The Dissertation Committee for Mariam Okhovat Certifies that this is the approved version of the following dissertation:

Molecular Regulation of Vasopressin Receptor Among (Mostly) Monogamous Prairie Voles

Committee:

Steven M. Phelps, Supervisor

Johann A. Hofmann

Steven A. Vokes

Mikhail Matz

Nigel Atkinson

Molecular Regulation of Vasopressin Receptor Among (Mostly) Monogamous Prairie Voles

by

Mariam Okhovat, B.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December 2016

Dedication

To my parents for their unconditional love and support and my husband for his endless kindness and patience

Acknowledgements

I have been extremely privileged to meet and work with countless intelligent and friendly people who have guided, supported and motivated me during my PhD studies. I am most grateful, to my supervisor Steven M. Phelps for holding my hand through the hurdles of graduate school, for lifting me up when I fell down and for always believing in me and cheering me on when I felt defeated. Steve gave me the freedom and support to explore my research interests and I will forever cherish all that I have learned from him about science, research, writing, and life in general.

I am also very thankful to the members of my PhD committee: Hans Hoffman, Misha Matz, Steven Vokes and Nigel Atkinson, all of whom were an invaluable source of guidance to me. I would like to specifically thank Hans for always supporting me and giving me great advice about graduation and post-graduation life, Steven Vokes for being available for scientific discussions, Misha, for providing insightful ideas and offering scientific support, and Nigel, for teaching me about the exciting world of epigenetics.

During my graduate studies, I have had many friends and colleague whom I have leaned on for scientific and emotional support. Specifically, I would like to thank Lauren O'Connell, who was truly a role model to me. Working in the lab with Lauren was a critical professional turn point for me and completely changed my perception of research. I would also like to express my gratitude to James Derry, who was always available and willing to provide programming support. I have also been privileged to work with many talented undergraduates, who made significant contributions to my research, Hamidat Momoh and Isabella Chen's assistance, in particular, has been extremely helpful during the last couple of years. I am also thankful to many other current and former members of

Molecular Regulation of Vasopressin Receptor Among (Mostly) Monogamous Prairie Voles

Mariam Okhovat, PhD

The University of Texas at Austin, 2016

Supervisor: Steven M. Phelps

Intraspecific variation in social behavior is common and often dramatic, but little is known about its underlying mechanisms. We use the prairie vole (*Microtus ochrogaster*) to examine how intraspecific variation in brain and behavior emerges as a result of genetic, epigenetic and environmental variation. Although prairie voles are socially monogamous, they vary in sexual fidelity; faithful prairie voles are described as intra-pair fertilizing (IPF), while unfaithful voles are extra-pair fertilizing (EPF). EPF males have large home-ranges and frequently mate with other females, but do so at the cost of being cuckolded. IPF males however, form small exclusive home-ranges, rarely intrude and are better at mate-guarding. These behavioral differences are predicted by abundance of the vasopressin receptor 1a (V1aR) in the retrosplenial cortex (RSC), a brain region implicated in spatial memory. We find that variation in RSC-V1aR and associated behaviors are predicted by two alternative *avpr1a* alleles. These "HI" and "LO" alleles are defined by four linked single nucleotide polymorphisms, one of which is a polymorphic CpG site (polyCpG) located within a putative intron enhancer. This polyCpG is weakly linked to several other polyCpGs in the enhancer. Since CpGs are targets for DNA methylation, polyCpGs may cause individual differences in DNA

methylation, gene regulation and environmental sensitivity. The unusually high number of polyCpGs within the intron enhancer drives *avpr1a* genotype differences in CpG density and methylation, which predict *avpr1a* expression and RSC-V1aR. Examination of *avpr1a* methylation among wild-caught voles also showed that RSC-V1aR correlated with enhancer methylation, possibly due to genotype differences in enhancer silencing or affinity for transcription factors, but not with promoter methylation. We also found that genotype differences in RSC-V1aR emerge in the first postnatal week, along with changes in enhancer methylation. Before this neurodevelopmental stage, the LO allele, which has more enhancer CpGs, is more sensitive to environmentally-induced changes in RSC-V1aR. These changes however, are not caused by alteration of enhancer methylation, suggesting additional regulatory elements contribute to genotype differences in RSC-V1aR regulation and its environmental sensitivity. Our findings show how genetic and epigenetic variation at a critical gene can shape intraspecific variation in brain and behavior.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Phenotypic variation is commonly found among individuals of the same species. These individual differences can range from morphological variation in form and function, such as color polymorphism in the rock pocket mouse (Nachman et al., 2003) or shell chirality in great pond snails (Sturtevant, 1923), to complex differences in social behavior, such as alternative mating strategies among male side-blotched lizards (Sinervo and Lively, 1996) or personality differences among humans (McCrea and Costa, 1999). Despite their prevalence, we know relatively little about the mechanisms that drive intraspecific phenotypic variation, especially in social behaviors. This is partially due to the complex nature of behavioral phenotypes, but also a result of systematic exclusion of genetic variation in inbred strains commonly used in neuroscience and the occasional disregard for intraspecific behavioral variation as experimental noise or non-adaptive deviation from the average species behavior (Lott, 1984). In many cases however, persistent individual differences in behavior are not random or aberrant, but rather results of adaptive diversity in the brain (Dall et al., 2004). Studying the mechanisms of intraspecific variation in brain and behavior, provide a unique opportunity for understanding how behaviors and their different forms arise from complex biological pathways within the brain, how these pathways interact at the molecular level, with each other and the environment, and why some outcomes are eliminated while others are favored by natural selection.

Social behavior often has significant fitness consequences and is directly targeted by natural selection (Dall et al., 2004). Evolutionary theories provide frameworks to explain these consequences and understand why selection may actively maintain intraspecific diversity in brain and social behavior. According to game theory (Smith,

1982), persistent differences in social behavior can coexist within a population when fitness payoffs dependent on competing strategies and the frequencies or density at which they exist in the population (density- and frequency-dependent selection). The optimum composition of the fixed social behaviors may however, be subject to change due to environmental variation. For example, the three mating variants of side-blotched lizards (*Uta stansburiana*) have been actively maintained by an evolutionary stable rock-scissor-paper game dynamic (Sinervo and Lively, 1996). But selection of multiple fixed strategies is not the only route to maintaining intraspecific behavioral differences. Phenotypic plasticity, which allows individuals to change tactic during their lifetimes -- based on their condition or environment circumstances-- is another mechanism that maintains intraspecific variation (Dall et al., 2004; Gross, 1996). For example, old male horseshoe crabs (*Limulus polyphemus*) in poor condition adopt a satellite mating tactic, while young and strong males pair with the female (Brockmann et al., 1994). Female barrow's goldeneye ducks (*Bucephala islandica*) adjust their investment into nesting versus parasitism in response to population density (Eadie and Fryxell, 1992). While such evolutionary frameworks provide invaluable insight into *why* behavioral diversity exists in brain and behavior, they do not inform us about *how* such differences emerge among individuals. To learn about the proximate drivers of intra-specific variation we need to examine the biological mechanisms involved in shaping behavioral phenotypes and how they vary among conspecifics.

Social behavior is shaped by the genetic and environmental factors that regulate brain function. Therefore, variation in brain and behavior often stems from differences in the genome and/or variation in the environment. Intraspecific differences in genetics are commonly found as single nucleotide polymorphisms (SNPs) in DNA sequence. Depending on the genetic context, DNA sequence polymorphisms can have

2

different consequences for individual variation. Most variation among individuals seems to arise from genetics differences in regulation of gene expression, rather than from changes in coding sequences (Robinson and Ben-Shahar, 2002; for discussion and examples see [Wray, 2007]). Sequence change within the coding regions can lead to global alteration of protein structure and function. However, sequence variation in non-coding regulatory regions –such as enhancers— can change regulation of genes expression in complex tissue-specific patterns (Gutierrez-Arcelus et al., 2015; Pastinen, 2010). In addition to genetic differences, variation in environmental experience –such as nutrition (Georgieff, 2007) and maternal care (Weaver et al., 2004)—can also alter neuronal gene expression and behavior. Such environmentally induced changes in gene regulation are often established via "epigenetic" marks (Feil and Fraga, 2012). Epigenetic modifications are structural alterations of chromatin that can have lasting impacts on the activity of genes without changing the corresponding DNA sequence (Bird, 2007). Epigenetic marks such as histone modifications and DNA methylation at CpG dinucleotides can permanently or transiently "reprogram" the genome of an individual (Reik, 2007). Interestingly, genetic differences among individuals can influence their susceptibility to such epigenetic modifications. For example, SNPs at CpG sites (polymorphic CpGs or polyCpGs) lead to individual differences in CpG availability and could result in differences in susceptibility to DNA methylation. Furthermore, DNA sequence variation can influence binding affinity of transcription factors or nucleosomes among individuals (Segal and Widom, 2009). Such genetic differences in sensitivity to environment –known as gene-by-environment interactions (G X E) – can have major implications for intra-specific variation in brain and behavior (Pigliucci, 2001). Thus, to fully understand mechanisms of variation in social behavior, we need to examine the genetic, epigenetics and environmental regulation of genes critical to brain function and

3

ask how they differ among individuals. Among such genes, the oxytocin- and vasopressin-related peptides, as well as their receptors, seem particularly important to social behavior (Donaldson and Young, 2008).

Oxytocin- and vasopressin-related nonapeptides play a crucial role in the modulation of brain function and social behaviors across taxa. In mammals, oxytocin and vasopressin are produced within the hypothalamus and then transferred to the pituitary gland for peripheral release or projection to other brain regions (Young and Wang, 2004). Oxytocin and vasopressin peptides have important physiological and behavioral roles in both males and females, but they often show sexual dimorphism in expression and behavioral effects (De Vries and Panzica, 2006). Oxytocin (OT) typically influences female-specific sociosexual behaviors. For example, in the periphery mammalian OT is involved in processes such as parturition and lactation and the in the brain, it regulates behaviors including female pair-bonding and maternal receptivity and attachment (Lee et al., 2009). Vasopressin (VP) however, typically influences male-specific behaviors. Across different mammalian species, peripheral VP is shown to modulate male erection and ejaculation, and in the brain, VP is involved in male-typical social behaviors such as aggression, paternal care and male pair-bonding (Donaldson and Young, 2008). Interestingly, the exact behavioral functions of OT and VP depend to a large extend on differences in the neuronal distribution and density of their receptors (Goodson and Bass, 2001). Unlike OT, which only has one receptor, vasopressin has two neuronal receptor subtypes: V1a and V1b and a peripheral receptor V2. Evidence suggests vasopressin's roles in social behavior are mostly mediated via the receptor V1a subtype (V1aR, [Lim and Young, 2006]). Much of what is known about the role of V1aR in social behaviors is acquired by studying the neurobiology of sociosexual behaviors among vole species

(*Microtus* genus), especially the socially monogamous prairie vole (*Microtus ochrogaster*, [McGraw and Young, 2010]).

Prairie voles are small North American rodents commonly used to study evolutionary and neurological basis of social monogamy (McGraw and Young, 2010), a social system present in less than 5% of mammalian species. Prairie voles form long-term pair bonds, provide bi-parental care to offspring, and aggressively defend their home range against intruders (Getz et al., 1993). Regulation of these behaviors is highly dependent on the vasopressin system. Various studies, including pharmacological and gene-transfer experiments, have shown that V1aR has a crucial role in mediating social attachment in male prairie voles (Lim et al., 2004; Young et al., 1999). Interestingly, distribution of V1aR in the prairie vole brain is dramatically different compared to related non-monogamous voles; comparative studies have shown that this difference, especially at the ventral pallidum within the reward circuitry, is causally linked to species differences in mating systems (Lim et al., 2004; Lim and Young, 2004; Young et al., 1999). Surprisingly however, neuronal V1aR also exhibits profound variation within the prairie vole species (Phelps and Young, 2003). The intraspecific V1aR variation is low in brain regions important to pair-bond formation, such as the ventral pallidum, but high within brain regions implicated in spatial memory, such as the laterodorsal thalamus (LDThal) and retrosplenial cortex (RSC, (Ophir et al., 2008b; Phelps and Young, 2003). Considering the roles of RSC in navigation and V1aR in male social behavior, it is perhaps not surprising that variation in abundance of V1aR in the RSC (RSC-V1aR) has important implications for individual differences in sexual and spatial behaviors among male prairie voles (Ophir et al., 2008b).

Like many other species, prairie voles vary in aspects of their social behavior. In fact, despite being socially monogamous, prairie voles vary in sexual fidelity (Ophir et

al., 2008a). These individual differences in sexual fidelity separate males into extra-pair fertilizing (EPF) and intra-pair fertilizing (IPF, [Ophir et al., 2008a]). EPF and IPF males face interesting reproductive and fitness tradeoffs (Okhovat et al., 2015; Ophir et al., 2008b). The unfaithful EPF males have large home-ranges that overlap with multiple other males and females. They often intrude into adjacent home-ranges and mate with other females, but at the cost of being cuckolded (Phelps and Ophir, 2009). The faithful IPF males however, have small exclusive home-ranges. They rarely intrude into other home-ranges and are more successful in mate-guarding (Okhovat et al., 2015; Ophir et al., 2008b). Interestingly, these differences in male sexual and spatial fidelity are predicted by V1aR abundance within the RSC (RSC-V1aR). The IPF males have significantly higher levels of V1aR in their RSC compared to EPF males (Okhovat et al., 2015; Ophir et al., 2008b). Despite much effort, the mechanisms that drive and maintain the extraordinary intraspecific variation of V1aR in the prairie vole brain are not fully known yet. Previous studies have attempted to understand the proximate mechanisms of prairie vole V1aR variation by linking it to genetic variation at the encoding gene, *avpr1a*. One prominent study suggested that length of a microsatellite upstream of *avpr1a* caused differences in neuronal V1aR (Hammock and Young, 2005; Young et al., 1999). However, the inconsistencies evident between different studies suggest this relationship is not causal. So far, the singular focus on microsatellite length has prevented intensive exploration of other *cis*-regulatory elements. As a result, the molecular basis of intraspecific RSC-V1aR variation and its behavioral consequences are not yet understood.

In this dissertation we take a candidate gene approach to extensively explore intraspecific genetic and epigenetic variation at the *avpr1a* locus. We ask how this variation and its interactions with the environment shape individual differences in RSC-

V1aR. Briefly, in chapter 2 we used lab-reared prairie voles in semi-natural enclosures to further examine the behaviors of IPF and EPF males and relate them to RSC-V1aR variation. We used the same animals along with an independent population of wild-caught prairie voles to find common SNPs and examined how well they predict RSC-V1aR. We identified two *avpr1a* alleles, "HI" and "LO". These alleles reliably predicted RSC-V1aR, as evident by a new controlled replication breeding experiment. We found that the HI and LO alleles differ in number of CpG sites within a putative intron enhancer, as a result of linked SNPs occurring at CpG sites (polyCpGs). Genotype differences in CpG density caused significant differences in putative intron enhancer methylation, which correlated with differences in *avpr1a* transcript abundance and RSC-V1aR. We also found that the reproductive tradeoffs between IPF and EPF reflected in the fitness of the HI and LO alleles and found evidence that the putative intron enhancer has been under balancing selection (Okhovat et al., 2015). In chapter 3 we used our wild-caught prairie voles to characterize the distribution of fixed and polymorphic CpGs across the *avpr1a* locus. We extended our DNA methylation measurements beyond the putative intron enhancer and assayed a ~3kb region of the *avpr1a* locus, including the promoter, first exon, half of the intron and intron enhancer. We found that the putative intron enhancer, which has an unusually high density of polyCpGs, is the only gene feature where individual differences in DNA methylation predict levels of RSC-V1aR. We showed that this association might be driven by local chromatin density differences at the putative enhancer or allele-differences in transcription factor binding. In chapter 4 we explore the early ontogeny of RSC-V1aR and find that HI and LO genotype differences in expression emerge during the first week of life, along with DNA methylation changes in the putative intron enhancer. We observe that the LO allele is more sensitive to early environmental changes as evident by RSC-V1aR change in response to pharmacological

manipulations at day 1. These allele differences in sensitivity are consistent with GxE effects caused by differences in enhancer CpG availability and methylation susceptibility. However, our findings show that methylation at the putative intron enhancer does not drive these GxE effects. Genome-wide analysis of HI and LO DNA methylation suggest more distal elements may contribute to *avpr1a* regulation in the RSC and its GxE effects early in life. Overall, our findings demonstrate how small genetic differences in sequences of a crucial gene can change individuals' brain and behavior by interacting with the epigenome and environment in complex ways.

Chapter 2: Sexual fidelity trade-offs promote regulatory variation in the prairie vole brain[1]

ABSTRACT

Individual variation in social behavior seems ubiquitous, but we know little about how it relates to brain diversity. Among monogamous prairie voles, vasopressin receptor (*avpr1a*) levels in brain regions related to spatial memory predict male space-use and sexual fidelity in the field. We find that trade-offs between the benefits of male fidelity and infidelity are reflected in patterns of territorial intrusion, offspring paternity, *avpr1a* expression and the evolutionary fitness of alternative *avpr1a* alleles. DNA variation at the *avpr1a* locus includes polymorphisms that reliably predict the epigenetic status and neural expression of *avpr1a*, while patterns of DNA diversity demonstrate *avpr1a* regulatory variation has been favored by selection. In prairie voles, trade-offs in the fitness consequences of social behaviors seem to promote neuronal and molecular diversity.

INTRODUCTION, RESULTS, DISCUSSION:

Social behavior emerges from the complex, dynamic, and often strategic interactions of individuals – a complexity that places it among the most challenging and interesting behaviors to study. Neuroscience has elucidated many mechanisms of social behavior (Gross, 1991; Young and Wang, 2004). In parallel, evolutionary biology has outlined how social interaction can promote variation within a species (Gross, 1991; Pfennig, 1992; Sinervo and Lively, 1996). Frequency- or density-dependent selection, for example, maintains individual differences in the parental care of sunfish (Gross, 1991),

---

[1] This chapter is a multi-authored published article. This is the author's version of the work. It is published here by permission of the AAAS for personal use, not for redistribution. The definitive version was published in Science Journal Title 350 (2015), doi:10.1126/science.aac5791. Full citation: Okhovat, M., Berrio, A., Wallace, G., Ophir, A.G., Phelps, S.M., 2015. Sexual fidelity trade-offs promote regulatory variation in the prairie vole brain. Science 350, 1371–1374.

territorial defense of lizards (Sinervo and Lively, 1996), and cannibalistic behavior of tadpoles (Pfennig, 1992). Among humans, similar forces have been proposed to explain differences in personality, resilience and psychiatric risk (Ellis and Boyce, 2008; Keller and Miller, 2006; Verweij et al., 2012). Given that social diversity is central to behavioral ecology, social psychology, and mental health, it is surprising we know so little about natural variation in the social brain, how it emerges from the interaction of genetic and epigenetic processes, or how it has been sculpted by evolutionary forces.

We explore individual differences in neuronal gene expression in the monogamous prairie vole, *Microtus ochrogaster*, a small North American rodent in which males and females form pair-bonds and share parental care (Getz et al., 1993). Prairie vole pair-bonding is governed by multiple modulators and brain regions (Carter et al., 1986; Cushing and Kramer, 2005; Young and Wang, 2004). Of these genes, the vasopressin 1a receptor (V1aR, encoded by *avpr1a*) is particularly well studied (Cushing and Kramer, 2005; De Vries and Panzica, 2006; Ophir et al., 2008b; Phelps and Ophir, 2009; Phelps and Young, 2003; Young and Wang, 2004). V1aR expression can vary profoundly across individual prairie voles (Phelps and Young, 2003), and its abundance in a spatial-memory circuit predicts sexual fidelity in males (Ophir et al., 2008b; Phelps and Ophir, 2009) but not females (Supplementary Results, Table 2.1) – a finding consistent with male-specific vasopressin effects in other contexts (De Vries and Panzica, 2006). We use the relationship between *avpr1a* expression and male fidelity to examine how social forces contribute to brain diversity. Specifically, we ask whether fitness consequences of male sexual fidelity promote genetic and epigenetic variation in *avpr1a*.

Although prairie voles are socially monogamous, they are not sexually exclusive (Ophir et al., 2008a). Approximately 25% of young are conceived outside a pair-bond (termed extra-pair fertilizations, or EPFs). Male fidelity is often thought to depend on

spatial strategies that balance the demands of mate-guarding against the value of mating multiply (Emlen and Oring, 1977; Kokko and Rankin, 2006). To examine the relationship between space-use and sexual fidelity among male prairie voles, we estimated the intensity of a male's space-use by fitting kernel density estimates to animal positions measured over several weeks by radiotelemetry (Figures 2.1A,B, 2.2). By overlaying these maps of space-use intensity, we could estimate how often males encounter other individuals either at home or in neighboring territories. We find that the spatial behavior of EPF-males differs from that of males who sire young only with a partner (intra-pair fertilizations, IPF). EPF-males have larger home-ranges (p<0.05; Figure 2.1C), and they more frequently encounter extra-pair females (p<0.0001; Figure 2.1D), intrude on territories (p<0.01; Figure 2.1E), and are intruded upon (p<0.01; Figure 2.1F). The rate at which a male intrudes on a neighbor's territory is correlated with the rate that he encounters extra-pair females (r=0.69, p<0.0001), but also with the rate he is intruded upon by other males (r=0.83, p<0.0001; Figure 2.1G). Overall, the data suggest that venturing away from a male's core home-range increases encounters with both extra-pair females and their aggressive mates; these intrusions may offer the opportunity for extra-pair paternity, but they also increase the rates at which a male's home-range is visited by neighboring males. This pattern is consistent with data suggesting pair-bonded EPF males are more likely to be cuckolded (Phelps and Ophir, 2009). Increasing extra-pair female encounter rate seems to come at the expense of intra-pair mate-guarding.

Among prairie voles, we find that neuropeptide receptors show profound variation in nodes of a spatial memory circuit including the hippocampus, laterodorsal thalamus (LDThal), and retrosplenial cortex (RSC; Figure 2.1H); remarkably, variation in each of these regions predicts aspects of space-use and paternity in the field (Ophir et al., 2008b; Phelps et al., 2010). The relationship between spatial memory and sexual fidelity is not

11

clear, but males with low V1aR in RSC or LDThal have been hypothesized to have a poor memory for locations of aggressive interactions, a cognitive strategy that could promote territorial intrusion and extra-pair encounters (Phelps and Ophir, 2009). In contrast, a male with abundant V1aR may better monopolize a mate, but might encounter fewer extra-pair females. To look for evidence of fitness trade-offs that could promote forebrain diversity, we examined the relationship between RSC-V1aR and our measures of space-use. As reported previously, faithful IPF males have more RSC-V1aR than EPF males (p<0.001, Figure 2.1I-K; [Ophir et al., 2008b]). Low levels of RSC-V1aR were also associated with high intrusion rates (RSC, p<0.01; pairing status, p<0.0001; RSC x status, p<0.05; Figure 2.1L) and poor mate guarding (male visits received: RSC, p<0.05; pairing status, p<0.0001; RSC x status, p>0.10). Interestingly, V1aR in another node in this circuit (LDThal) exhibited similar patterns, while brain regions associated with pair-bonding and aggression (ventral pallidum, lateral septum) did not predict space-use or sexual fidelity (Supplementary Results; [Ophir et al., 2008b]). These data suggest trade-offs between the fitness benefits of intra-pair and extra-pair paternity could contribute to diversity in this memory circuit.

In order for selection to have promoted neuronal diversity, such variation must be heritable. We asked whether single-nucleotide polymorphisms (SNPs) in *avpr1a* predicted individual differences in V1aR abundance. We sequenced ~8kb of the *avpr1a* locus (Figure 2.3, 2.4) from lab-reared males with substantial field data (Figure 2.1; [Ophir et al., 2008b]) and from wild-caught adults. Of 151 SNPs, 4 tightly linked polymorphisms predicted RSC-V1aR (Figure 2.4, multiple-test corrected α=5.4e-4). These SNPs were upstream of the coding sequence (SNP -1392, p=6.3e-6), in the intron (SNPs 2170 & 2676, p=4.7e-6), and in the second exon (SNP 3506, p=5.0e-5). We refer to the genotypes defined by these linked SNPs as HI- and LO-RSC alleles. Interestingly,

effects of HI and LO alleles were stronger among lab-reared animals (p<0.0001) than wild-caught animals (p<0.05; genotype x rearing p=0.002; Figure 2.5), suggesting population structure or developmental environment may influence cortical V1aR. Lastly, we found that a distinct SNP predicted V1aR in the LDThal (SNP 5168, p=3.6e-4), but none of the 151 SNPs predicted V1aR in the ventral pallidum or lateral septum (Figure 2.6). Thus, V1aR levels in regions implicated in spatial memory and sexual fidelity were linked to *avpr1a* sequence variation, while regions important in pair-bonding and aggression were not.

We examined the stability and specificity of the HI- and LO-RSC associations with a breeding design that controlled for potential confounds of our initial study. We obtained a new genetic stock from a third site >100mi from prior sites. Heterozygous HI/LO parents were crossed to produce siblings that shared a common genetic background, rearing environment and lack of sexual experience but differed in their genotypes. We again found that HI and LO alleles influenced V1aR in the RSC (p<0.0001; Figure 2.7A), but not in other brain regions (Figure 2.8). Thus our data demonstrate a replicable, robust and specific association between the HI-RSC allele and high RSC-V1aR expression. However, differences between wild-caught and lab-reared animals (Figure 2.5), as well as previously reported developmental manipulations (Bales et al., 2007), suggest epigenetic variation may also be at play.

If individual differences in RSC-V1aR abundance are due to differences in the regulation of *avpr1a*, then HI/HI and LO/LO genotypes should differ in *avpr1a* transcript abundance. We dissected the RSC of lab-crosses reported above and used qPCR to quantify *avpr1a* mRNA. Genotypes differed significantly in *avpr1a* transcript abundance (ΔCt vs. *β-actin*, p<0.001, Figure 2.7B). Moreover, individual differences in *avpr1a*

mRNA were strongly associated with RSC-V1aR protein (R^2=0.75, p<0.0001; Figure 2.7C).

To determine whether any RSC-associated SNPs were within DNA-sequences that might contribute to *avpr1a* regulation, we performed chromatin immunoprecipitation sequencing (ChIP-seq) targeting the histone modification H3K4me1, a marker for regulatory sequences known as enhancers (Heintzman et al., 2007). We dissected RSC samples from 8 novel lab-reared individuals. Within a 25kb sequence centered on the *avpr1a* translation start site (Figure 2.4D), the H3K4me1-mark was uniquely associated with two regions within the *avpr1a* locus (p<1e-7, q<0.0001; Supplementary Results). One putative enhancer was in the center of the intron, including both intron SNPs of the HI/LO alleles; the second overlapped the second exon and included the fourth linked SNP (Figure 2.4C). Remarkably, three of the polymorphisms that define the HI and LO alleles are within putative enhancer regions, while the fourth is within a conserved DNAse I hypersensitive site (Figure 2.4A, [Rosenbloom et al., 2013]). Thus all four RSC-associated SNPs coincide with markers of transcriptional regulation.

We next asked whether differences in RSC *avpr1a* transcript and V1aR protein abundance reflected differences in the epigenetic state of the *avpr1a* locus. We focused on the putative intron enhancer: this sequence had strong evidence of H3K4me1 enrichment and included the two SNPs most strongly linked to RSC-V1aR. SNP 2170 proved to be a G/T polymorphism that altered the presence of a CpG site, a common target of DNA methylation (Razin and Riggs, 1980). Moreover, this CpG/CpT polymorphism is linked to a cluster of CpG polymorphisms within the enhancer (Figure 2.7D). HI-RSC alleles have fewer CpG sites than LO alleles (p<0.0001, Figure 2.7E), suggesting fewer opportunities for methylation. We isolated DNA from the RSC, treated it with bisulfite and performed pyrosequencing of this enhancer. HI/HI animals had less

enhancer methylation than LO/LO animals (p<0.0001, Figure 2.7F). Genotypes also differed in enhancer methylation if we focused solely on non-variable CpG sites ($\mu\pm$SE, HI/HI 67.6±1.6%, LO/LO 75.6±1.3%; p=0.001). Moreover, *avpr1a* enhancer methylation is significantly associated with RSC-V1aR abundance (p<0.0001, Figure 2.7G). Methylation at non-coding CpG sites is known to recruit methyl-binding proteins, histone deacetylases and other silencing proteins (Nan et al., 1998); our data suggest that SNP 2170 and neighboring CpG polymorphisms may alter the function of an intron enhancer by changing the number of CpG sites available for methylation.

Our molecular data indicate that specific alleles are robust predictors of RSC-V1aR, and suggest mechanisms by which specific SNPs might exert influence on *avpr1a* expression. If genetic differences in RSC-V1aR are adaptive – a "balanced polymorphism" of the brain -- we might expect differences in how HI- and LO-RSC alleles gain fitness. Using data from lab-reared animals monitored in the field (Figure 2.1), we calculated the number of embryos each male sired either with a partner (IPFs) or non-partner (EPFs), and estimated the relative fitness of HI and LO alleles in each context. Although the alleles had similar fitness overall, selection favored HI alleles in the context of IPFs, and LO alleles in the context of EPFs (Figure 2.9A, p<0.05). Thus, fluctuations in the defensibility of females could profoundly influence the strength and direction of selection on HI and LO alleles. Interestingly, prairie voles exhibit wide fluctuations in population density, ranging from ~25-600 voles per hectare in a year (Getz et al., 2001); high densities increase the rate of extra-pair interactions (McGuire et al., 1990) and reduce the defensibility of prairie vole females (Streatfeild et al., 2011). Manipulative studies will be needed to test whether fluctuations in population density or allele frequency promote variation in *avpr1a* and related behaviors.

If genetic variation at *avpr1a* produces variation in memory regions, and this in turn influences space-use and sexual fidelity, then over time we expect selection to have influenced patterns of *avpr1a* nucleotide variation. We tested for a history of balancing selection by comparing the frequencies of SNPs at *avpr1a* to three putatively neutral nuclear loci among our original wild-caught samples. We found that the *avpr1a* locus was strongly skewed toward an excess of intermediate-frequency alleles, a classic signature of balancing selection (Figure 2.9B, LRT=120.3, df=4, p=4.7e-25). Similarly, *avpr1a* had a positive Tajima's D (p<0.05; [Tajima, 1989]), while our neutral loci had negative values (p>0.10, Figure 2.9C). Lastly, an HKA test (Hudson et al., 1987) comparing the number of within- and between-species differences indicates an excess of standing variation within regulatory regions (defined by H3K4me1 ChIP-seq and DNAse HS; p<0.01, Figure 2.9D). We conclude that balancing selection has actively maintained regulatory variation at the *avpr1a* locus. This regulatory variation seems to be specifically associated with brain regions related to space-use.

These data provide a remarkably coherent perspective on the origin and maintenance of diversity in the social brain. V1aR levels in spatial-memory structures predict whether males will intrude on neighbors and gain extra-pair paternity, or exclude intruders and improve intra-pair paternity. Nucleotide polymorphisms within regulatory sequences robustly and specifically predict V1aR variation in these same regions. Within the RSC, we find that low-expressing alleles differ in CpG abundance and methylation status. Because CpG sites can be gained or lost easily – ~25% of single nucleotide differences between humans and chimps, for example, consist of the gain or loss of a CpG site (Chimpanzee Sequencing and Analysis Consortium, 2005) – we hypothesize that CpG polymorphisms may often shape heritable variation in environmental sensitivity. Lastly, genetic markers for this neuronal phenotype exhibit strong evidence of

balancing selection. Together these data suggest that trade-offs in the fitness consequences of spatial behaviors promote diversity in the social brain. By focusing on what would seem the simplest of social phenotypes – the neural expression patterns of a single gene – we gain insights into the complex interplay of forces that shape both gene function and social evolution.

METHODS

All animal experiments were reviewed approved by Institutional Animal Care and Use Committees at the University of Memphis, the University of Florida and the University of Texas at Austin.

Field methods

Subjects

Eight replicate field enclosures were populated with 6 male and 6 female prairie voles, as previously described (Ophir et al., 2008b). This corresponds to a density of ~200 voles per hectare, a density higher than the density that which Getz et al. (2001) define as "medium density", but well within the ranges of densities (~25-600 voles/ha) documented from natural populations (Getz et al., 2001, 1993). Subjects were age- and weight-matched, sexually naïve and lab-reared. Four of the enclosures were derived from animals with origins in Champaign County, Illinois; the other four were derived from Shelby County, Tennessee. Animals were left in the field for 18-22 days, and paternity measures were derived from pups conceived in the field. While our analysis of field data is entirely new, as is all *avpr1a* genetic data, other analyses from these subjects have been published, including microsatellite-based paternity analyses (Ophir et al., 2008a), forebrain V1aR and simple measures of space use (e.g. #male home-range overlaps based on minimum convex polygons; [Ophir et al., 2008b]). Importantly, an extensive

examination of forebrain V1aR and field behavior reveals no population differences in either brain V1aR abundance or behavior in lab or field (Ophir et al., 2007).

Measuring space use

The enclosures were each 600m2; terrestrial mammals were excluded by a small electrified fence; avian and snake predators were not excluded. Manual radiotracking was performed ~2x per day for each animal. We used MATLAB to perform bivariate kernel density estimates from manual radiotracking data. This procedure uses a set of observed positions to estimate the probability that an animal will be at a given point in space. The left panel of Figure 2.2, for example, depicts the intensity of space use for a focal male as a solid surface whose height reflects the probability of observing him at each location in the enclosure. The frequency at which two individuals will encounter one another at a given location can be estimated by the product of their probability densities at that position, and the total rate of interaction between two individuals can be estimated as the sum of these products over space. We used this general logic to estimate 1) the overall rate of encountering same-sex animals (same-sex encounter rate, SSER); 2) the overall rate of encountering opposite-sex animals (OSER); 3) the rate at which a male encountered a territorial resident while intruding on that resident's territory (intrusion rate); 4) the rate at which a focal male's home-range was intruded upon by other males (male visits received); and 5) the rate at which a male encountered extra-pair females while on that female's home-range (extra-pair visits made).

To calculate same-sex encounter rate, we estimated how often a male encountered each other male at each location in the enclosure, and summed this over the enclosure. We next averaged this measure across all males that the focal male could interact with. OSER was calculated in the same way with females. To calculate the

remaining space-use metrics, we operationally defined an animal's home-range by the contour delimiting the region in which it was spending ~75% of time. In the left panel of Figure 2.2, for example, home-range space use is depicted for four resident males by blue contour plots; in the right panel, the home-range of a focal male is depicted in red contours. Thus each animal had a probability density function (pdf) that reflected overall space use (Figure 2.2, solid surfaces), and a second measure of space use that was set to 0 outside the home-range limits. To estimate a male's intrusion rate we examined the overlap between his overall pdf and the home-range of each territorial paired male within the enclosure (Figure 2.2, left). (Single males are not territorial, and so excursions into an unpaired male's home-range were not considered territorial intrusions.) This measure was averaged across the number of paired males in the enclosure; our "intrusion rate" measure is thus an estimate of how often a focal male would encounter a territorial male while intruding on that male's territory. To estimate how often a male was intruded upon (male visits received), we examined the overlap between the home-range of a focal male and the total space-use pdf of each single or paired male in the enclosure (Figure 2.2, right). We averaged this estimate across intruding males. Our measure of "male visits received" is thus an estimate of how often a focal male would encounter a particular intruder in his home-range. Lastly, we calculate "extra-pair visits made" by examining how often a focal male is expected to encounter a female that is not his mate while in her home-range. We focus on this measure of extra-pair interaction because it can be calculated for both paired and single males, since it is not confounded by the behavior of a focal male's potentially territorial female partner, and because it reflects the expected value of excursions into a neighbor's territory. Although our focus is on male behavior, it is worth noting that females exhibited substantial variation in space use and sexual

19

fidelity, and this has been examined elsewhere (Ophir et al., 2008a; Phelps et al., 2010; Zheng et al., 2013).

RSC-V1aR autoradiography

In addition to the animals from the above field study, we caught 32 wild adult males and females from Champaign County, IL. Brains and livers were fresh frozen on dry ice and stored at -80°C. Frozen brains were sectioned in 20-µm-thick slices at 100-µm intervals and mounted on SuperFrost slides (Fisher Scientific). The autoradiography was performed as previously described (Ophir et al., 2008b). Briefly, slides were quickly dried and lightly fixed in 0.1% paraformaldehyde. Slides were then washed in Tris and incubated with 50pM 125I-linear AVP for 60 minutes, followed by additional Tris washes. Finally, sections were dried rapidly under hot air. Sections were exposed to film for 68h with radiographic standards. Developed films were digitized using Microtek ScanMaker 5900. Forebrain V1aR abundance was scored using NIH ImageJ software.

V1aR abundance was measured in four different brain regions: the retrosplenial cortex (RSC), a brain region important for spatial and contextual memory (Vann et al., 2009); the laterodorsal thalamus (LDThal), which is connected to the RSC and also plays a role in spatial memory (Aggleton, 2014; van Groen et al., 2002); the ventral pallidum (VPall), a major reward center – V1aR in the VPall is known to be necessary for male pair-bonding (Lim and Young, 2004; Young and Wang, 2004); and the lateral septum (LS), a brain region extensively implicated in aggressive behavior, but which also seems to regulate male pair-bonding in prairie voles (Liu et al., 2001; Young and Wang, 2004). Background binding in non-expressing cortical regions of the same section as the structure of interest was used to correct for non-specific binding.

Analysis

We compared IPF to EPF males using t-tests, and calculated correlations between space use measures using a Pearson's correlation coefficient. The effect of region-specific V1aR on measures of individual space use was done using a general linear model in which pairing status (single, paired), V1aR abundance (dpm/mg tissue equivalents) and status*V1aR were the terms in the model. Analyses were performed using the statistical software JMP11 (SAS).

Association of SNPs with V1aR

DNA amplification and sequencing

We isolated genomic DNA from ~25mg tissue using the Qiagen DNeasy blood and tissue kit. All amplifications were carried out on a BioRad C1000 Thermocycler. We targeted 6 overlapping amplicons spanning a total of ~8kb, which ranged from approximately ~2.4kb 5' of the transcription start site to ~1kb 3' of the transcription stop site. This span included all sites known to exhibit significant mammalian conservation and *Mus* brain DNAse I hypersensitivity within 10kb of the transcription start site. A nested PCR strategy was used if locus-specific products could not be obtained with a single PCR reaction. Amplicons are shown schematically in Figure 2.3, primers in Table 2.2.

Sequencing and scoring

Following each amplification, samples were visualized on agarose gel to confirm correct size and specificity of amplification. Samples that exhibited non-specific amplification were purified using Qiaquick gel extraction kit (Qiagen), the rest were cleaned by Qiaquick PCR purification kit (Qiagen) according to the manufacturer's protocol. All amplicons were sequenced at the University of Texas at Austin Sequencing

Facility. Chromatograms from Sanger sequencing were analyzed in Geneious 5.5.7 (http://www.geneious.com/) and assembled to an *avpr1a* reference (AF069304.2, NCBI) using MAFFT v6.814b alignment (Katoh et al., 2002). Assembled sequences were then aligned to the reference to identify single nucleotide polymorphisms (SNPs) and insertions/deletions (indels). Variation present in only a single individual was considered a PCR or sequencing error. All variable sites are named based on their positions with respect to the translation start site in the reference sequence.

Association analyses

Because the two sources of animals were run in different autoradiographic assays, we calculated z-scores for each individual based on the mean and standard deviation of lab-reared or wild-caught animals, then pooled these data for association analyses. Using the quantitative trait option of the PLINK software package v1.07 (Purcell et al., 2007), we measured the main effects of all 151 SNPs on V1aR abundance in each of the four regions of interest: retrosplenial cortex (RSC), laterodorsal thalamus (LDThal), ventral pallidum (VPall), and lateral septum (LS). We report p-values uncorrected for multiple tests as a measure of association strength in Figure 2.4C of the main text and in Figure 2.5. We calculate α values corrected for multiple comparisons using permutation tests. The procedure is implemented in PLINK, and consists of randomly assigning observed genotypes (across all sites) to observed brain V1aR abundance without replacement, calculating associations for each SNP, then recording the highest test statistic (using Wald's T) observed across all SNPs for each of 10,000 replicate permutations. The critical value of the test statistic defined our corrected α for each brain region (RSC, α=5.4e-4; LDThal, α=4.0e-4; LS, α=5.1e-4; VPall, α=3.3e-5). These corrected α values are depicted in Figures 2.4C and 2.5. Lastly, to examine the effects of HI-RSC and LO-

RSC alleles more closely, we used a general linear model implemented in JMP11 software (SAS) to assess the effect of SNP 2170 genotype (HI-RSC vs. LO-RSC) and sample origin (lab-reared vs. wild-caught animals).

Replication of RSC-V1aR associations

We captured 27 new animals from a third population in Jackson County, Illinois, >100 miles from either of our other sites. We identified animals with the rarer HI-RSC allele and crossed them with LO/LO wild-caught animals to increase the frequency of the HI-RSC allele in the lab. Next we set up 5 breeding pairs with parents heterozygous for the HI-RSC and LO-RSC alleles to generate pups that were homozygous for either allele; these genotype-discordant littermates shared common parents, rearing environment and a lack of sexual experience. Using amplification and sequencing settings described above, we confirmed the phase of all four SNPs (-1392, 2170, 2676 and 3506) in our breeding pairs.

At weaning, tail clippings (~0.5 cm) were taken from the offspring. Genomic DNA was extracted from the samples using DNeasy blood and tissue kit (Qiagen). Samples were then genotyped using an allele-specific digestion protocol. Briefly, we used a nested PCR strategy to amplify 800bp of the intron around SNP 2170. PCR products were then digested with Bsh1236 I enzyme (ThermoScientific) following settings recommended by manufacturer. The 2170 SNP is a polymorphism that produces a CGCG or CGCT sequence, allowing Bsh1236 I to digest the low expressing allele (CGCG). Allele specific digestion allowed genotype determination via agarose gel imaging. Because all four sites were in phase in the parents, it was only necessary to genotype the pups at SNP 2170. Brains of homozygotes (n=14 LO/LO, n=12 HI/HI) were cryosectioned and subject to autoradiography as described above.

Epigenetic regulation of *avpr1a* expression

qPCR measures of RSC avpr1a mRNA abundance

After obtaining anterior sections from the RSC by cryosectioning for autoradiography, the remainder of RSC was micro-dissected from remaining brain tissues (n=24; two samples were lost). Total RNA was extracted with TRIZol reagent (Life technologies) following manufacturer's instructions. We used QuantiTect Reverse Transcription Kit (Qiagen), to construct cDNA libraries from 2ug total RNA according to manufacturer's protocol.

SYBR-green quantitative-PCR (qPCR) assays were designed to quantify abundance of RSC *avpr1a* transcript relative to endogenous controls *β-actin* and *gapdh*. All qPCR reactions were carried out on a ViiA Real Time PCR system (Life Technologies) in 10ul reactions consisting of 5ul of KAPA SYBR FAST qPCR master mix (2X) universal (Kapa Biosystems), 0.2ul ROX low, 200nM of *avpr1a* primer or 250nM control primers, and 1ul cDNA library. Amplifications were performed using the following settings: enzyme activation and DNA denaturation at 95°C for 1min, followed by 40 cycles of 1sec denaturation at 95°C, primer annealing and extension for 20sec at 60°C.

Following qPCR, amplification curves were analyzed in ViiA 7 software v1.2.1 (Life Technologies). The amplification cycle in which a significant fluorescence threshold was reached (Ct value) was used to quantify abundance. Each PCR reaction was performed in triplicate and the mean Ct value was used for comparison with endogenous controls. The ΔCt value was normalized by subtracting the mean reference *β-actin* and *gapdh* Ct values from the average target *avpr1a*-Ct.

Bisulfite pyrosequencing

We examined methylation levels in the putative intron enhancer focusing on a polymorphic CpG site (SNP 2170) and adjacent CpG sites. RSC tissue was dissected from frozen brain sections of the lab-crossed voles used above to examine RSC-V1aR and *avpr1a* transcript abundance. Using EpiTect Plus LyseAll Bisulfite Kit (Qiagen) we lysed cells, extracted and bisulfite-converted genomic DNA according to the manufacturer's instructions. A nested PCR strategy was used to produce specific amplicons for bisulfite pyrosequencing. The outer PCR reactions were set up in 25ul volume, consisting of 12.5ul of 2X KAPA HiFi Uracil+ mix (KAPA Biosystems), 300nM of each primer (Table 2.3) and 1.5ul of bisulfite converted genomic DNA. All PCR amplifications were carried out in a BioRad C1000 thermal cycler. Using Q24 PyroMark assay design software (Qiagen), two separate inner-PCR assays were designed to target smaller regions containing 3-4 intron CpG sites for pyrosequencing, and yielding methylation data for a total of 7 CpG sites. SNP 2170 and two other CpGs sites were polymorphic, while the remaining 4 CpG sites were not (Figure 2.7D). Following amplification, final PCR products were visualized on 2% agarose gel to ensure specific and sufficient amplification.

The biotin-labeled amplicons were sent to Epigendx (Hopkinton, MA) for bisulfite pyrosequencing. Sequencing primers were designed by Epigendx (Table 2.3). DNA methylation was reported as %(unconverted C/ [unconverted C+ converted T]) for each CpG site. We used t-tests to compare HI/HI and LO/LO offspring in *avpr1a* transcript abundance (ΔCt with respect to *β-actin*), #CpG sites, %methylation of all CpG sites (fixed + polymorphic), and %methylation of fixed CpG sites. Linear regressions were used to relate log(V1aR abundance) to either ΔCt or %methylation at focal intron CpG sites.

Chromatin Immunoprecipitation

We set up 4 additional heterozygous breeding pairs. Male offspring from each breeding pair were genotyped using the allele-specific digestion method described above. From each breeding pair, we selected one male from each homozygous genotype (4 breeding pairs x 2 homozygous genotypes = 8 total subjects). Subjects were euthanized post-weaning and their brains were immediately harvested. The RSC was micro-dissected from the fresh brain and fixed in 1% paraformaldehyde for 10min rotating at RT. The cross-linking process was terminated by addition of glycine (125mM final) and washed three times with PBS containing proteinase inhibitors (PI and PMSF). Tissue was homogenized using a manual douncer and washed again with PBS (+ PI and PMSF). Next, cells were lysed in buffer (5mM PIPES pH=8.0, 85mM KCl, 0.5% NP40) containing PI and PMSF on ice for 15min. After cell lysis, the nuclei were pelleted and lysed in nuclear lysis buffer (50mM Tris-HCl pH=8, 10mM EDTA, 1% SDS) containing PI and PMSF on ice for 10min. Chromatin was sonicated on ice by Q125 sonicator (Qsonica) to generate fragments with a size range of 100-300bp.

For immunoprecipitation, 150ul chromatin was pre-cleared with 20ul Protein A Dynabeads (Invitrogen) and 250ul dilution buffer (0.01% SDS, 1% Triton X-100, 2mM EDTA, 20mM Tris-Cl pH=8, 150mM NaCl) for 2 hours at 4°C. We stored 15ul aliquots of pre-cleared chromatin at 4°C for use as control INPUT DNA. The rest of the chromatin was incubated with 6ug of H3K4me1 antibody (Thermo Scientific, PIPA517418) rotating at 4°C overnight. After immunoprecipitation, beads were washed sequentially with fresh RIPA buffer (50mM Tris pH= 8.0, 150mM NaCl, 0.1% SDS, 0.5% Sodium Deoxycholate, 1% NP40, 1mM EDTA), high-salt buffer (50mM Tris pH=8.0, 500mM NaCl, 0.1% SDS, 0.5% Sodium Deoxycholate, 1% NP-40, 1mM EDTA), LiCl buffer (50mM Tris pH=8.0, 250mM LiCl, 1% NP-40, 0.5% Sodium

Deoxycholate and 1mM EDTA) and twice with 1x TE buffer. DNA was released from Dynabeads twice in fresh elution buffer (0.1M NaHCo3, 1% SDS) with 15 minute incubation at 65°C and occasional vortexing. Eluates were combined and INPUT was diluted in fresh elution buffer. Samples were then incubated at 65°C for at least 4 hours in presence of NaCl and RNase A (Fisher) to remove traces of RNA and reverse the protein-DNA cross-links. Next, samples were treated with Proteinase K (NEB) to remove protein. Finally, DNA was isolated according to standard phenol:chloroform procedure, precipitated and washed by EtOH and dissolved in 10ul DNase-free water.

Library preparation and sequencing

Using a KAPA Library Preparation Kit (Kapa Biosystems), 8-10ul of H3K4me1 ChIP and INPUT was end-repaired, adenylated and indexed with NEXTflex DNA barcodes (Bioo Scientific). Barcoded libraries were PCR amplified for 4-6 cycles. All steps were carried out according to manufacturer's instructions. To avoid sample loss, size selection was avoided. Following inspection of the fragment size distribution, we examined the concentration and quality of libraries on an Agilent 2200 TapeStation Instrument. Samples were sequenced on the Illumina Hi-Seq 2500 platform at University of Texas at Austin Genome Sequencing and Analysis Facility (GSAF, ~6 million 100XPE reads/per sample).

Raw read quality was evaluated and approved with FastQC (Andrews, 2010) for all samples. Since the *avpr1a* locus is absent in the published draft prairie vole genome assembly (http://www.broadinstitute.org/software/allpaths-lg/blog/?p=618), we manually added the BAC clone that includes the *avpr1a* locus (NCBI accession # DP001225) as a contig to the genome assembly. Next, H3K4me1-seq and INPUT reads were aligned to the prairie vole draft genome assembly using both bwa (Li and Durbin, 2009) and

Stampy (Lunter and Goodson, 2011) with default settings for paired-end reads. Since H3K4me1 identifies enhancers, but does not distinguish between active and inactive transcriptional states, data were pooled across genotypes. Aligned reads were filtered based on mapping quality to remove reads with poor or non-unique mapping (mapping quality<20). To eliminate noise and account for unequal total read numbers we used Model-based Analysis of ChIP-Seq (MACS2, [Zhang et al., 2008]) to find peaks with significant enrichment compared to the INPUT control (q-value cutoff=0.05). The output includes peak location range, summit, fold enrichment and q-value (p-value corrected for genome-wide comparisons).

Detecting selection on *avpr1a*

Estimating relative fitness HI and LO alleles

We estimated the fitness of HI and LO alleles by measuring the number of pups sired by males of each genotype (see [Ophir et al., 2008a, 2008b]) for details of paternity analysis). We then calculated the expected frequency of the alleles in the subsequent generation, and divided by the number of alleles present among adult males. Values were converted to relative fitness measures by dividing by the fitness of the fittest allele. We then repeated this calculation but limiting it to pups sired by intra-pair fertilization (IPF), or to pups sired through extra pair fertilization (EPF). To measure differences in allele fitness, we randomized the assignment of fertilization events (defined as all pups within a litter sired by the same father) and calculated the difference in relative fitness of the two alleles. For each randomization, we examined the i) difference between alleles in relative fitness across all fertilization contexts (total), and ii) the context-specific differences in relative fitness of the alleles, which we defined as $(HI-LO)_{IPF}-(HI-LO)_{EPF}$. We repeated this calculation 10,000 times to determine null distribution from which we could

calculate a p-value for the observed data. Results are reported in the main text (Figure 2.9A).

Population genetic analyses were conducted exclusively on DNA from the 32 wild-caught animals used in the association analysis. To examine the genetic structure of population variation in non-*avpr1a* loci, we sampled variation at noncoding sequences of three putatively neutral loci in 10 of our 32 individuals: LCAT (851bp), β-fibrinogen (589bp), and AP5 (423bp). These three loci were PCR amplified with the Promega chemistry and appropriate primers (Table 2.4) following manufacturer's recommendations. Targets were amplified by initial denaturing for 3min at 95°C, 35 cycles of denaturation at 94°C for 30sec, annealing at 60°C for 30sec, extension at 72°C for 1min, followed by one final extension step at 72°C for 5min. Sequencing reads for all loci were assembled and aligned in Geneious v6.1.7.

Calculation of population genetic summary statistics

Tajima's D is a traditional population genetics statistic that compares two measures of nucleotide diversity which, under neutral evolution and stable demography, should be equal (Tajima, 1989). Thus under neutral evolution and stable demography, the expected value is 0; a value above 2 is considered evidence of balancing selection at a locus, while a value of -2 or less is associated with positive or purifying selection. Demographic factors can also produce non-zero values of Tajima's D, so it is important to compare results to data from loci not hypothesized to be under strong selection. We used DNAsp (Librado and Rozas, 2009) with default parameter settings to calculate Tajima's D for the *avpr1a* locus and for the putatively neutral loci, and to perform coalescent simulations for estimating associated p-values for each.

To more directly compare the frequency spectrum of standing variation at the *avpr1a* locus and the putatively neutral loci, we used a likelihood ratio test. We first counted the number of derived polymorphic non-coding sites in each of five frequency bins, spanning from derived alleles that were rare (0.0-0.2 frequency) to derived alleles that were common (0.8-1.0). Derived and ancestral states were defined with respect to three outgroup taxa: *Microtus pennsylvanicus*, *Microtus pinetorum* and *Microtus richardsoni*. To exclude singletons that might be the result of sequencing error, sites at a frequency of less than 0.05 were excluded. (This corresponds to the frequency of a singleton in the neutral loci where 20 haplotypes were examined.) We consider this to be conservative, because it results in a lower estimate of low-frequency SNPs in the neutral data; the hypothesis we are testing posits a shift toward intermediate frequencies at the *avpr1a* locus.

To examine heterogeneity in the distribution of polymorphisms within the *avpr1a* locus, we calculated the number of within-species polymorphisms in regulatory regions (defined as regions within boundaries of the DNAse I hypersensitivity or H3K4me1 ChIP-seq peaks) or outside of these regulatory regions. We next calculated the number of fixed differences between *M. ochrogaster* and the closely related monogamous species *M. pinetorum* in these same regions. We compared within:between species differences in regulatory and non-regulatory regions with a Fisher's exact test (Figure 2.9D).

SUPPLEMENTARY RESULTS

Field results

IPF and EPF males differ in many but not all aspects of space use. EPF males have higher rates of same-sex encounters (p=0.008), but do not have higher rates of opposite-sex encounter rates than IPF males (p=0.12). EPF males are more likely to

intrude on a resident's territory (p=0.0025), more likely to encounter an extra-pair female while out of their own home-range (extra-pair visits made, p<0.0001), and receive visits from surrounding males more often (p=0.0031). As reported in the main text, the rate at which a male intrudes on another male's territory is correlated with both extra-pair visits made (r=0.69, p<0.0001), and with the rate at which neighboring males intrude upon a focal male's home-range (r=0.83, p<0.0001).

We previously reported that RSC-V1aR and LDThal-V1aR differ between IPF and EPF males, but VPall and LS do not (Ophir et al., 2008b). In our current data, we found that both RSC-V1aR and pairing status were significantly associated with male intrusion rates (RSC, p=0.0059; Status, p<0.0001; RSC x Status, p=0.034), as well as the rate at which males were intruded upon (RSC, p=0.013; Status, p<0.0001; RSC x Status, p=0.13), but not on any other measures of space use (all p>0.20).

Although the LDThal was not as robustly associated with IPF/EPF differences as the RSC (Ophir et al., 2008b), the overall pattern of association with space use was similar to that seen in the RSC. The LDThal was strongly associated with individual differences in intrusion rate (LDThal, p=0.0001; Status, p<0.0001; LDThal x Status, p=0.0002) and male visits received (LDThal, p=0.0068; Status, p<0.0001; LDThal x Status, p=0.010). It was not associated with other measures of space use (p>0.10). V1aR abundance in the VPall and LS were not significantly associated with any measures of space use (p>0.10 for all VPall effects, p>0.07 for all LS effects).

Association of SNPs with V1aR

Association analyses

As mentioned in the main text, we found a set of 4 highly linked SNPs that were strongly associated with RSC-V1aR and survived multiple-test corrections (α=5.3e-4).

Of these four linked SNPs, two intron SNPs (positions +2170, +2676) were perfectly linked with one another and exhibited the strongest linkage to RSC-V1aR in our sample (p=4.7e-6). An additional 29 sites were significant only before multiple test correction (5.4e-4<p<0.05). We repeated our association analysis conditioned on the genotype at SNP 2170. We found that even using a liberal, uncorrected criterion (α=0.05), only one SNP remained significant after accounting for the genotype at SNP 2170. Thus 31 of 32 observed associations could be explained by linkage to the HI-RSC and LO-RSC alleles. As mentioned in the main text, we found that the SNP 2170 genotype had a significant effect in both populations (lab-reared, p<0.0001; wild-caught, p<0.05), but the effect of genotype was significantly stronger in lab-reared animals (genotype x rearing environment, p=0.002; Figure 2.5).

We next looked at the strength of association of each SNP for three other brain regions – LDThal, a spatial memory region; VPall, extensively implicated in pair-bonding; LS, implicated in both pair-bonding and aggression in general. In the LDThal, we found 3 SNPs that were significant at the p<0.05 level, one of which remained significant after multiple test correction (SNP 5168, p=3.6e-4, Figure 2.6). Although SNPs in VPall (23 total) and in LS (4 total) were significant at the p<0.05 level, none approached significance after correcting for multiple tests (LS, α=5.1e-4; VPall, α=3.3e-5; Figure 2.6). Thus, two brain regions that predict sexual fidelity and space use in the field also exhibit significant *cis*-regulatory variation at the *avpr1a* locus. In contrast, brain regions that are important for pairing and aggression but do not predict fidelity or space use lack detectable *cis*-regulatory variation. Examination of the specific SNPs that are associated with LDThal and RSC-V1aR reveal that they are unlinked to one another (Figure 2.6).

Replication of RSC-V1aR associations

We used a mixed model to compare the RSC-V1aR abundance between homozygotes, with genotype entering as a main effect and parentage as a random effect. As reported in the main text, genotype was highly significant (p=0.0002), but REML parameter estimates for parents were not (all p>0.50). Overall the model explained 55% of the variation in RSC-V1aR abundance.

To assess the specificity of these associations, we measured V1aR in the LDThal, Vpall and LS and assessed whether HI-RSC and LO-RSC genotype also predicted expression in one of these other brain regions. We found it did not (p>0.25, Figure 2.8).

Epigenetic regulation of *avpr1a* expression

qPCR measures of RSC avpr1a mRNA abundance

The *avpr1a* assay was optimized to amplify 300bp of the *avpr1a* coding sequence (efficiency=93.7%, R^2= 0.999), with primers spanning the intron to avoid potential gDNA amplification. In order to correct for inter-sample variation we used both *β-actin* and *gapdh* transcripts as endogenous controls (*β-actin* assay efficiency=100.8%, R^2=0.999; *gapdh* assay efficiency=93.56%, R^2=0.992). Transcript abundances based on Ct values for *β-actin* and *gapdh* were highly correlated (*r*=0.91, p=1.1e-09). Genotype differences in *avpr1a* transcript abundance were evident whether *avpr1a* Ct was corrected with respect to *β-actin* (p=0.0002), *gapdh* (p=0.0003), or uncorrected (p=0.0003). Genotypes did not differ in β-actin (p=0.78) or *gapdh* (p=0.70) transcript abundance. For simplicity in the main text we report only ΔCt values with respect *to β-actin*.

ChIP-seq

Both H3K4me1 and INPUT libraries showed high quality reads (Phred>28), low duplication rate (36.16±15.5% and 18.18±3.3% respectively, mean±SE) and high mapping efficiency (79-88%) for both Stampy (Lunter and Goodson, 2011) and bwa (Li and Durbin, 2009), as estimated by SAMtools software package (Li et al., 2009). Results were equivalent, so only results from Stampy alignments are reported.

Using MACS2 (Zhang et al., 2008), we discovered a total of 36,238 statistically significant peaks (q<0.05) in the prairie vole genome. These peaks represent a catalog of putative RSC enhancers. Six of these peaks were located in the *avpr1a* BAC clone and only two were located at the *avpr1a* locus. Raw sequencing data and a complete list of all significant peaks can be found at NCBI's Gene Expression Omnibus (GEO, accession# GSE73670).

Detecting selection on *avpr1a*

Calculation of population genetic summary statistics

We examined the frequencies of a total of 96 non-coding SNPs for *avpr1a*, and 21 non-coding SNPs at our putatively neutral loci. Lastly, we used a multinomial distribution to calculate a likelihood ratio test comparing the frequency spectrum of non-coding SNPs at the *avpr1a* locus and control loci (df=4). Because we detected no high-frequency derived alleles in our control sites (Figure 2.9B), and a null frequency of zero is undefined in the LRT, we entered a non-zero value for our null expectation of high-frequency derived frequency classes 0.6-0.8 and 0.8-1.0. We used a value of .023, which corresponds to half the frequency associated with a observing a single SNP at the corresponding frequency. Since this increases the likelihood of the null hypothesis, we consider this to be a conservative estimate. If we limited analyses to just those SNPs that

34

fell in the first three frequency classes (0.0-0.6), the *avpr1a* locus still exhibited strong

evidence of an excess of intermediate frequency alleles (LRT=49.6, df=2, p=1.68e-11).

TABLES

Brain region	IPF	EPF	p-value
RSC	695±120	713±208	0.94
LDThal	2301±119	2207±563	0.88
LS	1269±77	1374±216	0.66
VPall	1975±121	2358±606	0.56

Table 2.1. Female V1aR abundance does not predict sexual fidelity. Means±SE in dpm/mg TE. IPF n=20, EPF n=6.

Amplicon	Size (kb)	Primer sequences	Outer primer sequences (if nested)
A. 5' non-coding seq.	3.4	F:TGTGGCACCCAGGTAAATGC R:GTAGCAGATGAAGCCATAGCAG	F:GCATGTGATTCTGGAATTTGTAAC R:ATAGTCTTCACGCTGCTGACA
B. Promoter + 5' UTR	1.7	F:AATAGACCAACGTTCTTAAG R:GCTCCTCGTTGCGTACATC	Not nested
C. First exon	1.2	F:CGGAAGCGGGAAGGAAGCAGCC R:CTCCCTCAGCCCATGATGCAG	F:GYGGTAGCCTAAACGCAGA R:GTTGGGATGRTTGAGAACCACA
D. Intron	2.5	F:CTACATCCTCTGCTGGGCTCC R:CATGTATATCCAGGGGTTGC	F:GCCTTGTGTCAGCAGCGTG R:TGTCTGTAGGCACCTTCTGTTCTG
E. Second exon	1.0	F:GCTGCTCTAACAGTGGTTGGTTTG R:CACATCACATGACTTAAACCAATC	F:GCCTTGTGTCAGCAGCGTG R:TGTCTGTAGGCACCTTCTGTTCTG
F. 3' UTR	0.6	F:CTACATCCTCTGCTGGGCTCC R:CATGTATATCCAGGGGTTGC	F:GCCTTGTGTCAGCAGCGTG R:TGTCTGTAGGCACCTTCTGTTCTG
G. 3' flanking	0.6	F:CGGACCATATAGAGATCATAAGAG R: GGGATAGAGGCAGAGACCCA	F: GTCCATTGTCTAAATCCGGACC R: GAACATGAGCAAAGAAGTCGG

Table 2.2. Amplicons and their corresponding primers (5' → 3') for characterizing *avpr1a* sequence variation.

Amplicon	Size (bp)	Primer sequences	Sequencing primer
Outer	1500	F:GGGGTTTTTGGTTAYGTTTTGTGTTAGTAG R:CACAAAAATCACCTAAAACCATCCTAAATTTCAA	
Inner1	245	F: AGATTATATTGTTAATAATARGGAATAAAGTAAAG R:/5Biosg/AAAACCACAACTATAAATCAATTAATACTATAT	TGGATCTAATTATTGAAATG
Inner2	251	F:GTGTTGTATATGTTGAGG TGTTTATTAA R:/5Biosg/ACCTAAAACCATCCTAAATTTCAAATATTTACA	ATAGTGGTGGTTTCTAATAA

Table 2.3. PCR and sequencing primers (5'→ 3') for two bisulfite pyrosequencing assays.

36

Target	Forward	Reverse
LCAT	AGAGGACTTCTTCACCATCTGGCT	TGTGCCCAATAAGGAAGACAGGCT
β-fibrinogen	GGCAATGATAAGATTAGCCAGCCAGCTCAC	AACGGCCACCCCAGTAGTATCTG
AP5	AATGCCCCATTCCACACAGC	GCAGAGACGTTGCCAAGGTG

Table 2.4. PCR primers (5'→ 3') for amplifying putatively neutral non-coding loci.
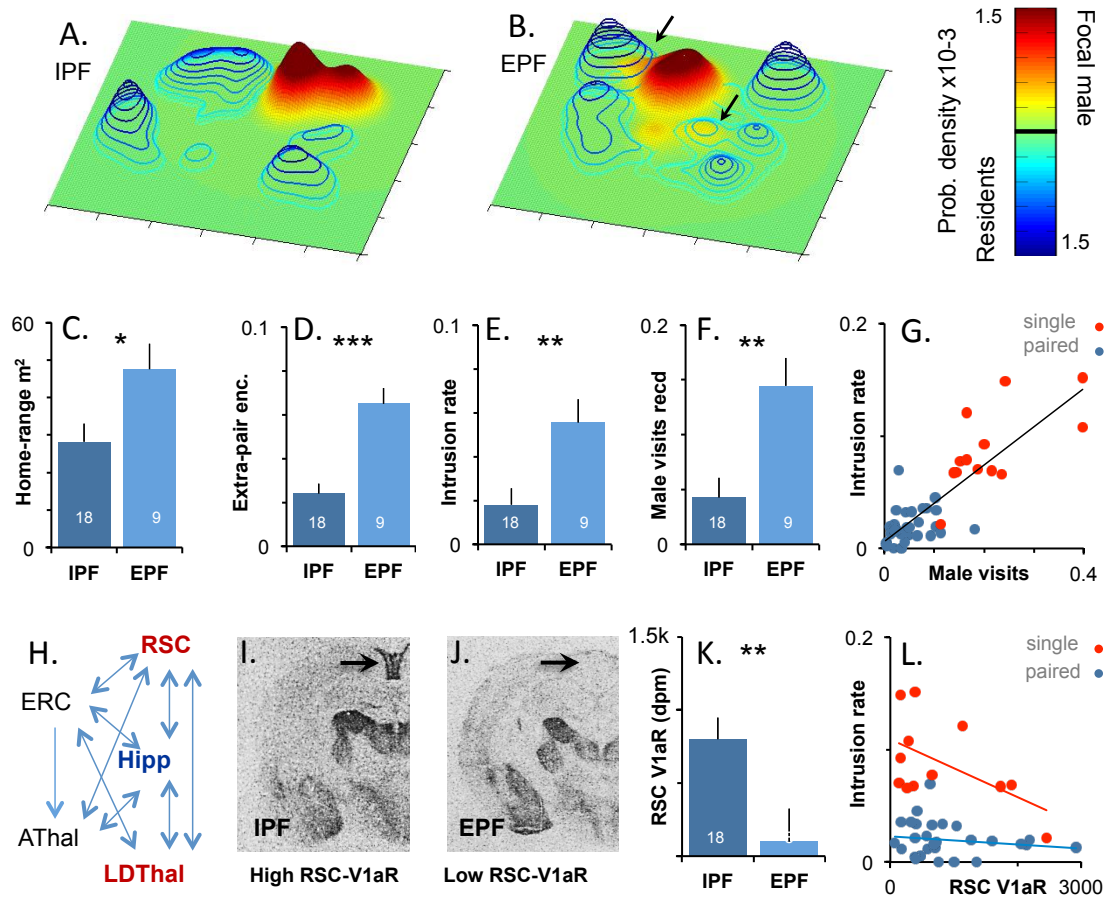
Figure 2.1. Male sexual fidelity predicted by patterns of space-use, social interaction and V1aR. A,B) Intensity of male space-use. X and Y axes are enclosure dimensions (20mx30m); height and color of peaks are probability densities. Focal male indicated in solid surface, non-focal males in blue contours. Single males not shown. Arrows indicate regions of likely intrusion by focal male. C-F) EPF and IPF males differ in space use. G) Rates of intrusion and of male visitation are correlated. H) Regions of a spatial-memory circuit (Aggleton, 2014) vary in receptors for vasopressin (red) or oxytocin (blue, [Ophir et al., 2008b; Phelps et al., 2010]). Abbreviations: RSC, retrosplenial cortex; ERC, entorhinal cortex; Hipp, hippocampus; AThal, anterior thalamus; LDThal, laterodorsal thalamus. I-K) Autoradiograms for vasopressin 1a receptor (V1aR) in retrosplenial cortex. RSC-V1aR (dissociations/min per mg tissue) predicts sexual fidelity and L) intrusion rate. All bars mean±SE. *p≤0.05, **p≤0.01, ***p≤ 0.001.
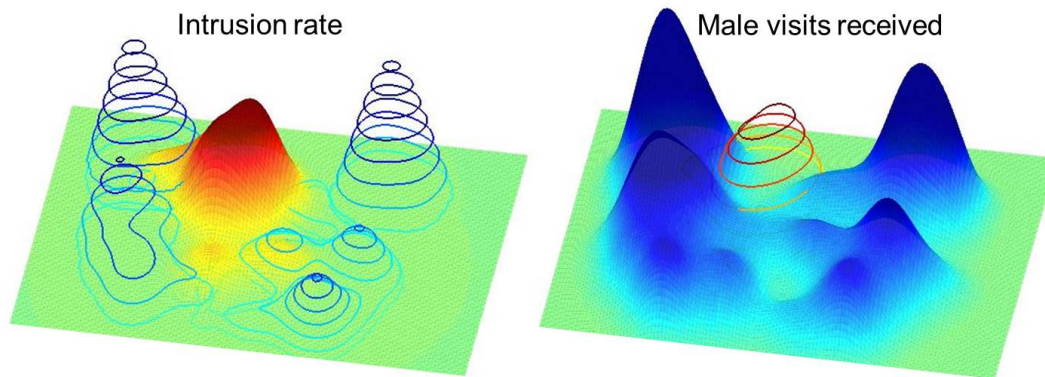
Figure 2.2. Calculating encounters between individuals from kernel density estimates. *Left panel* depicts the probability density function estimated from radiotracking data for a focal male (solid surface). Intensity of space use within territories of other paired males depicted as blue contours. The intrusion rate of the focal male is estimated by taking the product of his probability density with the corresponding density for each territory he could intrude upon and summing across space. *Right panel* depicts rate at which a focal male (red contours) is intruded upon. Solid surface (blue) corresponds to the sum of probability density estimates for all non-focal males, which in this case includes four paired males and one single male.
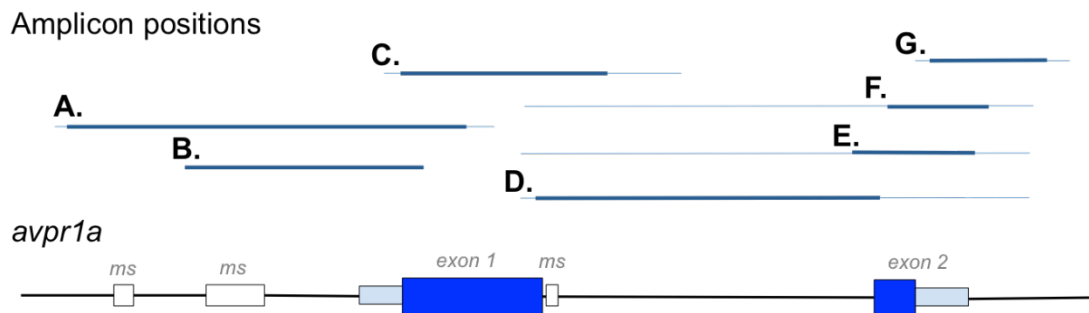
Figure 2.3. Schematic illustrating positions of amplicons used in sequencing. Dark blue lines correspond to sequenced amplicons. Thin blue lines correspond to outer amplicons in nested PCR amplifications. Primer sequences provided in Table 2.2
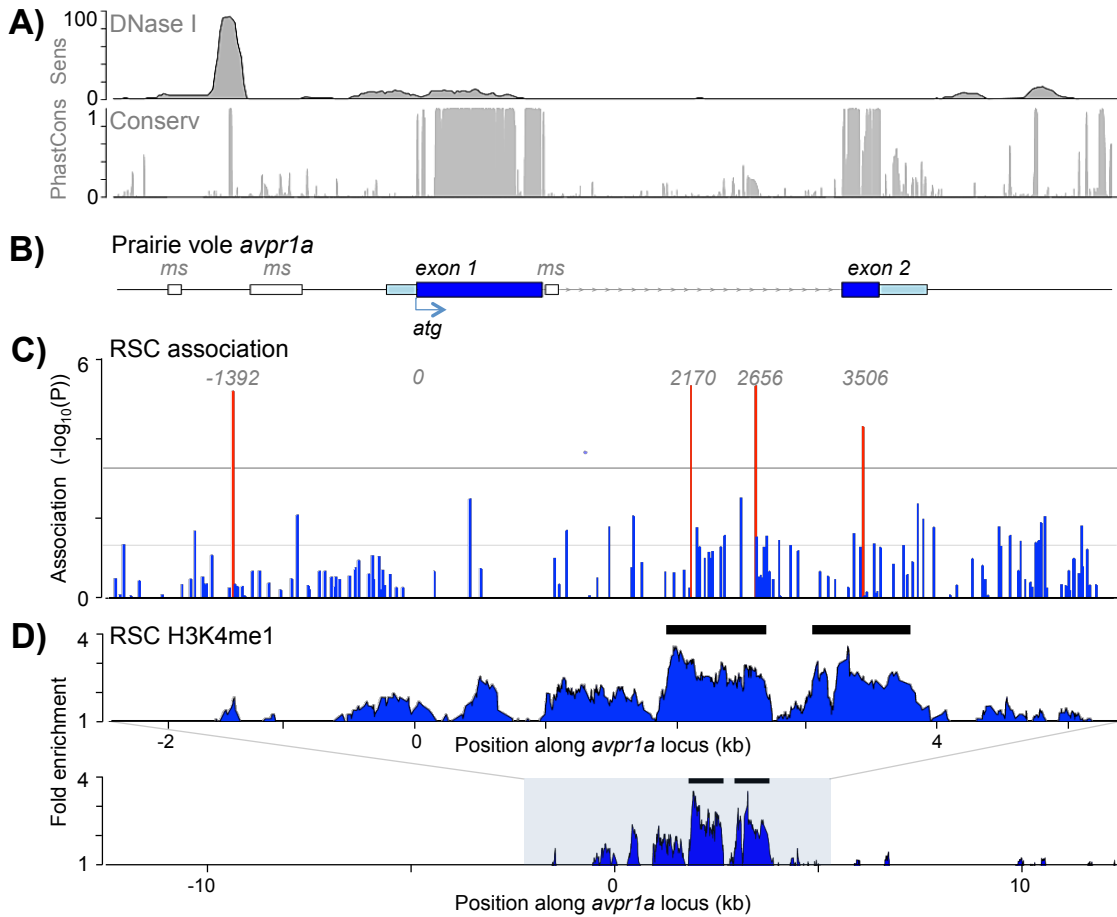
Figure 2.4. SNPs in regulatory regions of *avpr1a* locus predict RSC-V1aR. A) DNAse I hypersensitivity in *Mus* brain and mammalian conservation (Rosenbloom et al., 2013). B) Structure of prairie vole *avpr1a* locus (exons, blue; microsatellites, white). C) Association of *avpr1a* SNPs with RSC-V1aR abundance. Each bar is a SNP; X-axis depicts position along locus; Y-axis strength of association (-$\log_{10}$(P)). Lower horizontal line uncorrected α=0.05, upper line corrected α=0.00054. D) Fold enrichment by H3K4me1 ChIP-seq compared to input chromatin. Horizontal bars mark peaks corresponding to putative enhancers.
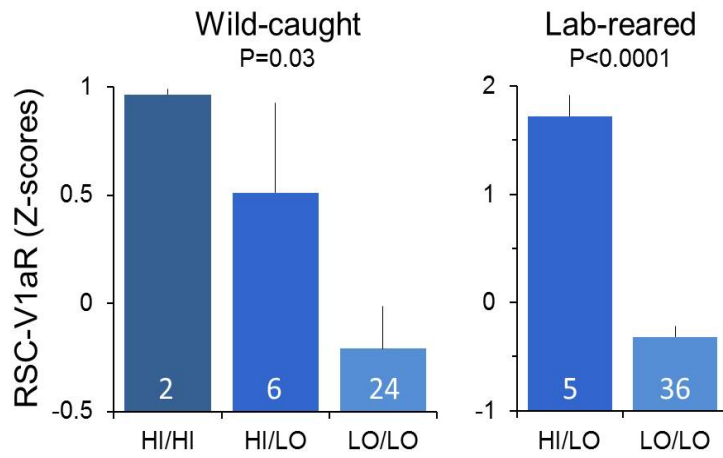
Figure 2.5. Sample differences in the effect of HI and LO alleles on RSC-V1aR. HI-RSC and LO-RSC alleles defined based on SNP 2170. Bars are mean+SE.
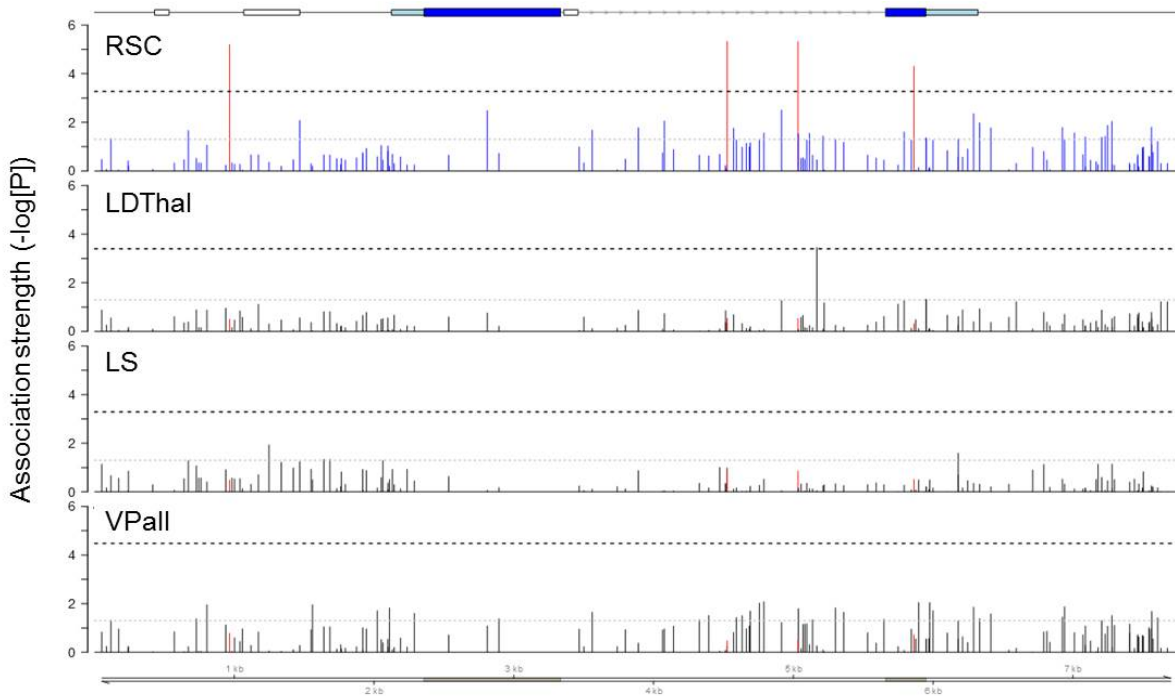
Figure 2.6. SNPs are associated with V1aR abundance in RSC and LDThal, but not VPall or LS. Association strength (-log$_{10}$(P)) for each SNP is plotted as a function of position at the *avpr1a* locus (top). Red vertical lines correspond to SNPs associated with RSC-V1aR abundance. Light gray horizontal lines correspond to uncorrected α=0.05, dark horizontal lines correspond to α after correcting for multiple tests by permutation simulations.

42
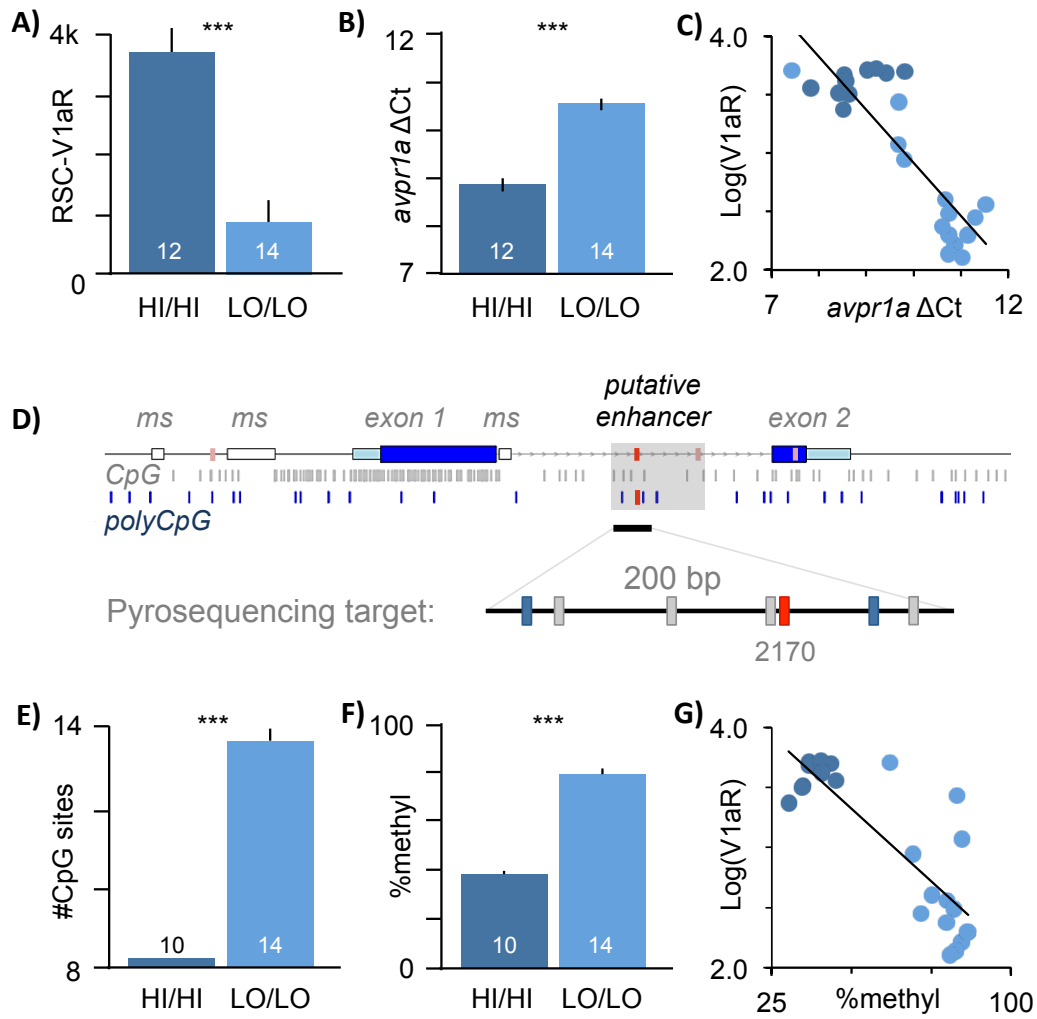
Figure 2.7. Genotype differences in regulation of *avpr1a*. A) Homozygotes differ in abundance of V1aR (dpm/mg) and B) *avpr1a* mRNA in RSC. C) RSC *avpr1a* transcript abundance correlates with V1aR protein. D) Fixed (gray) and polymorphic (blue) CpG sites along *avpr1a*. Red hatches are SNPs associated with RSC-V1aR. Shaded box is putative intron enhancer. A cluster of CpG sites were selected for pyrosequencing, including polymorphic CpG SNP 2170 in red. E) HI/HI males have fewer CpG sites in intron, and F) lower levels of enhancer methylation. G) RSC enhancer methylation correlates with V1aR abundance ($R^2=0.70$, $p<0.0001$). Bars are means±SE. *$p \leq 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$.

43
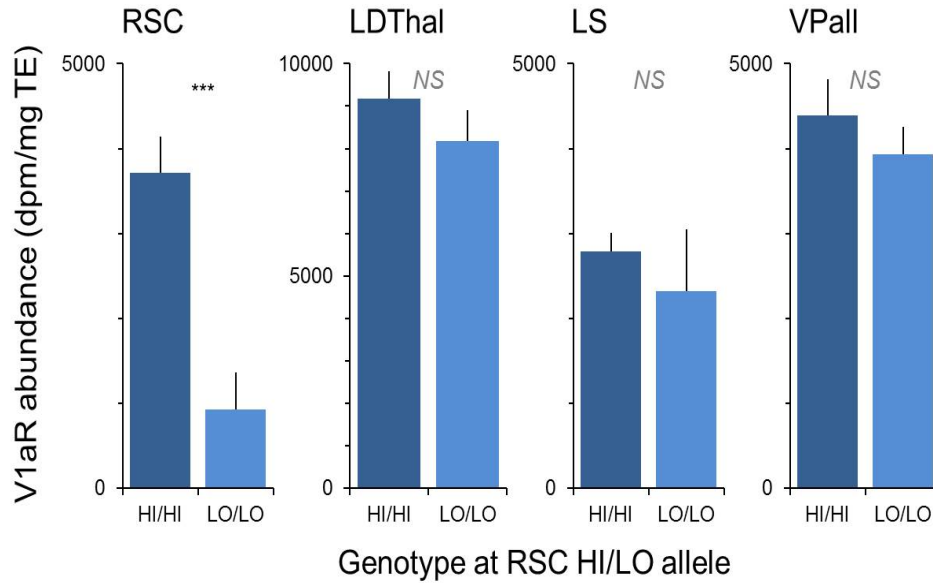
Figure 2.8. Replication of strength and specificity of SNPs associated with RSC-V1aR abundance. Crosses of parents heterozygous for RSC HI/LO alleles yields littermates homozygous for either allele (n=12 HI/HI, n=14 LO/LO). Results replicate association between SNPs and RSC abundance, as well as the specificity of these associations. V1aR abundance reported in dpm/mg TE, dissociations per minute per mg.
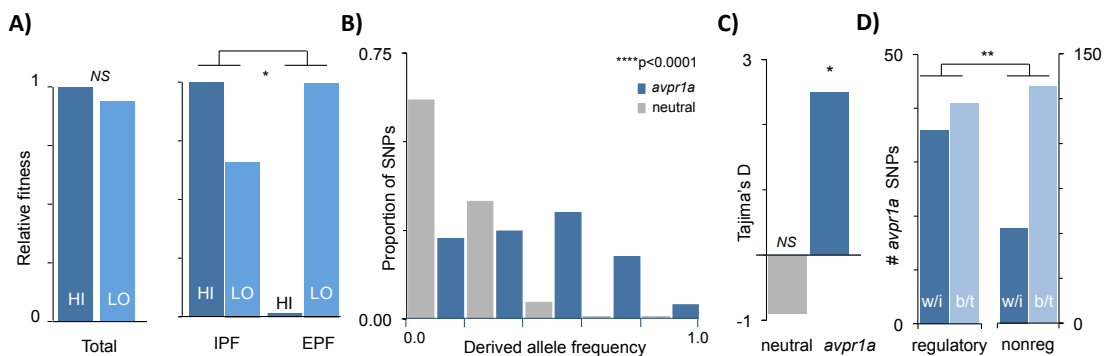
Figure 2.9. Selection maintains regulatory variation at *avpr1a*. A) Context-dependent selection on HI-RSC and LO-RSC alleles in field. B) *Avpr1a* has more intermediate frequency alleles than neutral markers. C) Tajima's D is significantly positive for *avpr1a*, but not neutral sites. D) Regulatory regions had higher ratios of within:between species differences than non-regulatory regions. *p<0.05, **p<0.01, ***p<0.001, *NS*=p>0.10

44

DISSERTATOR'S NOTES

This chapter is a multi-authored article (Okhovat et al. 2015), for which I was the lead author. I conducted most of the experimental work associated with identification of *avpr1a* alleles, all of the transcript measurements, DNA methylation quantification and ChIP-seq experiments. I participated in data analysis and writing of the manuscript.

Chapter 3: Methylation of *avpr1a* in the cortex of wild prairie voles: Effects of CpG position and polymorphism

Abstract

DNA methylation can cause stable changes in neuronal gene expression, but we know little about its role in individual differences in the wild. In this study we focus on the vasopressin 1a receptor (*avpr1a*), a gene extensively implicated in vertebrate social behavior, and explore natural variation in DNA methylation, genetic polymorphism and neuronal gene expression among 30 wild prairie voles (*Microtus ochrogaster*). Examination of CpG density across 8kb of the locus revealed two distinct CpG islands overlapping promoter and first exon, characterized by few CpG polymorphisms. We used a targeted bisulfite sequencing (bis-seq) approach to measure DNA methylation across ~3kb of *avpr1a* in the retrosplenial cortex, a brain region implicated in male space use and sexual fidelity. We find dramatic variation in methylation across the *avrp1a* locus, with pronounced diversity near the exon-intron boundary and in a genetically variable putative enhancer within the intron. Among our wild voles, differences in cortical *avpr1a* expression correlate with DNA methylation in this putative enhancer, but not with the methylation status of the promoter. We also find an unusually high number of polymorphic CpG sites (polyCpGs) in this focal enhancer. One polyCpG within this enhancer (polyCpG 2170) may drive variation in expression either by disrupting transcription factor binding motifs or by changing local DNA methylation and chromatin silencing. Our results contradict some assumptions made within behavioral epigenetics, but are remarkably concordant with genome-wide studies of gene regulation.

INTRODUCTION

Stable and persistent behavioral differences are common among conspecifics, and are thought to contribute to adaptive responses to diverse environments (Duckworth et al., 2015; Ledón-Rettig et al., 2013; McCrea and Costa, 1999; Pfennig, 1992; Sheriff et al., 2010; Sheriff and Love, 2013; Sinervo and Lively, 1996). Well-studied examples include the cannibalistic behavior of spadefoot toads (Pfennig, 1992), territorial defense of tree lizards (Sinervo and Lively, 1996), anti-predatory responses of snowshoe hares (Sheriff et al., 2010), and personality variation among humans (McCrea and Costa, 1999). The role that epigenetic factors play in the emergence of such behavioral diversity is an increasingly interesting and active area of work in ecology and evolution, with a variety of studies examining how developmental environments shape the behavior of adult offspring in the wild (Duckworth et al., 2015; Ledón-Rettig et al., 2013; Sheriff and Love, 2013). As behavioral ecologists seek to explore not only phenotypic variation and its consequences, but also its underlying mechanisms, they have begun to investigate how modifications of chromatin contribute to variation in gene expression and behavior (Ledón-Rettig et al., 2013; Simola et al., 2013). Of the many known chromatin modifications, DNA methylation at CpG dinucleotides is the most extensively investigated (Weaver et al., 2004). Despite the exciting prospects for behavioral epigenetics, it remains difficult to follow the relationship between DNA methylation, neuronal gene expression and behavior in the wild. These difficulties are in part due to the complex regulatory consequences of DNA methylation (Schübeler, 2015) and our limited understanding of how genetic and epigenetic variation interact to shape brain and behavior. In the current study, we examine how individual differences in sequence and methylation predict neuronal gene expression in the brains of wild prairie voles, *Microtus ochrogaster*.

Traditional studies of DNA methylation focus on CpG sites at a gene's promoter, where CpG methylation often silences gene expression (Tate and Bird, 1993). In contrast, methylation at CpG sites outside the promoter may be associated with either an increase or decrease in expression. For example, methylation within coding sequence can contribute to exon splicing and be associated with elevated expression (Laurent et al., 2010; Rauch et al., 2009). DNA methylation at more distal elements, such as enhancers and insulators, can either promote or inhibit gene expression (Jones, 2012; Jones and Takai, 2001). Thus, to understand the complex contributions of DNA methylation to gene expression, methylation should be studied across a gene's features. To understand gene regulation in natural settings, it is also critical to consider the genetic variation that could influence methylation and gene expression across these features.

In principle, genetic polymorphism at CpG sites can influence DNA methylation and gene expression by changing either the local density of CpG sites, or by altering specific binding sites for transcription factors (Figure 3.1). Though poorly understood, the overall density of CpG sites seems to be important for shaping the epigenetic status of a regulatory element. Short stretches of densely packed CpGs (~1kb) known as CpG islands (CPGi) can lead to stable de-methylation (Bird, 2002). In contrast, regions just outside CpG islands have lower CpG density, exhibit tissue-specific methylation, and are more likely to have single nucleotide polymorphisms (SNPs) within a CpG site (Irizarry et al., 2009; Tomso and Bell, 2003). A CpG polymorphism – for example, TG/CG or CA/CG – is referred to as a polyCpG. By altering local CpG density, such polymorphisms could change the likelihood of recruiting repressive proteins with methyl-binding domains (Figure 3.1A; [Hsieh, 1994; Tate and Bird, 1993]). PolyCpGs may also affect binding of a transcription factor that is sensitive to variation in motif sequence (Figure 3.1B), methylation, or both (Figure 3.1C; [Schübeler, 2015; Tate and Bird, 1993;

Tycko, 2010]). These alternatives reveal some of the complex ways in which CpG polymorphisms may interact with epigenetic mechanisms to produce differences in developmental sensitivity, plasticity and complex behaviors.

Although DNA methylation is present in a wide range of taxa (Bird, 1993) and CpG polymorphisms are common (Tomso and Bell, 2003), their contributions to natural neuronal and behavioral diversity are not well understood. Genetically diverse non-model species allow us to apply modern molecular techniques to examine natural variation in genetics and epigenetics, as well as their association with neuronal and behavioral variation. In the current study, we use the socially monogamous prairie vole, *Microtus ochrogaster*, to investigate the interaction between DNA methylation, CpG polymorphism and the expression of the vasopressin 1a receptor (*avpr1a*), a gene critical for social behavior in this and other species.

Prairie voles are socially monogamous rodents, but ~25% of the offspring are sired outside the pair (known as extra-pair fertilizations; [Ophir et al., 2008a]). Variation in prairie vole sexual fidelity is predicted by differences in space use that seem to be mediated by variation in *avpr1a* expression in the retrosplenial cortex (RSC-V1aR), a brain region important in spatial memory (Okhovat et al., 2015; Ophir et al., 2008b). Among lab-reared animals, the cortical expression of *avpr1a* is highly predicted by four single nucleotide polymorphisms that together define "HI" and "LO" alleles. Interestingly, one of the polymorphisms (SNP 2170) is a polymorphic CpG linked to several other polyCpGs. These polymorphisms occur within a short sequence identified as a putative enhancer by ChIP-seq targeting the histone mark H3K4me1 (Okhovat et al., 2015), and its methylation status predicts cortical V1aR abundance among lab-reared animals. Among wild-caught animals, we found that the relationship between genotype and phenotype was weaker, and speculated that this was due to increased variation in

49

developmental environment (Okhovat et al., 2015). In the present study, we ask whether methylation of the putative intron enhancer is also able to predict cortical expression of *avpr1a* among wild prairie voles. We expand on these finding by investigating sequence variation and methylation across a much broader expanse of the locus, allowing us to more systematically explore how genetic and epigenetic variation contribute to neuronal gene expression in the wild.

We first characterize the *avpr1a* locus by identifying CpG islands and examining the distribution of polyCpGs across the *avpr1a* locus. Next we validated a sequencing approach to estimate methylation at 122 CpG sites across ~3kb of *avpr1a*, spanning from promoter to the putative intron enhancer. We then use these data to examine the pattern of methylation across *avpr1a* features, to test how methylation in different features predicted cortical *avpr1a* expression, and to ask whether polymorphic CpG sites contribute to CpG density or sequence-specific effects of methylation. In the process, this study explores how previous results from genome-wide studies of methylation inform our understanding of individual differences in brain and behavior.

METHODS

Wild-caught samples and tissue processing

32 wild adult male (n=18) and female (n=14) prairie wild voles were collected from Champaign County, IL. Brains were frozen immediately on dry ice, stored at -80°C and later sectioned at 20μm thickness and 100μm intervals. V1aR autoradiography from these samples has been reported previously, and methodological details are provided there (Okhovat et al., 2015; Ophir et al., 2008b). An alternative set of fresh frozen brain sections was used as a source for genomic DNA in the Sanger sequencing of the locus.

To examine the methylation status of *avpr1a*, we dissected the retrosplenial cortex from a third set of alternative fresh-frozen sections. Fresh frozen sections were not available for 2 of 32 animals, which reduced our sample size to 30 individuals. We performed genomic DNA bisulfite conversion using the EpiTect Plus LyseAll Bisulfite Kit (Qiagen), following manufacturer's instructions.

Characterization of the *avpr1a* locus

Sequencing of the *avpr1a* was performed as described previously (Okhovat et al., 2015). Sequencing reads were aligned to *avpr1a* reference (AF069304.2, NCBI) in Geneious 5.5.7 software to find fixed and polymorphic CpGs (polyCpGs). PolyCpGs were defined as SNPs occurring at the C or G within a CpG dinucleotide. CpG polymorphisms present in only a single individual were disregarded, as they are too rare to be useful in examining associations.

To characterize CpG density across the locus, we calculated the CpG count in 300bp sliding windows across the reference *avpr1a* sequence. Also, we predicted the position of CpG islands at the *avpr1a* locus (AF069304.2, NCBI) using the online EMBOSS Cpgplot tool (Rice et al., 2000). We used a window size of 300bp and traditional CpG island algorithm criteria, including an island length>200bp, GC content>50% and ObsCpG/ExpCpG>0.60 (Gardiner-Garden and Frommer, 1987). Our CpG density analysis revealed two CpG islands, CpGi.1 is 5'of the transcription start site and includes parts of the *avpr1a* promoter, while CpGi.2 includes the first exon. CpGi.2 exhibited a distinct tri-modal pattern in CpG density. To capture this heterogeneity in CpG density, we subdivided CpGi.2 into three compartments defined by local minima in CpG density (Figure 3.2A). These features were the basis for the parsing of our analysis of methylation data across 3kb of the *avpr1a* locus described below.

DNA methylation measurements

Pyrosequencing

We used a nested PCR strategy to produce two pyrosequencing amplicons as described previously (Okhovat et al., 2015). The biotin-labeled PCR amplicons were sent to Epigendx (Hopkinton, MA) for DNA methylation pyrosequencing (assay IDs: Cluster1FS2 and Cluster 2FS2). DNA methylation at CpG sites was reported as %(unconverted C/ [unconverted C+ converted T]) for each CpG site.

Targeted bisulfite sequencing (bis-seq)

To examine individual differences in methylation across major gene features of the *avpr1a* locus, we generated a series of 5 amplicons spanning ~3kb from the promoter to the intron enhancer.

We used a semi-nested PCR approach to amplify 350bp upstream of the transcription start site (TSS) and the first exon (Table 3.1A). The outer PCR reaction included KAPA HiFi Uracil+ mix (KAPA biosystems), 300nM of each primer (Table 3.1A) and 1.5ul of bisulfite converted gDNA with the following settings: 3 min at 95°C, {20sec at 98°C, 30sec at 52°C and 90sec at 68°C}X36. Two following semi-nested inner PCR reactions consisted of HiFi Uracil+ mix (KAPA biosystems), 400nM of each primer and 2ul of undiluted outer amplicon. Amplifications were performed with the following settings: 2min at 95°C, {20sec at 98°C, 30sec at 55°C and 90sec at 68°C}X25.

Using primers provided in Table 3.1B, we amplified a 1.6kb amplicon around the exon1-intron boundary with PCR composition similar to the reaction described above and the following settings: 3 min at 95°C, {20sec at 98°C, 30sec at 58°C, 90sec at 68°C}X40.

We used a semi-nested PCR approach to amplify 1.5kb of the intron in a PCR reaction consisting of KAPA HiFi Uracil+ mix (KAPA biosystems), 9.5ul DNase free

water, 300nM of each primer (Table 3.1C) and 1.5ul of bisulfite converted gDNA with the following settings: 3min at 95°C, {20sec at 98°C, 30sec at 52°C and 90sec at 68°C}X36. Inner PCR reactions consisted of GoTaq Hot Start Colorless Master Mix (Promega), 200nM of each primer and 1ul undiluted outer amplicon. Amplifications were performed with the following settings: 3min at 93°C, {30sec at 93°C, 30sec at 55°C, 90sec at 70°C}X35, 2min at 70°C. All final PCR products were visualized on agarose gel and gel-extracted using Qiagen gel extraction kit (Qiagen).

Following PCR cleanup, DNA concentrations were measured on a Nanodrop 2000 Spectrophotometer (Thermo Fisher Scientific). For each individual, PCR amplicons were mixed in equimolar ratios and brought to a final volume of 500ul with 1xTE. Sample pools were then sonicated with Q125 sonicator (Qsonica) on ice for 25 cycles (10sec pulse, 10sec rest) at 50% amplitude. DNA was then precipitated with standard EtOH precipitation and eluted in 1xTE.

For each individual, 50ng of the sheared DNA pool was used to construct Illumina paired-end libraries using the Nextflex ChIP-Seq kit (BioScientific) according to manufacturer's instructions with minor modifications. Briefly, samples were end-repaired and size-selected to capture 300-400bp fragments. Size-selected fragments were adenylated and barcoded with Nextflex Illumina DNA barcodes (BioScientific). We used the KAPA library amplification kit (KAPA biosystems) to amplify the library for 5-6 cycles according to manufacturer's protocol. Libraries were sequenced on the MiSeq Illumina platform (2X250PE) at UT sequencing core facility (Austin, TX).

Reads were shortened to 130bp by trimming low quality 5' ends. Next, we used Trim-galore! (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) to remove remaining adaptor contamination, low quality reads (Phred<20), short reads (<16bp) and reads with a missing pair. The reference *avpr1a* sequence (AF069304.2,

53

NCBI) was modified to include known single nucleotide polymorphisms (SNPs). SNPs that involved CpG sites were left as CpG dinucleotides and the rest of SNPs were replaced by their corresponding ambiguous IUPAC symbol. These modifications allowed us to measure DNA methylation at both fixed and polymorphic CpG sites and avoid allelic bias in alignment. We used Bismark v0.7.7 (Krueger and Andrews, 2011) with bowtie2 (Langmead and Salzberg, 2012) for read alignment. Next, we used Bismark's Methylation extractor tool and a custom python script, to compile counts of methylated and unmethylated reads at each CpG site and determine percent CpG methylation. We also obtained non-CpG cytosine methylation within CHG and CHH contexts (H is A, C or T) from Bismark alignment reports. All methylation values were exported to R (http://www.r-project.org/) for further analysis.

To accommodate potential heterogeneity in methylation across the locus, we used the boundaries of *avpr1a* features defined above to partition our bis-seq target (Figure 3.2B). The first ~100bp of our bis-seq target corresponds to the 3' region of CpGi.1. A ~200bp region between the CpG islands includes the transcription start site and the 5' 18bp of the 5' UTR; we labeled this segment as Promoter. The labels CpGi.2a-c correspond to three local peaks in CpG density within CpGi.2. The label Intron refers to a ~1kb sequence from the end of CpGi.2 to the beginning of a putative intron enhancer. Lastly, our bis-seq target overlaps with the first ~300bp of a putative intron Enhancer identified by H3K4me1 ChIP-seq on prairie vole retrosplenial cortex (Okhovat et al., 2015).

Statistical analysis

Bis-seq technical validation

We used a linear model to examine the correlation between methylation values obtained at seven intron enhancer CpGs by targeted bis-seq to pyrosequencing data from the same sites. To determine the null distribution of the expected correlation, we randomly assigned pyrosequencing methylation values to individuals 1000 times and each time measured the Pearson correlation coefficient between pyrosequencing and bis-seq values. We used these randomized correlation coefficients to estimate a null distribution and resulting p-value.

CpG co-methylation within and between gene features and across avpr1a

We used a linear model to examine the relationship between co-methylation (Pearson correlation coefficient) and distance between the CpG pair within and between gene features. Significance of effects was determined by permutation analysis. We also used a heatmap to visualize heterogeneity in co-methylation between all pairs of CpGs across our bis-seq target.

Avpr1a alleles and enhancer CpG differences

We used sequence at the intron-enhancer SNP 2170 (T/T, T/G, G/G) to assign HI/HI, HI/LO and LO/LO *avpr1a* genotypes. We scored genotypes with values 0, 1 and 2 corresponding the number of HI alleles present. We ran ANOVA and Kendall's rank correlation analyses to compare V1aR abundance in the retroplenial cortex (RSC-V1aR), DNA methylation and enhancer CpG count among *avpr1a* genotypes. Data on genotype association with RSC-V1aR abundance (Figure 3.6B) were previously published (Okhovat et al., 2015), but are included here for completeness.

PolyCpG frequencies and distribution

We observed 30 polyCpG sites across the locus, one of which had three alternative alleles. For the 29 bi-allelic SNPs, we calculated the number of variants corresponding to each of six possible CpG polymorphisms: CpA, CpC, CpT, ApG, GpG and TpG. We performed a 6x2 chi-squared test comparing the observed SNP frequencies to a neutral expectation in which each polymorphism is equally likely.

To examine heterogeneity in the distribution of polyCpGs, we used a two-tailed Fisher's exact test to compare the ratio of polyCpGs:total CpGs within the CpG islands to the ratio at the rest of the locus. Similarly we compared polyCpGs:total CpGs and polyCpG:nucleotides in the enhancer to the rest of the locus. Fisher's exact and chi-squared tests were performed using the online GraphPad software (available at http://graphpad.com/quickcalcs/contingency1.cfm).

Sequence specific effects of polyCpGs and methylation

At the 8 polyCpG sites included in our bis-seq target, we used linear regression to test the association of RSC-V1aR with *total %DNA methylation* at each polyCpG, genotype, and with *%methylation per CpG* – a measure normalized for the number of CpG-containing alleles present at a specific polymorphic site. To be explicit, *total %DNA methylation* is defined as the proportion of reads that carry a methylated CpG at the site of interest, regardless of genotype. *Genotype* is the number of CpGs the individual possesses at a polymorphic site (0, 1 or 2). Lastly, for individuals homozygous for a CpG or alternative allele, *%methylation per CpG* equals *total %DNA methylation*, but for a heterozygous individual, it is 2*(*total %DNA methylation* at CpG site).

To predict transcription factor binding around polyCpG 2170 and to test if sequence differences between HI and LO affect their binding, we used the transcription factor affinity predictor web tool for SNP comparisons (sTRAP; [Manke et al., 2010]).

We used the HI and LO sequence in a 20bp window centered at polyCpG 2170 and selected transcription factor matrices from TRANSFAC (vertebrates-only) with a mouse-promoter background model. P-values were corrected using Benjamini-Hochberg corrections (Benjamini and Hochberg, 1995). Transcription factors that had significant (p<0.05) affinity to at least one of the genotypes at polyCpG 2170 were selected and ranked from highest to lowest genotype difference in affinity. Lastly, we examined the Allen Brain Atlas (Lein et al., 2007) to examine whether any of the identified transcription factors were expressed in the retrosplenial area of the mouse brain.

RESULTS

Characterization of the *avpr1a* locus

We sequenced and analyzed ~8kb of the *avpr1a* locus in 32 wild-caught prairie voles and found a total of 172 fixed CpG sites and 30 polymorphic CpGs (polyCpGs). We observed that CpG density was not homogeneous across the locus, with evidence of two CpG islands (Figure 3.2A). The first predicted CpG island (CpGi.1) is ~0.4kb long and starts ~0.6kb upstream of the *avpr1a* transcription start site (TSS). The second CpG island (CpGi.2) is 1.3kb long, and encompasses most of the 5'UTR, all of the first coding sequence and a short region of the intron. CpG density was variable within this CpGi, as evident by three local peaks of CpG density in a sliding-window analysis (Figure 3.2A).

The ~2kb of sequence that flanks either side of a CpG island are known as CpG island-shores or CpGi-shores. CpGi-shores have high methylation variation and show tissue-specific differential methylation (Irizarry et al., 2009). At the *avpr1a* locus, the CpGi-shores include a 2kb region upstream of CpGi.1 and a 2kb region downstream of CpGi.2, and includes most of the intron and all of a putative intron enhancer identified previously by H3K4me1 ChIP-seq (Okhovat et al., 2015). The CpG density is relatively

low within CpGi-shores and in features located outside the shore boundaries (e.g. second exon, Figure 3.2A).

DNA methylation measurements and bis-seq technical validation

To control for tissue differences in methylation, all our methylation measures were obtained directly from RSC dissections of wild-caught brains. While these methylation measures reflect averaged measures across multiple cell types, this approach is much more accurate than measuring methylation in the whole brain or in peripheral proxy tissues, such as blood (Walton et al., 2016).

We used bisulfite pyrosequencing to measure DNA methylation in the putative intron enhancer of 30 wild-caught animals. Our pyrosequencing assays measured methylation at 2 fixed and 5 polymorphic CpG (polyCpG) sites (Figure 3.2B). One of the polyCpGs (2170) has previously been shown to be highly predictive of V1aR abundance in the retrosplenial cortex (RSC-V1aR) in prairie voles (Okhovat et al., 2015). Of 30 samples, none failed standard QC measures (threshold for QC rejection: bisulfite conversion efficiency<93%) and genotypes were correctly captured at all polyCpG site.

We used a targeted bisulfite sequencing (bis-seq) approach to expand our DNA methylation measurements. Our bis-seq assay spanned from 300bp upstream of the *avpr1a* TSS to 2.3kb downstream, and covered 114 fixed and 8 polymorphic CpG sites (Figure 3.2C). Our assay generated single CpG resolution methylation measures and 100% coverage of all targeted CpG sites for all 30 wild-caught voles. To assess the accuracy of our bis-seq assay, we compared bis-seq DNA methylation measurements at each of the 7 putative enhancer CpG sites to pyrosequencing methylation measures at the same sites. Levels of methylation estimated by targeted sequencing were slightly higher but broadly similar to those we obtained by pyrosequencing. We regressed these

measures against one another and found that methylation measurements from the two techniques agree, especially at polyCpG sites (r=0.89, p<0.001; Figure 3.2B)

In addition to examining canonical CpG methylation, we used our bis-seq data to examine DNA methylation in CHG and CHH contexts (where H is A, C or T). Non-CpG methylation has previously been found in the mammalian adult brain, where it is negatively correlated with expression (Guo et al., 2014). Based on our bis-seq data however, %CHH and CHG methylation at the *avpr1a* locus were both very low (CHH:1.57$\pm$0.56%, CHG:1.30$\pm$0.61%, mean$\pm$SD; data not shown) and significantly correlated (r=0.45, p=0.01). We did not find any correlation between non-CpG methylation and *avpr1a* expression level (p>0.1), thus it is likely that our non-CpG methylation measurements merely reflect incomplete bisulfite conversion rather than true methylation. Based on this, we can estimate the rates of bisulfite conversion in our target by calculating 1- %non-CpG methylation. We estimate our bisulfite conversion rate to range from 94.1% to 98.7% (97.1$\pm$1.0%, mean$\pm$SD), which is consistent with the conversion rate estimates from pyrosequencing quality controls.

Patterns of CpG methylation across *avpr1a* and among wild-caught voles

We partitioned the bis-seq target with CpGi and prior gene annotations into sequence features we label as CpGi.1, Promoter, CpGi.2a, CpGi.2b, CpGi.2c, Intron and Enhancer (see Methods). Our bis-seq measurements show that DNA methylation varies greatly along the *avpr1a* locus and can differ dramatically among gene features (Figure 3.2C). Average DNA methylation was low in CpGi.1 (7.5$\pm$13.9%, mean$\pm$SD) and promoter (1.8$\pm$0.6%, mean$\pm$SD). However, along CpGi.2, DNA methylation appears much more variable. Average DNA methylation was low in the 5' end of CpGi.2, which includes CpGi.2a (1.2$\pm$1.3%, mean$\pm$SD) and CpGi.2b (5.1$\pm$6.4%, mean$\pm$SD), but

increased within CpGi.2c (44.6$\pm$26.6%, mean$\pm$SD) and toward the exon1-intron boundary. The increase in methylation at the border of CpGi.2b and CpGi.2c coincides with a mouse (*Mus musculus*) transcript start peak from cap analysis gene expression (CAGE) data (Figure 3.3; [The FANTOM Consortium and the RIKEN PMI and CLST (DGT), 2014]), suggesting this region may be involved in an unknown transcriptional function. Average CpG methylation was high throughout the intron (71.6$\pm$18.8%, mean$\pm$SD) and intron enhancer (72.5$\pm$17.0%, mean$\pm$SD; Figure 3.2C).

DNA methylation did not vary drastically among individuals at CpG sites within the CpGi.1, promoter, CpGi.2a and CpGi.2b. However, CpGi regions with higher average DNA methylation (i.e. CpGi.2c) and CpGi-shore features – such as the intron and putative enhancer – exhibited high inter-individual variation. Individual differences in methylation at the 5' end of CpGi.2c exist in the absence of CpG polymorphism. However, many CpG sites in the intron and the enhancer are polymorphic and it seems that methylation variation in this region was driven by genotype differences among individuals (Figure 3.2C).

CpG co-methylation across the *avpr1a* locus

We used our bis-seq data to examine the correlation of DNA methylation (co-methylation) between pairs of CpG sites across *avpr1a*. In general, the strongest methylation correlations (|r|>0.5) were found between close CpG pairs (<1kb; Figure 3.4A). We found a negative correlation between co-methylation and CpG distance. This correlation was significant for both same gene-feature (r=-0.14, p<10e-8) and between gene-feature co-methylation (r=-0.09, p<10e11). However, the correlation was stronger among CpG pairs within the same feature (distance x CpG feature position p<10e-07; Figure 3.4A).

Patterns of co-methylation were heterogeneous across the bis-seq target, as evident by three clusters of high positive correlation (Figure 3.4B). The first co-methylation cluster was found upstream of the *avpr1a* TSS, within the CpGi.1 and promoter. The second cluster was located at 3' end of the second CpG island and included some CpGs within CpGi.2b and CpG1.2c. The third cluster was found on the exon-intron boundary and included CpGs from both CpG.2c and the intron. CpGs in the latter cluster showed overall negative methylation correlation with many other CpGs located upstream the TSS (i.e. CpGi.1 and promoter) and the 5' side of the first exon (CpGi.2a and parts of CpGi.2b).

Avpr1a methylation and V1aR abundance in the RSC

We observed substantial variation in the abundance of RSC-V1aR among our wild-caught voles (Figure 3.5A). To examine the relationship between RSC-V1aR abundance and *avpr1a* methylation, we split individuals at the median value of RSC-V1aR (median=5669.5 dpm/mg TE) into high-expressing (high-exp) and low-expressing (low-exp; Figure 3.5A). We compared DNA methylation between the high-exp and low-exp wild voles at individual CpG sites and gene features.

In our single CpG comparisons, first we averaged DNA methylation of all individuals within high- and low-exp animals at each of the 122 CpG sites and calculated their difference (Figure 3.5B). Methylation differences were generally small (<10%), but many CpGs in the 3' end of CpG.i2 showed higher methylation in high-exp animals. These sites seem to correspond to cluster 2 in our co-methylation analysis. In contrast, CpGs within the enhancer showed lower methylation in high-exp animals. Using t-tests we found 4 CpG sites with different methylation between high and low-expressing animals (p<0.05; Figure 3.5B) however none survived false discovery rate corrections

(adjusted p>0.1; [Benjamini and Hochberg, 1995]). Three of these CpGs were in CpGi.2c, and one was in the intron enhancer region.

Examining average levels of methylation across features, we found a significant difference in DNA methylation between high- and low-exp animals in the putative enhancer (high-exp: 70.10±2.07%, low-exp: 77.02±1.40%, mean±SD, p=0.01; Figure 3.5C). Average CpGi.2c methylation was higher in the high-expressing animals (high-exp: 47.94±2.72%, low-exp: 42.60±1.60%, mean±SE; Figure 3.5C), but this difference was not statistically significant (p=0.10). None of the other gene features exhibited methylation differences between high- and low-exp animals (p>0.10; Figure 3.5C).

Avpr1a genotypes and the putative intron enhancer

Average %DNA methylation in the putative intron enhancer was negatively associated with RSC-V1aR among wild voles (*r*=-0.41, p=0.03; Figure 3.6A). As previously reported (Okhovat et al., 2015), we found 24 LO/LO, 6 heterozygous HI/LO and 2 HI/HI individuals, and these genotypes differ in RSC-V1aR abundance (ANOVA, F=4.99, p=0.03; Figure 3.6B; see also [Okhovat et al., 2015]). Here we find that these individuals also differ in average enhancer methylation (HI/HI 39.4±3.2%, HI/LO 55.4±4.3%, LO/LO 63.3±6.4%, mean±SD; ANOVA, F=20.23, p<0.0001; Figure 3.6C). Sequence differences between the HI and LO allele involve enhancer polyCpGs, which leads to genotype differences in numbers of CpG sites within the putative enhancer (HI/HI: 12.0±0.0, HI/LO: 15.5±0.3, LO/LO: 16.6±0.4, mean #CpG±SD; Kendall's tau=0.38, p=0.016; Figure 3.6D) but not across the whole *avpr1a* locus (HI/HI: 356.5±0.7, HI/LO: 364.2±3.4, LO/LO: 364.2±5.8, mean #CpG±SD; Kendall's tau=0.17, p=0.27; data not shown).

CpG polymorphisms

Among our wild voles, we found 30 polyCpGs across the *avpr1a* locus. We used the bi-allelic polyCpGs (n=29) to examine the frequency of each polyCpG variant. The frequency distribution of polyCpG variants was highly divergent from null expectations ($x^2_{(5, n=29)}$=30.37, p<0.0001; Table 3.2). More than half of the variants (79.2%) were G/A or C/T polymorphisms, which is consistent with the expected prevalence of methylation-induced deamination mutations and previous genome-wide characterizations of polyCpG frequencies (Tomso and Bell, 2003).

PolyCpGs were also non-homogeneously distributed across *avpr1a*. The two CpG islands, which together accommodate 72.1% of all the fixed *avpr1a* CpGs, only hold 3 polyCpGs; a significantly lower polymorphisms rate compared to the rest of the locus (Fisher's exact, p<0.0001; Figure 3.7A). In contrast, the 786bp enhancer has 7 polyCpGs and 5 fixed CpGs. The remaining 7,530bp of the *avpr1a* locus holds 23 polyCpGs and 167 fixed CpG sites (Figure 3.7A). Thus, a larger fraction of CpG sites are polymorphic within the enhancer than across the rest of the locus (58.3% vs. 12.1%, Fisher's exact, p=0.0004; Figure 3.7B). Similarly, polyCpG density is higher in the enhancer compared to the rest of the locus (%0.89 vs. %0.31, Fisher's exact, p=0.02; Figure 3.7C).

In our bis-seq assay, we captured 8 of the *avpr1a* polyCpG sites: five located within the putative intron enhancer and one in each of the CpGi.2b, CpGi.2c and intron features. We found that *total %DNA methylation* (*r*=-0.31, p=0.052) and *genotype* (*r*=-0.32, p=0.045) at polyCpG 2170 were associated with RSC-V1aR (Figure 3.7D). This polyCpG is one of the SNPs that define the HI and LO allele in both lab-reared and wild-caught animals (Okhovat et al., 2015). The 7 remaining polyCpG sites did not predict individual differences in RSC-V1aR (p>0.10).

63

Transcription factor affinity (sTRAP) analysis at polyCpG 2170 provided a list of candidate transcriptions factors predicted to bind to this sequence. These transcription factors are expected to show highly different affinity between the HI and LO allele (Table 3.3). Examination of the Allen Brain Atlas (Figure 3.7E; [Lein et al., 2007]) revealed that at least one of these transcription factors (GATA2) had strong expression in the mouse retrosplenial area. GATA2 binding is predicted to be much stronger to the LO allele sequence than to the HI allele sequence. Not all of the factors exhibited clear evidence of expression in the mouse RSC based on the Allen Brain Atlas, but the atlas is descriptive, and negative data are inconclusive.

DISCUSSION

In nature, individual differences arise as genetic and epigenetic forces interact to shape gene expression, cellular processes, and organismal phenotypes. In this study we explore DNA methylation and CpG distribution at *avpr1a*, the locus encoding the vasopressin 1a receptor (V1aR). We characterized CpG distribution across ~8kb of the *avpr1a* locus and found dramatic variation in CpG density (Figure 3.2A). The highest CpG density was found in two CpG islands that flanked the transcription start site (TSS, Figure 3.2A). Next we used high-throughput sequencing techniques and natural genetic variation among 30 wild prairie voles to examine the significance of DNA methylation and polymorphic CpGs (polyCpGs) in shaping cortical *avpr1a* expression associated with complex spatial and sexual behaviors.

We used a targeted bisulfite sequencing approach to characterize DNA methylation at 122 CpG sites across ~3kb of the *avpr1a* locus. Within the intron, we show high correlation between methylation measures obtained by traditional pyrosequencing and our targeted bisulfite sequencing (bis-seq) approach (Figure 3.2B).

The correlation was better among polymorphic intron CpGs compared to the fixed sites, but fixed intron CpGs had uniformly high levels of methylation. It also appears that bis-seq methylation measures are a little higher than pyrosequencing measures. The exact reason for this is not known, but we speculate that the higher GC content of methylated fragments may make them easier to amplify during the bis-seq library preparation. The main discrepancy, however, is at the 3' end of the first pyrosequencing assay (CpG 2113; Figure 3.2B), where pyrosequencing results are more error-prone (Huse et al., 2007). If so, we expect the bis-seq measures to be more accurate. Another issue worth noting is that our methylation measures have been collected and averaged over multiple cell types from RSC dissections. While this is much better than whole-brain or proxy tissue analyses, averaging across multiple cell types suggests a measure of caution. Nevertheless, on balance our technical validations suggest the targeted bis-seq approach is a useful means for exploring methylation variation across a targeted locus and among multiple individuals.

We found dramatic methylation changes across *avpr1a* gene features (Figure 3.2C). We observed low methylation at features with high CpG density (i.e. CpG islands and promoter), and substantially higher levels of methylation as CpG density declined near the end of the first exon and into the intron. Consistent with genome-wide studies (Eckhardt et al., 2006; Hodges et al., 2009), our co-methylation analysis revealed that CpG methylation was correlated at neighboring CpG sites (<1kb), especially between CpGs in the same gene feature (Figure 3.4A). Stronger co-methylation between CpGs within features suggests these labels capture meaningful dimensions of epigenetic regulation across individuals.

Remarkably, examination of V1aR abundance in the retrosplenial cortex (RSC-V1aR) revealed that the methylation status of the *avpr1a* promoter did not predict gene

expression (Figure 3.5C), because the *avpr1a* promoter remains uniformly unmethylated, even in individuals with low V1aR abundance. This is consistent with recent reports from genome-wide studies of mammalian brains (Lister et al., 2013) and multiple cell lines (Bird, 2002; Rollins et al., 2006), which find that CpG-rich promoters are often unmethylated. Indeed, recent work inserting randomized sequences into the mouse genome reveals that sequences with high GC content and high CpG abundance are sufficient to prevent CpG methylation (Wachter et al., 2014). In contrast, work in behavioral epigenetics often focuses more narrowly on individual differences in promoter methylation. For CpG-rich promoters, a lack of methylation seems to be necessary but not sufficient for gene expression. These results emphasize the need to look beyond promoter methylation to interpret epigenetic variation, either in a cell line or among individuals in the wild.

In contrast to the *avpr1a* promoter, gene features located in CpGi-shores had high methylation levels. Average methylation sharply increased around the first exon-intron boundary and remained high (>50%) throughout the intron and enhancer (Figure 3.2C). Sharp methylation transitions at the exon-intron boundary are thought to serve as a signal for regulation of transcription and mRNA splicing (Laurent et al., 2010). Interestingly, our analysis revealed heterogeneous patterns of co-methylation across the *avpr1a* locus, including two clusters of co-methylated CpGs around the exon-intron boundary (Figure 3.4B), suggesting these groups of CpGs are coherent regulatory units. We also noticed a trend towards higher methylation at CpGs immediately upstream of the exon-intron border in animals with elevated *avpr1a* expression (Figure 3.5B). These patterns of coding sequence methylation are all consistent with the hypothesized role of DNA methylation in the specification and splicing of exons during transcription.

66

In general, intron CpG sites were highly methylated and poorly predictive of RSC-V1aR abundance. However, within a previously identified putative intron enhancer (Okhovat et al., 2015), methylation levels were more varied and predictive. We found that high-expressing voles have lower methylation within the intron enhancer (Figure 3.5B, C). Similarly, individual differences in enhancer methylation were negatively correlated with RSC-V1aR (Figure 3.6A). The specificity of this relationship suggests that the lack of CpG methylation at the *avpr1a* promoter may be permissive, while methylation of the intron enhancer may inhibit RSC-V1aR expression. This is consistent with recent studies suggesting genes with CpG islands in the promoter have reliably low levels of methylation, while regulatory elements with low to intermediate CpG density are more likely to exhibit individual or tissue-specific methylation and regulation (Bock et al., 2008; Byun et al., 2009). Similarly, intron enhancers have been documented for a variety of genes (Arnold et al., 2013; Thurman et al., 2012), and loss of DNA methylation can activate such enhancers (Blattler et al., 2014).

We recently reported two *avpr1a* alleles (HI and LO) that predicted individual differences in RSC-V1aR and enhancer methylation among lab-reared prairie voles (Okhovat et al., 2015). Among our wild-caught voles, HI and LO genotypes show different levels of RSC-V1aR (Figure 3.6B; also [Okhovat et al., 2015]) and enhancer methylation (Figure 3.6C). Wild voles also differ in the total number of CpGs in the enhancer of each genotype (Figure 3.6D). Allelic differences in CpG density are caused by polyCpGs that are significantly more common in the intron enhancer than in the rest of the locus (Figure 3.7A,B,C). Polymorphic CpGs in the putative enhancer may drive RSC-V1aR variation by overall changes in CpG and methylation density, by disrupting transcription factor binding sites, or by some more complex combination of the two (Figure 3.1). If a given polyCpG were influencing expression by contributing to overall

67

levels of methylation, this may result in a correlation between *total %DNA methylation* and expression (Figure 3.1A). In contrast, if a CpG polymorphism influenced expression because only one of the alleles was recognized by a transcription factor, then we would expect to see an association between expression and *genotype* (Figure 3.1B). Lastly, if a methylation-sensitive transcription factor binds this site, we would expect to find a correlation between expression and the proportion of methylated CpG alleles (*%methylation per CpG*; Figure 3.1C). In 7 of 8 polyCpGs we found no associations between methylation measures or genotype and expression – these polyCpGs do not seem to shape transcription factor binding sites. They might, however, still contribute in aggregate to regulation through overall methylation. Interestingly, in one polymorphism (polyCpG 2170), we found that both *genotype* and *total %DNA methylation* predicted RSC-V1aR (Figure 3.7D). The CpG polymorphism at site 2170 is one of the defining SNPs of the HI and LO alleles – the only one within our bis-seq target, and its linkage to other sites complicates interpretation of our findings.

The genotype effect at polyCpG 2170 (Figure 3.7D) suggests a sequence-sensitive transcription factor may bind to this site (Figure 3.1B). Based on published position weight matrices, we identified three transcription factors that bind to the sequence containing this SNP (Table 3.3). Interestingly, all three transcription factors showed substantially higher affinity for the LO allele, but none favored the HI allele. At least one of these transcription factors, GATA2, is expressed in the mouse RSC (Figure 3.7E; [Lein et al., 2007]). While GATA2 often activates gene expression, it has been shown to silence expression as well (Schang et al., 2013). Thus GATA2, or some other transcription factor, could directly silence the LO *avpr1a* allele in the vole RSC. In this scenario, the genotype differences in overall enhancer methylation may actually be a downstream consequence of transcription factor-induced silencing. Alternatively, the

association between the site's total %DNA methylation and RSC-V1aR suggest that methylation at this site, possibly in aggregate with methylation at other linked polyCpGs, could suppress the LO allele by attracting methyl-binding proteins such as MeCP2 (Bird, 2002). Unlike the first scenario, this mechanism is not sequence-specific, as it does not depend on the exact sequence context of the polyCpGs. Unfortunately, these two interpretations cannot be distinguished with our current data. Approaches that characterize transcription factor binding to DNA *in vivo*, or that manipulate CpG density while leaving the SNP intact, could clarify the nature this interaction. In either case, our data demonstrate that attempts to link DNA sequence, methylation status and gene expression might do well to focus on enhancers rather than promoters. Such studies will be critical to understanding how genetic variation interacts with developmental environment to produce individual differences in complex behaviors.

In conclusion, we have used modern molecular tools to characterize how CpG distribution and polymorphism predict methylation and expression of the *avpr1a* locus in the brain of wild prairie voles. We find that a targeted bis-seq approach recapitulates traditional pyrosequencing methods, but allows characterization of a larger set of CpG sites. We find that the regulatory effects of *avpr1a* methylation are highly dependent on genetic context: enhancer methylation was associated with low expression while promoter status was not; similarly, methylation in the gene-body may shape transcription and splicing of *avpr1a*. Most polymorphic CpGs do not contribute to *avpr1a* expression by altering transcription-factor binding sites. Rather, allelic differences in methylation or transcription factor binding at polyCpG 2170, seem to shape the effects of the intron enhancer on cortical V1aR and its downstream behaviors. Future studies that target candidate transcription factors, or that modify DNA sequence and/or methylation, will be required to determine the precise mechanisms by which sequence variation influences

avpr1a expression. Overall, our results illustrate some of the complex ways that genetic and epigenetic variation can interact to shape brain and behavior in the wild. Such studies will prove critical to our understanding of plasticity, adaptation and evolution in the wild.

TABLES

Target	Primers for outer PCR	Primers for inner PCR (if nested)
A. Exon1 + promoter	F1:GAAAYGTTGGGTTTGGTGGATTAGTTAG R1:AAAATAATCTTCACRCTACTAACACAAAAC	F1: AAAYGTTGGGTTTGGTGGATTAGTTAG R2:AATACCCCAAAACTAAATAAAAATAACCCAAC --- F2:GGTTTTGTAGAGGAATTTAGGGAGTTTTTTTAG R1:AAAATAATCTTCACRCTACTAACACAAAAC
B. Exon1-intron boundary	F3:TAGTTTATGGTGGTTTTTGAGYGTTGAG R3:CTTACACAATAAACTCTAAAACRATTTCTA	–
C. Intron	F4:GGGGTTTTTGGTTAYGTTTTGTGTTAGTAG R4:CACAAAAATCACCTAAAACCATCCTAAATTTCAA	F4:GGGGTTTTTGGTTAYGTTTTGTGTTAGTAG R5:CCAAAAAAATATATCCATCCCTATCCTTA --- F5:GGGGTTAGGAGTTAGTATGTATGGATTATAT R4:CACAAAAATCACCTAAAACCATCCTAAATTTCAA

Table 3.1. PCR primers (5' -> 3') for bis-seq amplifications.

CpG polymorphism	Frequency
CpG/CpA	48.2% (14/29)
CpG/CpC	3.4% (1/29)
CpG/CpT	7.0% (2/29)
CpG/ApG	7.0% (2/29)
CpG/GpG	3.4% (1/29)
CpG/TpG	31.0% (9/29)

Table 3.2. Frequency of polyCpG variants across the *avpr1a* locus.

Difference log(p)	HI p-value	LO p-value	Matrix ID	Matrix name	Transcription Factor
-2.47	0.803	<0.00273	M00075	V$GATA1_01	GATA binding protein 1
-1.08	0.171	0.0143	M01082	V$BRCA_01	BRCA
-1.03	0.154	0.0145	M00076	V$GATA2_01	GATA binding protein 2

Table 3.3. Transcription factor affinity for the HI and LO alleles at polyCpG 2170.
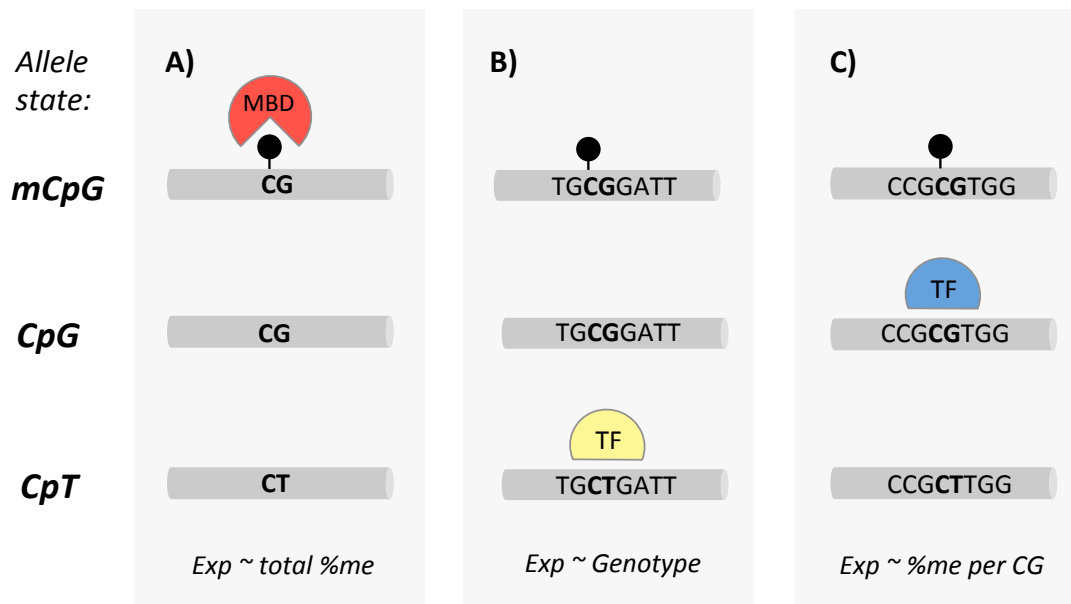
Figure 3.1. Effects of CpG methylation and polymorphism. Loci with polymorphic CpG sites (polyCpG) can be in several allele states (*left*). The CpG allele can be methylated (*top*, methylation depicted by black circle) or unmethylated (*middle*); but alternative allele (e.g. CpT, *bottom*) is always unmethylated. Depending on the context, these states may have different effects on expression. A) PolyCpGs can change local CpG density and susceptibility to DNA methylation. Methylated CpG allele may facilitate binding of methyl-binding domain (MBD) proteins and change gene expression. In this scenario, the strongest predictor of gene expression is amount of total %DNA methylation at polyCpG site. B) When polyCpG is located at the binding site of a transcription factor (TF) that only recognizes one of the alleles, expression is predicted by genotype at polyCpG. C) If polyCpG is located at the binding site of a methylation-sensitive TF, which only recognizes the CpG allele, expression is influenced by both sequence and methylation status, and is most strongly predicted by the fraction of methylated CpG alleles.
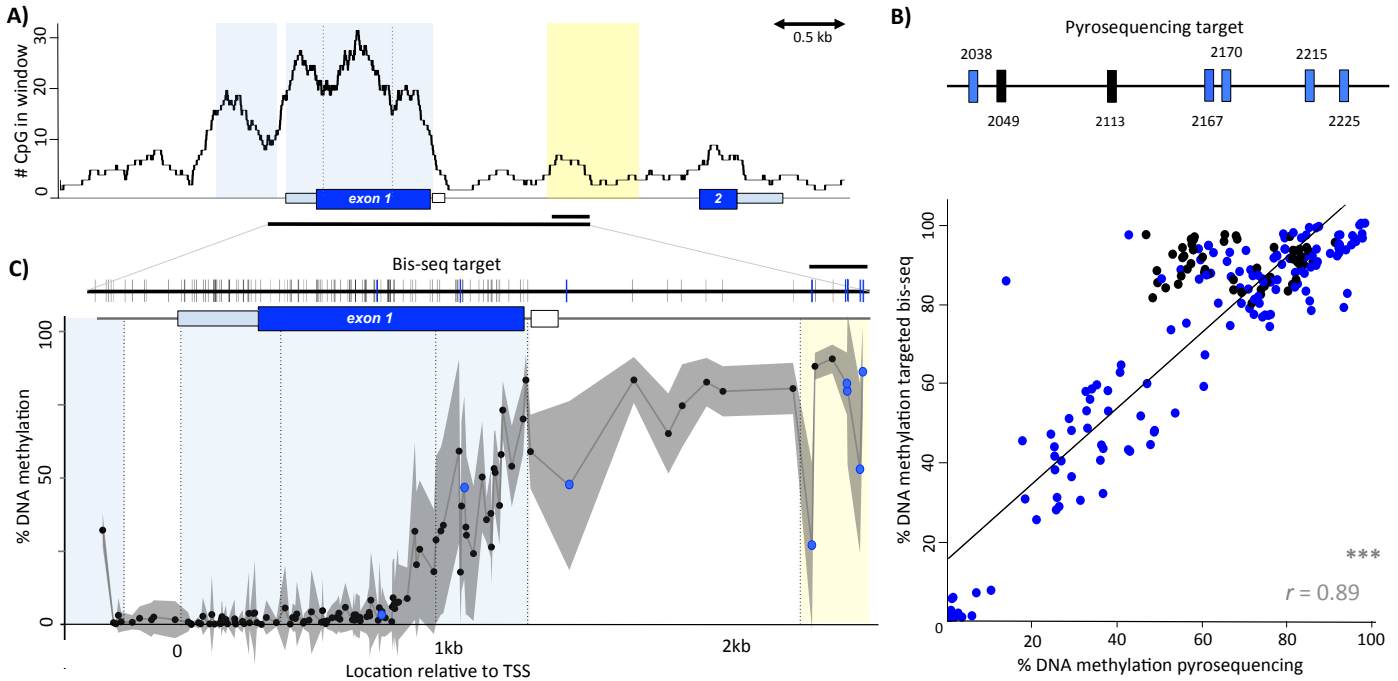
Figure 3.2. CpG distribution and DNA methylation across the *avpr1a* locus. A) A sliding window (window=300bp, step=1bp) of CpG count along 8kb of the *avpr1a* locus. Two predicted CpG islands are shaded light blue and the putative intron enhancer is yellow. The region covered by pyrosequencing (0.2kb) and bis-seq (3kb) marked by horizontal black bar belows. B) Pyrosequencing assay included 2 fixed CpGs (black) and 5 polyCpGs (blue) within the putative enhancer. Pyrosequencing methylation measures correlate with bis-seq results (*r*=0.89, p<0.001). C) *Top*, 113 fixed sites included in bis-seq target are represented by black vertical bars and 8 polymorphic CpGs (polyCpGs) are marked blue. *Bottom*, average %DNA methylation from bis-seq at fixed (black) and polymorphic (blue) CpGs across 3kb of *avpr1a* locus. Standard deviation depicted in gray shading. Gene feature borders are separated by dashed lines. ***p≤0.001.
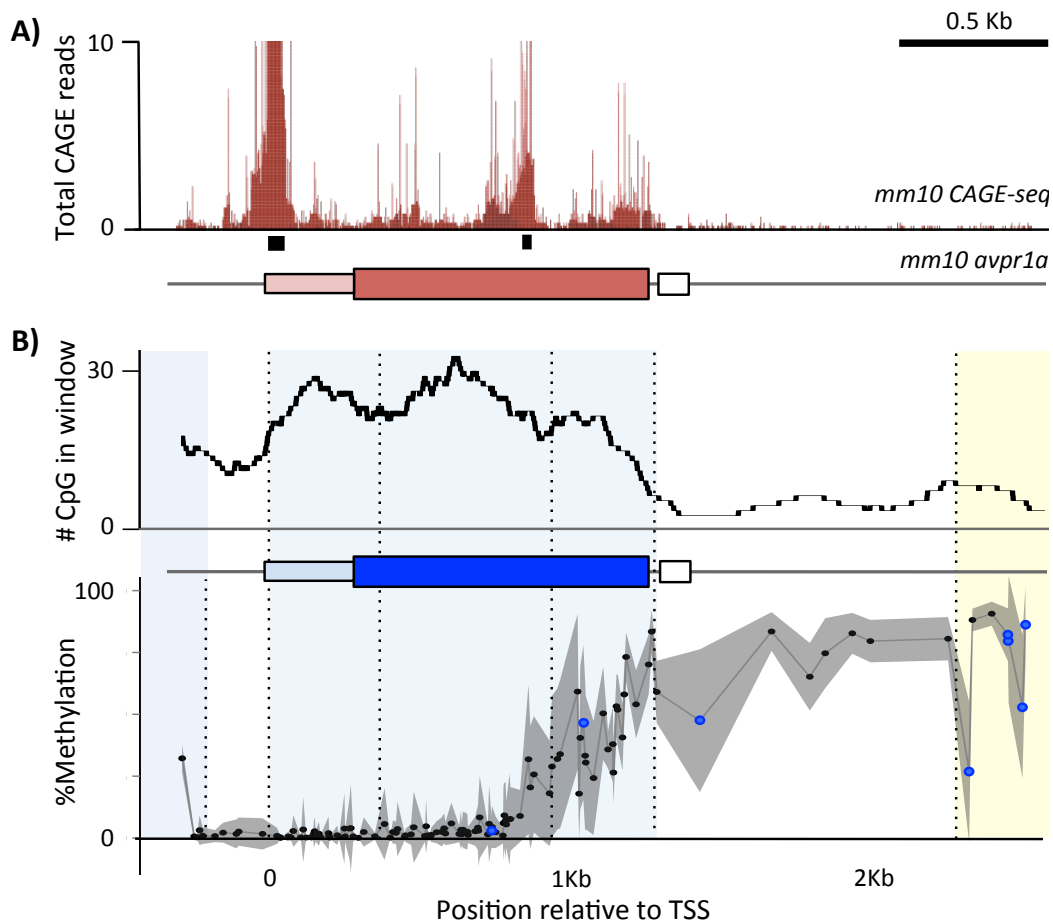
Figure 3.3. CAGE data reveal the 5' boundary of transcripts along the bis-seq target. A) Cap analysis gene expression (CAGE) sequencing reads from the FANTOM5 project (The FANTOM Consortium and the RIKEN PMI and CLST (DGT), 2014) are viewed at the mouse (mm10) *avpr1a* (http://genome.ucsc.edu), within a region homologous to the vole bis-seq target. Significant CAGE peaks are marked by black bars on the bottom of track. B) Modified figures from Figure 3.2, show CpG density and methylation across bis-seq gene-features. Gene-feature borders are depicted by dashed lines.
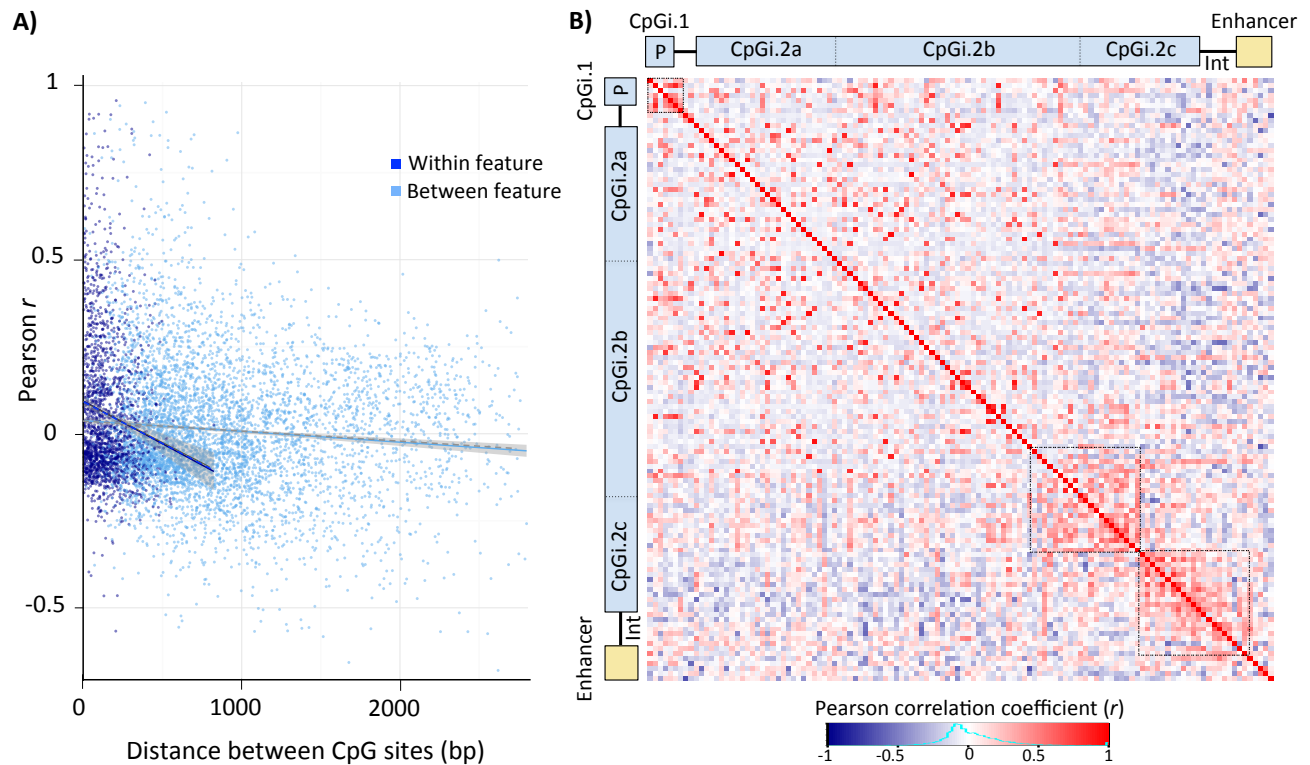
Figure 3.4. Patterns of CpG co-methylation across *avpr1a*. A) Co-methylation measures (Pearson correlation coefficient, *r*) are plotted against distance of CpG pairs within (dark blue, *r*=-0.14, p=0.0001) and between gene features (light blue, *r*=-0.09, p<0.0001). B) Co-methylation between 122 fixed and polymorphic CpG sites depicted in a heatmap. Corresponding gene features are schematized on top and left (abbreviations: P=Promoter, Int=Intron). Three clusters of positive co-methylation are outlined by dashed lines.
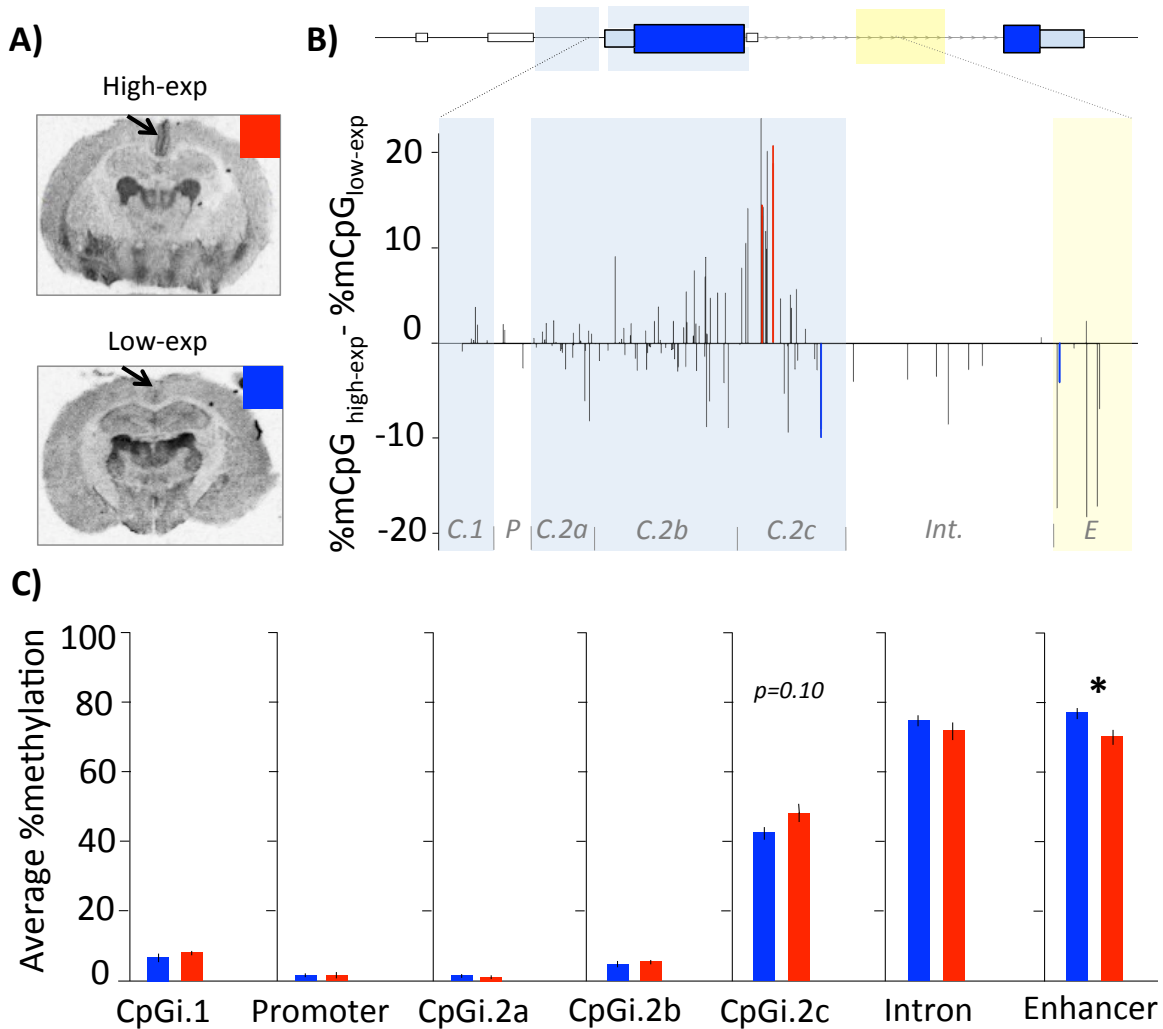
Figure 3.5. Relationship between *avpr1a* DNA methylation and RSC-V1aR. A) Autoradiograms of V1aR show dramatic variation in RSC of wild-caught voles. Median split divided individuals into high-exp (red) and low-exp (blue). B) CpG methylation differences between high-exp and low-exp at 122 CpG sites. Gene features marked on bottom of graph (abbreviations are as follows: C.1=CpGi.1, P=Promoter, C.2a=CpGi.2a, C.2b=CpGi.2b, C.2c=CpGi.2c, Int.=Intron, E=Enhancer). Red bars denote methylation higher among high-exp, blue bars lower methylation among high-exp (t-test, $p \leq 0.05$). C) Average feature methylation in high-exp (red) and low-exp (blue) individuals. Bars are means ± SE. *$p \leq 0.05$.
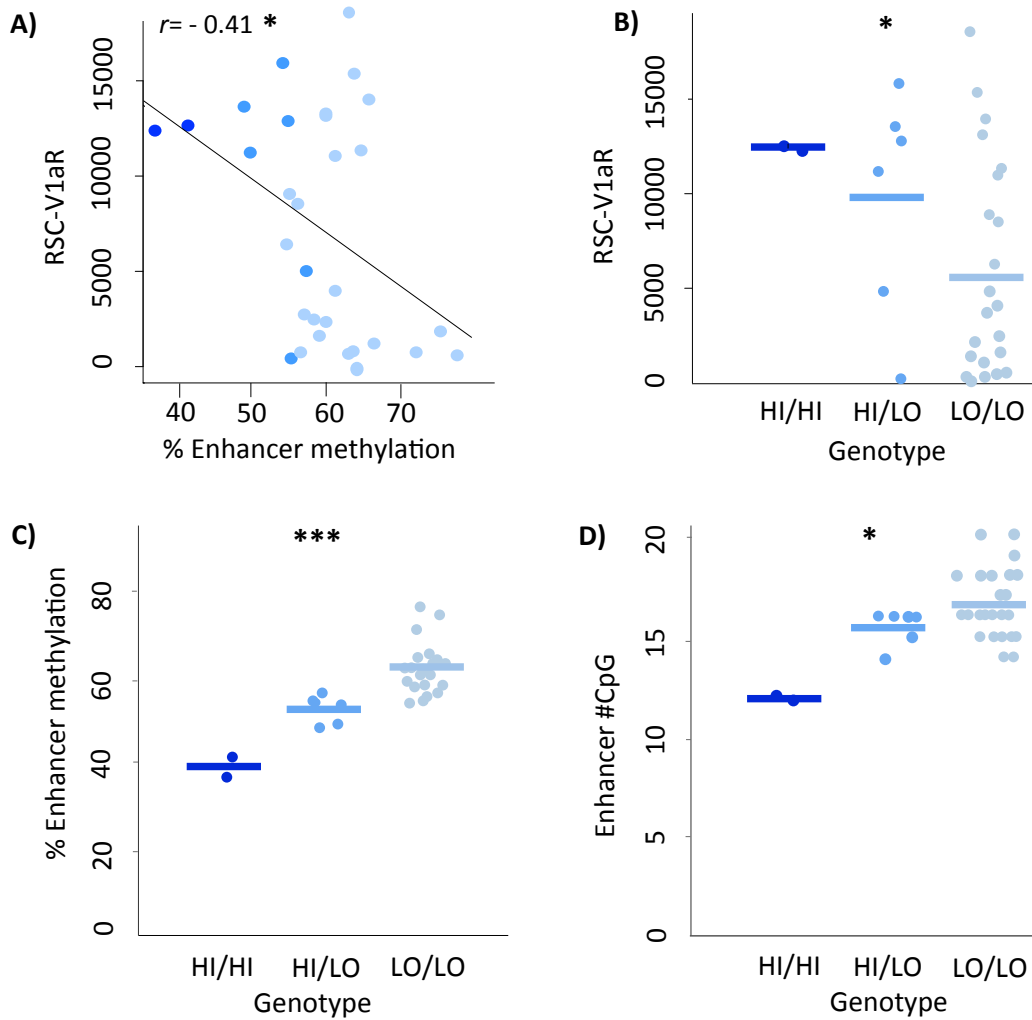
Figure 3.6. *Avpr1a* genotype differences in RSC-V1aR, enhancer methylation and CpG density. A) DNA methylation in the enhancer is negatively correlated with RSC-V1aR abundance. B) *Avpr1a* genotypes (HI/HI, HI/LO and LO/LO) differ in RSC-V1aR (Okhovat et al., 2015) and **C)** average enhancer methylation. D) CpG count within enhancer correlates with *avpr1a* genotype. All bars are means. *p≤0.05, ***p≤0.001
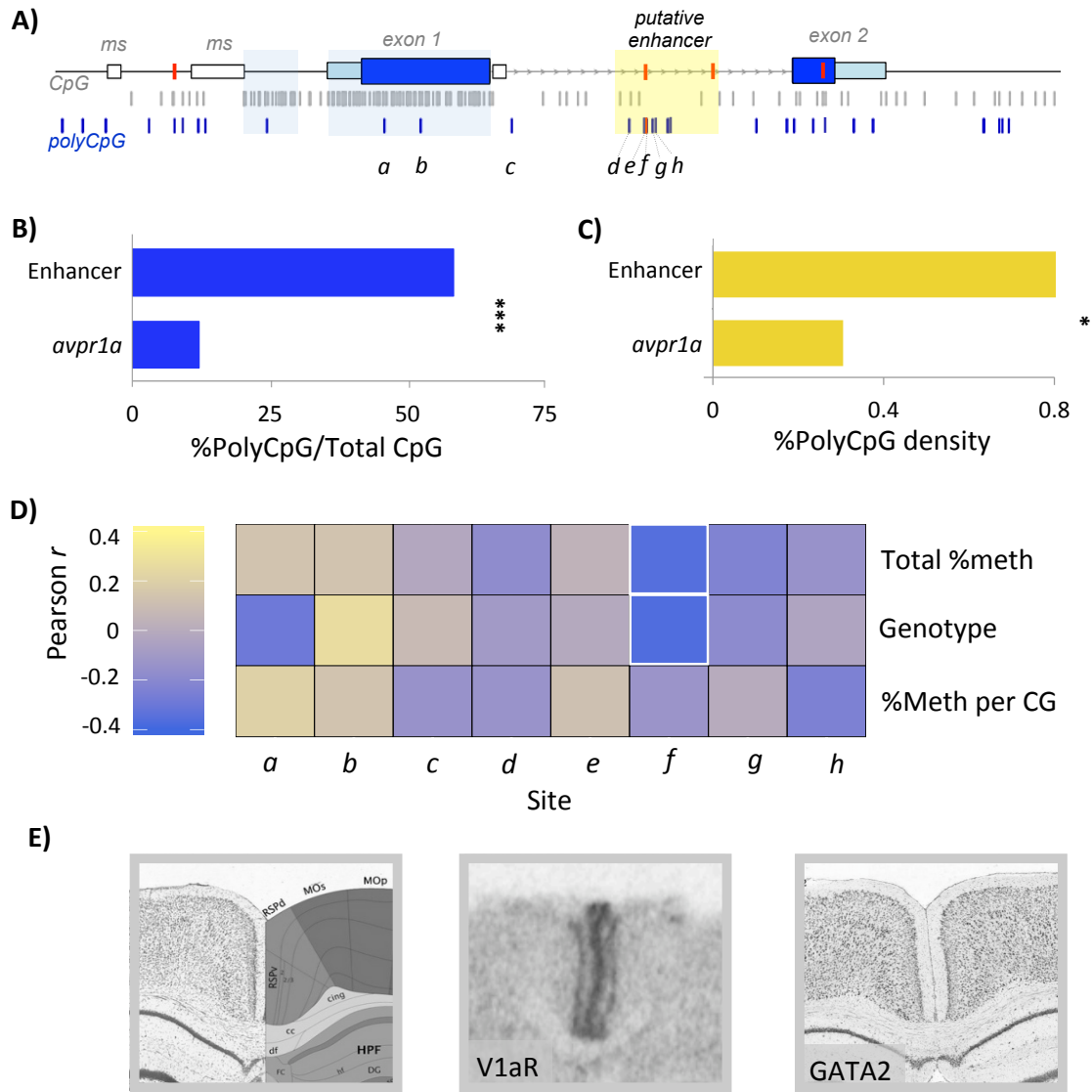
Figure 3.7. Distribution of polyCpGs and their sequence specific associations with RSC-V1aR. A) Distribution of fixed (gray) and polymorphic CpGs (blue) along *avpr1a* locus. The 4 linked SNPs that define HI and LO alleles are marked with red bars on the locus. The 8 polyCpGs covered in bis-seq assay are labeled *a-h* (modified from Okhovat et al., 2015). B) Percent polyCpGs/totalCpG in the enhancer is compared to the rest of the locus. C) Density of polyCpGs (polyCpGs per 100bp) within enhancer compared to rest of locus. D) For each polyCpG Pearson correlation coefficient is calculated between RSC-V1aR abundance and *total %DNA methylation* (*top*), *genotype* (*middle*), and *%methylation per CpG allele* (*bottom*). Cells with p≤0.05 are outlined with a white border. E) *Left*, Nissl image and atlas of mouse brain at the retrosplenial area (RSP). *Center*, prairie vole autoradiogram shows V1aR abundance at the retrosplenial cortex (RSC). *Right*, antisense RNA *in-situ* staining shows expression of GATA2 in the retrosplenial area of mouse (Image credit: Allen Institute. © 2015 Allen Institute for Brain Science. Allen Mouse Brain Atlas. Available from: http://mouse.brain-map.org/gene/show/14237). *p≤0.05 and ***p≤0.001.

DISSERTATOR'S NOTES

Chapter 3 is based on a multi-authored manuscript, currently (December 2016) in review at the journal of Royal Society Open Science. I am the lead author on this manuscript and co-authors are in order: S.M. Maguire and S.M. Phelps. I designed and carried out the molecular laboratory work, participated in data analysis and manuscript preparation.

ACKNOWLEDGEMENTS

Chapter 4: Genetic variation in the developmental regulation of cortical *avpr1a* among prairie voles

ABSTRACT

Early postnatal experiences can have enduring impacts on brain and behavior, but the strength of these effects can be influenced by genetic variation. In principle, polymorphic CpGs (polyCpGs) may contribute to such gene-by-environment interactions (GxE) by altering DNA methylation susceptibility. In this study, we investigate the influence of polyCpGs on the development of vasopressin receptor 1a expression in the retrosplenial cortex (RSC-V1aR) of prairie voles (*Microtus ochrogaster*). Two alternative alleles (HI and LO) predict RSC-V1aR abundance and sexual fidelity in adulthood, and differ in the abundance of CpG sites within a putative intron enhancer. We hypothesized that the elevated frequency of CpG sites in LO alleles would make LO/LO voles more sensitive to developmental perturbations. We found that genotype differences in RSC-V1aR abundance emerged early in ontogeny, and were accompanied by genotype differences in methylation of the putative enhancer. As predicted, postnatal treatment with an oxytocin receptor antagonist reduced RSC-V1aR abundance in LO/LO animals but not their HI/HI siblings. Similarly, methylation inhibition by zebularine increased expression in LO/LO animals, but not in HI/HI siblings. These data demonstrate a gene-by-environment interaction in RSC-V1aR. However, neither the oxytocin receptor antagonist nor zebularine altered the methylation of the putative enhancer, suggesting that differences in sensitivity could not be explained by CpG density at the enhancer alone. Methylated DNA immunoprecipiation-sequencing (MeDIP-seq) revealed additional differentially methylated regions between HI/HI and LO/LO voles. Future research should examine the role of these regions and other regulatory elements in the ontogeny of RSC-V1aR and its environmentally induced changes.

INTRODUCTION

Environmental experiences during early-postnatal development play a pivotal role in shaping an animal's neuronal and behavioral phenotypes later in life. Early-life experiences such as diet (Georgieff, 2007), maternal care (Weaver et al., 2004), stress (Lupien et al., 2009), and toxin exposure (Kundakovic and Champagne, 2011) can drastically change brain and behavior in adulthood. Most studies that use conventional animal models study the effects of early environment while controlling for genetic differences. In nature however, individuals often differ in their sensitivity and response to environmental experiences (Pigliucci, 2001). Genetic differences in environmental sensitivity or "phenotypic plasticity" (Debat and David, 2001) are known as gene-by-environment interactions (Pigliucci, 2005), and recent work has identified many single nucleotide polymorphisms (SNPs) associated with variation in developmental risk or resilience. However, to better understand the mechanisms of GxE, it is important to study this variation in conjunction with the epigenetic modifications that help relay environmental information to genes within the developing brain.

Among epigenetic marks, DNA methylation is the most intensely investigated, and it is often associated with neuronal reprogramming following early-life experiences (Szyf and Bick, 2013). DNA methylation is a stable epigenetic mark that can suppress gene expression by condensing chromatin, disrupting transcription factor binding or attracting methyl-binding proteins, such as MeCP2 (Bird and Wolffe, 1999). Eukaryotic DNA methylation is catalyzed by DNA methylatransferase (DNMT) enzymes, which add a methyl group to cytosines within a CG dinucleotide (Law and Jacobsen, 2010). Since DNA methylation occurs almost exclusively at CpG sites, gaining or losing a CpG may alter susceptibility to DNA methylation and sensitivity to the environment. Interestingly, SNPs commonly occur at CpG sites (Tomso and Bell, 2003) and such polymorphic CpGs

(polyCpGs) have been associated with GxE effects (Parnell et al., 2014); nevertheless, their role in mediating plasticity in the developing brain remains largely unexplored. In the current study we investigate the role of polyCpGs in environmental sensitivity and neuronal phenotype by focusing on the vasopressin receptor 1a (*avpr1a*), a gene implicated in the social behavior of male prairie voles (*Microtus ochrogaster*).

Prairie voles are socially monogamous rodents known for their capacity to form enduring pair-bonds (Getz et al., 1993). Although prairie voles form bonds, individuals vary in their sexual fidelity (Ophir et al., 2008; Phelps and Ophir, 2009). Male fidelity has been linked to individual differences in space-use, and to the expression of vasopressin receptor 1a in the retrosplenial cortex (RSC), a brain region widely implicated in spatial memory (Vann et al., 2009). Abundance of vasopressin 1a receptor in the RSC (RSC-V1aR) is predicted by two *avpr1a* alleles, known as "HI" and "LO". These two alleles are defined by four linked SNPs. One of the SNPs (SNP 2170) is a polyCpG located within a putative intron enhancer and linked to a few other adjacent polyCpGs. The LO allele has significantly more CpG sites in this intron enhancer, and the overall methylation of these sites predicts individual differences in expression (Figure 4.1; Okhovat et al., 2015). In this study, we investigate HI/HI and LO/LO genotype differences in development and sensitivity to pharmacological manipulations.

We explore the ontogeny of V1aR abundance and *avpr1a* enhancer methylation in the RSC of HI/HI and LO/LO prairie vole pups. We ask whether LO/LO voles are more sensitive to neonatal oxytocin receptor antagonist (OTA) exposure, a treatment that was previously shown to decrease RSC-V1aR in adulthood (Bales et al., 2007). We also compare genotype differences in sensitivity to zebularine, a DNMT inhibitor commonly used to disrupt DNA methylation (Cheng et al., 2003). We then examine how these treatments affect DNA methylation in the putative intron enhancer. Lastly, we use high-

throughput sequencing to explore more distal differentially methylated regions (DMRs) between HI/HI and LO/LO voles, and discuss the mechanisms that may underlie our findings.

METHODS

Animal subjects

All animals were lab-reared descendants of prairie voles captured in Jackson County, IL. Breeding pairs heterozygous for the HI and LO *avpr1a* alleles, were used to generate homozygous HI/HI and LO/LO animals. All breeding pairs were kept in 25x45x60cm cages in accordance to IACUC regulations and were given food and water *ad libitum*.

To examine the natural ontogeny of V1aR abundances and *avpr1a* methylation, unmanipulated pups were taken from 7 heterozygous breeding pairs on day 1, 7 or 14. Developmental manipulations were all performed on postnatal day 1 (P1), and brains were taken on P21. For all three experiments, brains and tail clippings were collected and frozen on dry ice following euthanasia. Frozen tissues were stored at -80°C until further processing.

Neonatal manipulations

At P1, litters from heterozygous breeding pairs received drug or one of two control treatments (saline injection or handling). Treatments were given in randomized orders across breeding pairs. For drug dose calculations we assumed pups weighed on average 3g at P1. All injections were done intraperitoneally with 30-gauge insulin syringes.

Our first manipulation involved injection of 0.1mg/kg oxytocin receptor antagonist (OTA, [d(CH2)5, Tyr(Me)2, Orn8]-vasotocin, Bachem) dissolved in 50μl

injectable saline. Control litters were either handled or injected with 50μl saline vehicle. All pups were returned to their home-cage after treatment, and remained in the cage undisturbed until weaning. At P21, animals were sexed and euthanized. The experiment was set up with 12 heterozygous breeding pairs, but individuals from two breeding pairs died before handling control litters were collected, reducing our sample size for the handling-only control.

To manipulate developmental methylation, we administered 400mg/kg fresh zebularine (1-β-D-Ribofuranosyl-2(1H)-pyrimidinone, Tocris Biosciences) in 50μl sterile saline. Control animals received 50μl sterile saline or were handled without injection. The manipulations were conducted on repeated litters from 7 heterozygous breeding pairs.

Genotyping

 Genomic DNA (gDNA) was extracted from tail clippings using the DNeasy blood and tissue kit (Qiagen) according to the manufacturer's protocol. All PCR amplifications were performed on a BioRad C1000 Thermal cycler (BioRad).

We determined sex of weanlings (P21) visually by inspecting anogenital distance. At P1, P7 and P14 we used a PCR assay targeting the *SRY* locus to determine sex. The PCR assay was validated by correctly predicting sex of 6 control male and female samples with known sex. We used primers designed to amplify a 214bp region of the prairie vole SRY locus (FN433505.1, NCBI). Primer sequences were: F:3'-GTGGTCTCGTGATCAGAGGCGCAAG-5'   and   R:3'-GGGTCTTGAGTCTCTGTGCCTCTTG-5'. PCR reactions were set up in 25ul reactions consisting of GoTaq Hot Start Colorless Master Mix (Promega), 200nM of each primer,

1.5ul gDNA and nuclease-free water. The PCR reaction condition was: 3min at 93°C, {30sec at 93°C, 30sec at 63°C and 10sec at 72°C}X32, 10sec at 72°C.

All individuals were genotyped for the *avpr1a* HI and LO allele using PCR amplification and allele-specific restriction digestion as described previously (Okhovat et al., 2015). Briefly, in a nested PCR assay we amplify 0.8Kb of the *avpr1a* intron, including the 2170 SNP that predicts HI and LO alleles. Next, we digest this amplicon with Bsh1236I restriction enzyme (ThermoScientific) using settings recommended by manufacturer. The 2170 SNP produces a CGCG or CGCT sequence in the LO and HI allele, respectively. Therefore Bsh1236I, which recognizes the CGCG sequence, will only digest the LO allele. Following digestion, we ran samples on agarose gel to visualize the banding pattern and determine genotype. Only homozygous subjects (HI/HI and LO/LO) were used for subsequent processing.

RSC-V1aR autoradiography

Frozen brains from homozygous HI/HI and LO/LO animals were sectioned in 20μm-thick slices at 100μm intervals and mounted on SuperFrost slides (Fisher Scientific) in four series. The autoradiography procedure has been described previously (Okhovat et al., 2015; Ophir et al., 2008). In brief, sections were lightly fixed in 1% paraformaldehyde solution following a quick drying step. Slides were then washed in Tris and incubated with 50pM of the radiolabeled 125I-linear arginine vasopressin receptor antagonist (NEX310010UC, Perkin Elmer) for 70min. Following incubation, slides were washed multiple times in Tris and rapidly dried under hot air. Sections were exposed to film for 68hrs, along with radiographic standards. Developed films were digitized using Epson perfection V800 Photo scanner. For each individual, V1aR abundance in the RSC was scored from three sections using FIJI software (Schindelin et

al., 2012) and averaged. Binding in non-expressing cortical regions of the same section was used to correct for non-specific binding.

DNA methylation

The RSC was dissected from alternative fresh frozen slides. We used the EpiTect Plus LyseAll Bisulfite Kit (Qiagen) to obtain bisulfite-converted gDNA from the dissected tissue. Next, we used a nested PCR approach to amplify two pyrosequencing-compatible fragments of the putative intron enhancer. Primer sequences and PCR settings have been described previously (Okhovat et al., 2015). PCR amplicons were sent to Epigendx (Hopkinton, MA) for bisulfite pyrosequencing.

DNA methylation was measured at 4 fixed and 3 polymorphic CpG sites (polyCpGs) and reported to us as %(unconverted C/[unconverted C+ converted T]) at each CpG site. Genotype at polyCpG 2170 was also reported along with the methylation measurements. Animals with conflicting genotypes from the pyrosequencing and in-house digestion assay were excluded from the data set (n=7). Total %DNA methylation was calculated by averaging methylation across all 7 CpG sites (fixed and polymorphic). At the 3 polymorphic CpG sites, some of the inter-individual variation in methylation arises as a result of sequence differences that abolish CG dinucleotides. To examine changes in methylation independent of sequence variation, %methylation at fixed CpG sites was calculated by averaging methylation at the 4 fixed (non-polymorphic) CpG sites. %Methylation at methylable CpGs was calculated by averaging methylation only at CpG sites that contained a CG dinucleotide (i.e. all fixed CpG sites and polyCpGs that contain a CG allele).

Statistical analysis

All data were analyzed in R (https://www.r-project.org). Linear mixed models were generated using the nlme package (Pinheiro et al., 2016). Single tail statistics were only used when direction was clearly predicted and justified based on previous data. Sex was excluded from analysis when significant effects were not detected.

Ontogeny of RSC-V1aR and avpr1a methylation

To characterize the ontogeny of RSC-V1aR at P1, P7 and P14 we generated a mixed model with genotype, age and genotype x age interaction as fixed effects and parentage as a random effect. Since the power to detect interaction is generally low, and we had clear *a priori* predictions for interaction effects, we also examined RSC-V1aR levels between the HI/HI and LO/LO genotype at each age using single-tailed Welch t-tests.

To characterize the relationship between early post-natal RSC-V1aR and *avpr1a* methylation we performed a simple linear regression between RSC-V1aR and all three methylation measures in the putative enhancer. Next, we examined the effects of age and genotype on *avpr1a* methylation by generating a mixed model with age, genotype and age x genotype interaction as fixed effects and parentage as a random effect. We compared DNA methylation measures between HI/HI and LO/LO genotypes at each age by single-tailed Welch t-tests.

Neonatal manipulations

We examined the effects of two different neonatal manipulations on the RSC-V1aR of HI/HI and LO/LO animals. The two control groups (saline and handling only) did not show any statistical difference in any cases and were thus combined into a single control group (CON) within each study.

First, we assessed effects of P1 oxytocin antagonist injections (OTA) on the RSC-V1aR at P21. We built a mixed model with sex, genotype, treatment and genotype x treatment as main effects and parentage as a random effect. Effects of OTA treatment were further examined within genotypes using mixed models with sex and treatment as fixed effects and parentage as a random effect.

Next, we examined effects of neonatal zebularine administration on P21 RSC-V1aR. Weight measures at P21 were used to test long-term cytotoxicity of zebularine. No significant effect of zebularine treatment was detected on weight, suggesting that long-term cytotoxic effects of zebularine are negligible at weaning. This observation is consistent with previous studies that demonstrate minimal cytotoxicity for zebularine (Cheng et al. 2003). We then generated a mixed model for RSC-V1aR with genotype, treatment and their interaction as main effects and parentage as a random effect. Next, we split the data set based on genotype and for each genotype we generated a mixed model with treatment as main effect and parentage as random effect.

For subjects in both the OTA and zebularine study, we examined the association of *avpr1a* intron enhancer methylation measures and RSC-V1aR using a linear regression. For each study, we built a mixed model for total %DNA methylation with genotype, treatment and genotype x treatment as main effects and parentage as random effect. For consistency, we next split each data set based on genotype and generated mixed models with treatment as main effect and parentage as a random effect. We also built models with genotype, treatment and genotype x treatment interaction as main effect and parentage as random effect for %methylation at fixed CpGs and %methylation at methylable CpGs. We did not further split the data as no significant effect was detected in these models.

Methylated DNA immunoprecipitation (MeDIP) validation

In order to explore methylation differences outside the targeted enhancer, we validated and used a methylated DNA immunoprecipitation (MeDIP) assay for gDNA from the vole RSC. First, gDNA was extracted using a standard phenol:chloroform procedure and incubated with RNase A (Fisher Scientific) for 45min at 37°C to remove all traces of RNA. A 20ug aliquot of the gDNA was brought to 450ul with 1xTE and sheared into 200-700bp fragments on ice, using Q125 sonicator ([5 sec pulse, 5 sec rest] X15 at 60% power, Qsonica). A 10% aliquot was taken from each sample as INPUT and the rest were used in MeDIP similar to previously described protocols (Mohn et al., 2009). Briefly, sheared DNA was denatured at 98°C for 10min and immediately transferred to ice for another 10min. We then added anti-5mC antibody (ab10805, Abcam) in a 1:1 mass ratio to DNA, and 50ul of 10X IP buffer (100mM Na-Phosphate pH=7.0, 1.4M NaCl, 0.5% Triton X-100) and allowed samples to rotate for 2hrs at 4°C. Following incubation, we added 40ul of cleaned Dynabeads M-280 Sheep Anti-Mouse IgG (Life technologies) in 0.1% PBS-BSA to each tube and incubated for another 2hrs at 4°C with rotation. Next, beads were collected on a magnetic stand and washed three times with 1X IP buffer. Cleaned beads were incubated with Proteinase K (Life Technologies) in 50mM Tris pH=8.0, 10mM EDTA and 0.5% SDS at 50°C for 3hrs with rotation. DNA was purified from beads using standard phenol:chloroform extraction and EtOH precipitation protocol.

To assess the sensitivity of the MeDIP assay we compared enrichment of native and in vitro methylated gDNA at the 5'UTR of *β-actin*, the promoter of *gapdh* and within the *avpr1a* intron. Prior to MeDIP, a 1μg aliquot of the fragmented gDNA was *in vitro* methylated using CpG Methyltransferase (M.SssI, New England Biolabs) according to manufacturer's protocol and cleaned using ZymoResearch Clean and Concentrator kit

(Zymo Research). Total amount of in vitro methylated gDNA was calculated based on concentration measures from nanodrop 2000 (Thermo Scientific) and an equal amount of the native fragmented gDNA was taken. A 10% INPUT aliquot was set aside from both native and *in-vitro* methylated gDNA and the remaining were immunoprecipitated in parallel according the MeDIP protocol described above. MeDIP outputs were examined using qPCR with following primers: $F_{\beta\text{-actin}}$: 5'-GGAGCGGCGGAGAAAGAGC-3', $R_{\beta\text{-actin}}$: 5'-GCGAGGCAGGTGAGTGAGC-3'; F_{gapdh}: 5'- GCCCAACCAGTCCCAGCAC-3', R_{gapdh}: 5'- ACGAGAGAGGTCCAGCTACTC-3' and F_{avpr1a}: 5'-GCCTCACACAGTTCCTCATGTTG-3', R_{avpr1a}: GTCACCTAAGCCCATCCTGAATTTC-3'. All qPCR amplifications were carried out on ViiA Real Time PCR system (Life Technologies) in 10ul reactions consisting of KAPA SYBR FAST qPCR master mix (ROX low, Kapa Biosystems), 200nM of each primer and 1ul DNA. Amplifications were performed in triplicate using the following settings: 1min at 95°C, {1sec at 95°C, 20sec at 60°C}X40. To adjust data based on amount of DNA input, we calculated %INPUT enrichment. Our qPCR technical replicates were used to calculate means and error bars in Figure 4.5A, as well as Welch t-tests comparisons between enrichment levels of native and in vitro methylated DNA.

To assess specificity of the MeDIP assay, we examined specific and non-specific antibody binding to DNA standards provided in the hMeDIP kit (Active Motif). DNA standards consisted of unmethylated, fully hydroxymethylated (5-hmC) or fully methylated (5-mC) amplicons of the human *APC* locus. We spiked equal aliquots of sheared vole gDNA with each of the *APC* standards, set aside 10% as INPUT and performed MeDIP according to the protocol described. To confirm absence of background pull-down and nonspecific qPCR amplification, we also carried out two parallel MeDIPs; one with unspiked vole gDNA and 5-mC antibody, the other with

unspiked vole gDNA and IgG antibody. For all assays, INPUT and MeDIP outputs were qPCR amplified in triplicates using *APC* primers provided in the hMeDIP kit (Active Motif). To normalize qPCR results for amount of starting material, we calculated %INPUT enrichment. Technical replicates were used to calculate means and error bars for Figure 4.5B and to perform pair-wise Welch t-tests between 5-mC spiked MeDIP and each negative control assay.

MeDIP-sequencing (MeDIP-seq) on the RSC of HI/HI and LO/LO voles

Genomic DNA was extracted from RSC of 4 HI/HI and 4 LO/LO sexually naive adult males. Each sample was treated with RNase, sheared and cleaned up as described above. DNA concentrations were then measured on a nanodrop 2000 (Thermo Scientific). We combined equal amounts of gDNA from each of the 4 HI/HI and 4 LO/LO individuals to generate one HI/HI and one LO/LO gDNA pool. From each of the two pools, 5μM DNA was end repaired, adenylated, adaptor-ligated and size-selected (250-700bp) using a KAPA LTP library prep kit (KAPA Biosystems) according to manufacturer's instructions. We stored 10% of each sample in -20°C as INPUT, and proceeded with the MeDIP process as described above. MeDIP outputs were PCR amplified for 5 cycles using KAPA LTP library prep kit (KAPA Biosystems). INPUT and MeDIP libraries were submitted for single-end sequencing on the Illumina HiSeq 4000 platform at genomic sequencing and analysis facility at University of Texas at Austin.

Sequencing reads from INPUT and MeDIP were aligned to the modified prairie vole draft genome assembly (Okhovat et al., 2015) using bwa (Li and Durbin, 2009) with default single-end settings. Alignments were improved by stampy (Lunter and Goodson, 2011) with default settings. Non-unique and low quality alignments (mapping

quality<20) were filtered out using samtools (Li et al., 2009). We used MACS2 (Zhang et al., 2008) to normalize MeDIP read counts and to generate fold enrichment tracks. Fold enrichment (FE) values at the *avpr1a* locus for the two genotypes were subtracted to visualize differentially methylated regions (DMRs) in the UCSC genome browser (https://genome.ucsc.edu/). Within each DMR the site with maximum FE difference between genotypes was considered the DMR summit.

RESULTS

Ontogeny of RSC-V1aR and *avpr1a* methylation

To characterize early postnatal changes in the abundance of vasopressin receptor 1a in the retrosplenial cortex (RSC-V1aR) we measured RSC-V1aR in unmanipulated homozygous HI/HI and LO/LO pups at postnatal day 1 (P1; $n_{HI/HI}$=6, $n_{LO/LO}$=7), P7 ($n_{HI/HI}$=7, $n_{LO/LO}$=8) and P14 ($n_{HI/HI}$=7, $n_{LO/LO}$=6). RSC-V1aR was undetectable at P1 in both genotypes; therefore we assigned zero RSC-V1aR to all P1 individuals. RSC-V1aR increased significantly with age within the first two weeks of life, especially among HI/HI animals. We found significant effects of age and genotype x age interaction on RSC-V1aR (genotype t_{30}=-0.04, p=0.97; age t_{30}=3.31, p=0.003; genotype x age t_{30}=2.54 p=0.02; Figure 4.2A,B). Our model explained 72% of all the RSC-V1aR variation. Although there was no main effect of genotype, the significant genotype by age interaction suggested this might be due to the lack of expression at P1 in both genotypes. Indeed, we found that HI/HI animals had significantly more RSC-V1aR compared to LO/LO animals at both P7 (LO/LO: 1280$\pm$341, HI/HI: 2911$\pm$656, mean$\pm$SE; Welch t-test: t=-2.21, p=0.03) and P14 (HI/HI: 5319$\pm$1008, LO/LO: 2757$\pm$669; Welch t-test: t=-2.12, p=0.03) but not at P1 (Figure 4.2A,B).

Across ages, we found a negative correlation between RSC-V1aR abundance and i) total %DNA methylation (r=-0.39, p=0.014), ii) %methylation at fixed CpG sites (r=-0.41, p=0.008) and iii) %methylation at methylable CpGs (r=-0.44, p=0.006). We also assessed effects of genotype, age and their interaction on each of the methylation measures. There was a significant effect of genotype on total %DNA methylation but no effect of age or genotype x age interaction (genotype t_{29}=-18.54, p<0.0001; age t_{29}=-0.20, p=0.84; genotype x age t_{29}=-1.33, p=0.19; Figure 4.2C). For %methylation at fixed CpG sites, we found no genotype or age effect, and only a weak trend was detected in the genotype x age interaction (genotype t_{29}=-0.29, p=0.78; age t_{29}=-0.10, p=0.92; genotype x age t_{29}=-1.61, p=0.12; Figure 4.2D). We found similar results when examining %methylation at methylable CpGs (genotype t_{29}=-0.61, p=0.54; age t_{29}=-0.14, p=0.89; genotype x age t_{29}=-1.61, p=0.12; Figure 4.2E).

Pair-wise genotype comparisons revealed that total %DNA methylation was higher in LO/LO pups compared to HI/HIs across all ages (Welch t-test, P1: t=17.36, p<0.0001; P7: t=28.07, p<0.0001; P14: t=16.48, p<0.0001; Figure 4.2C). At fixed CpG sites however, LO/LO animals had higher methylation at P7 (Welch t-test: t=2.34, p=0.02) and P14 (Welch t-test: t=1.96, p=0.045) but not at P1 (Welch t-test: t=0.54, p=0.30; Figure 4.2D). Similarly, genotype comparisons of %methylation at methylable CpGs showed that the LO/LO animals had higher methylation compared to HI/HI subjects at P7 (Welch t-test: t=3.51, p=0.002) and P14 (Welch t-test: t= 2.19, p=0.03) but not at P1 (Welch t-test: t=0.93, p=0.19; Figure 4.2E).

Effects of neonatal oxytocin receptor antagonist and zebularine injections on RSC-V1aR

We found a significant main effect of genotype, treatment and sex on RSC-V1aR following P1 oxytocin receptor antagonist (OTA) injections (genotype: t_{36}= 2.70, p=0.01;

treatment: t_{36}=-2.34, p=0.03; sex: t_{36}=-3.09, p=0.004; Figure 4.3A). This model accounted for 52% of the overall variation in RSC-V1aR. The genotype x treatment term was not significant (t_{36}=1.31, p=0.20). Nevertheless, RSC-V1aR levels were significantly lower in LO/LO voles that received OTA compared to control subjects (LO/LO$_{OTA}$: 669$\pm$100, LO/LO$_{CON}$: 1698$\pm$227, mean$\pm$SE; t_{19}=-2.78, p=0.01; Figure 4.3A), while HI/HI animals were unaffected (HI/HI$_{OTA}$: 2177$\pm$434, HI/HI$_{CON}$: 2538$\pm$285; t_{10}=-0.38, p=0.71; Figure 4.3A). In our *post hoc* models, effect of sex on RSC-V1aR was only found among HI/HI animals (t_{10}=-2.63, p=0.03).

In our zebularine study, we found an effect of both genotype and treatment, but no genotype x treatment interaction effect (genotype: t_{39}=4.52, p=0.0001, treatment t_{39}=1.98, p=0.055, genotype x treatment: t_{39}=-0.63, p=0.53; Figure 4.3B). This model explained 50% of the overall variation in RSC-V1aR. Splitting the data by genotype, however, revealed a significant main effect of zebularine treatment among LO/LO animals (LO/LO$_{ZEB}$: 6946$\pm$763, LO/LO$_{CON}$: 5584$\pm$1109, mean$\pm$SE; t_{16}=2.37, p=0.031), but not HI/HI subjects (HI/HI$_{ZEB}$: 12932$\pm$1138, HI/HI$_{CON}$: 11464$\pm$1373; t_{20}=1.09, p=0.29; Figure 4.3B).

Effects of neonatal manipulations on methylation of *avpr1a* enhancer

Among subjects in the OTA study, we found a significant linear relationship between total %DNA methylation within the *avpr1a* enhancer and RSC-V1aR (r=-0.47, p=0.0004; Figure 4.4A). However, we did not find a significant correlation between RSC-V1aR and total %DNA methylation at fixed CpGs (r=-0.14, p=0.30) or %methylation at methylable CpGs (r=-0.14, p=0.32).

In the OTA study, we found a significant effect of genotype on total %DNA methylation but no effect of treatment or genotype x treatment interaction (genotype:

t_{37}=-13.19, p<0.0001; treatment: t_{37}=1.038, p=0.31; genotype x treatment: t_{37}=-1.19, p=0.24; Figure 4.4B). This model accounted for 88% of the total variation in total %DNA methylation in the *avpr1a* enhancer. Effect of treatment on total %DNA methylation remained insignificant even after splitting subjects based on genotype (HI/HI: t_{11}=-1.50, p=0.16; LO/LO: t_{19}=1.11, p=0.28; Figure 4.4B). We found no effect of genotype, treatment or genotype x treatment interaction for either %methylation at fixed CpGs (genotype: t_{37}=1.40 p=0.17; treatment: t_{37}=0.19, p=0.85; genotype x treatment: t_{37}=-1.52, p=0.14) or %methylation at methylable CpGs (genotype: t_{37}=0.48 p=0.63; treatment: t_{37}=0.09, p=0.93, genotype x treatment: t_{37}=-1.43, p=0.16).

Among subjects in the zebularine study, we found a significant linear relationship between total %DNA methylation of the *avpr1a* enhancer and RSC-V1aR (r=-0.60, p<0.0001; Figure 4.4C). We did not find an association between RSC-V1aR and %methylation at fixed CpGs (r=-0.1, p=0.48), but we detected a trend between RSC-V1aR and %methylation at methylable CpGs (r=-0.24, p=0.09).

In the zebularine study, we found a significant main effect of genotype and treatment on total %DNA methylation but no genotype x treatment interaction (genotype: t_{39}=-5.65, p<0.0001; treatment: t_{39}=1.99, p=0.05; genotype x treatment: t_{39}=-1.59, p=0.11; Figure 4.4D). Our model accounted for 74% of the total variation in total %DNA methylation in the *avpr1a* enhancer. After splitting the data by genotype, no effect of treatment was found in either LO/LO (t_{16}=1.35, p=0.20; Figure 4.4D) or HI/HI subjects (t_{20}=-0.86, p=0.40, Figure 4.4D). We found no effect of genotype, zebularine treatment or genotype x treatment interaction on %methylation at fixed CpGs (genotype: t_{39}=1.39, p=0.17; treatment: t_{39}=-0.02, p=0.98; genotype x treatment: t_{39}=-0.55, p=0.59) or %methylation at methylable CpGs (genotype: t_{39}=0.03, p=0.98; treatment: t_{39}=0.59, p=0.56; genotype x treatment: t_{39}=-0.92, p=0.36).

MeDIP validation and MeDIP-seq of HI/HI and LO/LO RSC

We used methylated DNA immunoprecipitation-sequencing (MeDIP-seq) to find differentially methylation regions (DMRs) outside the immediate vicinity of the *avpr1a* locus. First, to validate the technique we performed parallel MeDIPs on *in-vitro* methylated and native vole genomic DNA (gDNA). The MeDIP output from native gDNA showed low %INPUT enrichments at *β-actin* (2.67$\pm$0.28%, mean$\pm$SD) and *gapdh* (1.83$\pm$0.03%, mean$\pm$SD), where we expect low methylation levels in native gDNA, but not at the *avpr1a* intron (15.02$\pm$0.46%, mean$\pm$SD), where CpGs are highly methylated in native gDNA (Okhovat et al., unpublished). Compared to native gDNA, MeDIP enrichment of *in-vitro* methylated DNA significantly increased at both *β-actin* (Welch t-test, t=-65.76, p=0.0002) and *gapdh* (Welch t-test, t=-19.00, p=0.002). In contrast, %INPUT enrichment decreased at the *avpr1a* enhancer (Welch t-test, t=4.77, p=0.01), most likely reflecting the higher genome-wide competition for antibody binding following *in-vitro* methylation (Figure 4.5A). Next, we tested the specificity of our assay by performing MeDIP on unmethylated, hydroxymethylated (5-hmC) or methylated (5-mC) DNA standards. Background precipitation and non-specific amplification was relatively low, with %INPUT measures ranging from only 0.21 to 31.03%. Significantly higher %INPUT enrichment was obtained by performing MeDIP on vole gDNA spiked with 5-mc standard DNA (387$\pm$17%, mean$\pm$SD; Welch t-test, all p<0.001; Figure 4.5B).

Across the *avpr1a* gene, fold enrichment (FE) of MeDIP-sequencing (MeDIP-seq) compared to input DNA revealed that both *avpr1a* genotypes had low DNA methylation at the transcription start site (TSS). Methylation levels increase towards the first exon-intron boundary, consistent with earlier descriptions of methylation at the *avpr1a* locus (Okhovat et al., unpublished). We also found elevated DNA methylation at the transition between the second exon and 3'UTR. In a 25kb window centered at the

avpr1a TSS, MeDIP-seq fold enrichment ranges were similar between the two genotypes, with LO/LO ranging from 0.00 to 4.08 and HI/HI from 0.00 to 3.53. However, within this window we found that LO/LO animals showed higher methylation levels in the putative intron enhancer compared to HI/HI animals (FE at summit, LO/LO = 1.64, HI/HI=0.07; Figure 4.5C,D). Additionally, we found two new DMRs located ~0.5 and 8.5kb upstream of the *avpr1a* TSS. At both DMRs, DNA methylation was higher in the LO/LO sample compared to HI/HI sample ($DMR_{0.5kb}$: LO/LO = 2.72, HI/HI =0.53; $DMR_{8.5kb}$: LO/LO =2.76, HI/HI =0.14; Figure 4.5D,E).

DISCUSSION

In this study we characterize developmental expression of vasopressin receptor 1a in the retrosplenial cortex (RSC-V1aR), a neuronal phenotype implicated in sexual fidelity and space use of prairie voles in adulthood (Figure 4.1; Okhovat et al., 2015; Ophir et al., 2008). We used the *avpr1a* locus to explore genetic differences in RSC-V1aR ontogeny and the role of CpG polymorphisms in developmental plasticity.

By visualizing brains of unmanipulated pups on postnatal day 1 (P1), P7 and P14, we found that RSC-V1aR levels changed drastically during the first two weeks of life in both HI/HI and LO/LO animals. All pups were born with undetectably low levels of RSC-V1aR that increased rapidly in the following two weeks (Figure 4.2A,B). These observations were generally in line with previous descriptions of V1aR ontogeny in the RSC of prairie voles (Wang et al., 1997). However, we observed that the postnatal rise in V1aR was steeper in HI/HI voles compared to LO/LO animals, as evident by a significant age x genotype interaction in a mixed model for RSC-V1aR. Our *post hoc* comparisons of HI/HI and LO/LO genotypes across age groups revealed that although there are no RSC-V1aR genotype differences at birth, HI/HI subjects had significantly more RSC-

V1aR than LO/LOs at both P7 and P14 (Figure 4.2A). Hence, the *avpr1a* genotype differences in RSC-V1aR of adult voles (Okhovat et al., 2015) emerge sometime during the first postnatal week.

Drastic changes in gene expression are common during the first few postnatal weeks in the rodent brain, and are often accompanied by dynamic changes in DNA methylation (Simmons et al., 2013). Here, we asked whether developmental changes in RSC-V1aR are associated with methylation in a putative enhancer in the intron of *avpr1a*. Consistent with our findings among adult voles (Okhovat et al., 2015), we found a significant negative relationship between total %DNA methylation and RSC-V1aR among all pups. Although total %DNA methylation may be a critical functional contributor to the epigenetic state of a regulatory element, it is highly influenced by sequence at polymorphic CpG sites (polyCpG). We calculated %methylation at fixed CpG sites and %methylation at methylable CpGs to examine non-genetic developmental changes in methylation. We again found a negative correlation between these methylation measures and RSC-V1aR. We compared methylation between genotypes across age groups and found that LO/LO pups, which have more CpG sites in the targeted intron enhancer, also had significantly higher total %DNA methylation compared to HI/HI throughout the first two postnatal weeks (Figure 4.2C). Interestingly however, %methylation at fixed CpG sites and %methylation at methylable CpGs was only higher among LO/LO animals compared to HI/HI pups at P7 and P14, and not at birth (Figure 4.2D,E). Thus, HI/HI and LO/LO genotype differences in overall enhancer methylation are present since birth, but additional non-genetic methylation differences at fixed and methylable CpGs appear sometime during the first postnatal week, which coincides with the emergence of genotype differences in RSC-V1aR abundance. These results indicate that the first postnatal week represents a critical neurodevelopmental stage for RSC-

V1aR and emergence of *avpr1a* genotype differences both in expression and enhancer methylation. While the exact timing may differ, such "critical" developmental stages are common in rodent neurodevelopment, and represent a period when the brain is highly sensitive to environmental or epigenetic perturbations (Roth and Sweatt, 2011). In our study, the synchronized emergence of genotype differences in RSC-V1aR and enhancer methylation suggests a role for *avpr1a* enhancer methylation in the development of RSC-V1aR expression.

To investigate whether allelic differences in CpG abundance resulted in differences in sensitivity to developmental silencing, we manipulated development with an oxytocin receptor antagonist (OTA), a treatment known to reduce RSC-V1aR (Bales et al., 2007). We predicted that LO/LO pups, which have more enhancer CpGs, would be more sensitive to OTA-induced silencing. We injected pups at P1, before genotype differences in RSC-V1aR are established, and measured RSC-V1aR at weaning (P21). Among our subjects, we found a significant main effect of genotype, treatment (OTA vs. control) and sex on RSC-V1aR (Figure 4.3A). Interestingly, *avpr1a* genotypes differed in their sensitivity to OTA; LO/LO animals showed significant decrease in RSC-V1aR while HI/HI animals were unaffected (Figure 4.3A). While systemic OTA exposure does not model any particular natural experience *per se*, it may mimic variation in parental care (Feldman et al., 2010). Interestingly, the presence of paternal care does seem to change RSC-V1aR among prairie voles in adulthood (Prounis et al., 2015, but see [Ahern et al., 2009]). Together these results suggest that the developing prairie vole brain –much like other rodents' (Weaver et al., 2004), is influenced by the quality and quantity of early parental care, and that this impact may vary among individuals.

If CpG abundance alters sensitivity to DNA methylation, we hypothesized that drugs that interfere with methylation should promote RSC-V1aR expression in LO/LO

animals, but not in HI/HI animals. Zebularine is a low toxicity DNA methyltransferase (DNMT) inhibitor commonly used to disrupt DNA methylation both *in-vitro* and *in-vivo* (Cheng et al., 2003). As predicted, neonatal zebularine injection increased RSC-V1aR, but only among LO/LO animals (Figure 4.3B). Thus for two very different developmental manipulations, LO/LO voles show increased sensitivity to environmental manipulations. Although the results were promising, pharmacological manipulations are inherently difficult to interpret because observed outcomes may result from effects elsewhere in the genome, the brain, or both. Thus, to elucidate whether changes in RSC-V1aR were directly mediated by methylation changes at the *avpr1a* intron enhancer, we measured enhancer methylation in subjects from the OTA and zebularine studies.

In both OTA and zebularine datasets, we found a negative correlation between total %DNA methylation and RSC-V1aR abundance (Figure 4.4A,C). To our surprise however, we found that for both datasets the majority of variation in enhancer methylation was explained solely by genotype. In the zebularine study, we also found a weak overall effect of treatment, with total enhancer methylation slightly increasing after zebularine exposure (Figure 4.4D). This outcome is in contrast to the expected de-methylating role of zebularine. Although developmental manipulations of methylation are common in behavioral epigenetics (REF), our results highlight some shortcomings of this approach. First, the long delay between treatment and brain collection suggest that effects of developmental manipulations are not likely to be due to the direct effects of the drug on the focal enhancer. Similarly, the global alteration of methylation could change function at other enhancers, or at other loci that contribute to regulation of our focal enhancer. While our data do not explain the exact mechanisms underlying OTA- and zebularine-induced changes in RSC-V1aR, it is clear that they are not solely due to enduring changes in methylation at the intron enhancer (Figure 4.4B,D), as we originally

hypothesized. This led us to explore additional regions with differential methylation between genotypes, which may provide targets for future study.

We found that a methylated DNA immunoprecipitation-sequencing (MeDIP-seq) assay was able to selectively enrich methylated regions of the vole genome (Figure 4.5A,B). We next compared methylation levels around the *avpr1a* locus in HI/HI and LO/LO voles (Figure 4.5D). The overall patterns of methylation were similar between the two genotypes, but the MeDIP assay suggests at least three differentially methylated regions (DMRs). First, it replicates our findings of differential methylation at the intron enhancer (Figure 4.5C,D,E). Second, it shows differential methylation ~500bp upstream of the transcription start site (Figure 4.5D,E) – a somewhat surprising finding, since promoters are often stably unmethylated, even when gene expression is suppressed (Weber et al., 2007). While this DMR is likely important to cortical *avpr1a* expression, there are no genetic differences between HI and LO alleles within this region (Okhovat et al., 2015), suggesting it is a downstream consequence of regulation at another site. However, the MeDIP revealed one additional DMR ~8.5kb upstream of the *avpr1a* locus which also exhibits higher levels of methylation among LO/LO animals (Figure 4.5D,E). To determine if this DMR contributes to *avpr1a* regulation and developmental sensitivity, future research should further characterize its genetic and epigenetic variation between HI/HI and LO/LO voles.

Involvement of additional genetic elements in regulation of *avpr1a* is complex and exciting, but perhaps not surprising. Regulation of eukaryotic gene expression is complicated and involves many regulatory regions and transcription factors. Often multiple regulatory elements regulate tissue-specific gene expression; these elements can be quite distal from their target gene (Bulger and Groudine, 2010). In fact, genes with G×E interactions are found to have disproportionally high association with

distal regulatory loci compared to other gene groups (Grishkevich and Yanai, 2013). Similarly, genetic contributions to variation in gene expression are often due to multiple regulatory variations. Each variant may have a modest effect on expression, but effect sizes may change in response to external stimuli, or as variants are inherited together (Corradin et al., 2014). Although we focus on a single SNP and linked CpG polymorphisms within an intronic enhancer, additional genetic differences occur within other regulatory regions (Okhovat et al., 2015), and genetic variation at more distal regulatory sequences has not been characterized. Hence, it is possible that additional linked variants at distal regulatory elements contribute to HI and LO differences in *avpr1a* regulation. If so, one or more of those additional sites may contribute to genotype differences in environmental sensitivity. Such effects may interact with CpG polymorphisms in the intron enhancer, or may operate through alternative mechanisms.

In addition to the potential effects of distal enhancers, it is plausible that methylation at the *avpr1a* intron enhancer and our pharmacological manipulations interact in other ways. For example, the methyl-binding protein MeCP2 regulates expression of a variety of genes, including neuropeptides such as vasopressin (*avp*; [Murgatroyd et al., 2009] and corticotrophin-releasing hormone (*crh*; [McGill et al., 2006]). Phosphorylation of MeCP2 can influence its affinity for methylated DNA, or even convert it into an activator of expression (Zimmermann et al., 2015). If our treatments alter MeCP2 phosphorylation -- as happens in mice exposed to early life stress, drug treatments and learning paradigms (Zimmermann et al., 2015)-- the relationship between methylation status and *avpr1a* expression might change in unexpected ways. Since the LO allele has higher methylation density, this allele is more likely to interact with methyl-binding effector proteins, such as MeCP2. As a result, we would expect LO/LO voles to be more susceptible to environmentally-induced

modifications to methyl-binding effector proteins. While our observations are consistent with this hypothesis, the exact mechanisms by which OTA and zebularine interact with *avpr1a* sequence variation remain to be determined. These results highlight the challenging complexity posed by GxE interactions, but also suggest that tools like MeDIP, ChIP-seq and sequence-targeted effector proteins (e.g. dCas9-MecP2 fusion proteins) provide exciting new means to meet this challenge.

The complex interplay between developmental processes and genetic variation shapes phenotypic diversity, and is increasingly important for our understanding of animal behavior, evolutionary biology and mental health (Caspi and Moffitt, 2006; Grishkevich and Yanai, 2013; Manuck and McCaffery, 2014). We asked whether an increased frequency of CpG sites would make an allele more sensitive to developmental perturbation. We tested this hypothesis by focusing on allelic differences in enhancer methylation and V1aR abundance within the retrosplenial cortex, an expression pattern associated with complex socio-spatial behaviors in the field. We found that genotype differences in RSC-V1aR abundance are absent at birth but emerge within the first postnatal week, and that these changes are accompanied by differences in methylation of a polymorphic *avpr1a* enhancer. We found that LO/LO animals were indeed more sensitive to developmental manipulations, but that this sensitivity is not simply due to differences in the methylation of our focal intron enhancer. This work highlights the complexity of interpreting pharmacological manipulations on gene expression, and the many genetic and epigenetic factors that come into play for even a single candidate gene. Lastly, despite the complexity of our results, this work highlights the utility of non-model organisms in better understanding genetic diversity. Such diversity is an essential component of individual and species differences in brain, behavior and evolution.
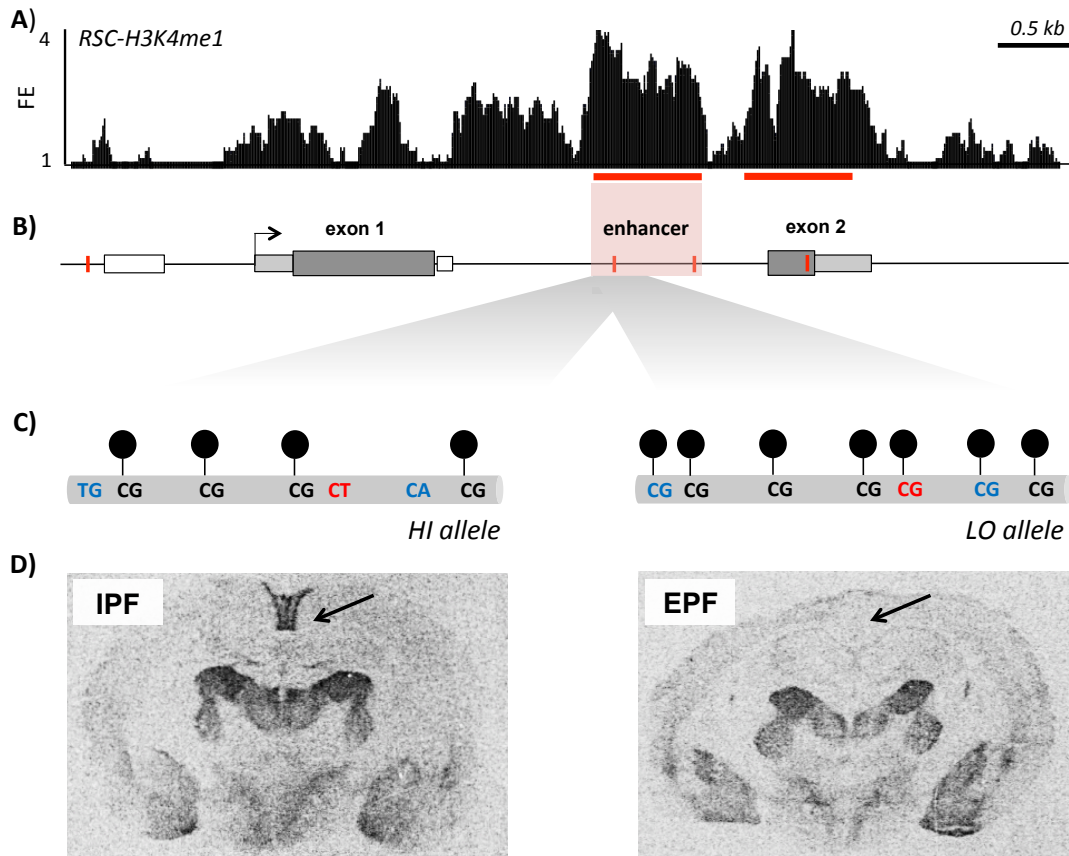
Figure 4.1. *Avpr1a* **genotype differences in enhancer CpG density and susceptibility to DNA methylation.** A) Fold enrichment (FE) values for H3K4me1 ChIP-seq on RSC of 8 male prairie voles is shown at the *avpr1a* locus. Putative enhancers are marked in red horizontal lines. B) A schematic view of the *avpr1a* locus. Allele-defining SNPs are marked with red bars. C) A schema of HI (*left*) and LO (*right*) allele differences in CpG and methylation density within the *avpr1a* putative intron enhancer. Sequence at SNP 2170 is shown in red, other polyCpGs in blue and fixed CpGs are black. Black circles depict 5-methyl at cytosines. D) V1aR autoradiograms of intra-pair fertilizing (IPF, *left*) and extra-pair fertilizing (EPF, *right*) males. Arrow shows the RSC. A, B and D modified from Okhovat et al. 2015.
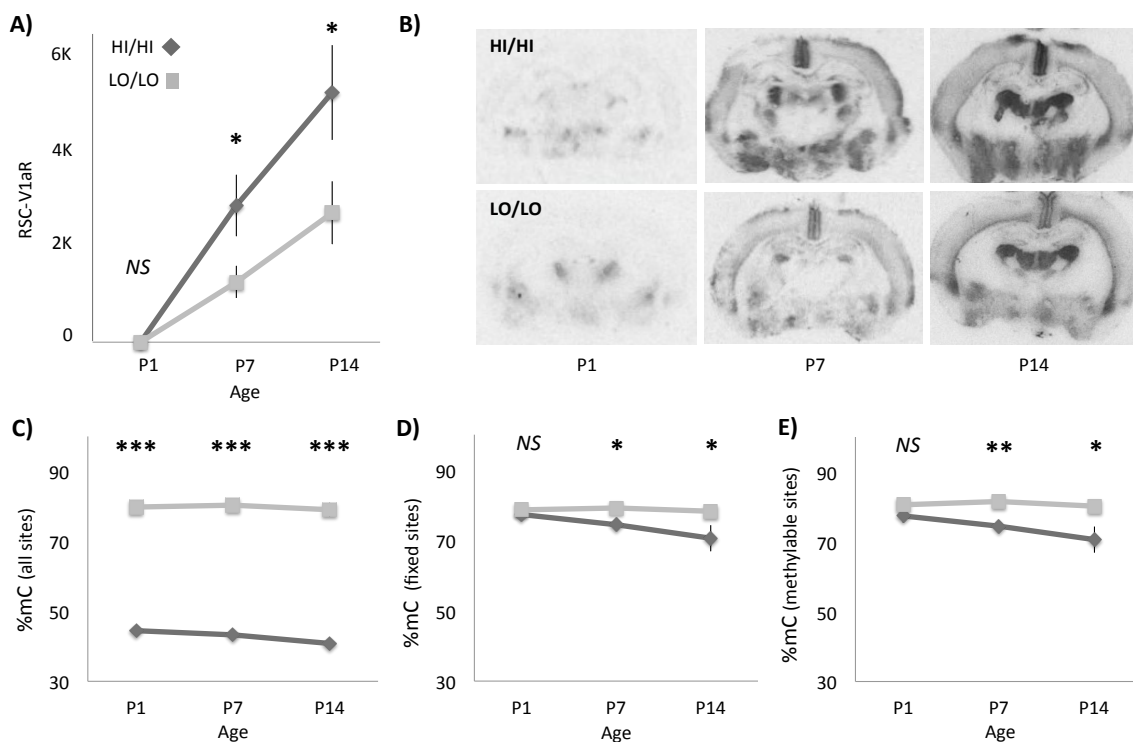
**Figure 4.2. Postnatal genotype differences in RSC-V1aR and intron enhancer
methylation**. A) Changes in RSC-V1aR abundance in first two postnatal weeks among
HI/HI and LO/LO pups. B) V1aR autoradiograms in HI/HI and LO/LO pups at P1, P7
and P14. C-E) Ontogeny differences between HI/HI and LO/LO animals in methylation
of the putative intron enhancer in developing RSC. Data presented as C) total %DNA
methylation D) %methylation at fixed CpGs, and E) %methylation at methylable CpGs.
Data points are mean ± SE. *p≤0.05, **p≤0.01, ***p≤0.001.
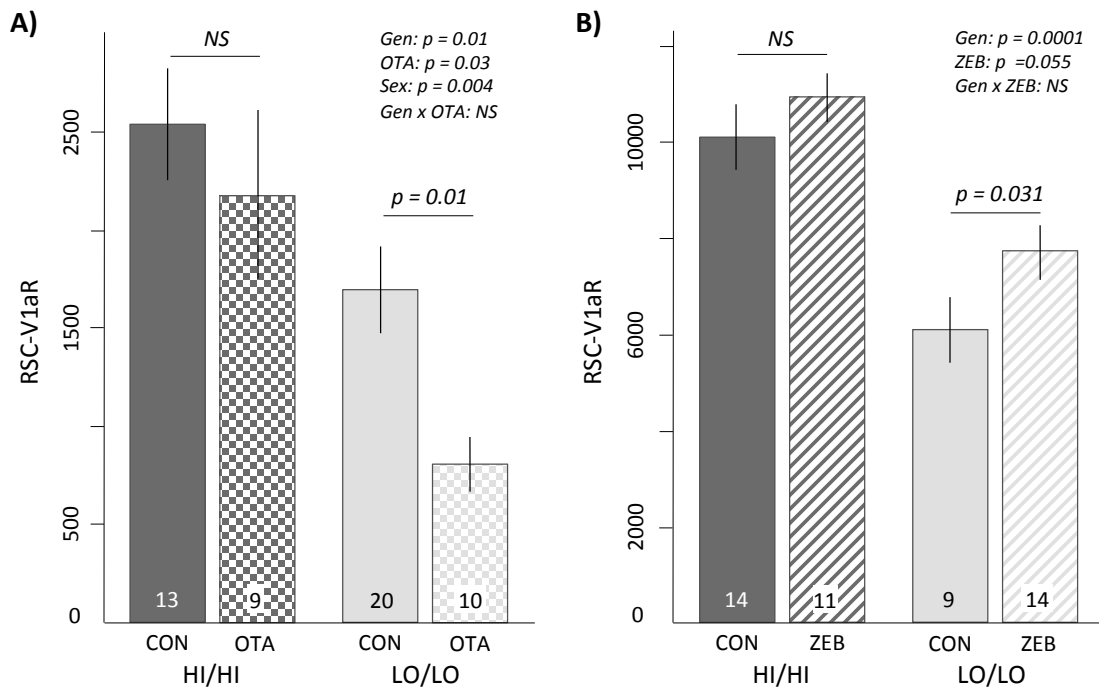
Figure 4.3. Genotype differences in sensitivity to neonatal manipulation. A) Abundance of RSC-V1aR at P21 is shown among subjects in OTA study. B) Autoradiogram RSC-V1aR measures are shown for subjects in zebularine study. Sample sizes are provided on each bar. Abbreviations are as follows: CON= control, OTA= oxytocin receptor antagonist, ZEB= zebularine. Bars are mean ± SE.
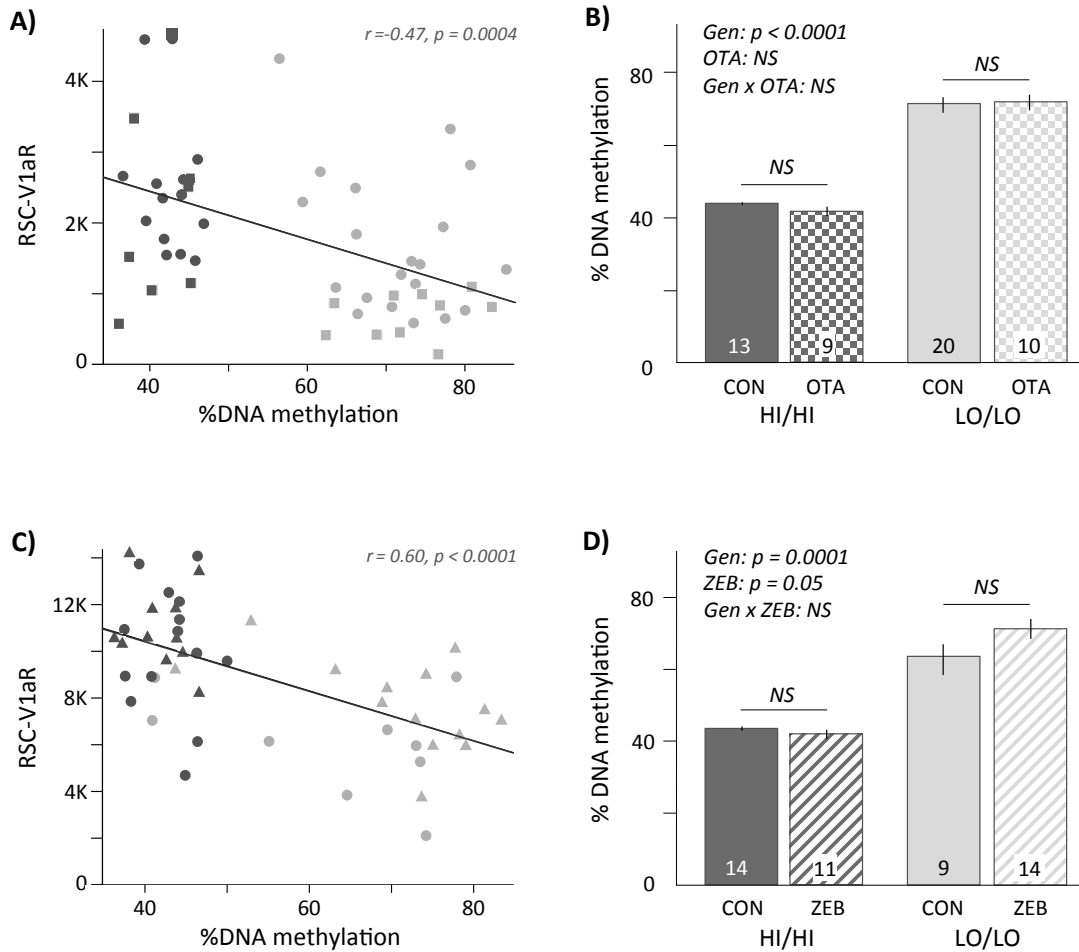
Figure 4.4. *Avpr1a* **enhancer methylation in the RSC of HI/HI and LO/LO voles.** A) Total %DNA methylation in the putative intron enhancer plotted against abundance of RSC-V1aR in HI/HI (dark gray) and LO/LO (light gray) voles receiving oxytocin receptor antagonist (squares) or control treatments (circles). B) Total %DNA methylation for control and oxytocin receptor antagonist treated HI/HI and LO/LO subjects. C) Total %DNA methylation plotted against RSC-V1aR of HI/HI (dark gray) and LO/LO (light gray) voles receiving control treatments (circles) or zebularine injections (triangles). D) Total %DNA methylation of control and zebularine-treated HI/HI and LO/LO voles. Sample sizes are provided on each bar and abbreviations are as follows: CON=control, OTA=oxytocin receptor antagonist, ZEB=zebularine. Bars are mean ± SE.

107
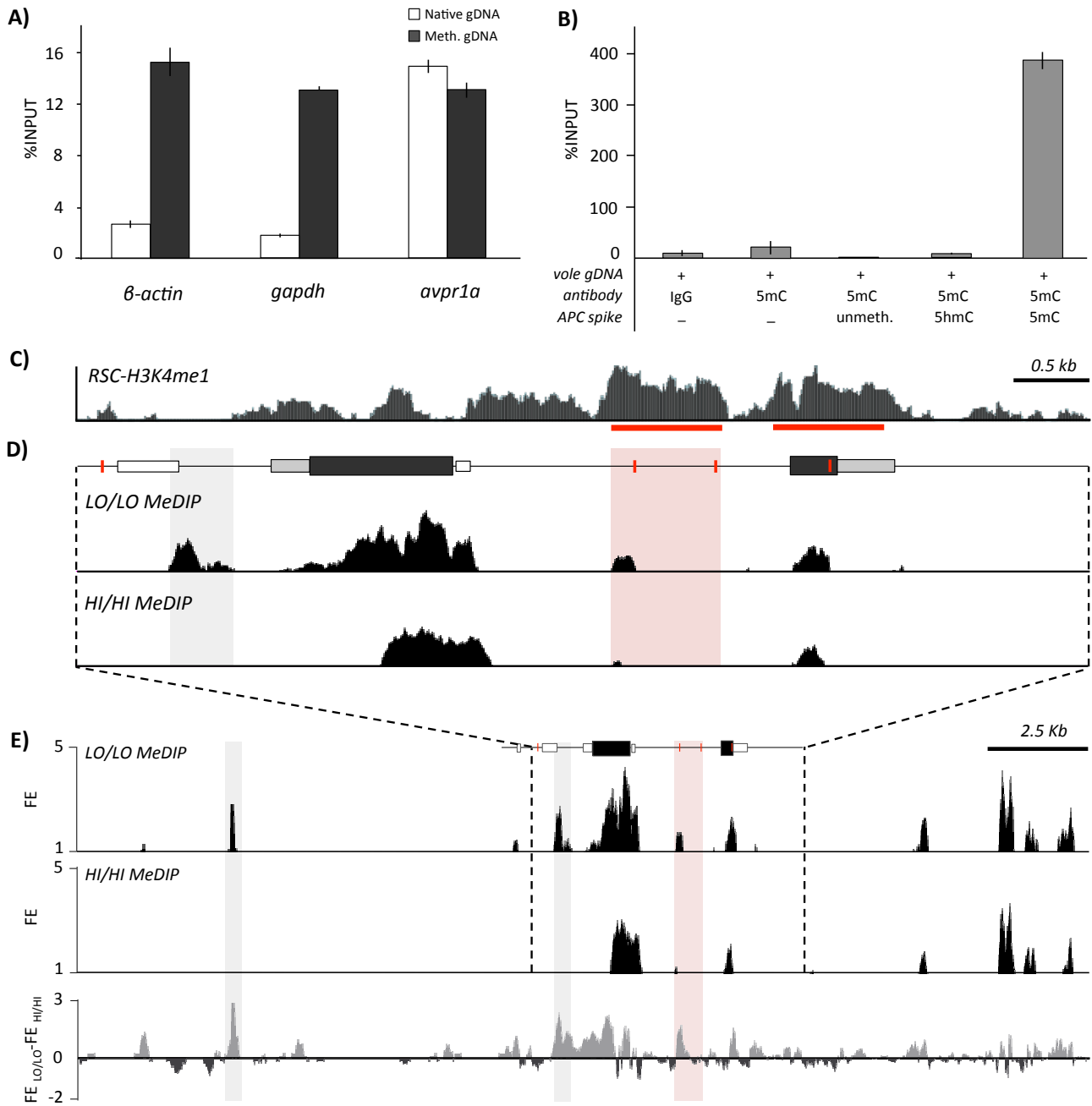
Figure 4.5. Methylated DNA immunoprecipitation of RSC from HI/HI and LO/LO animals.

Figure 4.5. cont. MeDIP enrichment (%INPUT) of native (white) and *in vitro* methylated prairie vole gDNA (black) at the *β-actin*, *gapdh* and the *avpr1a* intron enhancer. B) Ability of MeDIP to detect methylated control spike-in DNA (human *APC* locus, *APC*) in the presence of vole genomic DNA. Treatments from left to right include non-specific IgG antibody (IgG) but no *APC*; 5mC antibody but no *APC*; 5mC antibody and unmethylated *APC*; 5mC antibody + hydroxymethylated *APC* (5hmC-APC); and 5mC antibody + methylated *APC* (5mC-APC). Bars are mean ± SD. C) Fold enrichment track for H3K4me1 shown at the *avpr1a* locus. Significant peaks marked with red bars. Data from Okhovat et al. 2015. D) MeDIP fold enrichment tracks for RSC of LO/LO (*top*) and HI/HI animals (*bottom*) along the *avpr1a* locus. The putative intron enhancer is shaded pink. E) MeDIP fold enrichment (FE) track over a 25Kb window centered on the *avpr1a* transcription start site for LO/LO (*top*) and HI/HI animals (*middle)* and their difference (*bottom*). Differentially methylation regions (DMRs) are shaded.

DISSERTATOR'S NOTE

Chapter 4 is based on a multi-authored manuscript currently (December 2016) under review at Genes, Brain and Behavior. I am the lead author on this manuscript and co-authors are in order: I.C. Chen, Z. Dehghani, D.J. Zheng, J.E. Ikpatt, H. Momoh and S.M. Phelps. I have participated in design and execution of all wet lab experiments, performed the data analysis and participated in preparation of the manuscript.

ACKNOWLEDGEMENTS

Chapter 5: Summary, Future Directions, Conclusion and Significance

SUMMARY AND FUTURE DIRECTIONS

Individuals within a species often differ dramatically in aspects of their phenotype. These phenotypic differences can include morphological variation in form and function, such as the extraordinary feather diversity of rock pigeons (Darwin, 1868), or the alternative left- and right-handed feeding morphologies of scale-eating cichlid fish (Lee et al., 2012). Interestingly, intraspecific variation can also be found in behavioral phenotypes relevant to social interaction, as is the case in alternative male mating strategies of side-blotched lizards (Sinervo and Lively, 1996), in the density-dependent cannibalism of spadefoot toads (Pfennig, 1992), or in the more subtle personality differences evident among humans (McCrea and Costa, 1999). Understanding the mechanisms that underlie these intraspecific phenotypic differences, especially in the context of social behaviors, is one of the most challenging and interesting problems in biology.

Variation in social behavior emerges from individual differences in the brain. Such neuronal differences are either driven by variation in DNA sequence, by environmental and epigenetic influences, or by some combination of the two. The alternatives male mating strategies found in side-blotched lizards, for example, are rooted in highly heritable genetic differences (Sinervo and Zamudio, 2001). Such differences arise from DNA sequence variation that influences structure, function or regulation of genes in the brain. Behavioral variation may also arise due to environmentally induced phenotypic plasticity. For example, differences in feeding behavior and cannibalism among spadefoot-toad tadpoles emerge largely due to variation in pond longevity and food availability (Pfennig, 1992). At a molecular level, environmental variables, especially those experienced early in life, often modulate neuronal gene activity via

110

epigenetic modifications of chromatin state; the most commonly studied of which is DNA methylation at CpG dinucleotides (Feil and Fraga, 2012). Interestingly, genetic variation can alter individuals' susceptibility to such environmentally-induced changes, a phenomenon known as gene-by-environment interaction (GxE). For instance, SNPs located within CpG sites (polymorphic CpGs or polyCpGs) can alter local CpG availability and susceptibility to DNA methylation. GxE effects are prevalent (Pigliucci, 2001) and may have important consequences for intra-specific variation in brain and behavior. Therefore, to fully understand the molecular mechanisms of variation in social behavior, we need to ask how genetic and epigenetic variation interact to shape neuronal gene regulation and brain function. In this dissertation, I used the brain and behavior of prairie voles (*Microtus ochrogaster*) as a model to explore the molecular genetic and epigenetic basis of diversity in a socially relevant neuronal phenotype. Here I review our principal findings and outline their strengths and limitations. Finally, I discuss how our findings with the prairie vole *avpr1a* gene might serve as a model for larger questions about social cognition, evolution, and gene-by-environment interactions.

Perhaps one of the most dramatic and best-documented variation in social behavior comes in the form of alternative male reproductive tactics, a phenomenon documented in taxa including crustaceans, insects, fishes, birds, reptiles, amphibians and mammals (see Taborsky et al., 2008 for a review). Many of these alternative male reproductive behaviors involve males that specialize in monopolizing a female through mate-guarding, while others try to mate more opportunistically in what is known as scramble competition. For example, in giant freshwater prawns, some males opt a precopulatory mate-guarding tactic, while others practice sneak mating (Ra'Anan and Sagi, 1985). Also, in many monogamous bird species, such as the Mediterranean blue tit, a subset of paired males mate exclusively with their mate, but others gain extra pair

copulation from neighboring females (García-Navas et al., 2014). Our study species, the socially monogamous prairie vole, exhibits significant variation in space-use associated with sexual fidelity. Sexually unfaithful males gain extra-pair fertilizations (EPFs) by establishing large overlapping home-ranges and intruding into other territories, but do so at the cost of being cuckolded (Phelps and Ophir, 2009). In contrast, the sexually faithful intra-pair fertilizing (IPF) males, maintain small exclusive home-ranges, intrude rarely into other territories, and are more successful at securing paternity by guarding their mate (Okhovat et al., 2015; Ophir et al., 2008a; Ophir et al., 2008b; Figure 5.1A).

Prairie voles are widely used to study the behavioral, neurological and molecular basis of social behavior, making them an excellent species for investigating the mechanisms underlying variation in social behavior (Young and Wang, 2004). Among the mechanisms implicated in the study of alternative reproductive tactics in prairie voles and other species, the vasopressin system (or the non-mammalian homolog, vasotocin) is consistently involved in male social and reproductive behaviors (Donaldson and Young, 2008; Oldfield et al., 2015); Young and Wang, 2004).

Vasopressin mediates various social behaviors such as social learning, aggression, territoriality and stress, via its main neuronal receptor, V1aR (Lim and Young, 2006). Intraspecific variation in neuronal abundance, distribution and function of V1aR has been linked to variation in social behaviors, especially in males (Donaldson and Young, 2008). For instance, blockade of neuronal V1aR impairs social recognition in adult mice and rats (Bielsky et al., 2004; Veenema et al., 2012) and alters aggressive behavior among Syrian hamsters (Gutzler et al., 2010) and zebra finches (Goodson et al., 2004). Among vole species, variation in distribution of neuronal V1aR in regions implicated in pair-bonding, such as the ventral-pallidum (VPall), directly contributes to species differences in the ability to form pair-bonds (Lim et al., 2004; Lim and Young, 2004). Interestingly,

neuronal abundance of V1aR exhibits drastic variation among individuals within the prairie vole species, as well. This diversity is however, generally absent from brain regions implicated in pair-bonding—such as the VPall or lateral septum (LS). In contrast, the retrosplenial cortex (RSC), which contributes to spatial and contextual memory (Vann et al., 2009), varies tremendously in V1aR abundance across individual prairie voles (Ophir et al., 2008b; Phelps and Young, 2003). Interestingly, variation in V1aR abundance in the RSC (RSC-V1aR) is highly predictive of prairie vole male differences in sexual and spatial fidelity; EPF males have low RSC-V1aR levels, while IPF males have high V1aR abundance in their RSC (Okhovat et al., 2015; Ophir et al., 2008b; Figure 5.1B). Based on the role of the RSC in memory, and V1aR in male social behavior, we speculate that intraspecific RSC-V1aR differences drive male variation in memory for social encounters across space. This memory variation may be responsible for shaping IPF vs. EPF tactics, since intrusions are – in principle – more likely if a male forgets home-range borders, their overlaps, and where he previously encountered an aggressive resident. Future studies that directly examine the memory of males with different RSC-V1aR levels will help evaluate this hypothesis. Furthermore, RSC lesion-studies and targeted manipulation of V1aR abundance using viral vectors can help elucidate other important behavioral and cognitive consequences of RSC-V1aR in male prairie voles. Critically, these behavioral findings suggest that by understanding how genetic and developmental forces interact to shape cortical *avpr1a* expression, we will further our understanding of how changes in genome function can influence a complex behavior.

The dramatic intraspecific variation in neuronal V1aR has motivated many researchers to study its molecular mechanisms. A prominent study in prairie voles, for example, suggested that length of a microsatellite upstream of *avpr1a*, the gene that

encodes V1aR, is causally linked to neuronal V1aR (Young et al. 1999, Hammock et al., 2005). Since then, other studies in chimpanzee (Hopkins et al., 2012) and humans (Walum et al., 2008) have also found associations between *avpr1a* microsatellite length polymorphism and variation in social behavior and V1aR abundance. However, the inconsistencies found in these and other studies (Hammock et al. 2005; Mabry et al., 2011; Ophir et al., 2008; Solomon et al., 2009) suggested that DNA microsatellites may not cause variation in expression, but rather may be linked to other variants that govern *avpr1a* regulation. Therefore, we examined the association between RSC-V1aR and intraspecific genetic variation along the whole *avpr1a* locus. We found four strongly linked single nucleotide polymorphisms (SNPs) that formed two *avpr1a* alleles, HI and LO (Figure 5.1B). These alleles predicted *avpr1a* transcription and V1aR abundance in the RSC, but not other socially relevant brain regions, such as the laterodorsal thalamus (LDThal), VPall or LS. Three of the predictive SNPs co-localized with two putative RSC enhancer regions, which we identified by chromatin immunoprecipitation sequencing (ChIP-seq) for a general enhancer mark, histone 3 lysine 4 monomethylation (H3K4me1; Okhovat et al., 2015; Figure 5.1B). Future studies could use a combination of reporter gene-assays, an assay for transposase-accessible chromatin sequencing (ATAC-seq; Buenrostro et al., 2015), or methods for examining the causal roles of putative regulatory regions (e.g. STARR-seq; Arnold et al., 2013), to assess whether our identified sequences are driving variation in RSC-V1aR. Nevertheless, our ChIP-seq and association data suggest that at least some subset of SNPs within these putative enhancer sequences may directly shape *avpr1a* expression. At other loci, enhancer SNPs have been shown to have drastic regulatory consequences – for example, they can cause changes in DNA methylation (Izzi et al., 2016), transcription factor binding or chromatin conformation (Visser et al., 2012).

114

Interestingly, the DNA sequence variation we found at the *avpr1a* locus and within its putative intron enhancer did not account for all the variation we observed in RSC-V1aR. This was especially evident among wild-caught prairie voles, which compared to lab-reared voles, exhibited higher RSC-V1aR diversity but weaker association between RSC-V1aR and *avpr1a* genotypes. This observation indicated that RSC-V1aR variation might be partially due to phenotypic plasticity in response to the environmental and developmental diversity that voles are naturally exposed to in the wild (e.g. population and resource fluctuations; Getz et al., 2001). In fact, previous work on prairie voles (Bales et al., 2007; Prounis et al., 2015), rats (Francis et al., 2002) and mice (Lukas et al., 2010) have demonstrated that developmental manipulations and variation in early rearing environment can alter V1aR regulation in the RSC or other brain regions. While the exact molecular mechanisms for these neuronal changes are not known, environmentally induced changes in neuronal gene expression are often mediated by molecular epigenetic modifications, such as DNA methylation (Szyf and Bick, 2013). Thus, environmental and developmental alterations in RSC-V1aR may be associated with changes in *avpr1a* DNA methylation, especially at the putative intron enhancer. To examine the interaction of genetic and epigenetic variation more closely, we examined allelic differences DNA methylation within a putative enhancer we found in the intron of *avpr1a*.

One of the *avpr1a* allele-defining SNPs (SNP 2170) was a polymorphism that altered the presence/absence of a CpG site located within a putative intron enhancer. This site was weakly linked to additional polyCpGs within the same enhancer, leading to significant genotype difference in CpG density and opportunity for methylation (Okhovat et al., 2015; Figure 5.1C). We also found a negative association between enhancer methylation and *avpr1a* transcription, suggesting that such methylation lowers RSC-

V1aR by reducing *avpr1a* transcription, consistent with widely reported silencing effects of DNA methylation (Nan et al., 1998). We found a similar association between enhancer methylation and RSC-V1aR among wild-caught prairie voles. Surprisingly however, a broad examination of the locus revealed that promoter methylation, which is often closely associated with gene activity (Tate and Bird, 1993), was dissociated from variation in RSC-V1aR among wild-caught voles (Okhovat et al., unpublished; Figure 5.1C). Although unexpected, this finding is in line with a growing number of cell-line (Rollins et al., 2006) and tissue specific (Lister et al., 2013) studies that suggest methylation and sequence variation in regulatory elements outside of the promoter area, especially within enhancer sequences, may be a better predictor of expression. Our findings in lab-reared and wild-caught prairie voles suggest that an intron enhancer regulates RSC-V1aR, and is likely to be affected by changes in both genetic sequence and epigenetic status.

A detailed analysis suggested at least two mechanisms by which sequence variation and epigenetic mechanisms might interact at the *avpr1a* enhancer. First, allelic differences in recruitment of repressive methyl-binding proteins – such as MeCP2 (Bird, 2002) – may account for differences in expression. In this scenario, repressive proteins are expected to preferably bind and silence the LO allele enhancer, which has higher methylation levels. Alternatively, sequence specific binding of transcription factors may be influenced by SNPs and generate genotype differences in RSC-V1aR. Targeted in-vivo manipulation of sequence and methylation state of the putative intron enhancer (e.g. using CRISPR/Cas9-fusion protein techniques; Cong et al., 2013) can help determine whether the regulatory function of the putative intron enhancer is shaped by sequence differences, methylation variation or both. These manipulations may also be followed by chromatin immunoprecipitation studies, such as ChIP-seq or reverse-ChIP (Rusk, 2009) which can help determine if transcription factors or methyl-binding proteins bind to the

putative enhancer sequence, and how their binding may be influenced by variation in sequence and methylation. While more detailed molecular mechanisms remain to be elucidated, we next sought to characterize the emergence of genotype differences in RSC-V1aR abundance, and to test whether the enhancer CpG polymorphisms that characterize HI and LO alleles could generate differences in sensitivity to developmental manipulations.

In prairie voles, neuronal V1aR abundance undergoes drastic changes postnatally (Wang et al., 1997). Our pups were born with no RSC-V1aR, but V1aR abundance increased rapidly during the next two weeks (Figure 5.1D). These changes were broadly similar to previous documentation of V1aR development in voles (Wang et al., 1997), rats (Tribollet et al., 1991) and the Brazilian opossum (Kuehl-Kovarik et al., 1997). We found that the HI and LO genotype differences in RSC-V1aR were absent at birth, but emerged during the first postnatal week. Interestingly, genotype differences in *avpr1a* enhancer methylation also emerged around the same time, suggesting that enhancer methylation is involved in early-life regulation of RSC-V1aR. Rodent brain undergoes periods of dramatic developmental change in gene expression and methylation, which may represent developmental "critical periods" when neuronal gene expression is highly responsive to environmental variation in parental care, diet or stress (Roth and Sweatt, 2011). In the case of RSC-V1aR, however, due to GxE effects, individuals may vary in their sensitivity to these early developmental perturbations; some individuals may exhibit long-term effects on brain and behavior, while others are completely unaffected. Surprisingly few studies have documented such variation at the scale of the current work. Thus, we tested for GxE interactions (or their pharmacological equivalents) by manipulating developmental conditions in specific ways.

117

We first manipulated development using an oxytocin receptor antagonist, a manipulation that is sometimes considered analogous to poor parenting, and that has been shown to alter adult RSC-V1aR of voles (Bales et al., 2007). Oxytocin receptor antagonist injections on the first postnatal day reduced RSC-V1aR later at weaning age, demonstrating that *avpr1a* regulation is sensitive to early developmental and environmental perturbations. This sensitivity was however, only detected in LO/LO pups, and not their HI/HI siblings. We found similar results when manipulating development using the global inhibitor of methylation, zebularine (Cheng et al., 2003), which only increased RSC-V1aR in LO/LO pups (Figure 5.1D). Overall, these data presented a remarkably coherent picture in which the high CpG density of LO alleles made them both more sensitive to the silencing effects of the oxytocin receptor antagonist, and the demethylating effects of zebularine. Examination of the methylation state of the putative intron enhancer, however, suggests a more complex story. As in our prior studies, individual differences in enhancer methylation were associated with RSC-V1aR expression, but enhancer methylation was not influenced by our developmental manipulations.

Considering the global nature of our pharmacological treatments, as well as the delay between treatment and measurements, it was hard to interpret our negative results. However, our findings strongly suggested that genotype differences in developmental sensitivity were not due to CpG density differences in the putative intron enhancer alone. One possibility is that genetic differences in the enhancer are inherited along with genetic variation at additional enhancers we have not examined. Involvement of additional – often-distal– regulatory elements is common in eukaryotic tissue-specific gene regulation (Bulger and Groudine, 2010), where variation in expression is often due to multiple regulatory variants (Corradin et al., 2014). To address this, we performed a genome-wide

methylation assay, which revealed a promising site located ~8kb upstream of the *avpr1a* gene. This region was highly methylated in the RSC of LO/LO subjects, but not HI/HI animals. Future research that characterizes genetic and epigenetic variation at this site can determine if it is involved in *avpr1a* regulation and genotype differences in developmental sensitivity. Additional distal enhancers that interact with the *avpr1a* locus may be identified by chromatin conformation techniques, such as circularized conformation capture sequencing (4C-seq; Zhao et al., 2006) or Hi-C (Belton et al., 2012). These studies may also shed light on distal genetic and epigenetic variants that contribute to individual differences in *avpr1a* regulation and developmental sensitivity.

Although there are many follow up studies to be done on the interaction between genetic and epigenetic variation at the *avpr1a* locus, it is worth mentioning that this is neither the only neuromodulator, nor the only brain region, to shape monogamy and social cognition in prairie voles. For example the corticotrophin-releasing factor (Lim et al., 2007), corticosterone (Carter et al., 1995), estrogen (Cushing and Kramer, 2005), oxytocin (Williams et al., 1994), opioid (Resendez et al., 2012) and dopamine systems (Aragona et al., 2006) are all known to alter aspects of pair-bonding and its related behaviors. In addition, future studies should examine neuronal diversity in other brain regions which may be involved in male sociosexual fidelity, for example those tightly connected to the RSC (e.g. hippocampus or the anterior and laterodorsal thalamus (Vann et al., 2009), or those functionally related to the mating circuitry (e.g. bed nucleus of the stria terminalis [BNST]; Young and Wang, 2004). Genome-wide comparisons of EPF and IPF gene transcription in these brain regions using RNA-seq can help identify neuronal and genomic networks that are involved in mediation of male sexual and spatial fidelity.

Another exciting but under-explored topic is female variation in spatial and sexual fidelity. Similar to males, female prairie voles can be categorized into EPF and IPF based on their sexual fidelity. These reproductive differences however, are not predicted by RSC-V1aR, but rather appear to be associated with variation in oxytocin receptor (OXTR) abundance in the hippocampus (Hip-OXTR; Phelps et al., 2010), another brain region widely implicated in spatial memory (Bird and Burgess, 2008). The genetic and epigenetic basis of Hip-OXTR diversity has not been determined, but a previous study found that a polyCpG drives variation in neuronal OXTR abundance and behavior in prairie voles (King et al., 2016). Thus, similar mechanisms may be at play in regulation of Hip-OXTR and female sexual fidelity. Future studies could provide invaluable insight into the mechanistic differences between regulation of brain and behavior in females and males, or across alternative loci and brain regions.

Conclusions and Significance

This dissertation provides invaluable insight into the architecture of complex social traits and how variation can percolate across biological scales, from DNA sequence all the way to complex behaviors. Our findings show that small nucleotide polymorphisms, which are often used merely as proxy markers for causal genetic variants, may in fact act as drivers of intraspecific phenotypic variation in brain and behavior. As demonstrated in this dissertation, these regulatory consequences can emerge when genetic variation occurs within enhancer sequences that regulate neuronal gene expression. In higher eukaryotes, genes with complex tissue-specific expression patterns often have several enhancers, each of which drives expression in particular cell-types or developmental period (Rubinstein and de Souza, 2013). Therefore, individual differences

in function or activity of each enhancer, for example due to genetic variation, may drive phenotypic change only in one or a few tissues. The modular regulatory consequences of enhancer variation have major implications for intraspecific variation in brain and behavior, since they allow for neuronal changes in specific brain regions without major pleiotropic effects. Enhancers that help regulate behaviors critical for survival and reproduction may therefore be highly preserved while others, which regulate less critical behaviors can show higher intraspecific polymorphism. For example, in this dissertation, SNPs within a putative *avpr1a* enhancer drive V1aR differences in the RSC, without changing expression in other brain regions, such as the VPall. As a result, male prairie voles differ in spatial and sexual fidelity; but not their capability to form a pair-bond, which improves their reproductive success.

In addition to enhancer sequence polymorphisms found among conspecifics, a growing number of studies find that enhancer differences may also be responsible for some of the phenotypic differences found among species (Rubinstein and de Souza, 2013). While most of these studies examine non-behavioral phenotypes such as wing formation in bats (Booker et al., 2016) or lactase tolerance in humans (Swallow, 2003), we believe that similar mechanisms could drive species-specific patterns of brain and behavior. Intraspecific studies on the relationship between enhancer variation and neuronal phenotype can provide insight on the mechanisms of enhancer evolution and interspecific differences in behavior. Ultimately however, a fully comprehensive understanding will only be achieved by series of studies focused on particular loci and their genetic differences in several species or populations. This information will help clarify evolutionary processes and molecular processes that lead to behavioral variation among species.

Identifying enhancer regions is crucial for understanding the exact genetic variations that drive differences in brain and behavior within and between species. Finding regulatory sequences – especially those with tissue-specific activity – has been extremely challenging, because their function and activity depends highly on the genomic, cellular and tissue context (Inoue et al., 2016). Enhancers themselves are also relatively short, on the order of a few hundred bases, and must be found among the megabase or more of sequence flanking a gene. Similarly, showing that a putative regulatory region could causally influence gene expression required vector-based reporter gene assays in cell lines or transgenic organisms (Rubinstein and de Souza, 2013). While these studies have provided significant insight into mechanisms of gene regulation, they examine regulatory function outside of the native biological context and may therefore generate unreliable results (Inoue et al., 2016). For example, previous transcription reporter assays suggested that *avpr1a* microsatellite length differences could drive variation in neuronal expression in prairie voles (Hammock and Young, 2005). However, the effects of these microsatellites depended on cell types, and were not always in the direction predicted by individual or species differences (Hammock et al., 2005; Hammock and Young, 2005). Similarly, later examinations of *avpr1a* regulation within the brains of prairie voles yielded inconsistent findings (Hammock and Young, 2005; Hammock et al., 2005; Mabry et al., 2011; Phelps and Ophir, 2009; Solomon et al., 2009). As demonstrated in this dissertation, modern high-throughput techniques, such as ChIP-seq, can overcome some of these challenges by providing genome-wide maps of transcription-factor binding and epigenetic marks associated with enhancers within a brain region of interest. Other modern technologies such as CRISPR/Cas9 (Cong et al., 2013), which allow targeted and tissue specific in-vivo manipulations of DNA sequence, will allow direct examination of regulatory function. The growing appreciation for these

modern techniques among behavioral scientists promises new and exciting discoveries that may revolutionize our understanding of how brains vary, how their variation shapes differences in behavioral phenotypes, and how such differences are sculpted by evolutionary forces.

Another critical insight of this dissertation is that variation in CpG density may drive differences in brain and behavior by influencing neuronal DNA methylation and expression. Intraspecific variation in CpG abundance often occurs due to random single nucleotide polymorphisms (SNPs) at CpG dinucleotides. These polymorphisms – which we call polyCpGs – are the most common type of SNPs in the genome (Tomso and Bell, 2003). Considering their prevalence, polyCpGs may represent a major source of heritable behavioral diversity. However, their functional role in regulation of gene expression in the brain is poorly understood. Like any other SNP, polyCpGs can influence gene regulation by changing the sequence of regulatory elements. However, what sets polyCpGs apart from other SNPs is that by disrupting CG dinucleotides – which are the main targets for DNA methylation – they can change the local epigenetic environment and add a secondary layer of tissue-specific variation. Even one or few polyCpGs within an enhancer can influence transcription factors binding by altering the binding motif 's sequence and methylation (Schübeler, 2015). Presence of multiple linked regulatory polyCpGs may also influence recruitment of methyl-binding proteins, such as MeCP2, and change local chromatin structure (Bird, 2002). Since methylation is generally stable and tissue-specific, polyCpGs may lead to persistent and specific changes in neuronal phenotypes and the behaviors they modulate.

Considering the critical role of CpG methylation in environmentally induced neuronal change, polyCpGs may also represent a major mechanism for gene by environment (GxE) effects in brain and behavior. While metadata analysis suggests

polyCpGs are associated with some GxE effects (Parnell et al., 2014), their role in natural neuronal and behavioral diversity and plasticity are largely unexplored. To consider how our hypothesis of polyCpG-mediated GxE interactions could influence behavioral regulators, it is worth considering a well-characterized example, individual differences in stress reactivity (Cohen and Hamrick, 2003; Meaney, 2001). In rats, variation in maternal care drives changes in methylation and expression of glucocorticoid receptor in the hippocampus of pups, thereby changing their stress reactivity later in adulthood (Weaver et al., 2004). This effect has been attributed to specific CpG sites within the corresponding promoter (Weaver et al., 2004), thus a polyCpG at one of these sites would presumably cause individual differences in susceptibility to methylation, and in principle, variation in both neuronal and behavioral sensitivity to maternal care. Such genotype differences in developmental sensitivity can have important behavioral and evolutionary consequences for species such as the snowshoe hare, whose stress and vigilance is influence by environmental cycles of predator abundance (Sheriff et al., 2010). Indeed, maternal effects of hare stress reactivity are thought to explain complex patterns of population cycling, such as the trans-generational delay in reproduction after predator population crash (Sheriff et al., 2010). In the future, the mechanisms of such environmentally induced changes and their possible associated GxE effects, should be studied by combining modern genetic and epigenetic techniques such as RNA- and methylated DNA immunoprecipitation-sequencing (MeDIP-seq), which allows simultaneous examination of genome-wide sequence polymorphism, methylation and gene expression in different brain regions and individuals. We hope that our work will serve as a model that organismal biologists can use to guide work on GxE interactions, and the broader topic of heritable variation in phenotypic plasticity.

In conclusion, we explored some of the molecular mechanisms that underlie intraspecific variation in brain and social behavior of the prairie voles. Although we focused on a single gene and its expression in a single brain region, our findings provide valuable insights into the relationship between genetic and epigenetic variation, and how they can interact to shape individual differences in complex behavior. We explored how minute sequence differences within tissue-specific enhancers may promote dramatic but adaptive variation in brain and behavior. We examined how polymorphic CpG sites may contribute to both heritable and plastic aspects of brain and behavior. We encourage behavioral scientists to take advantage of the increasingly accessible molecular techniques to explore the detailed mechanisms of behavioral diversity. These studies might help us finally understand how alternative mating strategies emerge among side-blotched lizards, or how toad tadpoles translate pond longevity into the decision to cannibalize their conspecifics. More broadly however, these studies will help us connect the dots between genetic diversity, epigenetic variation, and individual differences in brain and behavior. Such work will enrich our understanding of the molecular mechanisms that underlie complex phenotypic diversity both within and between species.

Figure 5.1. Graphical summary.

Figure 5.1. cont. A) Male prairie voles are classified into IPF and EPF, based on patterns of sexual fidelity and space-use. B) IPF and EPF behaviors seem to be mediated by RSC-V1aR, which is predicted by HI and LO *avpr1a* alleles. A and B Adapted from the illustration by K. SUTLIFF/SCIENCE from Gene E. Robinson SCIENCE 350: 1310 (2015), with permission from AAAS. C) HI and LO alleles differ in CpG density and methylation in a putative intron enhancer, but not at the promoter. D) Genotype differences in RSC-V1aR differences emerge in the first postnatal week, in this period the LO allele was sensitive to environmental manipulations, but the HI allele was not.

References

Adolphs, R., 2003. Cognitive neuroscience of human social behaviour. Nat. Rev. Neurosci. 4, 165–178.

Aggleton, J.P., 2014. Looking beyond the hippocampus: old and new neurological targets for understanding memory disorders. Proc. R. Soc. Lond. B Biol. Sci. 281, 20140565.

Ahern, T.H., Young, L.J., Ahern, T.H., Young, L.J., 2009. The impact of early life family structure on adult social attachment, alloparental behavior, and the neuropeptide systems regulating affiliative behaviors in the monogamous prairie vole (*Microtus ochrogaster*). Front. Behav. Neurosci. 3, 17.

Andrews, S., 2010. FastQC: a quality control tool for high throughput sequence data (available at http://www.bioinformatics.babraham.ac.uk/projects/fastqc/).

Aragona, B.J., Liu, Y., Yu, Y.J., Curtis, J.T., Detwiler, J.M., Insel, T.R., Wang, Z., 2006. Nucleus accumbens dopamine differentially mediates the formation and maintenance of monogamous pair bonds. Nat. Neurosci. 9, 133–139.

Arnold, C.D., Gerlach, D., Stelzer, C., Boryń, Ł.M., Rath, M., Stark, A., 2013. Genome-wide quantitative enhancer activity maps identified by STARR-seq. Science 339, 1074–1077.

Bales, K.L., Plotsky, P.M., Young, L.J., Lim, M.M., Grotte, N., Ferrer, E., Carter, C.S., 2007. Neonatal oxytocin manipulations have long-lasting, sexually dimorphic effects on vasopressin receptors. Neuroscience 144, 38–45.

Belton, J.M., McCord, R.P., Gibcus, J.H., Naumova, N., Zhan, Y., Dekker, J., 2012. Hi-C: a comprehensive technique to capture the conformation of genomes. Methods 58, 268–276.

Bird, A., 2007. Perceptions of epigenetics. Nature 447, 396–398.

Bird, A., 2002. DNA methylation patterns and epigenetic memory. Genes Dev. 16, 6–21.

Bird, A.P., 1993. Functions for DNA methylation in vertebrates. Cold Spring Harb. Symp. Quant. Biol. 58, 281–285.

Bird, A.P., Wolffe, A.P., 1999. Methylation-induced repression— belts, braces, and chromatin. Cell 99, 451–454.

Bird, C.M., Burgess, N., 2008. The hippocampus and memory: insights from spatial processing. Nat. Rev. Neurosci. 9, 182–194.

Blattler, A., Yao, L., Witt, H., Guo, Y., Nicolet, C.M., Berman, B.P., Farnham, P.J., 2014. Global loss of DNA methylation uncovers intronic enhancers in genes showing expression changes. Genome Biol. 15, 469.

Bock, C., Walter, J., Paulsen, M., Lengauer, T., 2008. Inter-individual variation of DNA methylation and its implications for large-scale epigenome mapping. Nucleic Acids Res. 36, e55.

Booker, B.M., Friedrich, T., Mason, M.K., VanderMeer, J.E., Zhao, J., Eckalbar, W.L., Logan, M., Illing, N., Pollard, K.S., Ahituv, N., 2016. Bat accelerated regions identify a bat forelimb specific enhancer in the HoxD locus. PLoS Genet. 12.

Brockmann, H.J., Colson, T., Potts, W., 1994. Sperm competition in horseshoe crabs (*Limulus polyphemus*). Behav. Ecol. Sociobiol. 35, 153–160.

Buenrostro, J., Wu, B., Chang, H., Greenleaf, W., 2015. ATAC-seq: a method for assaying chromatin accessibility genome-wide. Curr. Protoc. Mol. Biol. 109, 1–9.

Bulger, M., Groudine, M., 2010. Enhancers: the abundance and function of regulatory sequences beyond promoters. Dev. Biol., Special Section: Gene Expression and Development 339, 250–257.

Byun, H.M., Siegmund, K.D., Pan, F., Weisenberger, D.J., Kanel, G., Laird, P.W., Yang, A.S., 2009. Epigenetic profiling of somatic tissues from human autopsy specimens identifies tissue- and individual-specific DNA methylation patterns. Hum. Mol. Genet. 18, 4808–4817.

Carter, C., Devries, C.A., Getz, L.L., 1995. Physiological substrates of mammalian monogamy: the prairie vole model. Neurosci. Biobehav. Rev. 19, 303–314.

Carter, C.S., Getz, L.L., Cohenparsons, M., 1986. Relationships between social-organization and behavioral endocrinology in a monogamous mammal. Adv. Study Behav. 16, 109–145.

Caspi, A., Moffitt, T.E., 2006. Gene–environment interactions in psychiatry: joining forces with neuroscience. Nat. Rev. Neurosci. 7, 583–590.

Cheng, J.C., Matsen, C.B., Gonzales, F.A., Ye, W., Greer, S., Marquez, V.E., Jones, P.A., Selker, E.U., 2003. Inhibition of DNA methylation and reactivation of silenced genes by zebularine. J. Natl. Cancer Inst. 95, 399–409.

Chimpanzee Sequencing and Analysis Consortium, 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. Nature 437, 69–87.

Cohen, S., Hamrick, N., 2003. Stable individual differences in physiological response to stressors: implications for stress-elicited changes in immune related health. Brain. Behav. Immun. 17, 407–414.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., Zhang, F., 2013. Multiplex genome engineering using CRISPR/Cas systems. Science 339, 819–823.

Corradin, O., Saiakhova, A., Akhtar-Zaidi, B., Myeroff, L., Willis, J., Cowper-Sal lari, R., Lupien, M., Markowitz, S., Scacheri, P.C., 2014. Combinatorial effects of

multiple enhancer variants in linkage disequilibrium dictate levels of gene expression to confer susceptibility to common traits. Genome Res. 24, 1–13.

Cushing, B.S., Kramer, K.M., 2005. Mechanisms underlying epigenetic effects of early social experience: the role of neuropeptides and steroids. Neurosci. Biobehav. Rev. 29, 1089–1105.

Dall, S.R.X., Houston, A.I., McNamara, J.M., 2004. The behavioral ecology of personality: consistent individual differences from an adaptive perspective. Ecol. Lett. 7, 734–739.

Debat, V., David, P., 2001. Mapping phenotypes: canalization, plasticity and developmental stability. Trends Ecol. Evol. 16, 555–561.

De Vries, G.J., Panzica, G.C., 2006. Sexual differentiation of central vasopressin and vasotocin systems in vertebrates: different mechanisms, similar endpoints. Neuroscience 138, 947–955.

Donaldson, Z.R., Young, L.J., 2008. Oxytocin, vasopressin, and the neurogenetics of sociality. Science 322, 900–904.

Duckworth, R.A., Belloni, V., Anderson, S.R., 2015. Cycles of species replacement emerge from locally induced maternal effects on offspring behavior in a passerine bird. Science 347, 875–877.

Eadie, J.M., Fryxell, J.M., 1992. Density dependence, frequency dependence, and alternative nesting strategies in goldeneyes. Am. Nat. 140, 621–641.

Eckhardt, F., Lewin, J., Cortese, R., Rakyan, V.K., Attwood, J., Burger, M., Burton, J., Cox, T.V., Davies, R., Down, T.A., Haefliger, C., Horton, R., Howe, K., Jackson, D.K., Kunde, J., Koenig, C., Liddle, J., Niblett, D., Otto, T., Pettett, R., Seemann, S., Thompson, C., West, T., Rogers, J., Olek, A., Berlin, K., Beck, S., 2006. DNA methylation profiling of human chromosomes 6, 20 and 22. Nat. Genet. 38, 1378–1385.

Ellis, B.J., Boyce, W.T., 2008. Biological sensitivity to context. Curr. Dir. Psychol. Sci. 17, 183–187.

Emlen, S.T., Oring, L.W., 1977. Ecology, sexual selection, and the evolution of mating systems. Science 197, 215–223.

Feil, R., Fraga, M.F., 2012. Epigenetics and the environment: emerging patterns and implications. Nat. Rev. Genet. 13, 97–109.

Feldman, R., Gordon, I., Schneiderman, I., Weisman, O., Zagoory-Sharon, O., 2010. Natural variations in maternal and paternal care are associated with systematic changes in oxytocin following parent–infant contact. Psychoneuroendocrinology 35, 1133–1141.

Francis, D.D., Young, L.J., Meaney, M.J., Insel, T.R., 2002. Naturally occurring differences in maternal care are associated with the expression of oxytocin and vasopressin (V1a) receptors: gender differences. J. Neuroendocrinol. 14, 349–353.

García-Navas, V., Ferrer, E.S., Bueno-Enciso, J., Barrientos, R., Sanz, J.J., Ortego, J., 2014. Extrapair paternity in Mediterranean blue tits: socioecological factors and the opportunity for sexual selection. Behav. Ecol. 25, 228–238.

Gardiner-Garden, M., Frommer, M., 1987. CpG islands in vertebrate genomes. J. Mol. Biol. 196, 261–282.

Georgieff, M.K., 2007. Nutrition and the developing brain: nutrient priorities and measurement. Am. J. Clin. Nutr. 85, 614S–620S.

Getz, L.L., Hofmann, J.E., McGuire, B., Dolan, T.W., 2001. Twenty-five years of population fluctuations of *Microtus ochrogaster* and *M. pennsylvanicus* in three habitats in east-central Illinois. J. Mammal. 82, 22–34.

Getz, L.L., McGuire, B., Pizzuto, T., Hofmann, J.E., Frase, B., 1993. Social organization of the prairie vole (*Microtus ochrogaster*). J. Mammal. 74, 44–58.

Goodson, J.L., Bass, A.H., 2001. Social behavior functions and related anatomical characteristics of vasotocin/vasopressin systems in vertebrates. Brain Res. Brain Res. Rev. 35, 246–265.

Goodson, J.L., Lindberg, L., Johnson, P., 2004. Effects of central vasotocin and mesotocin manipulations on social behavior in male and female zebra finches. Horm. Behav. 45, 136–143.

Grishkevich, V., Yanai, I., 2013. The genomic determinants of genotype × environment interactions in gene expression. Trends Genet. 29, 479–487.

Gross, M.R., 1996. Alternative reproductive strategies and tactics: diversity within sexes. Trends Ecol. Evol. 11, 92–98.

Gross, M.R., 1991. Evolution of alternative reproductive strategies: frequency-dependent sexual selection in male Bluegill sunfish. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 332, 59–66.

Guo, J.U., Su, Y., Shin, J.H., Shin, J., Li, H., Xie, B., Zhong, C., Hu, S., Le, T., Fan, G., Zhu, H., Chang, Q., Gao, Y., Ming, G., Song, H., 2014. Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. Nat. Neurosci. 17, 215–222.

Gutierrez-Arcelus, M., Ongen, H., Lappalainen, T., Montgomery, S.B., Buil, A., Yurovsky, A., Bryois, J., Padioleau, I., Romano, L., Planchon, A., Falconnet, E., Bielser, D., Gagnebin, M., Giger, T., Borel, C., Letourneau, A., Makrythanasis,

P., Guipponi, M., Gehrig, C., Antonarakis, S.E., Dermitzakis, E.T., 2015. Tissue-specific effects of genetic and epigenetic variation on gene regulation and splicing. PLOS Genet 11, e1004958.

Gutzler, S.J., Karom, M., Erwin, W.D., Albers, H.E., 2010. Arginine-vasopressin and the regulation of aggression in female Syrian hamsters (*Mesocricetus auratus*). Eur. J. Neurosci. 31, 1655–1663.

Hammock, E.A.D., Lim, M.M., Nair, H.P., Young, L.J., 2005. Association of vasopressin 1a receptor levels with a regulatory microsatellite and behavior. Genes Brain Behav. 4, 289–301.

Hammock, E.A.D., Young, L.J., 2005. Microsatellite instability generates diversity in brain and sociobehavioral traits. Science 308, 1630–1634.

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., Wang, W., Weng, Z., Green, R.D., Crawford, G.E., Ren, B., 2007. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat. Genet. 39, 311–318.

Hodges, E., Smith, A.D., Kendall, J., Xuan, Z., Ravi, K., Rooks, M., Zhang, M.Q., Ye, K., Bhattacharjee, A., Brizuela, L., McCombie, W.R., Wigler, M., Hannon, G.J., Hicks, J.B., 2009. High definition profiling of mammalian DNA methylation by array capture and single molecule bisulfite sequencing. Genome Res. 19, 1593–1605.

Hopkins, W.D., Donaldson, Z.R., Young, L.J., 2012. A polymorphic indel containing the RS3 microsatellite in the 5 flanking region of the vasopressin V1a receptor gene is associated with chimpanzee (*Pan troglodytes*) personality. Genes Brain Behav. 11, 552–558.

Hsieh, C.L., 1994. Dependence of transcriptional repression on CpG methylation density. Mol. Cell. Biol. 14, 5487–5494.

Hudson, R.R., Kreitman, M., Aguadé, M., 1987. A test of neutral molecular evolution based on nucleotide data. Genetics 116, 153–159.

Huse, S.M., Huber, J.A., Morrison, H.G., Sogin, M.L., Welch, D.M., 2007. Accuracy and quality of massively parallel DNA pyrosequencing. Genome Biol. 8, R143.

Inoue, F., Kircher, M., Martin, B., Cooper, G.M., Witten, D.M., McManus, M.T., Ahituv, N., Shendure, J., 2016. A systematic comparison reveals substantial differences in chromosomal versus episomal encoding of enhancer activity. bioRxiv 061606.

Irizarry, R.A., Ladd-Acosta, C., Wen, B., Wu, Z., Montano, C., Onyango, P., Cui, H., Gabo, K., Rongione, M., Webster, M., Ji, H., Potash, J.B., Sabunciyan, S., Feinberg, A.P., 2009. The human colon cancer methylome shows similar hypo-

and hypermethylation at conserved tissue-specific CpG island shores. Nat. Genet. 41, 178–186.

Izzi, B., Pistoni, M., Cludts, K., Akkor, P., Lambrechts, D., Verfaillie, C., Verhamme, P., Freson, K., Hoylaerts, M.F., 2016. Allele-specific DNA methylation reinforces PEAR1 enhancer activity. Blood 128, 1003–1012.

Jones, P.A., 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat. Rev. Genet. 13, 484–492.

Jones, P.A., Takai, D., 2001. The role of DNA methylation in mammalian epigenetics. Science 293, 1068–1070.

Katoh, K., Misawa, K., Kuma, K., Miyata, T., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30, 3059–3066.

Keller, M.C., Miller, G., 2006. Resolving the paradox of common, harmful, heritable mental disorders: which evolutionary genetic models work best? Behav. Brain Sci. 29, 405–452.

King, L.B., Walum, H., Inoue, K., Eyrich, N.W., Young, L.J., 2016. Variation in the oxytocin receptor gene predicts brain region–specific expression and social attachment. Biol. Psychiatry, Autism and the Social Brain 80, 160–169.

Kokko, H., Rankin, D.J., 2006. Lonely hearts or sex in the city? Density-dependent effects in mating systems. Philos. Trans. R. Soc. B Biol. Sci. 361, 319–334.

Krueger, F., Andrews, S.R., 2011. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics 27, 1571–1572.

Kuehl-Kovarik, M.C., Iqbal, J., Jacobson, C.D., 1997. Autoradiographic localization of arginine vasopressin binding sites in the brain of adult and developing Brazilian opossums. Brain. Behav. Evol. 49, 261–275.

Kundakovic, M., Champagne, F.A., 2011. Epigenetic perspective on the developmental effects of bisphenol A. Brain. Behav. Immun. 25, 1084–1093.

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359.

Laurent, L., Wong, E., Li, G., Huynh, T., Tsirigos, A., Ong, C.T., Low, H.M., Sung, K.W.K., Rigoutsos, I., Loring, J., Wei, C.L., 2010. Dynamic changes in the human methylome during differentiation. Genome Res. 20, 320–331.

Law, J.A., Jacobsen, S.E., 2010. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat. Rev. Genet. 11, 204–220.

Ledón-Rettig, C.C., Richards, C.L., Martin, L.B., 2013. Epigenetics for behavioral ecologists. Behav. Ecol. 24, 311–324.

Lee, H.J., Kusche, H., Meyer, A., 2012. Handed foraging behavior in scale-eating cichlid fish: its potential role in shaping morphological asymmetry. PLOS ONE 7, e44670.

Lee, H.J., Macbeth, A.H., Pagani, J., Young, W.S., 2009. Oxytocin: the great facilitator of life. Prog. Neurobiol. 88, 127–151.

Lein, E.S., Hawrylycz, M.J., Ao, N., Ayres, M., Bensinger, A., Bernard, A., Boe, A.F., Boguski, M.S., Brockway, K.S., Byrnes, E.J., Chen, L., Chen, L., Chen, T.M., Chi Chin, M., Chong, J., Crook, B.E., Czaplinska, A., Dang, C.N., Datta, S., Dee, N.R., Desaki, A.L., Desta, T., Diep, E., Dolbeare, T.A., Donelan, M.J., Dong, H.W., Dougherty, J.G., Duncan, B.J., Ebbert, A.J., Eichele, G., Estin, L.K., Faber, C., Facer, B.A., Fields, R., Fischer, S.R., Fliss, T.P., Frensley, C., Gates, S.N., Glattfelder, K.J., Halverson, K.R., Hart, M.R., Hohmann, J.G., Howell, M.P., Jeung, D.P., Johnson, R.A., Karr, P.T., Kawal, R., Kidney, J.M., Knapik, R.H., Kuan, C.L., Lake, J.H., Laramee, A.R., Larsen, K.D., Lau, C., Lemon, T.A., Liang, A.J., Liu, Y., Luong, L.T., Michaels, J., Morgan, J.J., Morgan, R.J., Mortrud, M.T., Mosqueda, N.F., Ng, L.L., Ng, R., Orta, G.J., Overly, C.C., Pak, T.H., Parry, S.E., Pathak, S.D., Pearson, O.C., Puchalski, R.B., Riley, Z.L., Rockett, H.R., Rowland, S.A., Royall, J.J., Ruiz, M.J., Sarno, N.R., Schaffnit, K., Shapovalova, N.V., Sivisay, T., Slaughterbeck, C.R., Smith, S.C., Smith, K.A., Smith, B.I., Sodt, A.J., Stewart, N.N., Stumpf, K.R., Sunkin, S.M., Sutram, M., Tam, A., Teemer, C.D., Thaller, C., Thompson, C.L., Varnam, L.R., Visel, A., Whitlock, R.M., Wohnoutka, P.E., Wolkey, C.K., Wong, V.Y., Wood, M., Yaylaoglu, M.B., Young, R.C., Youngstrom, B.L., Feng Yuan, X., Zhang, B., Zwingman, T.A., Jones, A.R., 2007. Genome-wide atlas of gene expression in the adult mouse brain. Nature 445, 168–176.

Librado, P., Rozas, J., 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25, 1451–1452.

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079.

Lim, M.M., Liu, Y., Ryabinin, A.E., Bai, Y., Wang, Z., Young, L.J., 2007. CRF receptors in the nucleus accumbens modulate partner preference in prairie voles. Horm. Behav. 51, 508–515.

Lim, M.M., Wang, Z., Olazábal, D.E., Ren, X., Terwilliger, E.F., Young, L.J., 2004. Enhanced partner preference in a promiscuous species by manipulating the expression of a single gene. Nature 429, 754–757.

Lim, M.M., Young, L.J., 2006. Neuropeptidergic regulation of affiliative behavior and social bonding in animals. Horm. Behav. 50, 506–517.

Lim, M.M., Young, L.J., 2004. Vasopressin-dependent neural circuits underlying pair bond formation in the monogamous prairie vole. Neuroscience 125, 35–45.

Lister, R., Mukamel, E.A., Nery, J.R., Urich, M., Puddifoot, C.A., Johnson, N.D., Lucero, J., Huang, Y., Dwork, A.J., Schultz, M.D., Yu, M., Tonti-Filippini, J., Heyn, H., Hu, S., Wu, J.C., Rao, A., Esteller, M., He, C., Haghighi, F.G., Sejnowski, T.J., Behrens, M.M., Ecker, J.R., 2013. Global epigenomic reconfiguration during mammalian brain development. Science 341, 1237905.

Liu, Y., Curtis, J.T., Wang, Z., 2001. Vasopressin in the lateral septum regulates pair bond formation in male prairie voles (*Microtus ochrogaster*). Behav. Neurosci. 115, 910–919.

Lott, D.F., 1984. Intraspecific variation in the social systems of wild vertebrates. Barlow, G., Bateson, P.P.G., Oppenheim, R.W. (Eds.), Cambridge University Press, Cambridge, UK, pp. 1-5.

Lukas, M., Bredewold, R., Neumann, I.D., Veenema, A.H., 2010. Maternal separation interferes with developmental changes in brain vasopressin and oxytocin receptor binding in male rats. Neuropharmacology, Neuropeptides 58, 78–87.

Lunter, G., Goodson, M., 2011. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. Genome Res. 21, 936–939.

Lupien, S.J., McEwen, B.S., Gunnar, M.R., Heim, C., 2009. Effects of stress throughout the lifespan on the brain, behaviour and cognition. Nat. Rev. Neurosci. 10, 434–445.

Mabry, K.E., Streatfeild, C.A., Keane, B., Solomon, N.G., 2011. *Avpr1a* length polymorphism is not associated with either social or genetic monogamy in free-living prairie voles. Anim. Behav. 81, 11–18.

Manke, T., Heinig, M., Vingron, M., 2010. Quantifying the effect of sequence variation on regulatory interactions. Hum. Mutat. 31, 477–483.

Manuck, S.B., McCaffery, J.M., 2014. Gene-environment interaction. Annu. Rev. Psychol. 65, 41–70.

McCrea, R.R., Costa, P.T. Jr., 1999. A five-factor theory of personality. In handbook of personality: theory and research. 2nd ed. Pervin L.A., John O.P., Elsevier.

McGill, B.E., Bundle, S.F., Yaylaoglu, M.B., Carson, J.P., Thaller, C., Zoghbi, H.Y., 2006. Enhanced anxiety and stress-induced corticosterone release are associated with increased Crh expression in a mouse model of Rett syndrome. Proc. Natl. Acad. Sci. U. S. A. 103, 18267–18272.

McGraw, L.A., Young, L.J., 2010. The prairie vole: an emerging model organism for understanding the social brain. Trends Neurosci. 33, 103.

McGuire, B., Pizzuto, T., Getz, L.L., 1990. Potential for social interaction in a natural population of prairie voles (*Microtus ochrogaster*). Can. J. Zool. 68, 391–398.

Meaney, M.J., 2001. Maternal care, gene expression, and the transmission of individual differences in stress reactivity across generations. Annu. Rev. Neurosci. 24, 1161–1192.

Mohn, F., Weber, M., Schübeler, D., Roloff, T.C., 2009. Methylated DNA immunoprecipitation (MeDIP), Methods in Mol. Biol. 507, 55–64.

Murgatroyd, C., Patchev, A.V., Wu, Y., Micale, V., Bockmühl, Y., Fischer, D., Holsboer, F., Wotjak, C.T., Almeida, O.F.X., Spengler, D., 2009. Dynamic DNA methylation programs persistent adverse effects of early-life stress. Nat. Neurosci. 12, 1559–1566.

Nachman, M.W., Hoekstra, H.E., D'Agostino, S.L., 2003. The genetic basis of adaptive melanism in pocket mice. Proc. Natl. Acad. Sci. 100, 5268–5273.

Nan, X., Ng, H.H., Johnson, C.A., Laherty, C.D., Turner, B.M., Eisenman, R.N., Bird, A., 1998. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. Nature 393, 386–389.

Okhovat, M., Berrio, A., Wallace, G., Ophir, A.G., Phelps, S.M., 2015. Sexual fidelity trade-offs promote regulatory variation in the prairie vole brain. Science 350, 1371–1374.

Okhovat, M., Maguire, S.M., Phelps, S.M., unpublished. Methylation of *avpr1a* in the cortex of wild prairie voles: effects of CpG position and polymorphism. Roy. Soc. Open Science.

Oldfield, R.G., Harris, R.M., Hofmann, H.A., 2015. Integrating resource defence theory with a neural nonapeptide pathway to explain territory-based mating systems. Front. Zool. 12, S16.

Ophir, A.G., Phelps, S.M., Sorin, A.B., Wolff, J.O., 2008a. Social but not genetic monogamy is associated with greater breeding success in prairie voles. Anim. Behav. 75, 1143–1154.

Ophir, A.G., Phelps, S.M., Sorin, A.B., Wolff, J.O., 2007. Morphological, genetic, and behavioral comparisons of two prairie vole populations in the field and laboratory. J. Mammal. 88, 989–999.

Ophir, A.G., Wolff, J.O., Phelps, S.M., 2008b. Variation in neural V1aR predicts sexual fidelity and space use among male prairie voles in semi-natural settings. Proc. Natl. Acad. Sci. U. S. A. 105, 1249–1254.

Parnell, L.D., Blokker, B.A., Dashti, H.S., Nesbeth, P.D., Cooper, B.E., Ma, Y., Lee, Y.C., Hou, R., Lai, C.Q., Richardson, K., Ordovás, J.M., 2014. CardioGxE, a catalog of gene-environment interactions for cardiometabolic traits. BioData Min. 7, 21.

Pastinen, T., 2010. Genome-wide allele-specific analysis: insights into regulatory variation. Nat. Rev. Genet. 11, 533–538.

Pfennig, D.W., 1992. Polyphenism in spadefoot toad tadpoles as a logically adjusted evolutionarily stable strategy. Evolution 46, 1408–1420.

Phelps, S.M., Campbell, P., Zheng, D.J., Ophir, A.G., 2010. Beating the boojum: comparative approaches to the neurobiology of social behavior. Neuropharmacology 58, 17–28.

Phelps, S.M., Ophir, A.G., 2009. Monogamous brains and alternative tactics: neuronal V1aR, space use, and sexual infidelity among male prairie voles, in: Dukas, R., Ratcliffe, J.M. (Eds.), Cognitive Ecology II. University of Chicago press, pp. 156-176.

Phelps, S.M., Young, L.J., 2003. Extraordinary diversity in vasopressin (V1a) receptor distributions among wild prairie voles (*Microtus ochrogaster*): patterns of variation and covariation. J. Comp. Neurol. 466, 564–576.

Pigliucci, M., 2005. Evolution of phenotypic plasticity: where are we going now? Trends Ecol. Evol. 20, 481–486.

Pigliucci, M., 2001. Phenotypic plasticity: beyond nature and nurture, in Johns Hopkins Press, Baltimore.

Pinheiro, J., Bate, D., DebRoy, S., Sarkar, D., R Core Team, 2016. nlme: linear and nonlinear mixed effects models. R package version 3.1-128.

Prounis, G.S., Foley, L., Rehman, A., Ophir, A.G., 2015. Perinatal and juvenile social environments interact to shape cognitive behaviour and neural phenotype in prairie voles. Proc R Soc B 282, 20152236.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., Sham, P.C., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81, 559–575.

Ra'Anan, Z., Sagi, A., 1985. Alternative mating strategies in male morphotypes of the freshwater prawn *Macrobrachium rosenbergii* (De Man). Biol. Bull. 169, 592-601.

Rauch, T.A., Wu, X., Zhong, X., Riggs, A.D., Pfeifer, G.P., 2009. A human B cell methylome at 100−base pair resolution. Proc. Natl. Acad. Sci. U. S. A. 106, 671–678.

Razin, A., Riggs, A.D., 1980. DNA methylation and gene function. Science 210, 604–610.

Reik, W., 2007. Stability and flexibility of epigenetic gene regulation in mammalian development. Nature 447, 425–432.

Resendez, S.L., Kuhnmuench, M., Krzywosinski, T., Aragona, B.J., 2012. ϰ-Opioid receptors within the nucleus accumbens shell mediate pair bond maintenance. J. Neurosci. 32, 6771–6784.

Rice, P., Longden, I., Bleasby, A., 2000. EMBOSS: the European molecular biology open software suite. Trends Genet. 16, 276–277.

Robinson, G.E, 2015. Dissecting diversity in the social brain. Science. 350, 1310-1312.

Robinson, G.E., Ben-Shahar, Y., 2002. Social behavior and comparative genomics: new genes or new gene regulation? Genes Brain Behav. 1, 197–203.

Rollins, R.A., Haghighi, F., Edwards, J.R., Das, R., Zhang, M.Q., Ju, J., Bestor, T.H., 2006. Large-scale structure of genomic methylation patterns. Genome Res. 16, 157–163.

Rosenbloom, K.R., Sloan, C.A., Malladi, V.S., Dreszer, T.R., Learned, K., Kirkup, V.M., Wong, M.C., Maddren, M., Fang, R., Heitner, S.G., Lee, B.T., Barber, G.P., Harte, R.A., Diekhans, M., Long, J.C., Wilder, S.P., Zweig, A.S., Karolchik, D., Kuhn, R.M., Haussler, D., Kent, W.J., 2013. ENCODE data in the UCSC genome browser: year 5 update. Nucleic Acids Res. 41, D56–63.

Roth, T.L., Sweatt, J.D., 2011. Epigenetic mechanisms and environmental shaping of the brain during sensitive periods of development. J. Child Psychol. Psychiatry 52, 398–408.

Rubinstein, M., de Souza, F.S.J., 2013. Evolution of transcriptional enhancers and animal diversity. Philos. Trans. R. Soc. B Biol. Sci. 368.

Rusk, N., 2009. Reverse ChIP. Nat. Methods 6, 187–187.

Schang, A.L., Granger, A., Quérat, B., Bleux, C., Cohen-Tannoudji, J., Laverrière, J.N., 2013. GATA2-induced silencing and LIM-homeodomain protein-induced activation are mediated by a bi-functional response element in the rat GnRH receptor gene. Mol. Endocrinol. Baltim. Md 27, 74–91.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.Y., White, D.J., Hartenstein, V., Eliceiri, K., Tomancak, P., Cardona, A., 2012. Fiji: an open-source platform for biological-image analysis. Nat. Methods 9, 676–682.

Schübeler, D., 2015. Function and information content of DNA methylation. Nature 517, 321–326.

Segal, E., Widom, J., 2009. From DNA sequence to transcriptional behaviour: a quantitative approach. Nat. Rev. Genet. 10, 443–456.

Sheriff, M.J., Krebs, C.J., Boonstra, R., 2010. The ghosts of predators past: population cycles and the role of maternal programming under fluctuating predation risk. Ecology 91, 2983–2994.

Sheriff, M.J., Love, O.P., 2013. Determining the adaptive potential of maternal stress. Ecol. Lett. 16, 271–280.

Simmons, R.K., Stringfellow, S.A., Glover, M.E., Wagle, A.A., Clinton, S.M., 2013. DNA methylation markers in the postnatal developing rat brain. Brain Res. 1533.

Simola, D.F., Ye, C., Mutti, N.S., Dolezal, K., Bonasio, R., Liebig, J., Reinberg, D., Berger, S.L., 2013. A chromatin link to caste identity in the carpenter ant Camponotus floridanus. Genome Res. 23, 486–496.

Sinervo, B., Lively, C.M., 1996. The rock–paper–scissors game and the evolution of alternative male strategies. Nature 380, 240–243.

Sinervo, B., Zamudio, K.R., 2001. The evolution of alternative reproductive strategies: fitness differential, heritability, and genetic correlation between the sexes. J. Hered. 92, 198–205.

Smith, M., 1982. Evolution and the theory of games. Cambridge University Press, Cambridge, UK.

Solomon, N.G., Richmond, A.R., Harding, P.A., Fries, A., Jacquemin, S., Schaefer, R.L., Lucia, K.E., Keane, B., 2009. Polymorphism at the *avpr1a* locus in male prairie voles correlated with genetic but not social monogamy in field populations. Mol. Ecol. 18, 4680–95.

Streatfeild, C.A., Mabry, K.E., Keane, B., Crist, T.O., Solomon, N.G., 2011. Intraspecific variability in the social and genetic mating systems of prairie voles, *Microtus ochrogaster*. Anim. Behav. 82, 1387–1398.

Sturtevant, A.H., 1923. Inheritance of direction of coilling in Limnaea. Science 58, 269–270.

Swallow, D.M., 2003. Genetics of lactase persistence and lactose intolerance. Annu. Rev. Genet. 37, 197–219.

Szyf, M., Bick, J., 2013. DNA methylation: a mechanism for embedding early life experiences in the genome. Child Dev. 84, 49–57.

Tajima, F., 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123, 585–595.

Tate, P.H., Bird, A.P., 1993. Effects of DNA methylation on DNA-binding proteins and gene expression. Curr. Opin. Genet. Dev. 3, 226–231.

The FANTOM Consortium and the RIKEN PMI and CLST (DGT), 2014. A promoter-level mammalian expression atlas. Nature 507, 462–470.

Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernot, B., Garg, K., John, S., Sandstrom, R., Bates, D., Boatman, L., Canfield, T.K., Diegel, M., Dunn, D., Ebersol, A.K., Frum, T., Giste, E., Johnson, A.K., Johnson, E.M., Kutyavin, T., Lajoie, B., Lee, B.K., Lee, K., London, D., Lotakis, D., Neph, S., Neri, F., Nguyen, E.D., Qu, H., Reynolds, A.P., Roach, V., Safi, A., Sanchez, M.E., Sanyal, A., Shafer, A., Simon, J.M., Song, L., Vong, S., Weaver, M., Yan, Y., Zhang, Z., Zhang, Z., Lenhard, B., Tewari, M., Dorschner, M.O., Hansen, R.S., Navas, P.A., Stamatoyannopoulos, G., Iyer, V.R., Lieb, J.D., Sunyaev, S.R., Akey, J.M., Sabo, P.J., Kaul, R., Furey, T.S., Dekker, J., Crawford, G.E., Stamatoyannopoulos, J.A., 2012. The accessible chromatin landscape of the human genome. Nature 489, 75–82.

Tomso, D.J., Bell, D.A., 2003. Sequence context at human single nucleotide polymorphisms: overrepresentation of CpG dinucleotide at polymorphic sites and suppression of variation in CpG islands. J. Mol. Biol. 327, 303–308.

Tribollet, E., Goumaz, M., Raggenbass, M., Dreifuss, J.J., 1991. Appearance and transient expression of vasopressin and oxytocin receptors in the rat brain. J. Recept. Res. 11, 333–346.

Tycko, B., 2010. Allele-specific DNA methylation: beyond imprinting. Hum. Mol. Genet. ddq376.

Van Groen, T., Kadish, I., Wyss, J.M., 2002. The role of the laterodorsal nucleus of the thalamus in spatial learning and memory in the rat. Behav. Brain Res. 136, 329–337.

Vann, S.D., Aggleton, J.P., Maguire, E.A., 2009. What does the retrosplenial cortex do? Nat. Rev. Neurosci. 10, 792–802.

Veenema, A.H., Bredewold, R., De Vries, G.J., 2012. Vasopressin regulates social recognition in juvenile and adult rats of both sexes, but in sex- and age-specific ways. Horm. Behav. 61, 50–56.

Verweij, K.J.H., Yang, J., Lahti, J., Veijola, J., Hintsanen, M., Pulkki-Råback, L., Heinonen, K., Pouta, A., Pesonen, A.K., Widen, E., Taanila, A., Isohanni, M., Miettunen, J., Palotie, A., Penke, L., Service, S.K., Heath, A.C., Montgomery, G.W., Raitakari, O., Kähönen, M., Viikari, J., Räikkönen, K., Eriksson, J.G., Keltikangas-Järvinen, L., Lehtimäki, T., Martin, N.G., Järvelin, M.R., Visscher, P.M., Keller, M.C., Zietsch, B.P., 2012. Maintenance of genetic variation in human personality: testing evolutionary models by estimating heritability due to common causal variants and investigating the effect of distant inbreeding. Evolution 66, 3238–3251.

Visser, M., Kayser, M., Palstra, R.J., 2012. HERC2 rs12913832 modulates human pigmentation by attenuating chromatin-loop formation between a long-range enhancer and the OCA2 promoter. Genome Res. 22, 446–455.

Wachter, E., Quante, T., Merusi, C., Arczewska, A., Stewart, F., Webb, S., Bird, A., 2014. Synthetic CpG islands reveal DNA sequence determinants of chromatin structure. eLife 3, e03397.

Walton, E., Hass, J., Liu, J., Roffman, J.L., Bernardoni, F., Roessner, V., Kirsch, M., Schackert, G., Calhoun, V., Ehrlich, S., 2016. Correspondence of DNA methylation between blood and brain tissue and its application to schizophrenia research. Schizophr. Bull. 42, 406–414.

Walum, H., Westberg, L., Henningsson, S., Neiderhiser, J.M., Reiss, D., Igl, W., Ganiban, J.M., Spotts, E.L., Pedersen, N.L., Eriksson, E., Lichtenstein, P., 2008. Genetic variation in the vasopressin receptor 1a gene (*AVPR1A*) associates with pair-bonding behavior in humans. Proc. Natl. Acad. Sci. 105, 14153–14156.

Wang, Z., Young, L.J., Liu, Y., Insel, T.R., 1997. Species differences in vasopressin receptor binding are evident early in development: comparative anatomic studies in prairie and montane voles. J. Comp. Neurol. 378, 535–546.

Weaver, I.C.G., Cervoni, N., Champagne, F.A., D'Alessio, A.C., Sharma, S., Seckl, J.R., Dymov, S., Szyf, M., Meaney, M.J., 2004. Epigenetic programming by maternal behavior. Nat. Neurosci. 7, 847–854.

Weber, M., Hellmann, I., Stadler, M.B., Ramos, L., Pääbo, S., Rebhan, M., Schübeler, D., 2007. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nat. Genet. 39, 457–466.

Williams, J.R., Insel, T.R., Harbaugh, C.R., Carter, C.S., 1994. Oxytocin administered centrally facilitates formation of a partner preference in female prairie voles (*Microtus ochrogaster*). J. Neuroendocrinol. 6, 247–250.

Wray, G.A., 2007. The evolutionary significance of cis-regulatory mutations. Nat. Rev. Genet. 8, 206–216.

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

Young, L.J., Nilsen, R., Waymire, K.G., MacGregor, G.R., Insel, T.R., 1999. Increased affiliative response to vasopressin in mice expressing the V1a receptor from a monogamous vole. Nature 400, 766–768.

Young, L.J., Wang, Z., 2004. The neurobiology of pair bonding. Nat. Neurosci. 7, 1048–1054.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., Liu, X.S., 2008. Model-based analysis of ChIP-seq (MACS). Genome Biol. 9, R137.

Zhao, Z., Tavoosidana, G., Sjölinder, M., Göndör, A., Mariano, P., Wang, S., Kanduri, C., Lezcano, M., Singh Sandhu, K., Singh, U., Pant, V., Tiwari, V., Kurukuti, S., Ohlsson, R., 2006. Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. Nat. Genet. 38, 1341–1347.

Zheng, D.J., Larsson, B., Phelps, S.M., Ophir, A.G., 2013. Female alternative mating tactics, reproductive success and nonapeptide receptor expression in the social decision-making network. Behav. Brain Res. 246, 139–147.

Zimmermann, C.A., Hoffmann, A., Raabe, F., Spengler, D., 2015. Role of Mecp2 in experience-dependent epigenetic programming. Genes 6, 60–86.