Acta Endocrinol. (Copenh.)* **86**, 634–640 (1977).

224. Jensen, C. C. Quantitative determination of urinary pregnanediol and allopregnanediol for clinical use. *Eur. J. Endocrinol.* **18**, 281–287 (1955).

225. Choi, H. M. T. *et al.* Third-generation in situ hybridization chain reaction: Multiplexed, quantitative, sensitive, versatile, robust. *Development* **145**, dev165753 (2018).

226. Kuehn, E. *et al.* Segment number threshold determines juvenile onset of germline cluster expansion in *Platynereis dumerilii*. *J. Exp. Zoolog. B Mol. Dev. Evol.* **338**, 225–240 (2022).

227. Renier, N. *et al.* IDISCO: A simple, rapid method to immunolabel large tissue samples for volume imaging. *Cell* **159**, 896–910 (2014).

228. Picelli, S. *et al.* Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res.* **24**, 2033–2040 (2014).

229. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

230. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

231. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).

232. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

233. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).

234. Turro, E. *et al.* Haplotype and isoform specific expression estimation using multi-mapping RNA-seq reads. *Genome Biol.* **12**, R13 (2011).

235. Kirilenko, B. M. *et al.* Integrating gene annotation with orthology inference at scale. *Science* **380**, eabn3107 (2023).

236. Detlefsen, A. J., Wangtrakuldee, P. & Penning, T. M. Characterization of the major single nucleotide polymorphic variants of aldo-keto reductase 1C3 (type 5 17β-hydroxysteroid dehydrogenase). *J. Steroid Biochem. Mol. Biol.* **221**, 106121 (2022).

237. Tobiansky, D. J. *et al.* Testosterone and Corticosterone in the Mesocorticolimbic System of Male Rats: Effects of Gonadectomy and Caloric Restriction. *Endocrinology* **159**, 450–464 (2018).

238. Bright, D. P. & Smart, T. G. Methods for recording and measuring tonic $GABA_A$ receptor-mediated inhibition. *Front. Neural Circuits* **7**, 193 (2013).

239. Broman, K. W., Wu, H., Sen, Ś. & Churchill, G. A. R/qtl: QTL mapping in experimental crosses. *Bioinformatics* **19**, 889–890 (2003).

240. Lapiedra, O., Schoener, T. W., Leal, M., Losos, J. B. & Kolbe, J. J. Predator-driven natural selection on risk-taking behavior in anole lizards. *Science* **1020**, 1017–1020 (2018).

241. Bedford, N. L. & Hoekstra, H. E. Peromyscus mice as a model for studying natural variation. *eLife* **4**, e06813 (2015).

242. Calhoon, G. G. & Tye, K. M. Resolving the neural circuits of anxiety. *Nat. Neurosci.* **18**, 1394–1404 (2015).

243. Lezak, K. R., Missig, G. & Carlezon Jr, W. A. Behavioral methods to study anxiety in rodents. *Dialogues Clin. Neurosci.* **19**, 181–191 (2017).

244. Pellow, S., Chopin, P., File, S. E. & Briley, M. Validation of open : closed arm entries in an elevated plus-maze as a measure of anxiety in the rat. *J. Neurosci. Methods* **14**, 149–167 (1985).

245. Baier, F. *et al.* The neural basis of defensive behaviour evolution in Peromyscus mice. *bioRxiv* 2023.07.04.547734 (2023) doi:10.1101/2023.07.04.547734.

246. Yalcin, B. *et al.* Genetic dissection of a behavioral quantitative trait locus shows that Rgs2 modulates anxiety in mice. *Nat. Genet.* **36**, 1197–1202 (2004).

247. Goodson, M. *et al.* Cofilin-1: A Modulator of Anxiety in Mice. *PLoS Genet.* **8**, (2012).

248. Parker, C. C., Sokoloff, G., Leung, E., Kirkpatrick, S. L. & Palmer, A. A. A large QTL for fear and anxiety mapped using an F2 cross can be dissected into multiple smaller QTLs. *Genes Brain Behav.* **12**, 714–722 (2013).

249. Nadeau, J. H., Singer, J. B., Matin, A. & Lander, E. S. Analysing complex genetic traits with chromosome substitution strains. *Nat. Genet.* **24**, 221–225 (2000).

250. Shao, H. *et al.* Genetic architecture of complex traits: Large phenotypic effects and pervasive epistasis. *Proc. Natl. Acad. Sci.* **105**, 19910–19914 (2008).

251. Corbett-Detig, R. & Nielsen, R. *A Hidden Markov Model Approach for Simultaneously Estimating Local Ancestry and Admixture Time Using Next Generation Sequence Data in Samples of Arbitrary Ploidy*. *PLoS Genetics* vol. 13 (2017).

252. Andreas F. Kautt *et al.* Evolution of gene expression across brain regions in behaviorally divergent deer mice. *bioRxiv* 2023.09.28.559844 (2023) doi:10.1101/2023.09.28.559844.

253. Hoshina, N. *et al.* Protocadherin 17 regulates presynaptic assembly in topographic corticobasal Ganglia circuits. *Neuron* **78**, (2013).

254. Chang, H. *et al.* The protocadherin 17 gene affects cognition, personality, amygdala structure and function, synapse development and risk of major mood disorders. *Mol. Psychiatry* **23**, 400–412 (2018).

255. Hyde, C. L. *et al.* Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. *Nat. Genet.* **48**, 1031–1036 (2016).

256. Wray, N. R. *et al.* Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* **50**, 668–681 (2018).

257. Jansen, P. R. *et al.* Genome-wide analysis of insomnia in 1,331,010 individuals identifies new risk loci and functional pathways. *Nat. Genet.* **51**, 394–403 (2019).

258. Watanabe, K. *et al.* Genome-wide meta-analysis of insomnia prioritizes genes associated with metabolic and psychiatric pathways. *Nat. Genet.* **54**, 1125–1132 (2022).

259. Groza, T. *et al.* The International Mouse Phenotyping Consortium: comprehensive knockout phenotyping underpinning the study of human disease. *Nucleic Acids Res.* **51**, D1038–D1045 (2023).

260. Stark, J. E. *et al.* Juvenile OLFM4-null mice are protected from sepsis. *Am. J. Physiol. - Ren. Physiol.* **318**, F809–F816 (2020).

261. Kulkarni, N. H., Karavanich, C. A., Atchley, W. R. & Anholt, R. R. Characterization and differential expression of a human gene family of olfactomedin-related proteins. *Genet. Res.* **76**, 41–50 (2000).

262. Anholt, R. R. H. Olfactomedin proteins: Central players in development and disease. *Front. Cell Dev. Biol.* **2**, 1–10 (2014).

263. Liu, R. *et al.* Depletion of OLFM4 gene inhibits cell growth and increases sensitization to hydrogen peroxide and tumor necrosis factor-alpha induced-apoptosis in gastric cancer cells. *J. Biomed. Sci.* **19**, 38 (2012).

264. Eshed, Y. *et al.* Gliomedin mediates Schwann cell-axon interaction and the molecular assembly of the nodes of Ranvier. *Neuron* **47**, 215–229 (2005).

265. Lee, J.-A., Anholt, R. R. H. & Cole, G. J. Olfactomedin-2 mediates development of the anterior central nervous system and head structures in zebrafish. *Mech. Dev.* **125**, 167–181 (2008).

266. Nakaya, N. *et al.* Deletion in the N-terminal half of olfactomedin 1 modifies its interaction with synaptic proteins and causes brain dystrophy and abnormal behavior in mice. *Exp. Neurol.* **250**, 205–218 (2013).

267. Sultana, A. *et al.* Deletion of olfactomedin 2 induces changes in the AMPA receptor complex and impairs visual, olfactory, and motor functions in mice. *Exp. Neurol.* **261**, 802–811 (2014).

268. Li, Q., Liu, A., Gu, X. & Su, Z. Olfactomedin domain-containing proteins: evolution, functional divergence, expression patterns and damaging SNPs. *Mol. Genet. Genomics* **294**, 875–885 (2019).

269. Dall'Aglio, L., Lewis, C. M. & Pain, O. Delineating the Genetic Component of Gene Expression in Major Depression. *Biol. Psychiatry* **89**, 627–636 (2021).

270. Lein, E. S. *et al.* Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445**, 168–176 (2007).

271. Steuernagel, L. *et al.* HypoMap—a unified single-cell gene expression atlas of the murine hypothalamus. *Nat. Metab.* **4**, 1402–1419 (2022).

272. Nicod, J. *et al.* Genome-wide association of multiple complex traits in outbred mice by ultra-low-coverage sequencing. *Nat. Genet.* **48**, 912–918 (2016).

273. Yazdani, N. *et al.* Hnrnph1 Is A Quantitative Trait Gene for Methamphetamine Sensitivity. *PLoS Genet.* **11**, (2015).

274. Stylianou, I. M. *et al.* Genetic complexity of an obesity QTL (Fob3) revealed by detailed genetic mapping. *Mamm. Genome* **15**, 472–481 (2004).

275. Bryant, C. D., Kole, L. A., Guido, M. A., Sokoloff, G. & Palmer, A. A. Congenic dissection of a major QTL for methamphetamine sensitivity implicates epistasis. *Genes Brain Behav.* **11**, 623–632 (2012).

276. Wooldridge, T. B. *et al.* An enhancer of Agouti contributes to parallel evolution of cryptically colored beach mice. *Proc. Natl. Acad. Sci. U. S. A.* **119**, e2202862119 (2022).

277. Hager, E. R. *et al.* A chromosomal inversion contributes to divergence in multiple traits between deer mouse ecotypes. *Science* **377**, 399–405 (2022).

278. Hendaoui, I. *et al.* Tenascin-C is required for normal Wnt/β-catenin signaling in the whisker follicle stem cell niche. *Matrix Biol. J. Int. Soc. Matrix Biol.* **40**, 46–53 (2014).

279. Imhof, T. *et al.* Pivotal Role of Tenascin-W (-N) in Postnatal Incisor Growth and Periodontal Ligament Remodeling. *Front. Immunol.* **11**, 608223 (2020).

280. Scherberich, A. *et al.* Tenascin-W is found in malignant mammary tumors, promotes alpha8 integrin-dependent motility and requires p38MAPK activity for BMP-2 and TNF-alpha induced expression in vitro. *Oncogene* **24**, 1525–1532 (2005).

281. Harburger, D. S. & Calderwood, D. A. Integrin signalling at a glance. *J. Cell Sci.* **122**, 159–163 (2009).

282. Schuster, D. J. *et al.* Biodistribution of adeno-associated virus serotype 9 (AAV9) vector after intrathecal and intravenous delivery in mouse. *Front. Neuroanat.* **8**, 42 (2014).

283. Radford-Smith, D. E. & Anthony, D. C. Prebiotic and Probiotic Modulation of the Microbiota–Gut–Brain Axis in Depression. *Nutrients* **15**, 1880 (2023).

284. Ishida, M. *et al.* Reproductive phenotypes in mice with targeted disruption of the 20α-hydroxysteroid dehydrogenase gene. *J. Reprod. Dev.* **53**, 499–508 (2007).

285. Kaludjerovic, J. & Ward, W. E. The interplay between estrogen and fetal adrenal cortex. *J. Nutr. Metab.* **2012**, (2012).

286. Stocco, C. O. *et al.* Prostaglandin F2α-induced expression of 20α-hydroxysteroid dehydrogenase involves the transcription factor NUR77. *J. Biol. Chem.* **275**, 37202–37211 (2000).

287. Oakley, R. H. & Cidlowski, J. A. The biology of the glucocorticoid receptor: New signaling mechanisms in health and disease. *J. Allergy Clin. Immunol.* **132**, 1033–1044 (2013).

288. Bale, T. L. Stress sensitivity and the development of affective disorders. *Horm. Behav.* **50**, 529–533 (2006).

289. Haller, J., Halasz, J., Makara, G. B. & Kruk, M. R. Acute effects of glucocorticoids: Behavioral and pharmacological perspectives. *Neurosci. Biobehav. Rev.* **23**, 337–344 (1998).

290. Pivonello, R. *et al.* Neuropsychiatric disorders in cushing's syndrome. *Front. Neurosci.* **9**, 1–6 (2015).

291. Simpson, E. R. & Davis, S. R. Minireview: aromatase and the regulation of estrogen biosynthesis--some new perspectives. *Endocrinology* **142**, 4589–4594 (2001).

292. Niepoth, N. & Bendesky, A. How Natural Genetic Variation Shapes Behavior. *Annu. Rev. Genomics Hum. Genet.* **21**, 1–27 (2020).

Appendix

Supplementary Figures and Tables

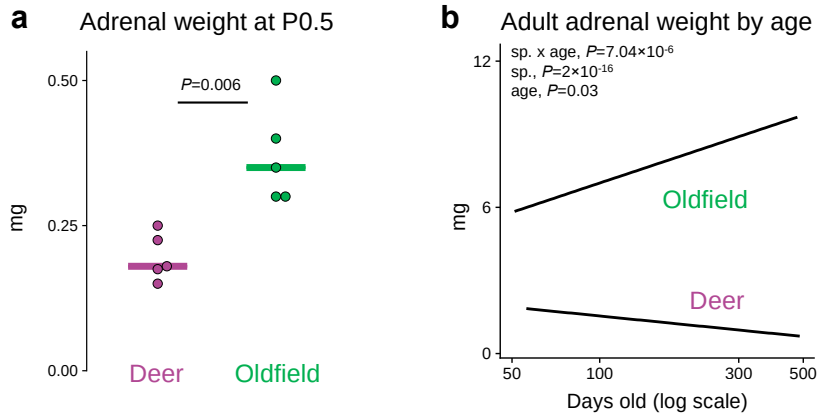

Fig. A1.1: Oldfield mice adrenal glands are larger than deer mice glands from birth and continue to grow throughout adulthood. a, Adrenal weight at postnatal day 0.5 (species P=0.006, t-test; N=5 per species; line at median). **b**, Adult adrenal weight by age (species P=2×10^{-16}, age P=0.03, species × age P=7.04×10^{-6}, generalized linear model; deer: N=114, oldfield: N=69.

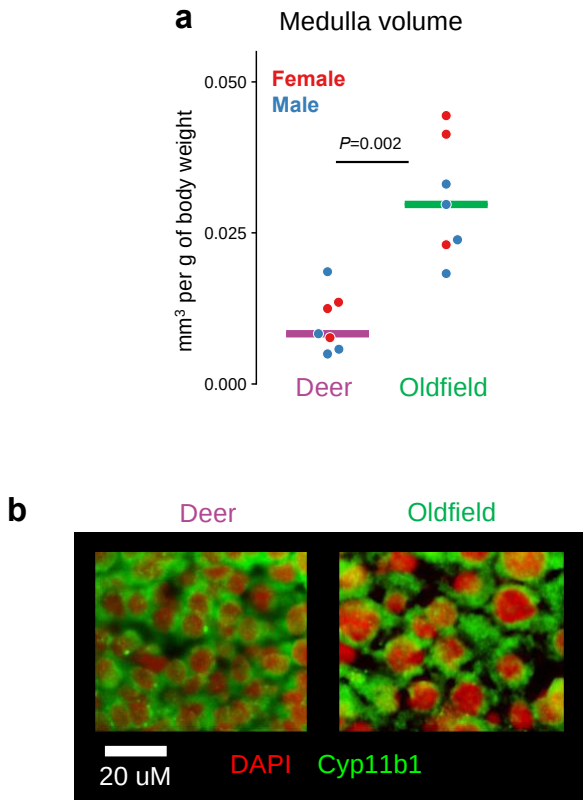

Fig. A1.2: Oldfield mice have a larger adrenal medulla and larger cortex cells. a, Adrenal medulla volume (species *P*=0.002, sex *n.s.*, species × sex *n.s.*, generalized linear model; N=7 per species; line at median). **b**, Representative adrenal section from the *zona fasciculata* (zF) of the deer- and oldfield mouse adrenal cortex. Cyp11b1 (green) labeled by *in situ* hybridization and counterstained with DAPI (red).

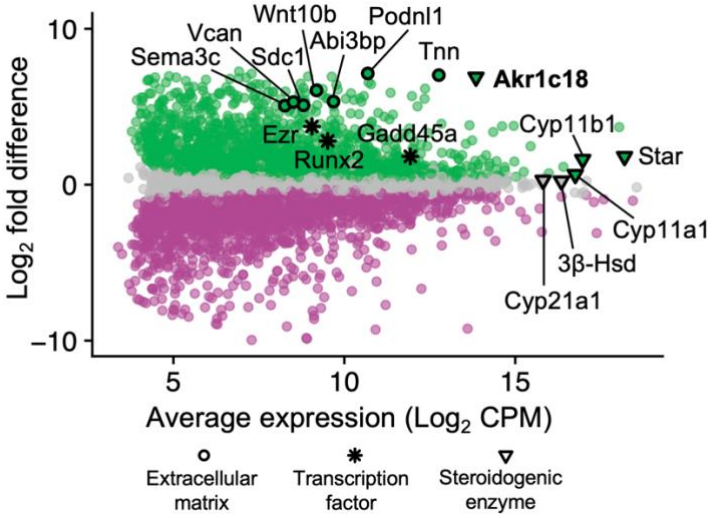

Fig. A1.3: Differential gene expression in the adrenals of deer- and oldfield mice from bulk RNA-seq. Plot of differential adrenal gene expression between deer and oldfield mouse (purple: higher in deer mouse, green: higher in oldfield mouse, grey: *n.s.*)

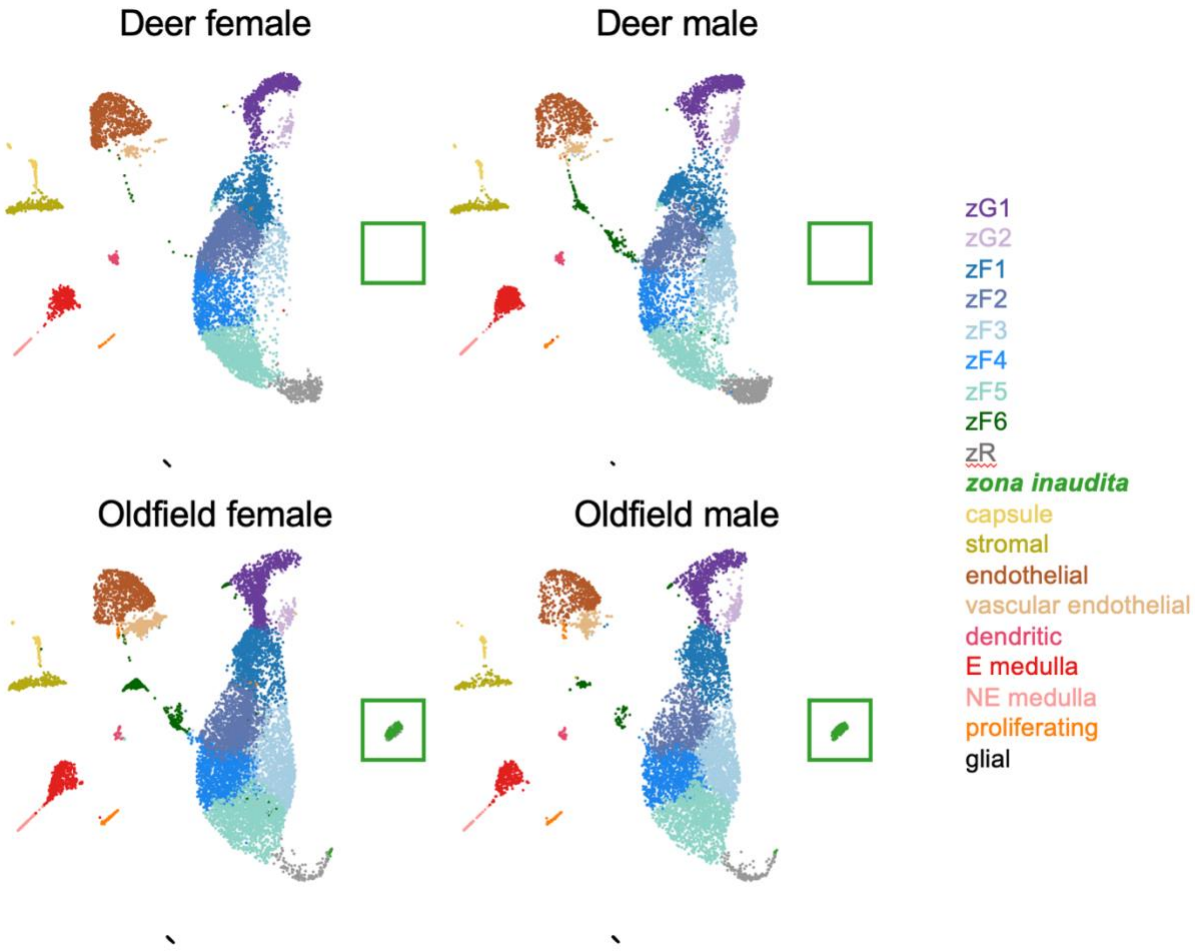

Fig. A1.4: UMAP visualization of the *Peromyscus* adrenal gland split by species and sex. UMAP of adrenal snRNA-seq after reciprocal PCA integration of deer mouse and oldfield mouse cells, split by species and sex.

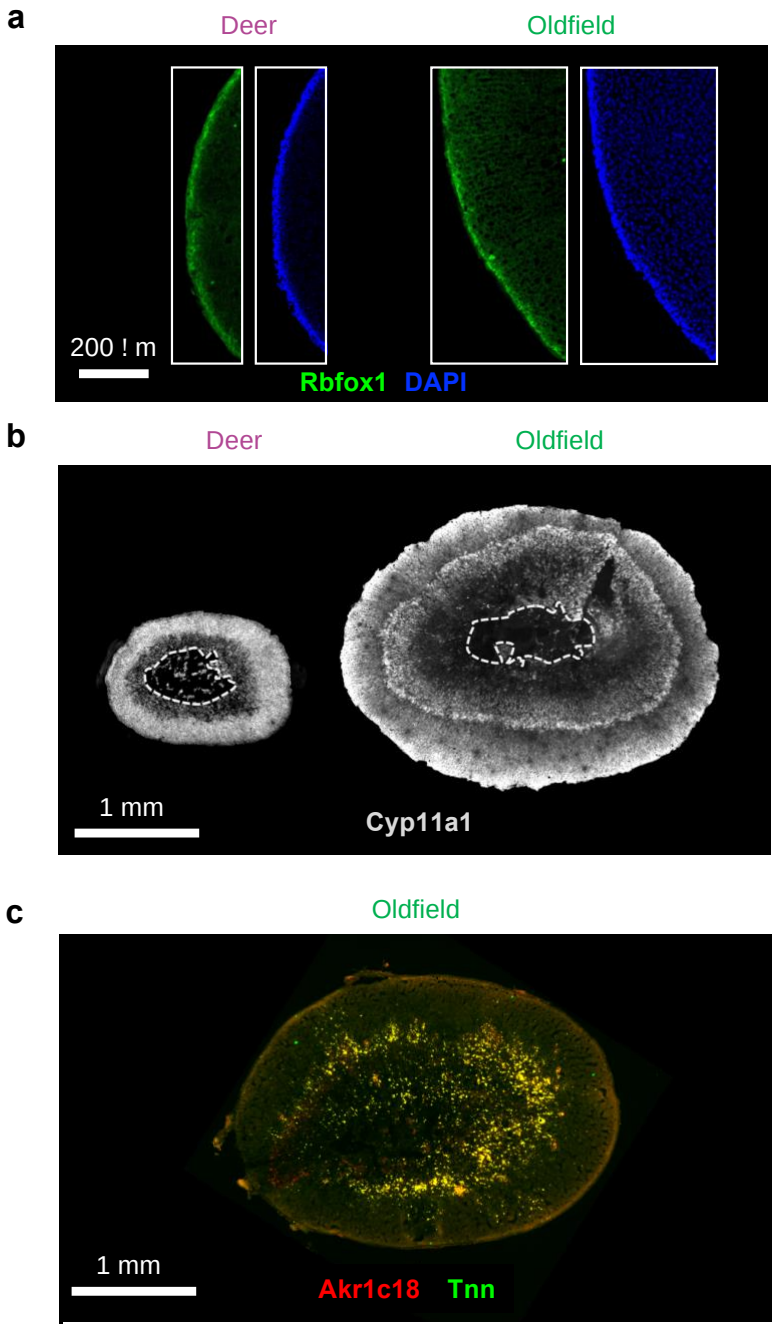

Fig. A1.5: Defining cell types by histology. a, Antibody staining of Rbfox1 (green) labels the *zona glomerulosa*, a cell-dense zone as shown by DAPI nuclear stain (blue). **b**, *In situ* hybridization of Cyp11a1 labels all steroidogenic cortical cell types of the adrenal (high expression in the *zona glomerulosa* and *zona fasciculata*; low expression in the *zona reticularis*). **c**, Co-localization of expression of Akr1c18 and Tnn, markers of the oldfield-specific *zona inaudita*.

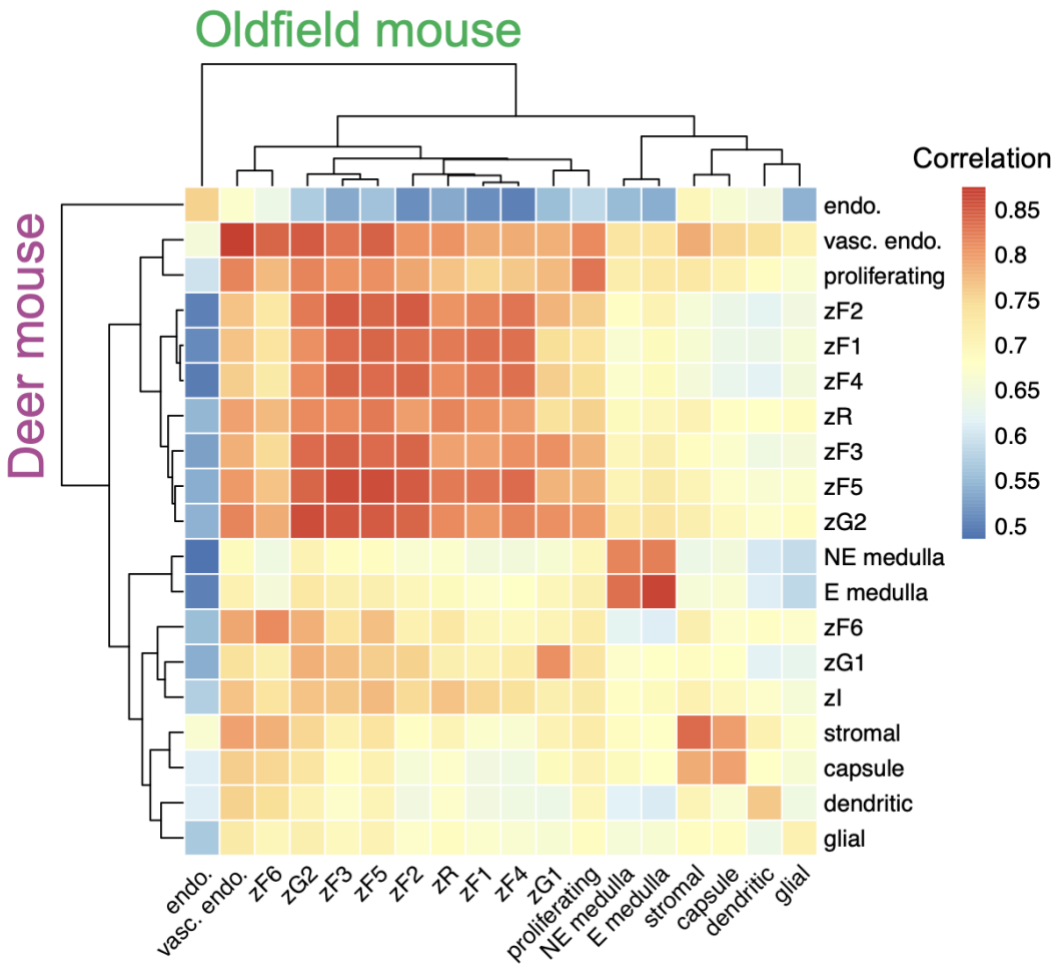

Fig. A1.6: Hierarchical clustering of cell types from the deer- and oldfield adrenal glands. Hierarchical clustering of deer mouse cell types (columns) and oldfield cell types (rows) based on average expression of all adrenal genes. Color: spearman correlation of gene expression of top 2000 variable features; rows: deer cell types, columns: oldfield cell types.

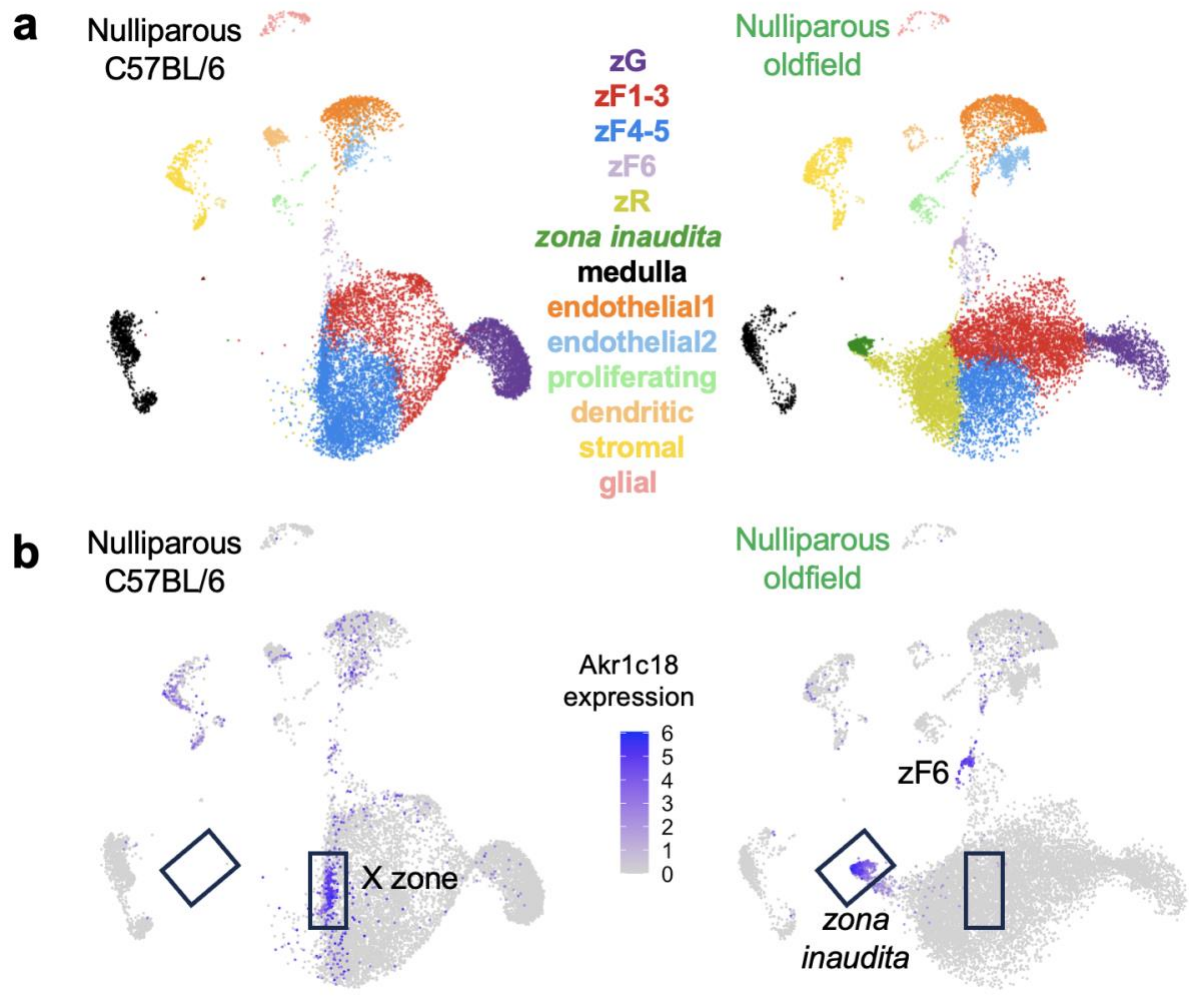

Fig. A1.7: The oldfield *zona inaudita* is molecularly distinct from the house mouse X zone. **a**, UMAP visualization of the adrenal cells from adult nulliparous C57BL/6 females (left) and adult nulliparous oldfield females (right) after integration of C57BL/6, deer, and oldfield mice cells. **b**, Expression of Akr1c18 in the nulliparous C57BL/6 and oldfield mice demarcates the X-zone in house mice, zF6 cells in oldfield mice, and the *zona inaudita*.

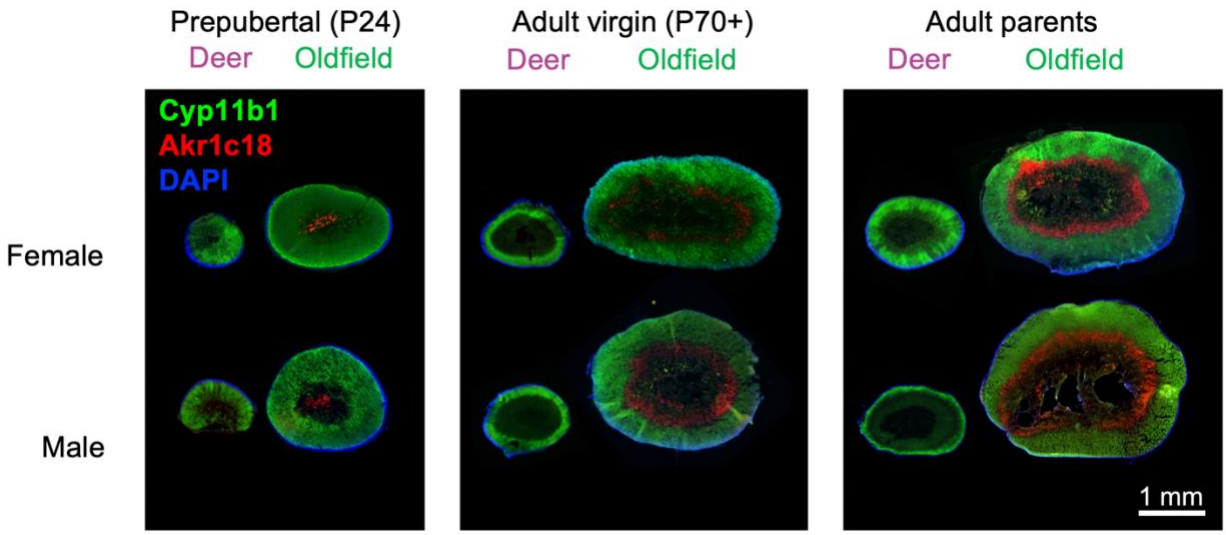

Fig. A1.8: Akr1c18+ *zona inaudita* cells arise around postnatal day 24 and persist into adulthood. Representative adrenal sections from deer- and oldfield mice before puberty (postnatal day 24, left), after puberty (>70 days old, middle), and after mating and parturition (right); stained for Cyp11b1 (green), Akr1c18 (red), DAPI (blue) by *in situ* hybridization

Fig. A1.9: Akr1c18 is highly expressed in the ovaries of deer- and oldfield females. Expression of Akr1c18 in female adrenal glands of deer and oldfield mice (left, *P*=0.005) and in ovaries; *P*-values by *t*-test; N=4 per species; lines at median.

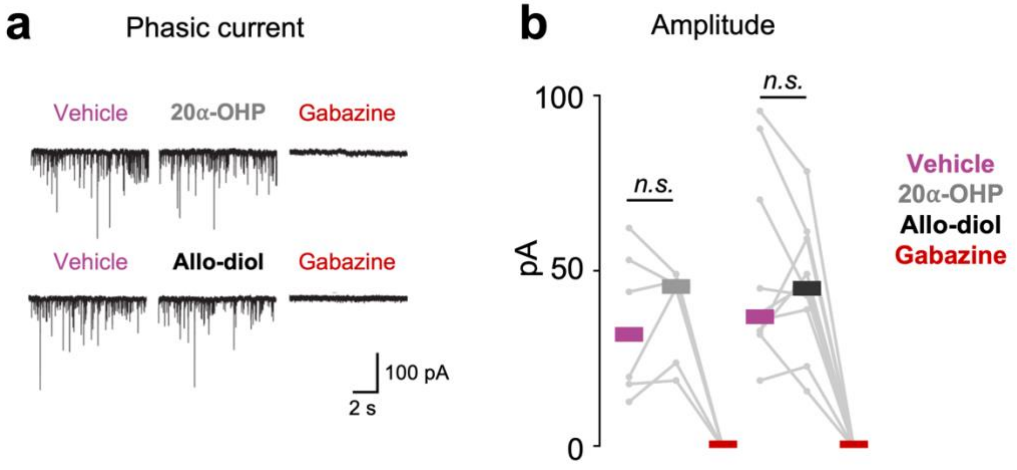

Figure A1.10: No effect of 20α-OHP or allo-diol on phasic GABAergic currents
a, Representative phasic GABAergic current trace recorded from house mouse molecular layer interneurons at baseline (vehicle), after addition of 1uM 20α-OHP or 1uM allo-diol, and after addition of 50uM Gabazine, a $GABA_A$ receptor antagonist. **b**, Amplitude of phasic GABAergic current by treatment; lines at median. Experiments conducted by Stefano Lutzu and Stephanie Rudolph.

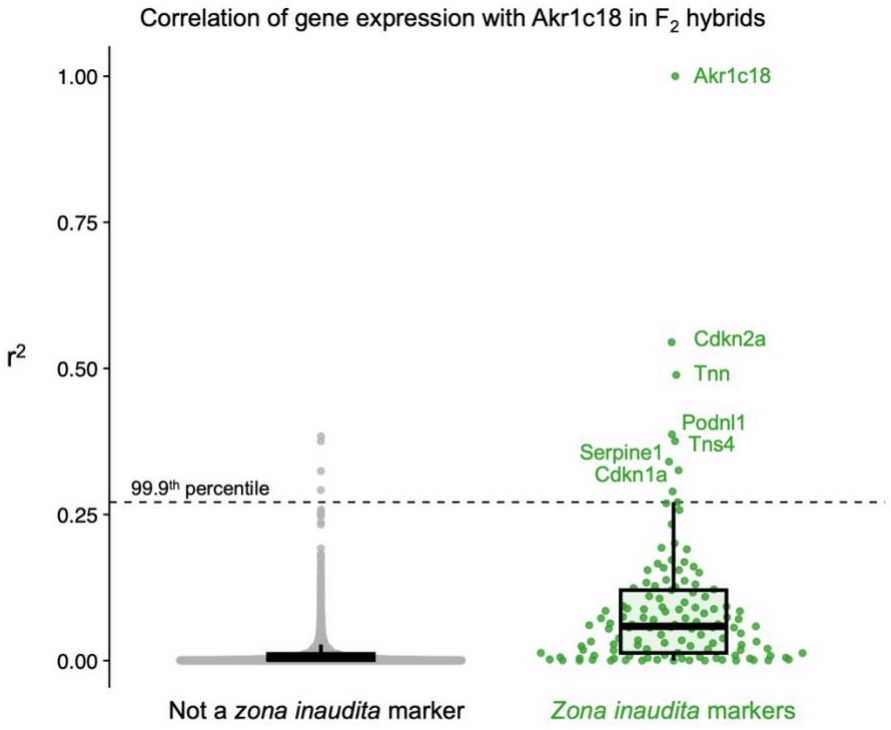

Fig. A1.11: *Zona inaudita* markers are particularly highly correlated with Akr1c18 expression in F2 hybrids. Correlation with Akr1c18 expression in F_2 hybrids by whether the gene is more

highly expressed in the *zona inaudita* (green, right) compared to other cells of oldfield adrenal cortex (grey, left).

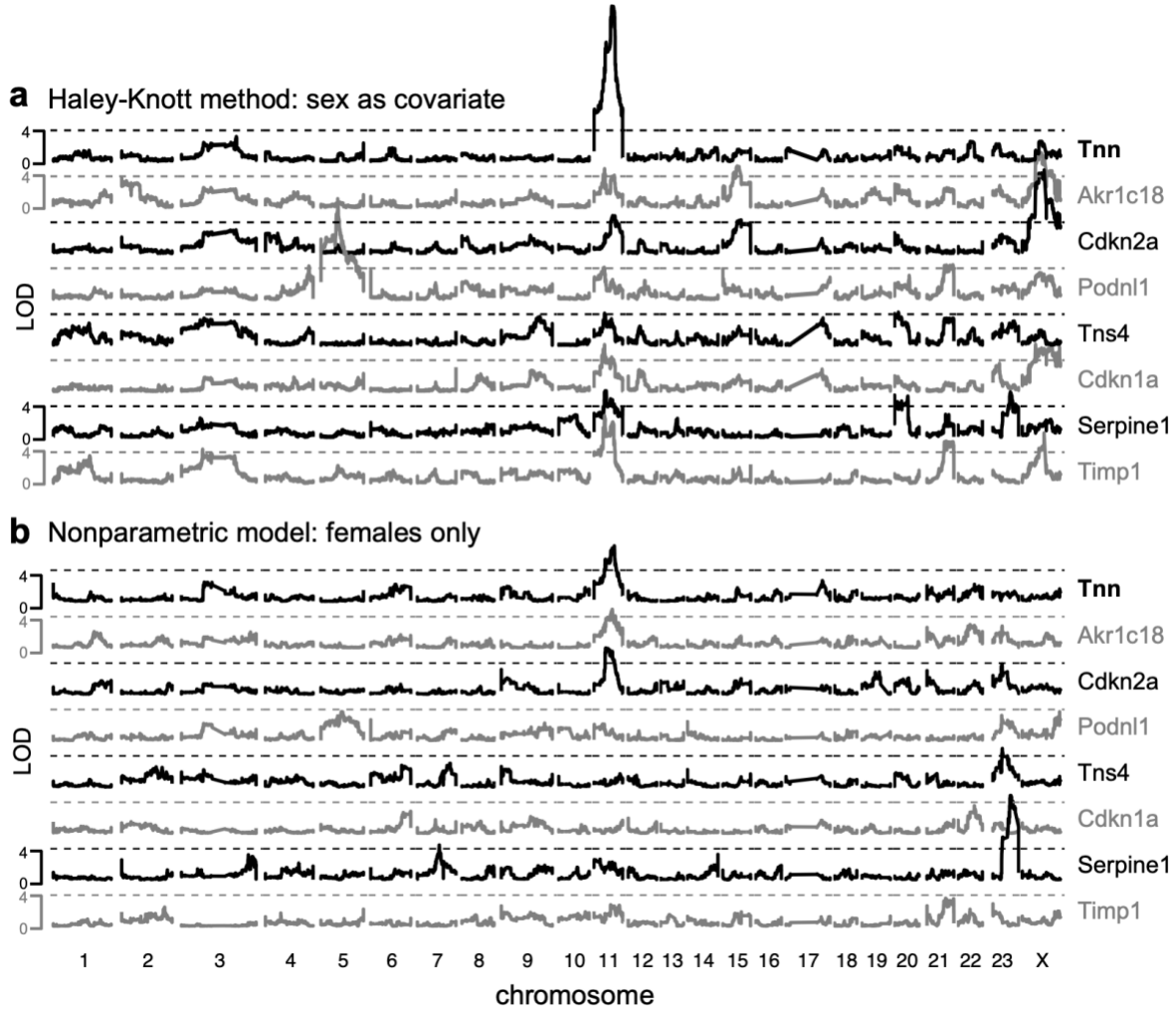

Fig. A1.12: Expression quantitative trait locus mapping of *zona inaudita* genes in both sexes and in females. Logarithm of the odds (LOD) of adrenal expression of *zona inaudita* marker genes correlated with F_2 adrenal Akr1c18 expression, by genotype across the genome in F2 hybrids. Dashed lines denote genome- wide threshold of significance (α=0.05). **a**, Haley-Knott regression of all 706 F_2 hybrids with sex as a covariate. **b**, Nonparametric scan in females only.

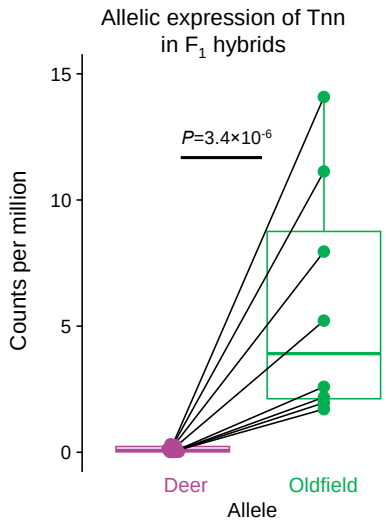

Fig. A1.13: *Cis*-regulatory variation underlies gain of Tnn expression in oldfield adrenals. Allele-specific expression of Tnn in F_1 hybrids of deer- and oldfield mice (P=3.4×10-6, paired t-test; N=8). Box: interquartile range (IQR); whiskers extend to largest value no further than 1.5×IQR from the first or third quartile respectively.

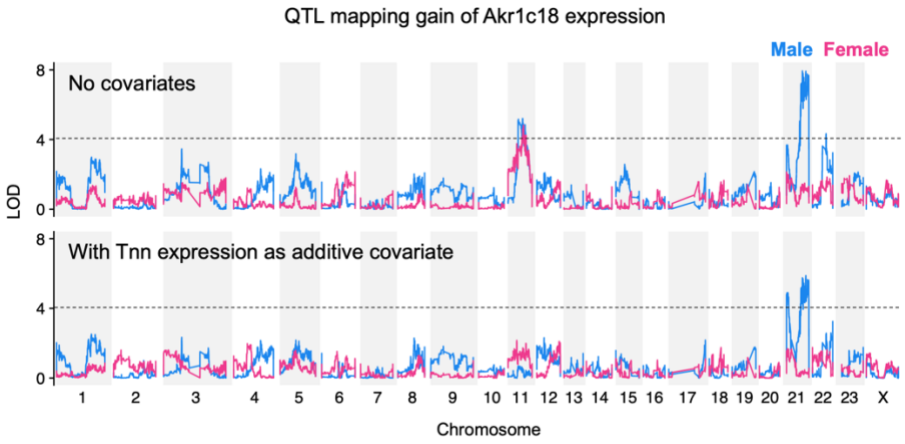

Fig. A1.14: Mediation analysis indicates that Tnn expression explains the contribution of the chromosome 11 eQTL to the gain of Akr1c18 expression. QTL mapping presence vs. absence of Akr1c18 expression with no added covariates or expression of Tnn (bottom). LOD: logarithm of the odds. Dashed lines represent genome-wide threshold of significance (α=0.05).

Table 1.1: Markers of the *zona inaudita* compared to cortex cell types of deer- and oldfield mice

Gene symbol	Gene name	Average log_2 fold change	Adj. p-value	ECM	TF
LOC102910062	aldo-keto reductase family 1 member C18	4.55888694	0	FALSE	FALSE
Serpine1	serpin family E member 1	4.47711104	0	TRUE	FALSE
Runx2	RUNX family transcription factor 2	4.19662766	0	FALSE	TRUE
Tnn	tenascin N	4.19288853	0	TRUE	FALSE
LOC102922445	ABI family member 3 binding protein	4.12305952	0	TRUE	FALSE
Pdzrn4	PDZ domain containing ring finger 4	3.47703179	0	FALSE	FALSE
LOC102912527	haptoglobin	3.20019876	0	FALSE	FALSE
Vim	vimentin	3.0930133	0	FALSE	FALSE
LOC102922906	retinal dehydrogenase 1	3.08824454	0	FALSE	FALSE
LOC102918527	dehydrogenase/reductase SDR family member 7-like	3.08456582	0	FALSE	FALSE
Igf1	insulin like growth factor 1	3.05216543	0	TRUE	FALSE
Podnl1	podocan like 1	3.04622143	0	TRUE	FALSE
LOC102911743	multidrug resistance protein 2	2.90028048	0	FALSE	FALSE
Fkbp5	FKBP prolyl isomerase 5	2.85094989	0	FALSE	FALSE
Dennd4a	DENN domain containing 4A	2.82393749	0	FALSE	FALSE
Apoe	apolipoprotein E	2.76908456	0	FALSE	FALSE
Efna5	ephrin A5	2.75904745	0	FALSE	FALSE
Tacc2	transforming acidic coiled-coil containing protein 2	2.74490372	0	FALSE	FALSE
Dhrs9	dehydrogenase/reductase 9	2.62058939	0	FALSE	FALSE
Htra1	HtrA serine peptidase 1	2.61015298	0	TRUE	FALSE
Kcnt2	potassium sodium-activated channel subfamily T member 2	2.55038724	0	FALSE	FALSE
Gpx3	glutathione peroxidase 3	2.54371798	0	FALSE	FALSE
LOC102914915	cyclin-dependent kinase inhibitor 2A-like	2.52673399	0	FALSE	FALSE
LOC121822425	uncharacterized LOC121822425	2.49209482	0	FALSE	FALSE
Gadd45a	growth arrest and DNA damage inducible alpha	2.44427573	0	FALSE	TRUE
LOC121829429	uncharacterized LOC121829429	2.35333738	0	FALSE	FALSE
Pde4b	phosphodiesterase 4B	2.32158298	0	FALSE	FALSE

Timp1	TIMP metallopeptidase inhibitor 1	2.28701424	0	TRUE	FALSE
Batf	basic leucine zipper ATF-like transcription factor	2.28680923	0	FALSE	TRUE
Lrrtm3	leucine rich repeat transmembrane neuronal 3	2.25141982	0	FALSE	FALSE
Plek	pleckstrin	2.22395283	0	FALSE	FALSE
Rgcc	regulator of cell cycle	2.21279992	0	FALSE	FALSE
Serp1	stress associated endoplasmic reticulum protein 1	2.18169399	0	FALSE	FALSE
Igfbp4	insulin like growth factor binding protein 4	2.1785408	0	TRUE	FALSE
F11r	F11 receptor	2.1603736	0	FALSE	FALSE
Grn	granulin precursor	2.14036545	0	FALSE	FALSE
Arhgap31	Rho GTPase activating protein 31	2.13513095	0	FALSE	FALSE
Rftn1	raftlin, lipid raft linker 1	2.07620798	0	FALSE	FALSE
Fam78b	family with sequence similarity 78 member B	2.06872824	0	FALSE	FALSE
Ctnna3	catenin alpha 3	2.01380302	0	FALSE	FALSE
Ptgfr	prostaglandin F receptor	1.95215111	0	FALSE	FALSE
Ntng1	netrin G1	1.93600549	0	TRUE	FALSE
LOC102928503	NADPH-dependent 3-keto-steroid reductase HSD3B3-like	1.89840111	0	FALSE	FALSE
LOC102926742	neutrophilic granule protein-like	1.89401691	0	FALSE	FALSE
Adrb2	adrenoceptor beta 2	1.85768888	0	FALSE	FALSE
LOC102916382	tubulin alpha-1C chain	1.76612365	0	FALSE	FALSE
LOC102922647	cyclin-dependent kinase 4 inhibitor B	1.76392669	0	FALSE	FALSE
Wnt10b	Wnt family member 10B	1.75424298	0	TRUE	FALSE
Amdhd1	amidohydrolase domain containing 1	1.74421824	0	FALSE	FALSE
Fstl4	follistatin like 4	1.73181357	0	FALSE	FALSE
LOC107402659	uncharacterized LOC107402659	1.70580253	0	FALSE	FALSE
Tbc1d1	TBC1 domain family member 1	1.68570229	0	FALSE	FALSE
Creg1	cellular repressor of E1A stimulated genes 1	1.66974653	0	FALSE	FALSE
Atp2b4	ATPase plasma membrane Ca2+ transporting 4	1.65972206	0	FALSE	FALSE
Megf11	multiple EGF like domains 11	1.62787996	0	TRUE	FALSE
Cd59	CD59 molecule (CD59 blood group)	1.60791817	0	FALSE	FALSE
LOC121821259	uncharacterized LOC121821259	1.59263022	0	FALSE	FALSE
Tns4	tensin 4	1.56563755	0	FALSE	FALSE
Emilin2	elastin microfibril interfacer 2	1.5442396	0	TRUE	FALSE
LOC102904443	glutathione S-transferase Mu 1	1.53725167	0	FALSE	FALSE
Mmp23b	matrix metallopeptidase 23B	1.47884638	0	TRUE	FALSE

Gas7	growth arrest specific 7	1.4743802	0	FALSE	FALSE
Arap2	ArfGAP with RhoGAP domain, ankyrin repeat and PH domain 2	1.44368268	0	FALSE	FALSE
Spag16	sperm associated antigen 16	1.42771998	0	FALSE	FALSE
S100a11	S100 calcium binding protein A11	1.3851866	0	TRUE	FALSE
Rps27l	ribosomal protein S27 like	1.31774315	0	FALSE	FALSE
Adamts2	ADAM metallopeptidase with thrombospondin type 1 motif 2	1.27446392	0	TRUE	FALSE
Gjb4	gap junction protein beta 4	1.27308398	0	FALSE	FALSE
Emp3	epithelial membrane protein 3	1.26519102	0	FALSE	FALSE
Lgals1	galectin 1	1.19333714	0	TRUE	FALSE
Nt5e	5'-nucleotidase ecto	2.05264924	9.79E-301	FALSE	FALSE
Sfxn1	sideroflexin 1	1.54274725	2.11E-296	FALSE	FALSE
LOC102917248	NADPH-dependent 3-keto-steroid reductase HSD3B3-like	1.29740885	7.51E-296	FALSE	FALSE
Ipmk	inositol polyphosphate multikinase	2.35400777	2.71E-293	FALSE	FALSE
Ezr	ezrin	1.80016262	2.23E-292	FALSE	TRUE
Cemip2	cell migration inducing hyaluronidase 2	1.82676538	5.78E-292	FALSE	FALSE
Fam126a	family with sequence similarity 126 member A	1.5635505	7.01E-292	FALSE	FALSE
Serpine2	serpin family E member 2	1.36092205	2.26E-288	TRUE	FALSE
Hsd11b1	hydroxysteroid 11-beta dehydrogenase 1	1.53384739	1.23E-275	FALSE	FALSE
Mtarc1	mitochondrial amidoxime reducing component 1	1.99718969	1.44E-269	FALSE	FALSE
Gjb3	gap junction protein beta 3	1.30759166	9.29E-266	FALSE	FALSE
LOC107402876	phosphatidylcholine translocator ABCB4	2.62044808	1.10E-252	FALSE	FALSE

Sv2c	synaptic vesicle glycoprotein 2C	1.22136293	2.84E-250	FALSE	FALSE
Bst2	bone marrow stromal cell antigen 2	1.4832337	3.31E-246	FALSE	FALSE
Mvb12a	multivesicular body subunit 12A	1.64089454	2.03E-244	FALSE	FALSE
Igfbp3	insulin like growth factor binding protein 3	2.08955985	7.96E-237	TRUE	FALSE
Ankh	ANKH inorganic pyrophosphate transport regulator	1.71518858	2.00E-220	FALSE	FALSE
Tshz3	teashirt zinc finger homeobox 3	1.58155758	2.00E-220	FALSE	FALSE
Nedd4l	NEDD4 like E3 ubiquitin protein ligase	2.24321572	6.26E-218	FALSE	FALSE
Idh1	isocitrate dehydrogenase (NADP(+)) 1	2.2398565	2.04E-217	FALSE	FALSE
Mrpl33	mitochondrial ribosomal protein L33	1.82832873	1.44E-215	FALSE	FALSE
Aig1	androgen induced 1	2.13371006	7.85E-212	FALSE	FALSE
Angpt1	angiopoietin 1	1.60792973	9.56E-205	TRUE	FALSE
Ttll7	tubulin tyrosine ligase like 7	1.4434443	1.32E-203	FALSE	FALSE
LOC121828709	uncharacterized LOC121828709	1.96993609	8.01E-201	FALSE	FALSE
Flvcr2	FLVCR heme transporter 2	1.59952677	6.68E-200	FALSE	FALSE
LOC121823529	uncharacterized LOC121823529	1.98706797	5.15E-195	FALSE	FALSE
Hif1a	hypoxia inducible factor 1 subunit alpha	2.2322526	3.83E-193	FALSE	TRUE
LOC121827286	uncharacterized LOC121827286	1.46940716	1.13E-190	FALSE	FALSE

Phldb2	pleckstrin homology like domain family B member 2	1.92713326	7.82E-186	FALSE	FALSE
Ftl	ferritin light chain	1.35804487	3.01E-185	FALSE	FALSE
LOC121832414	uncharacterized LOC121832414	1.97168064	1.34E-184	FALSE	FALSE
Cfh	complement factor H	2.07191609	1.19E-183	FALSE	FALSE
Arhgap24	Rho GTPase activating protein 24	1.07967511	2.43E-183	FALSE	FALSE
Nhs	NHS actin remodeling regulator	1.01975514	1.64E-180	FALSE	FALSE
Atp10a	ATPase phospholipid transporting 10A (putative)	1.66329418	3.23E-179	FALSE	FALSE
Diaph2	diaphanous related formin 2	2.09737882	2.18E-174	FALSE	FALSE
Camk2d	calcium/calmodulin dependent protein kinase II delta	2.09984071	3.78E-174	FALSE	FALSE
Sparc	secreted protein acidic and cysteine rich	1.78431221	4.03E-174	TRUE	FALSE
Klf6	Kruppel like factor 6	1.25750784	1.12E-173	FALSE	TRUE
Rps21	ribosomal protein S21	1.11280648	1.84E-172	FALSE	FALSE
LOC121822385	uncharacterized LOC121822385	1.57385316	2.03E-171	FALSE	FALSE
Cd63	CD63 molecule	1.29382324	2.80E-166	FALSE	FALSE
Kiaa0513	KIAA0513 ortholog	1.38604962	4.05E-162	FALSE	FALSE
Pkm	pyruvate kinase M1/2	1.46461582	5.72E-162	FALSE	TRUE
Esd	esterase D	1.08818155	4.51E-159	FALSE	FALSE

LOC107402015	uncharacterized LOC107402015	1.89235163	2.10E-155	FALSE	FALSE
Pik3c2g	phosphatidylinositol-4-phosphate 3-kinase catalytic subunit type 2 gamma	1.57422476	3.31E-155	FALSE	FALSE
Supt3h	SPT3 homolog, SAGA and STAGA complex component	2.0544753	2.02E-151	FALSE	FALSE
Lrrk2	leucine rich repeat kinase 2	1.41088963	8.82E-147	FALSE	FALSE
Vwa8	von Willebrand factor A domain containing 8	2.5127083	6.08E-145	FALSE	FALSE
Thsd4	thrombospondin type 1 domain containing 4	1.78171846	3.15E-141	FALSE	FALSE
Fam13a	family with sequence similarity 13 member A	2.22948713	8.23E-140	FALSE	FALSE
Xdh	xanthine dehydrogenase	1.55175471	7.22E-129	FALSE	FALSE
Ccdc68	coiled-coil domain containing 68	1.73497411	1.53E-127	FALSE	FALSE
LOC121830727	uncharacterized LOC121830727	1.17497397	5.45E-124	FALSE	FALSE
Pdia6	protein disulfide isomerase family A member 6	1.20324828	8.28E-121	FALSE	FALSE
Rdx	radixin	1.4178459	6.32E-120	FALSE	FALSE
Btg1	BTG anti-proliferation factor 1	1.40752324	1.17E-117	FALSE	FALSE
LOC102909637	sodium-dependent phosphate transporter 1	1.62823059	1.44E-117	FALSE	FALSE
Tsc22d3	TSC22 domain family member 3	1.51403085	8.03E-117	FALSE	FALSE
Glce	glucuronic acid epimerase	1.45892253	4.43E-111	FALSE	FALSE
LOC102915907	3 beta-hydroxysteroid dehydrogenase/Delta 5-->4-isomerase type 1-like	1.34653643	7.10E-111	FALSE	FALSE

Sgpl1	sphingosine-1-phosphate lyase 1	1.40731712	5.63E-110	FALSE	FALSE
Map1b	microtubule associated protein 1B	1.18109511	4.18E-106	FALSE	FALSE
LOC102906375	aldehyde oxidase 4-like	1.37840678	4.26E-106	FALSE	FALSE
Ctsd	cathepsin D	1.23458313	3.46E-104	TRUE	FALSE
LOC121826345	uncharacterized LOC121826345	1.32493741	2.83E-101	FALSE	FALSE
Ror1	receptor tyrosine kinase like orphan receptor 1	1.43791565	6.81E-95	FALSE	FALSE
Enah	ENAH actin regulator	1.51551024	4.61E-94	FALSE	FALSE
Asap1	ArfGAP with SH3 domain, ankyrin repeat and PH domain 1	1.26578678	1.66E-93	FALSE	FALSE
Crot	carnitine O-octanoyltransferase	1.79712953	4.61E-93	FALSE	FALSE
Frmd6	FERM domain containing 6	1.125405	1.32E-91	FALSE	FALSE
Tiam1	TIAM Rac1 associated GEF 1	1.1443771	1.78E-91	FALSE	FALSE
Rgl1	ral guanine nucleotide dissociation stimulator like 1	1.19441135	4.14E-91	FALSE	FALSE
LOC102913629	3 beta-hydroxysteroid dehydrogenase/Delta 5-->4-isomerase type 1-like	1.15092855	1.15E-90	FALSE	FALSE
Lipa	lipase A, lysosomal acid type	1.43828273	1.40E-90	FALSE	FALSE
Rhoq	ras homolog family member Q	1.21459351	2.26E-90	FALSE	FALSE
Gclc	glutamate-cysteine ligase catalytic subunit	1.8801854	3.66E-90	FALSE	FALSE
St6galnac3	ST6 N-acetylgalactosaminide alpha-2,6-sialyltransferase 3	1.38171825	7.25E-89	FALSE	FALSE
Dhcr24	24-dehydrocholesterol reductase	1.36291956	2.52E-87	FALSE	FALSE
Pdzrn3	PDZ domain containing ring finger 3	1.18248617	2.90E-87	FALSE	FALSE
Ifngr1	interferon gamma receptor 1	1.77998273	3.33E-87	FALSE	FALSE
Dtx2	deltex E3 ubiquitin ligase 2	1.16609693	1.13E-86	FALSE	FALSE
Notch2	notch receptor 2	1.12334629	8.04E-83	FALSE	FALSE
Grk4	G protein-coupled receptor kinase 4	1.35518085	9.29E-83	FALSE	FALSE
Ywhah	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein eta	1.13215475	9.32E-82	FALSE	FALSE
Akap13	A-kinase anchoring protein 13	1.4191422	9.34E-82	FALSE	FALSE
Anxa11	annexin A11	1.06866413	1.01E-81	TRUE	TRUE

LOC121826786	uncharacterized LOC121826786	1.03009198	3.13E-80	FALSE	FALSE
LOC102920651	cytochrome P450 2B2-like	1.20835789	1.97E-74	FALSE	FALSE
Chd9	chromodomain helicase DNA binding protein 9	1.2357405	1.59E-71	FALSE	FALSE
Cdyl	chromodomain Y like	1.22434911	2.43E-71	FALSE	FALSE
LOC121827330	uncharacterized LOC121827330	1.3420524	1.78E-69	FALSE	FALSE
Klf13	Kruppel like factor 13	1.23029441	4.99E-69	FALSE	FALSE
Lhfpl2	LHFPL tetraspan subfamily member 2	1.12859886	7.08E-68	FALSE	FALSE
Trps1	transcriptional repressor GATA binding 1	1.06865161	1.81E-67	FALSE	TRUE
Lpin2	lipin 2	1.06139796	4.19E-67	FALSE	FALSE
Timp2	TIMP metallopeptidase inhibitor 2	1.10312288	4.89E-67	FALSE	FALSE
Cd47	CD47 molecule	1.09726058	3.12E-66	FALSE	FALSE
Ascc3	activating signal cointegrator 1 complex subunit 3	1.20372482	1.21E-64	FALSE	FALSE
Foxn3	forkhead box N3	1.13767953	7.93E-64	FALSE	TRUE
Scp2	sterol carrier protein 2	1.15421265	1.70E-62	FALSE	FALSE
Appbp2	amyloid beta precursor protein binding protein 2	1.3089289	1.63E-61	FALSE	FALSE
Prim2	DNA primase subunit 2	1.16184014	8.19E-60	FALSE	FALSE
Ston2	stonin 2	1.07302137	1.08E-59	FALSE	FALSE
Mtor	mechanistic target of rapamycin kinase	1.54913349	3.04E-56	FALSE	FALSE
Osbpl9	oxysterol binding protein like 9	1.26738687	4.91E-56	FALSE	FALSE
LOC102926693	aldehyde oxidase 4	1.22189729	8.63E-56	FALSE	FALSE
Prickle1	prickle planar cell polarity protein 1	1.08914898	2.36E-53	FALSE	FALSE
Cblb	Cbl proto-oncogene B	1.03244516	1.51E-50	FALSE	FALSE
LOC102915951	zinc finger protein 431-like	1.13431927	1.75E-49	FALSE	FALSE
Adam10	ADAM metallopeptidase domain 10	1.2788183	1.56E-48	TRUE	FALSE
Slc6a6	solute carrier family 6 member 6	1.10733007	6.34E-48	FALSE	FALSE
Rbbp6	RB binding protein 6, ubiquitin ligase	1.12567645	3.87E-42	FALSE	FALSE
Nudt4	nudix hydrolase 4	1.15817624	5.31E-40	FALSE	FALSE
Vps35	VPS35 retromer complex component	1.05313822	1.01E-39	FALSE	FALSE
Parvb	parvin beta	1.4700028	4.40E-34	FALSE	FALSE