Analysis of candidate genes for behavioral differences in mice

Dissertation

zur Erlangung des Doktorgrades der Mathematisch-Naturwissenschaftlichen Fakultät der Christian-Albrecht-Universität zu Kiel

vorgelegt von

Maryam Keshavarz

Plön, July, 2018

First referee

Prof. Dr. Diethard Tautz

Second referee

Prof. Dr. Hinrich Schulenburg

Date of oral examination

10.10.2018

Contents

Summary

Organisms have evolved different behavioral strategies for better survival and reproduction. However, the genetic basis for such traits remains still as a longstanding fascinating question in evolutionary biology. Mate choice strategy is one of the behavioral traits which can play an important role in the life history of organisms. A previous study had shown that mate choice preference between two populations of the Western house mouse (*M. m. domesticus*) is influenced by the genetic background of the fathers. Transcriptome analysis in a follow up study revealed an imprinted cluster on Chromosome 7, known as Prader-Willi Syndrome (PWS) region, and also *Peg13* on Chromosome 15 as loci that are highly differentiated between mouse populations and therefore have been suggested as potential regions which may regulate this paternal preference in the house mouse. The present thesis was aimed to investigate the functional role of these two imprinted regions in Western house mice behavior.

In the first chapter, I investigated the role of the PWS region on Chromosome 7 through a variety of techniques, including copy number variation analysis, behavioral correlations and transcriptomics. I found that two paternally imprinted tandemly repeated regulatory RNA coding genes (SNORD115 and SNORD116) are of special interest. Their copy number evolves very fast and correlates highly with personality traits between individuals. Further I found that the copy number variation influences the expression of more than 130 genes including genes involved in serotonin regulation, vocalization and bone development. The findings suggest a molecular mechanism for the generation and variability of personality traits in mammals.

The second chapter focuses on the analysis of paternally imprinted *Peg13* gene, which has so far not been functionally studied. This gene has been known as non-coding gene, but data from ribosomal profiling analysis revealed that this gene could indeed produce a small peptide and it is not a simple non-coding gene. By using knock out mice, I showed even a partial deletion at 3' part of *Peg13* could significantly change expression of hundreds genes in the brain and could thus influence various mouse behaviors. The phenotypic analysis showed a significant effect on mouse sexual and parental behavior. Since RNAseq analysis from mouse brain development timeline showed that this gene has highest expression in day

12.5, I propose that *Peg13* may play important role in preoptic area development with possible direct role in sexual and mate choice behavior.

Overall the work presented in this thesis describes the first major molecular mechanism underlying mouse personality traits. It presents also the first functional study on the *Peg13* gene in mice, which highlights its possible role in mouse brain development and sexual behavior.

Zusammenfassung

Alle Organismen haben Strategien für ihr Überleben und ihre Reproduktion entwickelt. Allerdings ist die genetische Grundlage für diese Merkmale bisher weitgehend ungeklärt und damit eine der großen faszinierenden Fragen der Evolutionsbiologie. Partnerwahl ist eine der Verhaltensstrategien die eine wichtige Rolle im Lebenslauf eines Organismus spielt. Eine frühere Studie hatte gezeigt, dass die Partner Präferenz zwischen zwei Populationen der Westlichen Hausmaus (*M. m. domesticus*) durch den genetischen Hintergrund des Vaters beeinflusst wird. Transkriptomanalysen in einer weiteren Studie hatten nahe gelegt, dass eine uniparental exprimierte Genregion auf Chromosom 7, bekannt als Prader-Willi-Syndrom Region (PWS), sowie auch der Locus *Peg13* auf Chromosom 15, als Loci in Frage kommen, die die väterliche Prägung vermitteln könnten und zwischen den Populationen hoch differenziert sind. Die vorliegende Doktorarbeit hatte zum Ziel die funktionelle Rolle dieser beiden uniparental exprimierten Regionen bei Mäusen zu untersuchen.

Im ersten Kapitel untersuchte ich die Rolle der PWS Region auf Chromosom 7 mittels verschiedener Techniken, nämlich Analyse der Variation in der Kopien Anzahl, Korrelation mit dem Verhalten und Transkriptom Analysen. Ich konnte zeigen, dass zwei paternal exprimierte und in tandem wiederholte, regulatorische RNA kodierende Gene (SNORD115 und SNORD116) von speziellem Interesse sind. Sie evolvieren schnell in Bezug auf ihre Kopien Anzahl und die Kopien Anzahl korreliert mit Persönlichkeitsmerkmalen zwischen den Individuen. Ich konnte zeigen, dass die Kopien Anzahl Variation die Genexpression von mehr als 130 Genen beeinflusst, u.a. Gene die an der Serotonin Regulation, Vokalisation und Knochen Entwicklung beteiligt sind. Die Ergebnisse legen einen molekularen Mechanismus für die Generierung der Variabilität der Persönlichkeit bei Säugetieren nahe.

Das zweite Kapitel befasst sich mit der Analyse des paternal exprimierten Gens *Peg13*, für das bisher keine funktionellen Studien vorlagen. Das Gen war bisher als nicht-kodierend beschrieben worden, aber Daten von ribosomalen Profil Analysen legen nahe, dass das Gen für ein kurzes Protein kodieren könnte. Durch die Untersuchung einer knock-out Maus konnte ich zeigen, dass bereits eine partielle Deletion des 3´-Endes von *Peg13* die Expression von hunderten Genen im Gehirn verändert und damit auch Einfluss auf das Verhalten haben könnte. Die phänotypische Analyse zeigte in der Tat eine Veränderung des sexuellen und parentalen Verhaltens. Da die RNA Expressions-Analyse die höchste Expression in der Embryonal Entwicklung (Tag 12.5) zeigt, schlage ich vor, dass *Peg13* eine wichtige Rolle in

der Entwicklung präoptischen Region im Gehirn hat, die eine direkte Funktion für die sexuelle Entwicklung und das Paarungsverhalten hat.

Die Ergebnisse dieser Doktorarbeit ergeben den ersten größeren Einblick in die molekularen Mechanismen die Persönlichkeits-Merkmale bei Mäusen beeinflussen. Sie beinhalten auch die ersten funktionellen Studien zum *Peg13* Gen bei Mäusen, mit Hinweisen auf eine Rolle in der Gehirnentwicklung und sexuellem Verhalten.

General introduction

Behavior and evolution

A core function of behavioral traits is to help an organism to find food resources, shield against predators and find the best mates for successful reproduction (Waddington 1960). However, the evolution of behavioral traits has been subject to many discussions, including problems of definition and the role of genetics versus plasticity (Plotkin 1988; Duckworth 2008).

Behavior can act as driver of evolution where the mutations cause changes in an organism's behavior. In this case behavioral changes in an individual may change the way of its interaction with the environment and could thus trigger evolutionary responses of morphology, physiology and/or life history of an organism. (Mayr 1963; Piaget 1978; Wyles et al 1983; Huey et al. 2003; Sol et al. 2005). Behavior can also act as inhibitor of evolution where changes in external environmental conditions cause the behavioral reaction in an organism. These behavioral reactions, also known as plasticity, will allow an organism to find either new way of interaction with the environment or even move to a new environment to have less stressful condition and protect from strong directional selection (Losos et al. 2004; Badyaev 2005). Therefore exploring the genetic mechanisms underlying behavioral traits would be necessary to understand how behavior can evolve.

Mate choice

Mate choice is a key behavior with direct evolutionary consequences for a population. In case of the absence of mate choice, one would have random mating, a situation in which any individual with any given genotype would have an equal chance of mating with other individuals. This would be the situation of an ideal Hardy-Weinberg equilibrium, in which allele frequencies do not change. However, in natural populations, there is almost always some mate choice bias (Hedrick 2017).

Mate selection occurs usually according to some form of attractiveness of an individual's phenotype. This phenotype could be morphological features such as body size, skin coloration and pigmentation or behavioral and even sensory traits (Ryan et al. 1990; Vincent & Sadler 1995; Jones & Ratterman 2009; Xu et al. 2016; Fernald 2017).

There have been several attempts to explain the causality of mate choice preference. The first model is the condition-dependent indicator theory (Mead & Arnold 2004; Jones & Ratterman 2009). According to this model, a certain phenotypic trait which causes attractiveness in mating behavior may reflect how well an individual can cope with the environmental conditions, including defense against parasites. This model proposes that a female chooses her mate according to a special phenotypic trait of a male which may reflect the genetic quality to cope with the environment and thus have a better condition. So females tend to mate with a male who has better alleles for a given environment and may allow a female to produce more offspring (Zahavi 1975; Jones & Ratterman 2009).

The second model, which is called the condition-independent indicator theory, differs from the first model in a way that the ability of having a special phenotypic trait in a male is no longer dependent on his condition. In this model one assumes a genetic correlation between a special phenotypic trait and male viability. Therefore, such a genetic correlation will give the chance to females to produce offspring with higher fitness (Smith 1991; Jones & Ratterman 2009).

In the third model, the focus is on genetic compatibility. This model suggests that females choose males who complement their own genome and by this way avoid any genomic incompatibility which may cause problems for the next generation (Wright 1975; Ehrlich & Raven 1969; Jones & Ratterman 2009). There are also several other theories which basically suggest that females may exhibit special mate choice preference simply because they are predisposed to do so, and the preference may have evolved as a results of evolutionary mechanisms such as natural selection or drift, unrelated to sexual selection (Clegg 1995; Fuller 2005; Jones & Ratterman 2009).

Assortative mating

Assortative mating is a form of non-random mating which has been defined differently in the literature. In evolutionary biology, it has been defined as a mechanism of premating reproductive isolation between divergent populations or distinct species (Johannesson 1995; Jiang et al. 2013). In behavioral biology, assortative mating has been described as a specific form of mate choice where individuals choose their mates based on the phenotypic similarity to themselves (Crespi 1989; Shine et al. 2001). It could be positive assortative mating if individuals with similar phenotypes mate with one another more frequently than would be

expected under a random mating system. It could be negative, also known as disassortative mating, where the tendency is to mate with less phenotypically similar individuals (Hooper & Miller 2008; Jiang et al. 2013; Hedrick 2017; Fargevieille et al. 2017). The third definition proposed a general concept for assortative mating and describes it as a pattern of non-random mating, regardless of any specific assumption related to behavioral aspect or evolutionary mechanism (Lewontin et al. 1968; Kondrashov & Shpak 1998).

Positive assortative mating can significantly reduce intermediated phenotypes and shift the population to extreme phenotypes. This assortment increases homozygosity within loci and also linkage disequilibrium between loci (Lynch & Walsh 1998; Jiang et al. 2013; Hedrick 2017). Oppositely, negative assortative mating by avoiding inbreeding can decrease the extreme phenotype within the population, increases heterozygosity and keeps the variation within the population (Kirkpatrick & Ravigné 2002). Therefore both negative and positive assortative mating could break Hardy-Weinberg equilibrium and change allele frequencies in a population (Redden & Allison 2006).

Assortative mating has the potential to act as evolutionary agent and may trigger sympatric speciation. Sympatric speciation is defined as an evolutionary process when new species evolve from a single ancestral species without any geographical or physical barrier in the environment. This kind of speciation may happen when strong assortative mating and weak disruptive selection operate on the same phenotypic trait (Xue et al. 2014).

Assortative mating could be also detectable where two allopatric populations, which already adapted to different ecological niches, meet each other. In this case the mate choice would mostly occur between individuals from similar environments. It implies that prezygotic isolation has been already evolved within each population to help individuals mate truly with their own species and origin. This prezygotic isolation may also contribute to sympatric speciation under secondary contact conditions (Vines & Schluter 2006). Therefore assortative mating can play a major role in population differentiation and speciation.

House mouse and paternal mate choice preference

The house mouse (*Mus musculus)* is used as an important model organism in genetics and biomedical research (Harr et al. 2016). At the same time, it is also considered as the most successful invasive mammal after humans, implying a high adaptability to new

environmental conditions (Lowe et al. 2000). This success is partly related to their commensal interaction with humans, which have allowed them to spread across the world (Lowe et al. 2000). There are three subspecies of *Mus musculus*: the Western house mouse *Mus musculus domesticus*, the Eastern house mouse *Mus musculus musculus* and the Southeast-Asian house mouse *Mus musculus castaneus* (Harr et al. 2016) but there are also further as yet not well defined lineages, especially in Iran (Hardouin et al. 2015).

A previous study in our research group found that mate choice decisions between two allopatric populations of the Western house mouse (*M. m. domesticus*), one from Germany (CB) and one from France (MC), appear to be influenced by paternally transmitted cues. In the respective study, an experiment which allowed the free choice of mates in a semi-natural environment was conducted. Mating success was assessed through molecular paternity analysis. Mattings between a hybrid and an animal of pure origin showed a significant preference for matching with the paternal side of the hybrid (Montero et al. 2013).

Phylogeographic and fossil analysis suggest that these two populations of the Western house mouse are derived from animals that have colonized Western Europe only about 3,000 years ago (Cucchi et al. 2005). Despite their very recent time of divergence, it was found that these two populations exhibit considerable genetic (Bryk et al. 2013; Ihle 2006; Lorenc et al. 2014; Staubach et al. 2012; Teschke et al. 2008) as well as behavioral differences (Krebs 2018; von Merten et al. 2014; Montero et al. 2013). I provide here a short summary of these behavioral differences and their role in mate choice.

1-Competitive Ability

The competitive potential of the individuals is one of the factors which can influence mate choice behavior (Singleton & Hay 1983; Zack & Stutchbury 1992; Wroblewski et al. 2009). Recent study from our group on the competitive ability of German and French house mice showed a strong biased toward German male mice that showed a higher propensity of competitive ability than French male mice. This behavior consequently resulted also in higher reproductive success (Linnenbrink et al. 2018). These results clearly suggest that these two populations are behaviorally well differentiated, but it does not provide clues which could explain the paternal mate choice preference between these mouse populations.

2-Personality

Personality is another factor which could influence mate choice behavior. Personality can be defined as those characteristics of individuals which are repeatable across time and context (Gosling 2001; Sih et al. 2004). For many years, it was assumed that family environment, including parenting quality, played a causal role in personality development (Bell 1968). However later, the dynamic interactionistic paradigm showed that children were not simply the products of parental behavior (Patterson 1982; Sameroff 1983; Magnusson 1990; Caspi & Shiner 2006). South et al. (2008) proposed that family environment is not the only component influencing individual offspring and instead suggest a moderate relationship between parent personality and offspring. There are also some studies which put the parent-adolescent relationship to its extreme position, suggesting that parents have little or even no impact on adolescent personality (Harris 1995, 1998). Hence, the general evidence suggests that the personality relationship between parent and offspring could be varied from high to no similarity.

Several studies have shown that personality traits have a heritable component (Van Oortmerssen et al. 1981; Benus et al.1991; Koolhaas et al. 1999; Dingemanse et al. 2002; Veenema et al. 2003; van Oers et al. 2011) and genetics should play important role. However, despite considerable efforts since a long time, the genetic variants that influence personality are largely unknown and still there is no molecular mechanism which could explain this trait (Sanchez-Roige 2017; Lo et al. 2017).

A recent study of the role of personality in mate choice (Krebs 2018) showed that these two populations of Western house mice (CB and MC) show overall differences in different aspects of their personality such as activity, anxiety and curiosity. The results of this study also suggested individual's personality can significantly influence general reproductive behavior and mate choice decision in mice (Krebs 2018). But this study provided no further clues on the prevalence and mechanism of paternally determined mate choice.

3-Ultrasonic Vocalization

Mate preferences can be influenced by learning, and mate recognition may not be entirely genetic (Svensson et al. 2010). Although in mice, olfactory cues play a major role in their mate choice behavior, ultrasonic vocalization (USV) could also help them to recognize preferred mates (Nyby 1983; Holy & Guo 2005; Hammerschmidt et al. 2009; Musolf et al.

2010). House mice are social animals and they use USV also in different social contexts such as pup-mother interaction (Branchi et al. 1998) or adult communication (Panksepp et al. 2007). CB and MC house mouse populations show different patterns in emitted USV. In sexual interaction, MC mice show higher syllable rates than CB mice (von Merten 2014). However, the study showed also that females show more vocalization when they are among females, suggesting a role for social communication rather than mate choice (von Merten 2014). Thus, it is unlikely that the paternal mate choice preference is based on vocalization patterns from their fathers.

Genomic imprinting

Genomic imprinting refers to genes that are expressed from only one of the two parental alleles in a parent-of-origin-specific manner. Imprinted genes in mammals have arisen with the development of the placenta, possibly more than 125 million years ago (Barlow & Bartolomei 2014).

The distribution of imprinted genes in the mammalian genome is not random and they are mostly found to be in clusters. Mono-allelic expression of multiple genes within an imprinted cluster is coordinately regulated by imprinting control (IC) region. IC region silences expression of specific prenatal allele through different epigenetic mechanisms such as DNA methylation, histone methylation or antisense noncoding transcript (Suzuki et al. 2011; Barlow & Bartolomei 2014). Therefore these epigenetic processes allow imprinted genes to be expressed from only one of the two parental alleles. This epigenetic phenomenon is established during gametogenesis and maintained throughout development of an organism (Suzuki et al. 2011).

Genomic imprinting and brain development

The association of genomic imprinting to the brain development and function was first found in the context of the study of a human brain disorder, known as Prader–Willi syndrome (PWS). This was shown to be linked to paternally transmitted mutations on human Chromosome 15 (15q11-13) (orthologous region on Chromosome 7 of the mouse) (Nicholls et al. 1989).

PWS is a neurodevelopmental disorder which is much studied, but still not fully understood. PWS individuals show several developmental abnormalities in the frontal white matter, left dorsomedial thalamus, posterior limb of the internal capsule bilaterally, and the splenium of the corpus callosum (Yamada et al. 2006). Subsequently these abnormal developments in specific areas of the brain change significantly their behavior such as anxiety and intellectual ability (Jauregi et al. 2007). Therefore these symptoms from PWS patients indicated the important role of imprinted genes in brain development and function.

Eight years after the discovery of the PWS imprinted locus, Keverne et al. 1996 discovered that chimeric embryos that carried two maternal genomes (gynogenetic) showed abnormal brain development by having large brains. The gynogenetic cells were found mostly in the cortex, striatum and hippocampus (Keverne et al. 1996). Interestingly in opposition to what they observed for gynogenetic embryos, chimeric embryos with two paternal genomes (androgenetic) showed small brains and the androgenetic cells were found mostly in hypothalamic regions (Keverne et al. 1996). This fact that gynogenetic and androgenetic were found in different parts of the brain, reflected distinct roles for maternally and paternally inherited information in different brain regions.

Later, several studies discovered many other imprinted genes with widespread expression throughout the brain (Wilkinson et al. 2007; Gregg et al. 2010). By generating specific mutations on imprinted genes, researchers could indeed highlight the central role of imprinted genes in neurogenesis, neuronal migration, axonal and dendritic outgrowth (Schmidt-Edelkraut et al. 2014; Ferrón 2015; Bando 2014; Störchel et al. 2015) and therefore supported the crucial role for parent-of-origin information in brain development.

These observations all together also suggested the brain as a main target of genomic imprinting which can affect different parts of brain function and subsequently mammalian behavior (Perez et al. 2016).

Imprinted gene and mate choice preference in house mouse

Since the F1 hybrids in the study of Montero et al. (2013) shared the same autosomal genome combinations, but still showed paternal mate choice behavior, it suggested that genomically imprinted loci mediated this special mate choice preference and these inherited cues must

have diverged between the two populations to allow the population-specific decision (Montero et al. 2013).

Therefore a follow-up study was conducted to identify imprinted loci which have paternal expression and have highly diverged between these two Western mouse house populations. The study focused on the organs that are of special relevance for mouse mate choice behavior: the vomeronasal organ (VNO), because it senses signaling peptides, the hypothalamus, because it governs general behavior and the liver, since it synthesizes major urinary proteins involved in mate recognition (Lorenc et al. 2015).

To first identify candidate transcripts at a genome-wide scale, they used reciprocal crosses between *M. m. domesticus* and *M. m. musculus* inbred strains and then conducted RNA sequencing analysis of the respective tissues. The largest number of imprinted transcripts was found in the hypothalamus; fewer were found in the VNO, and the least were found in the liver. To assess molecular differentiation of imprinted transcripts in the wild-derived *M. m. domesticus* populations, they sequenced the RNA of the hypothalamus from individuals of these populations. This confirmed the presence of the above identified transcripts also in wild populations and allowed them to search for those that show a high genetic differentiation between these populations (Lorenc et al. 2015).

Their results identified two imprinted regions. The first region was *Ube3a–Snrpn* imprinted cluster on Chromosome 7 as a region that encompasses the largest number of paternal expression bias and also has been highly differentiated between German and French house mouse populations. As mentioned previously, this region has been implicated in Prader–Willi syndrome (PWS) in human and therefore it is known as PWS region in mammalian genomes (Lorenc et al. 2015). PWS patients show several abnormalities in their cognitive behaviors such as social communication, speech, anxiety, intellectual ability, and decision making (Jauregi et al. 2007; Woodcock et al. 2009). Several studies on mouse also have suggested an important role of this region in mouse behavior. Chromosomal deletion from *Ube3a* to *Gabrb3* changed numerous behaviors such as anxiety, activity, and ultrasonic vocalization in mice (Jiang et al. 2010). Paternal duplication of this region impaired brain development and altered mouse social interaction, vocalization, and anxiety (Nakatani et al. 2009; Takumi 2010; Urraca et al. 2013). Paternal deletion from *Snrpn* to *Ube3a* in the mouse caused hypotonia, growth deficiency with partial lethality compared to wildtype mice (Tsai et al.

1999). Therefore the PWS region is known to play important role in both human and mouse behavior.

Lorenc et al. (2015) also found the paternally expressed PEG13 transcript within the *Trappc9* gene region on Chromosome 15 as a second region which has been highly differentiated. Mutation of this locus have been already reported in several human brain disorders such as autism-spectrum disorder, Birk-Barel mental retardation syndrome and autosomal-recessive intellectual disability in human (Mochida et al. 2009; Mir et al. 2009; Kakar et al. 2012; Marangi et al. 2013). However the role of this locus in mouse brain development and behavioral phenotypes still remains unknown.

So Lorenc et al. (2015) suggested these two genomically imprinted regions as potential candidates for influencing the population-specific mate-choice; however the functional role of these regions during mouse evolution and their relation to mate choice preference remains unknown so far.

Aim of the project

The present project focuses on the *Ube3a–Snrpn* imprinted cluster on Chromosome 7 (the first chapter), as well as on the PEG13 transcript within the *Trappc9* gene region on Chromosome 15 (the second chapter). The aim was to elucidate how these two imprinted regions could be associated to mate choice decisions in mice and how they could influence their paternal mate choice preference.

This research thus addresses the general question how genetic mechanisms could regulate behavioral preference and identify the molecular mechanisms which lead to evolutionary divergence between populations.

Chapter1

Involvement of the Prader-Willi Syndrome (PWS) region genes in mouse behavior

Introduction

The Prader-Willi Syndrome (PWS) region includes several genes and transcripts, among them *Ube3a* and *Snrpn*. (Fig. 1). The evolutionary emergence of *Ube3a* can be traced to the genomes of basal metazoans (Marín 2010), and to the fungi such as *Mortierella verticillata* (Grau-Bové et al. 2013). In older evolutionary lineages, *Ube3a* is not an imprinted gene and shows bi-allelic expression. Imprinting of this gene can be found only after the emergence of the neighboring gene *Snrpn* and the imprinting center that lies upstream of *Snrpn* in the eutherians genomes (Sato 2017) (Fig. 1). As mentioned in the introduction, both *Ube3a* and *Snrpn* have been functionally studied and were suggested to play a role in mouse behavior (Tsai et al. 199; Jiang et al. 2010).

The imprinted PWS region expresses also two small nucleolar RNA (snoRNA) gene families which are organized in large, tandemly repeated clusters known as SNORD115 and SNORD116 (Cavaillé 2017) (Fig. 1). The origin of these SNORDs seems to be related to the emergence of *Snrpn* in the eutherian genomes (Sato 2017) (Fig. 2). The expression of both SNORD115 and SNORD116 is strictly brain-specific and restricted to the alleles on the paternal chromosome (Bortolin-Cavaillé et al. 2012).

Small nucleolar RNAs (snoRNAs) are part of a large group of small, metabolically stable RNAs which regulate post-transcriptional modification of their target genes (Lui et al. 2013). Imprinted snoRNAs of the PWS region have been found in all eutherian genomes tested. Unlike other snoRNAs, which show broad expression across different tissues, the PWS associated imprinted snoRNAs are mainly expressed in the brain and were suggested to play important roles in brain function (Zhang et al. 2014).

Based on common sequence motifs involved in the assembly of their target genes, snoRNAs are categorized into two major classes, the C/D box and H/ACA box groups (Hirose et al.2001; Brown et al. 2003). These snoRNAs are called ‘guide’ snoRNAs if they account for the 2′-O-ribose methylation and pseudouridylation modifications of rRNA or snRNA and called ‘orphan’ snoRNAs if they do not target to any rRNA or snRNA. Functionally, these orphan snoRNAs have been implicated in the formation of miRNAs, regulation of chromatin structure, alternative splicing and modulation of cell survival under oxidative stress (Falaleeva & Stamm 2013; Dupuis-Sandoval et al. 2015). Hence, there is a broad potential variety of the biological functions of SNORDs (Falaleeva & Stamm 2013).

SNORD115 and SNORD116 are classified into the C/D box class of snoRNAs. C/D box snoRNAs contain two short conserved sequence motifs, C (RUGAUGA) and D (CUGA), located near the 5' and 3' ends of the snoRNA, respectively (Fig. 3). Short regions (~ 5 nucleotides) located upstream of the C box and downstream of the D box are usually base complementary and form a stem-box structure, which brings the C and D box motifs into close proximity (Fig. 3). This stem-box structure has been shown to be essential for correct snoRNA synthesis and nucleolar localization (Samarsky 1998). Many C/D box snoRNAs also contain an additional less-well-conserved copy of the C and D motifs (referred to as C' and D') located in the central portion of the snoRNA molecule (Samarsky 1998) (Fig. 3).

The imprinted SNORD115 and SNORD116 have evolved very fast during mammalian evolution and show species-specific copy number (Zhang et al. 2014). Accordingly, they are good candidates for a possible role of SNORD115 and SNORD116 families in the fast divergence of mate choice preference and behavior in the house mouse.

Figure 1: Schematic view of the imprinted Prader-Willi Syndrome (PWS) locus. Blue color represents paternally expressed genes, red color shows the maternally expressed gene and black represents the regulatory imprinting center (IC). The arrows indicate the direction of transcription.

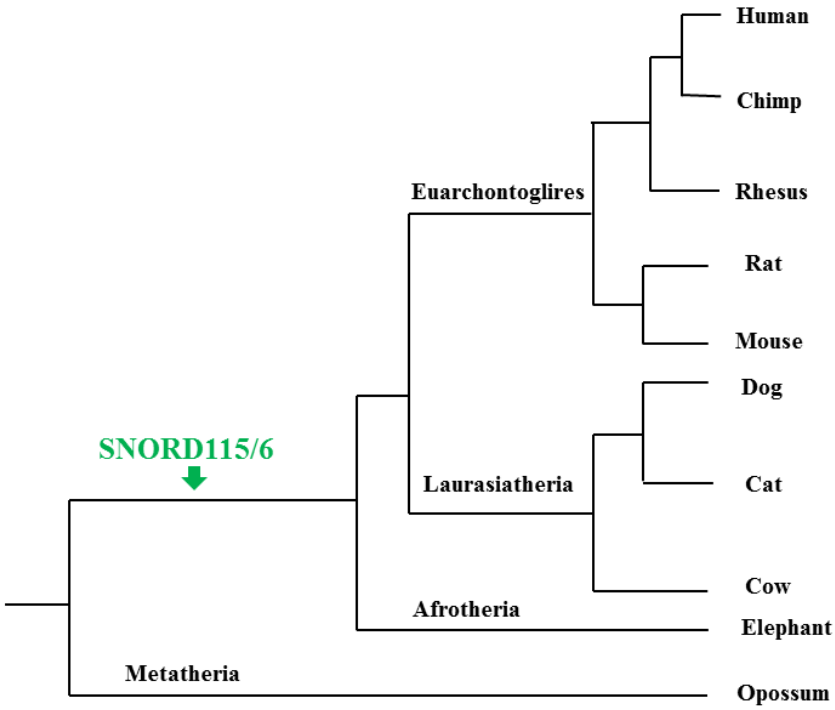


Figure 2: Timing of the emergence of SNORD115/116 during mammalian evolution. The green arrowhead represents the acquisition of SNORDs and SNRPN according to Sato 2017.

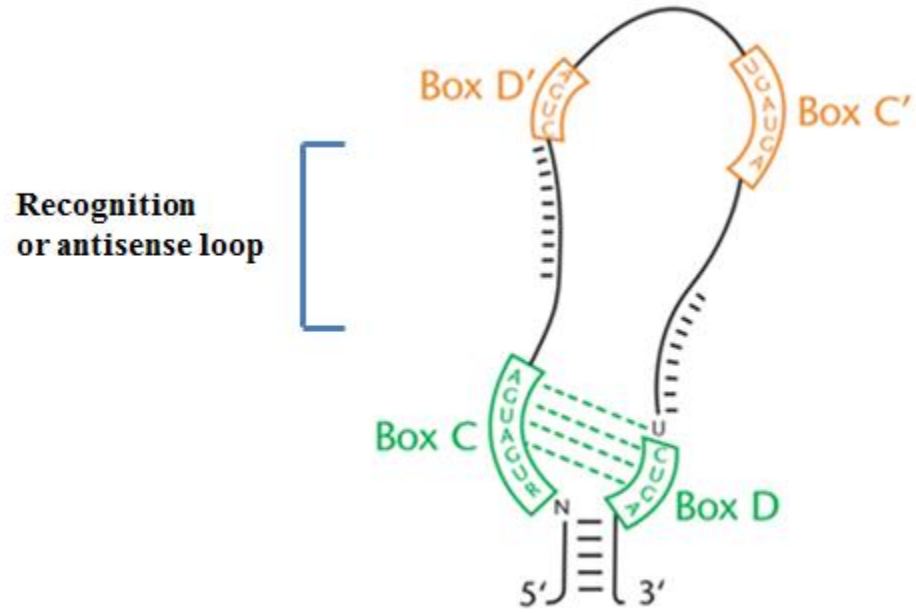


Figure 3: C/D box snoRNA structure. C/D box snoRNAs contain two short conserved sequence motifs, C (RUGAUGA) and D (CUGA), located near the 5' and 3' ends of the snoRNA, respectively. Short regions (~ 5 nucleotides) located upstream of the C box and downstream of the D box are usually base complementary and form a stem-box structure, which brings the C and D box motifs into close proximity. This stem-box structure has been shown to be essential for correct snoRNA synthesis and nucleolar localization (Samarsky 1998). Many C/D box snoRNAs also contain an additional less-well-conserved copy of the C and D motifs (referred to as C' and D') located in the central portion of the snoRNA molecule. The figure was modified from Falaleeva M., et al 2016.

Analysis of SNORD Copy Number Variation

Copy number variation (CNVs) is a major contributor to genetic variation with important consequences for different aspects of evolutionary processes including adaptation, genetic load, the evolution of novel genes and biological innovations, genome evolution and speciation (Bergthorsson et al. 2016). Hence, a comparative study of CNVs in populations could yield insights into such processes. However, it has long been hampered by the limited availability of suitable technology to measure broad ranges of copy number variation in a reproducible way. Droplet digital PCR (ddPCR) has now emerged as a powerful method to measure gene copy numbers across a broad range of copies of a given genomic region (Härmälä et al. 2017). This method is much faster and less error prone than real-time qPCR, because quantification is obtained without the need for standardization assays (Mazaika et al.

2014). Therefore here I applied ddPCR to determine SNORD115 and 116 copy number variation between populations and species.

My first test was a comparison between animals from the German (CB) and French (MC) house mouse populations that are kept under outbreeding conditions in the mouse house in Plön. I included 23 randomly chosen animals (all males) from each population. DNA was isolated from a small part of ear and the ddPCR was run and analyzed as described in the methods section. Primers were designed based on regions in SNOR115/116 which are conserved between these mouse populations. To find these conserved regions, I generated a small RNA library and sequenced it (see Methods). Reads were mapped to the mm10 mouse reference genome (Genome Sequencing Consortium 2002) using NextGenMap based on the best match option (Sedlazeck et al. 2013). The PWS region on Chromosome 7 was extracted from the bam file. To extract SNORD115/116 copies from the PWS region, CDseeker was run. CDseeker is part of snoSeekerNGS package which can find CD box snoRNA from next generation sequencing data (Zheng et al. 2016). The extracted data from this part was exported to Geneious 9.0.5 (Kearse et al. 2012) for further analysis of the conserved parts.

The results of the droplet PCR studies on copy number variation showed that although there is a large variation of copy numbers among individuals, with an overlap in the spread of SNORD copy number between the CB and MC mouse populations, there is also a significant difference in their averages (Fig. 4a-b).

Next I tested whether this copy number variation leads also to corresponding differences in expression. A positive correlation of gene copy number and its expression level is not always the case. Multiple copies of a gene can interact to co-suppress each other (Flavell et al. 1994). Even genes that are duplicated could be expressed at equal levels when compared to a single one, due to a feedback regulation (Adams et al. 2005).

To check how the variation in SNORD copy number (CN) affects their expression levels, SNORDs transcript analysis was performed. To this end, total RNA which is enriched in small RNA was extracted and ddPCR was used to quantify SNORD115/116 expression (see Methods).

Figure 5 shows, in both populations for both SNORD families that there is indeed a positive correlation between SNORD copy number and their respective expression (Fig. 5),

suggesting that the measurement of genomic copy number should indeed reflect the relative activity of the genes.

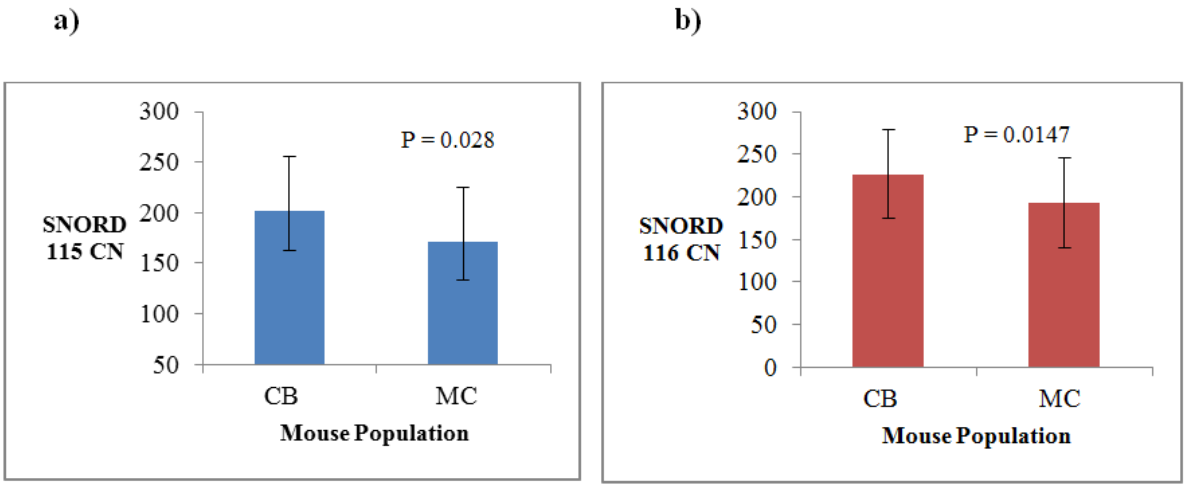


Figure 4: SNORD CNV between German (CB) and French (MC) house mouse populations. a) SNORD115 Copy Number (CN) b) SNORD116 Copy Number (CN).t-test was used to determine whether there is a significant difference between the means of SNORDs CN of German (N=23) and French (N=23) *M. m. domesticus* populations. Error bar represents the variability of SNORDs CN within each mouse population.

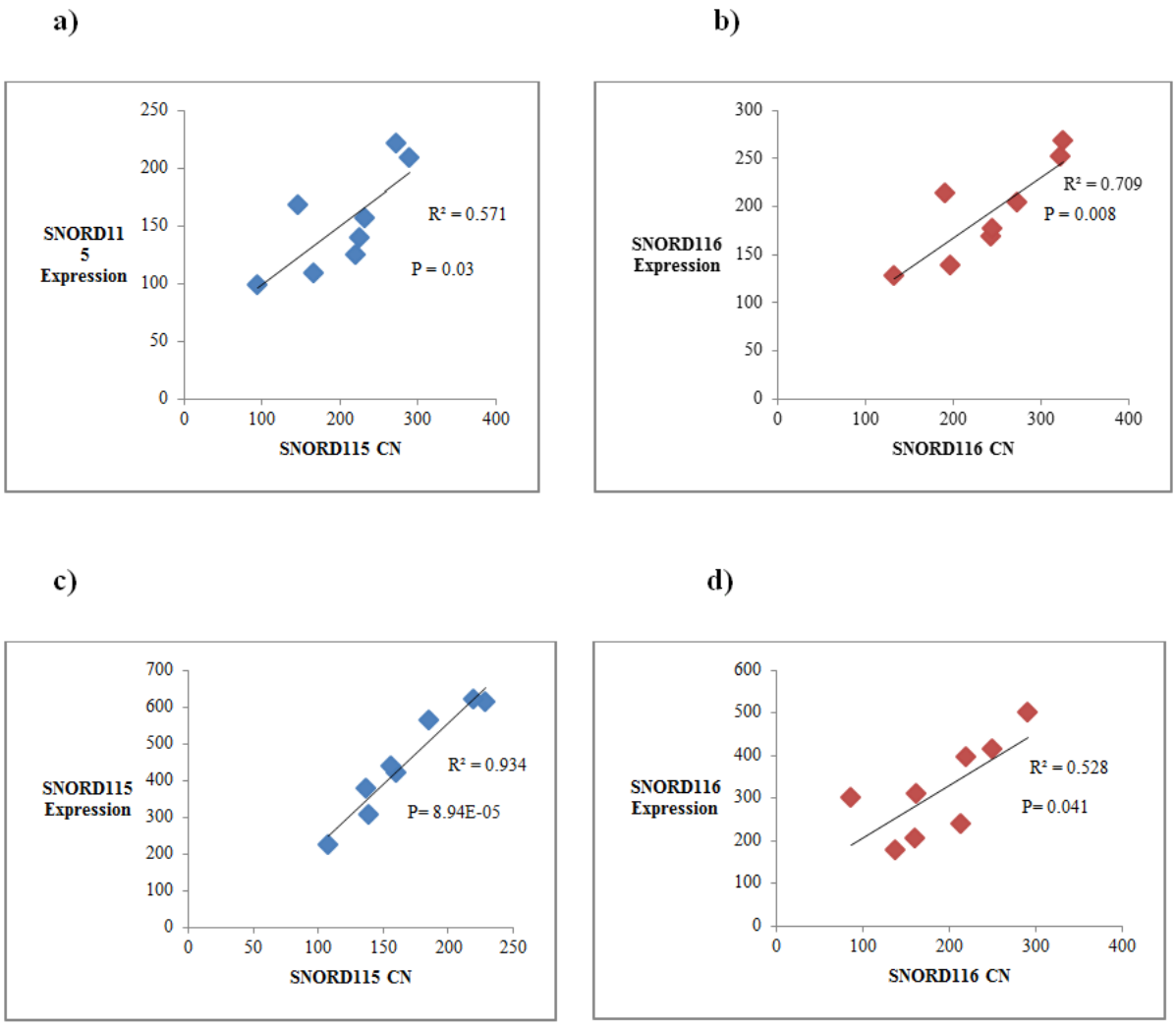


Figure 5: Positive correlation between SNORD genomic CN and their expression. Correlation between SNORD115 (a, c) and SNORD116 (b, d) CN and their relative expression in the CB population (a,b) and the MC population (c, d). Regression analysis has been done by using lm function in R.

SNORD115 and 116 target genes

SNORD115 lacks any complementarity to an rRNA, snRNA, or tRNA (Cavaillé 2017). However the antisense sequence element of SNORD115 exhibits complementarity to the alternatively spliced exon Vb of the serotonin receptor 5-HT2C, which is a seven-transmembrane receptor located on the X Chromosome (Kishore 2006) (Fig. 6 a). Vb encodes the second intracellular loop of the receptor, which is crucial for G protein binding. Skipping of exon Vb causes a frame shift, resulting in a receptor that is truncated after the third transmembrane domain (Wang 2000). At least five sites (A to E) in exon Vb can be edited from A to I. Exon Vb editing promotes its inclusion and there is a positive correlation between SNORD115 expression and exon Vb usage (Kishore 2006).

The 5-HT2C receptor is one of the many binding sites for serotonin. Activation of this receptor by serotonin inhibits dopamine and norepinephrine release in certain areas of the brain and significantly regulates mood, anxiety, feeding and reproductive behavior. *5-Ht2c* knockout mice exhibit consistent and significant reductions in anxiety-like behavior (Alex 2005) and in parallel overexpression of 5-HT2C receptors in forebrain lead to elevated anxiety in mice (Kimura 2009).

SNORD116 also does not show any complementarity sequence to an rRNA, snRNA, or tRNA (Cavaillé 2017). Instead SNORD116 shows a complementary sequence to exon X of ANKRD11-202 in human and mouse (Fig. 6 b) (Bazeley et al. 2008). However, this is only a computational prediction, but it has not been experimentally studied so far.

Ankrd11 is a chromatin regulator implicated in neural development and autism spectrum disorder (ASD) (Marshall et al. 2008; Sirmaci et al. 2011; Lo-Castro et al. 2013). Cell culture studies indicate that ANKRD11 is a large nuclear protein that regulates transcription, potentially by binding chromatin modifying enzymes like histone deacetylases (HDACs) (Zhang et al. 2004; Zhang et al. 2007a, 2007b; Li et al. 2008; Neilsen et al. 2008). Modification of chromatin structure by histone acetylation is essential for nervous system development and function and plays an important role in neural precursors (Lilja et al. 2013; Rudenko & Tsai 2014; Castelo-Branco et al. 2014).

Ankrd11 mutations in mice disrupt neural development and cause abnormal anxiety like behavior (Gallagher et al. 2015). Interestingly, mice with a SNORD116 deletion showed also abnormal anxiety like behavior, but by an unknown mechanism (Ding et al. 2008).

To check how SNORD CNVs affect their target exons, expression analysis on both SNORD families and their targets was performed. Expression analysis of SNORD was done as described above. To analyze expression of their target gene, I used the same RNA samples from above, because the RNA samples were total RNA with an enrichment of small RNA. Therefore each RNA sample has all different kinds of RNA. However, for this analysis cDNA was synthesized by a different kit, MMLV High Performance Reverse Transcriptase kit, which has been designed to synthesize cDNA from polyadenylated mRNA. To quantify expression of their target genes, primers were designed specifically for their target exons (Fig. 6) and expression level was analysed by using ddPCR.

I found that, both SNORD115 and 116 CN show indeed a positive correlation to the expression of their targets, i.e. higher expression of these SNORDs leads to higher usage of their target exons as well. This pattern is significant in both, the CB (Fig.7 a-b) and the MC (Fig.7 c-d) mouse populations.

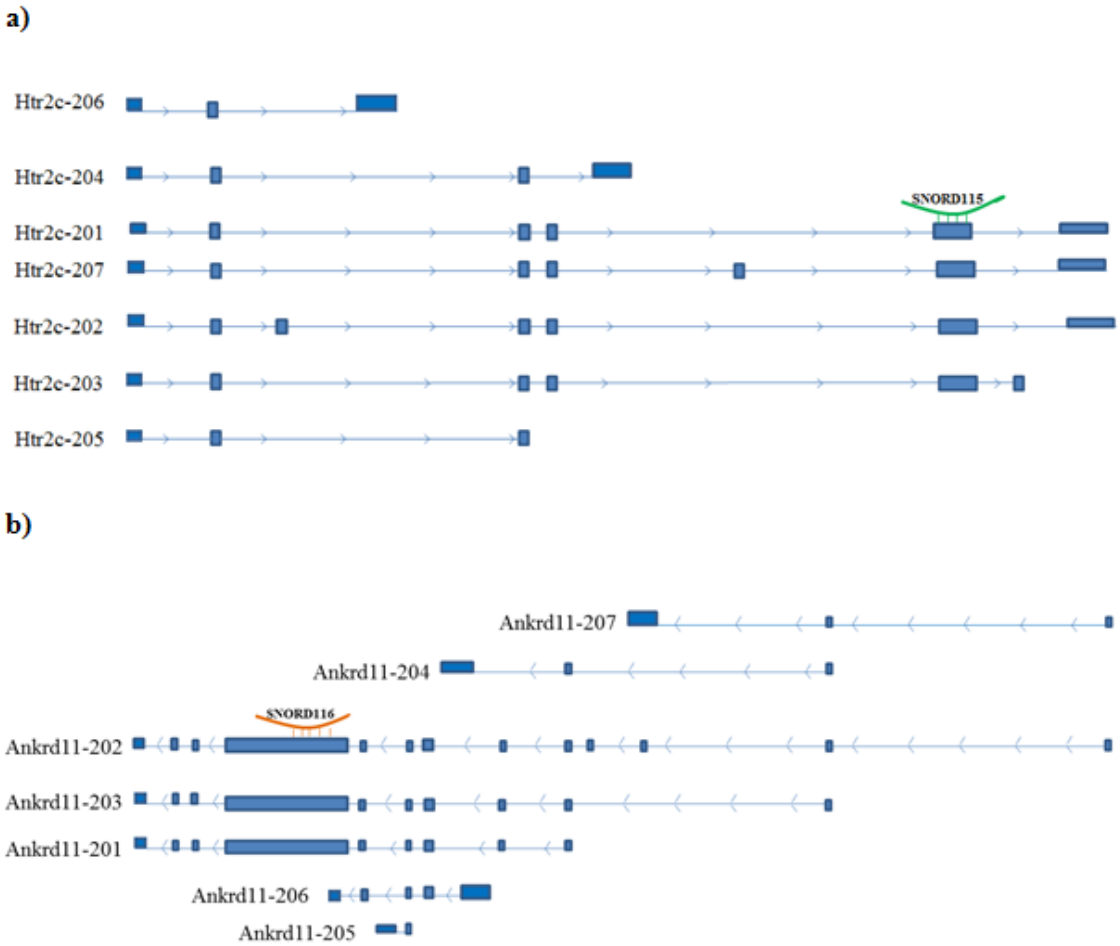


Figure 6: Schematic view of SNORD115/116 and their exons targets. a) 7 splice variants of serotonin receptor 5-HT2C. The antisense sequence element SNORD115 exhibits complementarity to the alternatively spliced exon Vb of the 5-HT2C. b) 7 splice variants of ANKRD11 that SNORD116 shows a complementary sequence to exon X of ANKRD11-202.

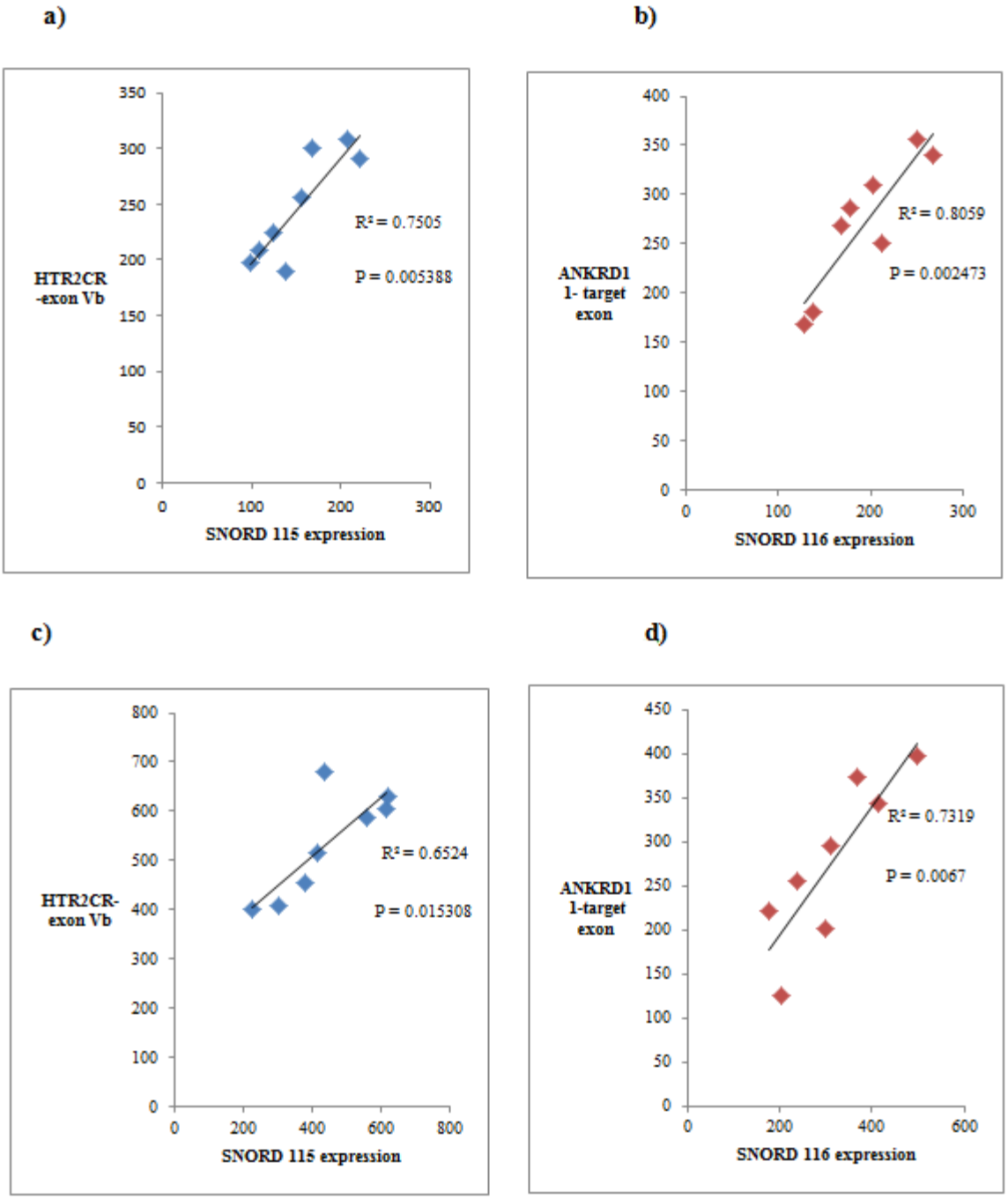


Figure 7: Correlations between SNORD expression and their target genes. Correlation between SNORD115 (a, c) and SNORD116 (b, d) expression and their target exons in the CB population (N=8) (a,b) and the MC population (N=8) (c, d). Regression analysis has been done by using lm function in R.

Link between SNORD copy number and personality

Anxiety is one of the most fundamental emotions required to survive or to overcome potential threatening stimuli (Kotov et al. 2010). A link between personality and level of anxiety has been hypothesized since the time of the ancient Greeks (Hohoff 2009). Neuroticism which reflects mood, feeling and anxiety is one of five big personality traits in humans (Thompson 2008). In mice, anxiety is one of the repeatable behaviors which is considered as one of the mouse personality traits (Lewejohann et al. 2011; Krebs 2018).

So, according to my results from the previous part, namely that SNORD115 and 116 target the genes which are directly or indirectly involved in regulation of mood, feeling and anxiety and also the known link between personality and anxiety, I tested whether there is any correlation between SNORD copy number and mouse individual's personality.

For these experiments, I used a set of personality tests that were developed by Rebecca Krebs in a parallel thesis project at the institute (Krebs 2018). These personality tests (Open Field Test, Dark/Light Box and Elevated Plus Maze) were carried out on 23 individuals (all male) from each population every 4 weeks for three times. Measurements taken for these tests were tested for repeatability over the course of the experiment using intra-class correlation coefficients. Those found to be repeatable were clustered to form the main category of behaviors. These behavioral clusters were used in a principle component analyses, and the first principle component was used to combine measurements into one behavioral score for anxiety-like behaviors (see Methods section). The procedure was validated in the thesis of Rebecca Krebs.

To check a possible relationship between SNORD copy number and personality, a correlation analysis for 23 randomly picked animals (all males) from each population was performed. The results showed a significant positive correlation between behavioral score and SNORD copy number, SNORD expression and also expression of their target genes in both the CB (Fig. 9a-f) and the MC (Fig.10g-l) mouse populations.

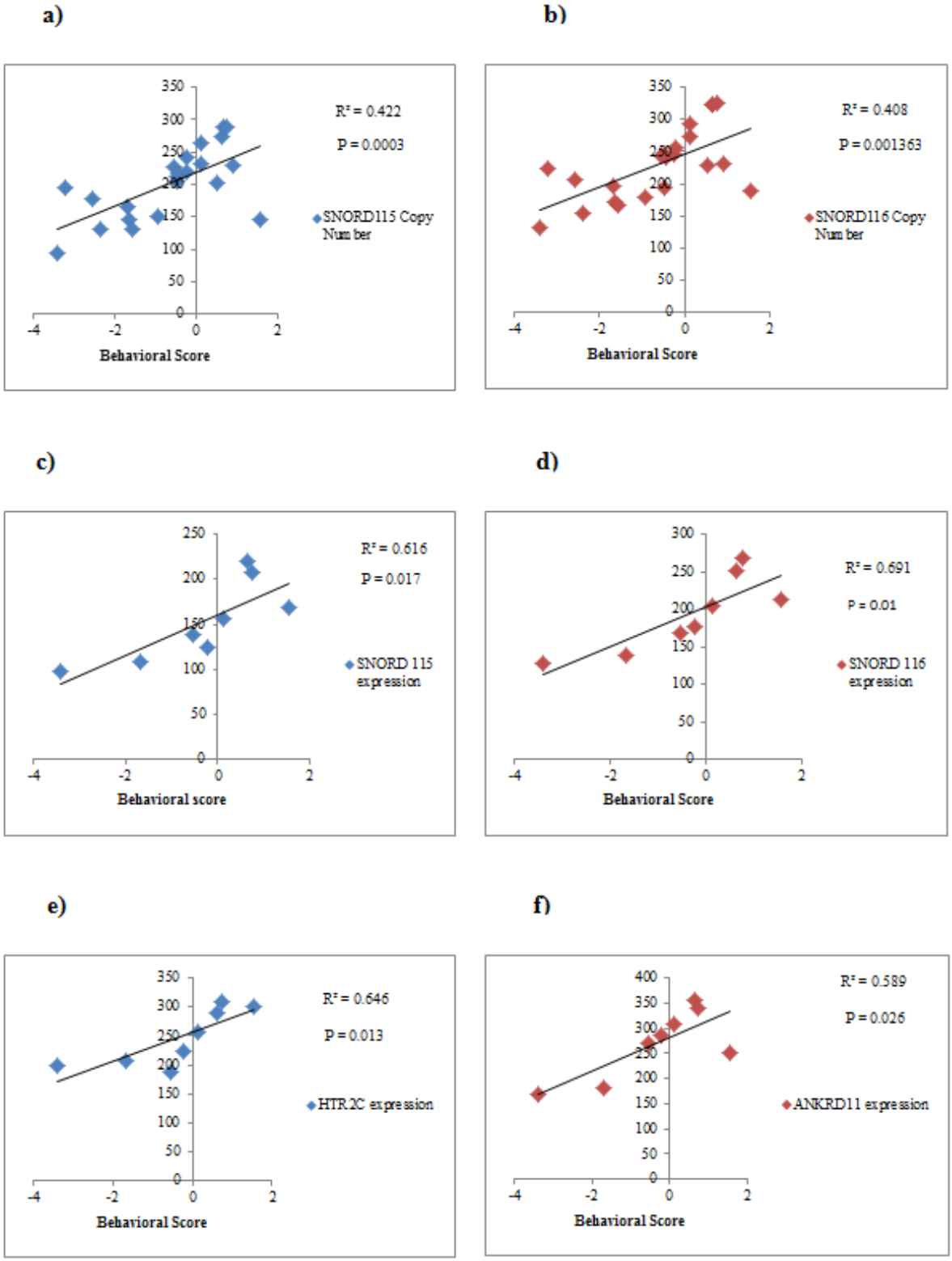


Figure 9: Correlations between SNORD measures and Behavioral Score in the CB population. a) Correlation between SNORD115 CN and Behavioral Score (N=23). b) Correlation between SNORD116 CN and Behavioral Score (N=23). c) Correlation between SNORD115 expression and Behavioral Score (N=8). d) Correlation between SNORD116 expression and Behavioral Score (N=8). e) Correlation between HTR2C expression and Behavioral Score (N=8). f) Correlation between ANKRD11 expression and Behavioral Score.

For SNORDS CN correlation analysis (a, b) 23 individuals were used but expression analysis was performed on 8 individuals (c-f) by using ddPCR. The y-axis in c-f shows the absolute concertation which has been calculated by ddPCR after normalization for each individual as expression value. β-catenin was used to normalize the data from HTR2C and ANKRD11 expression and SNORD66 was used as reference to normalize the data from SNORD115/6 expression. Regression analysis has been done by using lm function in R.

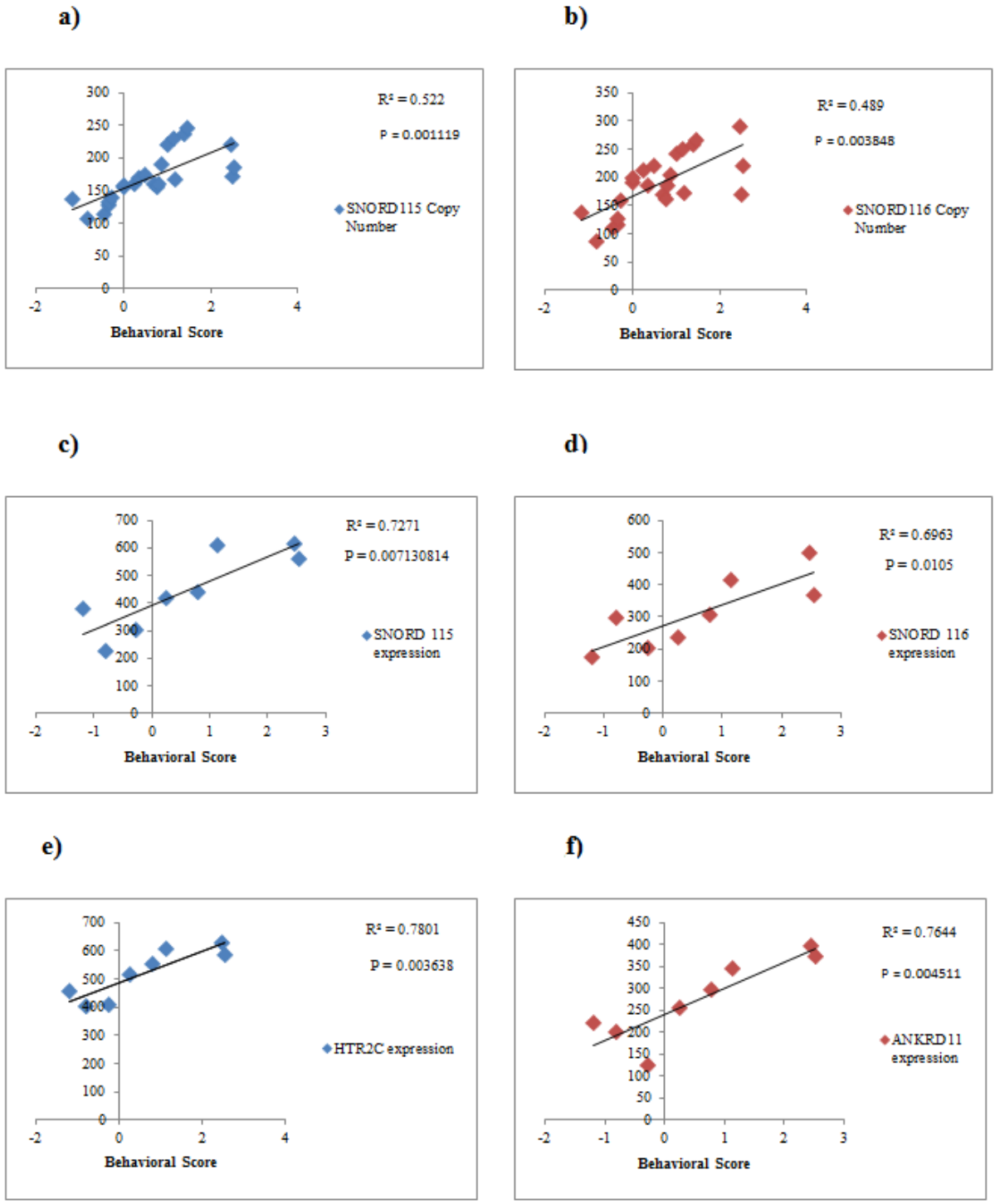


Figure 10: Correlations between SNORD measures and Behavioral Score in the MC population. a). Correlation between SNORD115 CN and Behavioral Score (N=23). b) Correlation between SNORD116 CN and Behavioral Score (N=23). c) Correlation between SNORD115 expression and Behavioral Score (N=8). d) Correlation between SNORD116 expression and Behavioral Score (N=8). e) Correlation between HTR2C expression and Behavioral Score (N=8). f) Correlation between ANKRD11 expression and Behavioral Score. For SNORDS CN correlation analysis (a, b) 23 individuals were used but expression analysis was performed on 8 individuals (c-f) by using ddPCR. The y-axis in c-f shows the absolute concertation which has been calculated by ddPCR after normalization for each individual as expression value. β-catenin was used to normalize the data

from HTR2C and ANKRD11 expression and SNORD66 was used as reference to normalize the data from SNORD115/6 expression. Regression analysis has been done by using lm function in R.

To test how specific this mechanism works, I analyzed the expression of other transcripts (splice variants) of HT2CR and ANKRD11. For this analysis, I used the same cDNA which has been synthesized in previous part. Both *Ht2c* and *Ankrd11* have 7 splice variants which are shown in Figure 11. HT2CR-204 and HT2CR-206 were tested as non-specific targets for SNORD115 and ANKRD11-204 and ANKRD11-207 as non-specific targets for SNORD116 (Fig11). Primers were designed specifically for these transcripts. In Figure 11 the red mutual arrows show exons which were targeted with ddPCR primers as non-specific target. Both SNORD115 and 116 did not show any significant correlation to their non-specific targets in both the CB (Fig. 12) and the MC (Fig. 13) mouse populations. Correspondingly, none of them also showed a significant correlation to the behavioral scores in both the CB (Fig. 14) and the MC (Fig. 15) mouse populations.

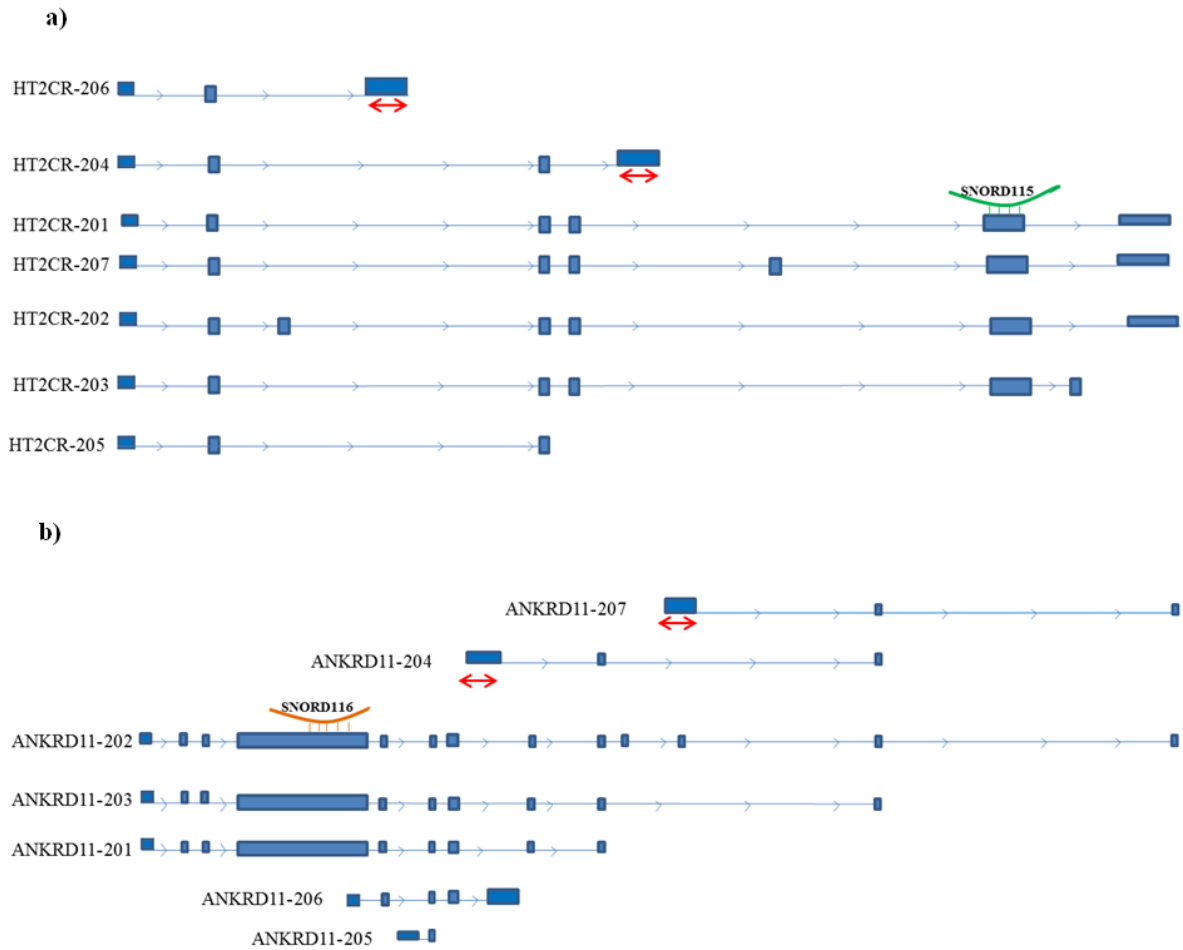


Figure 11: SNORD115 and 116 non-target transcripts. a) 7 splice variants of serotonin receptor 5-HT2C. b) 7 splice variants of ANRKRD11. Red mutual arrows show exons which were targeted with ddPCR primers as non-specific targets.

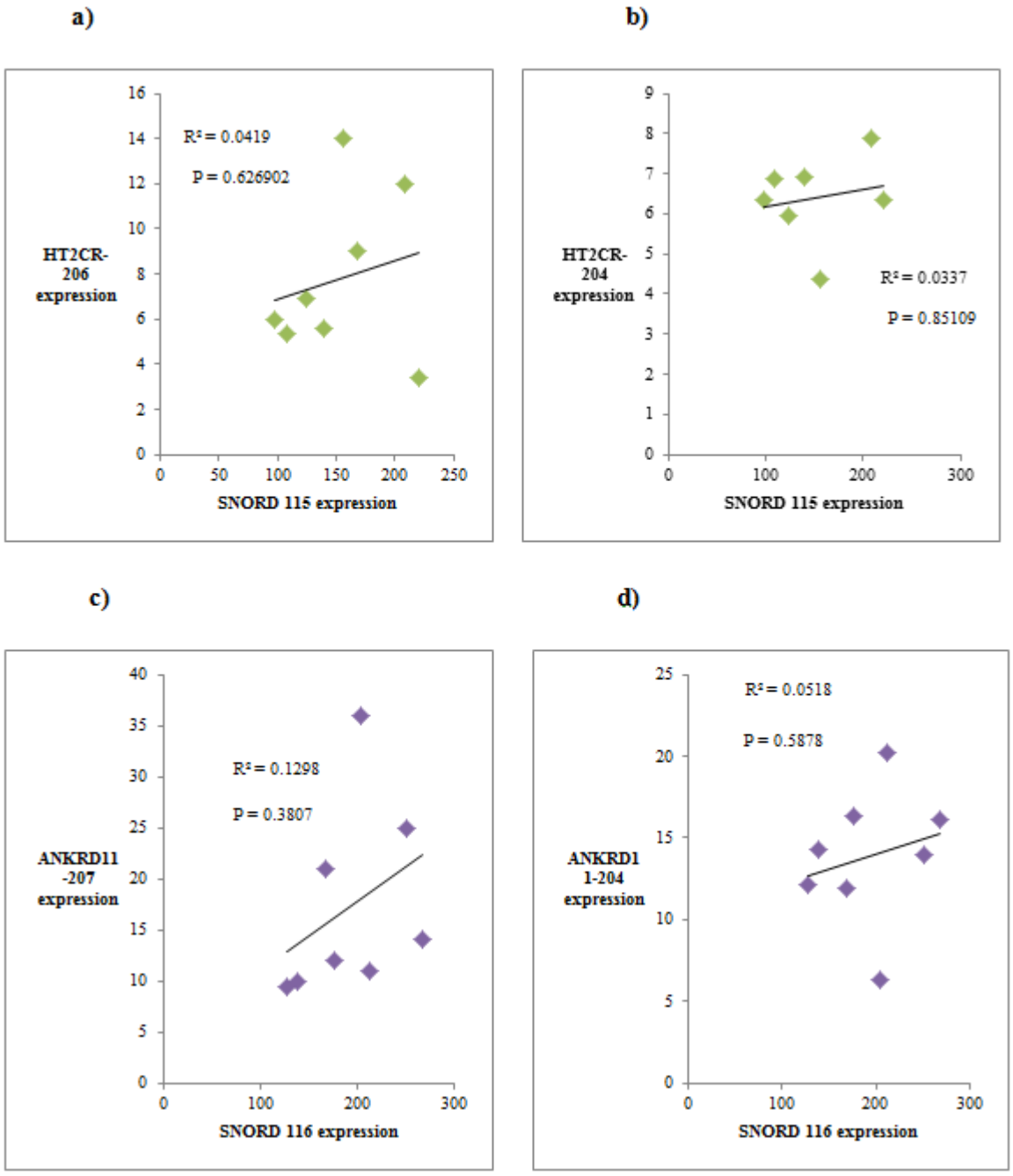


Figure 12: Correlation analysis between SNORD expression and their non-target exons in the CB house mouse population. (a-b) SNORD115 (c-d) SNORD116. None of the correlations is significant. This expression analysis was performed on 8 individuals by using ddPCR. The y-axis shows the absolute concertation which has been calculated by ddPCR after normalization (using β-catenin expression as housekeeping gene) for each individual as expression value. Regression analysis has been done by using lm function in R.

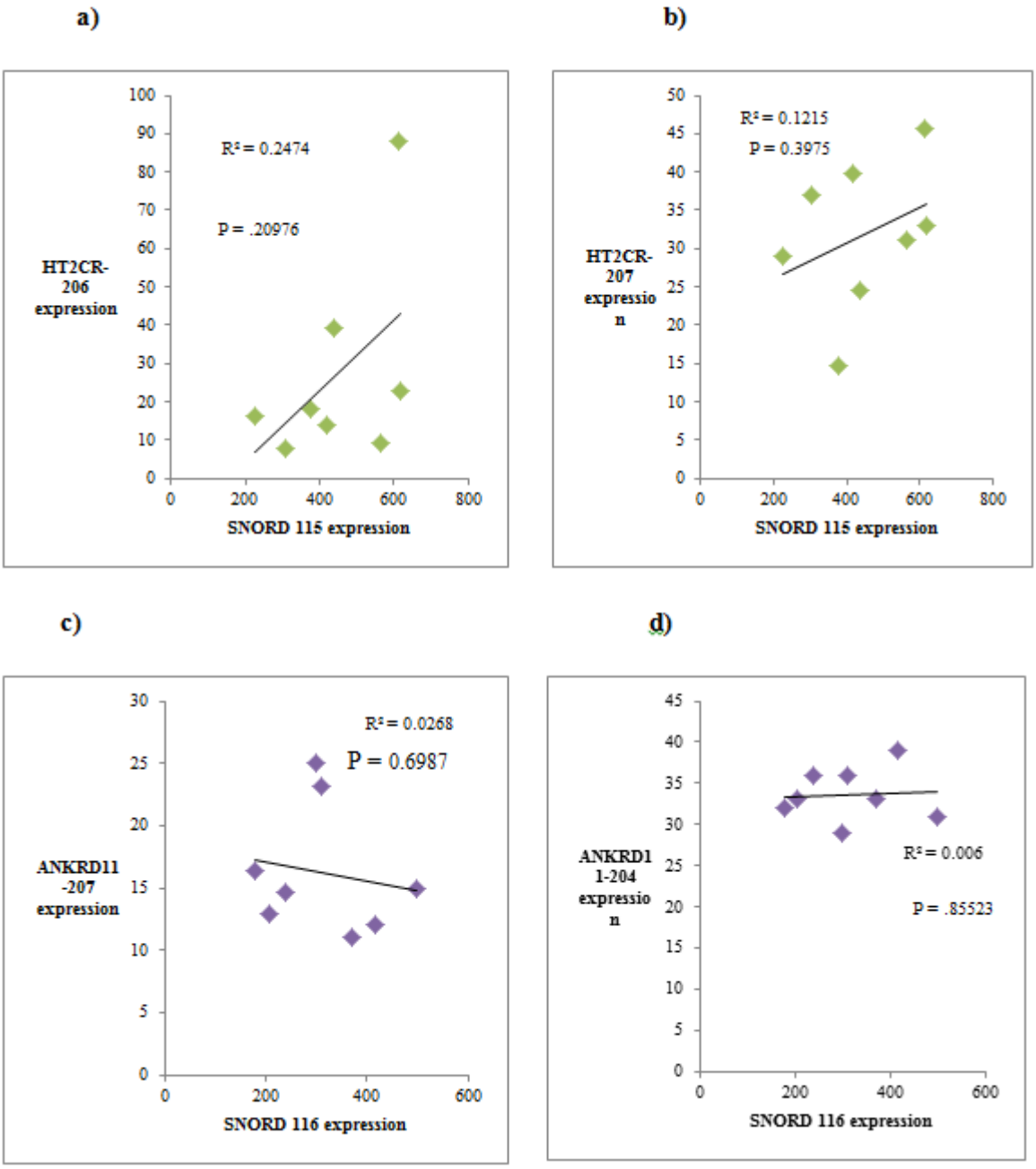


Figure 13: Correlation analysis between SNORD measures and their non-target exons in the MC house mouse population. (a-b) SNORD115 (c-d) SNORD116. None of the correlations is significant. (a-b) SNORD115 (c-d) SNORD116. None of the correlations is significant. This expression analysis was performed on 8 individuals by using ddPCR. The y-axis shows the absolute concertation which has been calculated by ddPCR after normalization (using β-catenin expression as housekeeping gene) for each individual as expression value. Regression analysis has been done by using lm function in R.

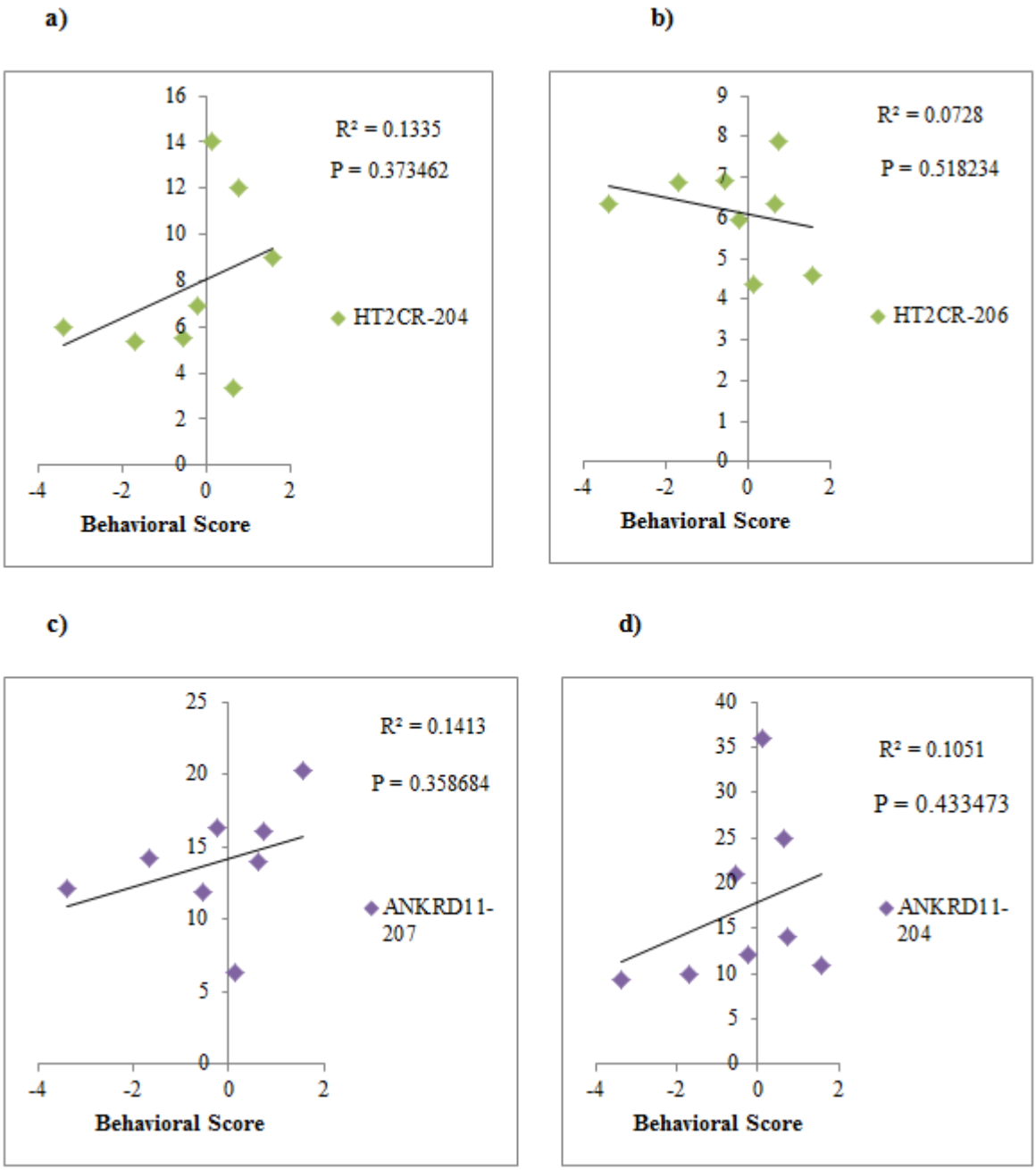


Figure 14: Correlation between SNORDs non-specific target and behavioral score in the CB popuation. a) Correlation between HTR2C-204 and Behavioral Score. b) Correlation between HTR2C-206 and Behavioral Score. c) Correlation between ANKRD11-207 and Behavioral Score. d) Correlation between ANKRD11-204 and Behavioral Score. None of the correlations is significant. This expression analysis was performed on 8 individuals by using ddPCR. The y-axis shows the absolute concertation which has been calculated by ddPCR after normalization (using β-catenin expression as housekeeping gene) for each individual as expression value. Regression analysis has been done by using lm function in R.

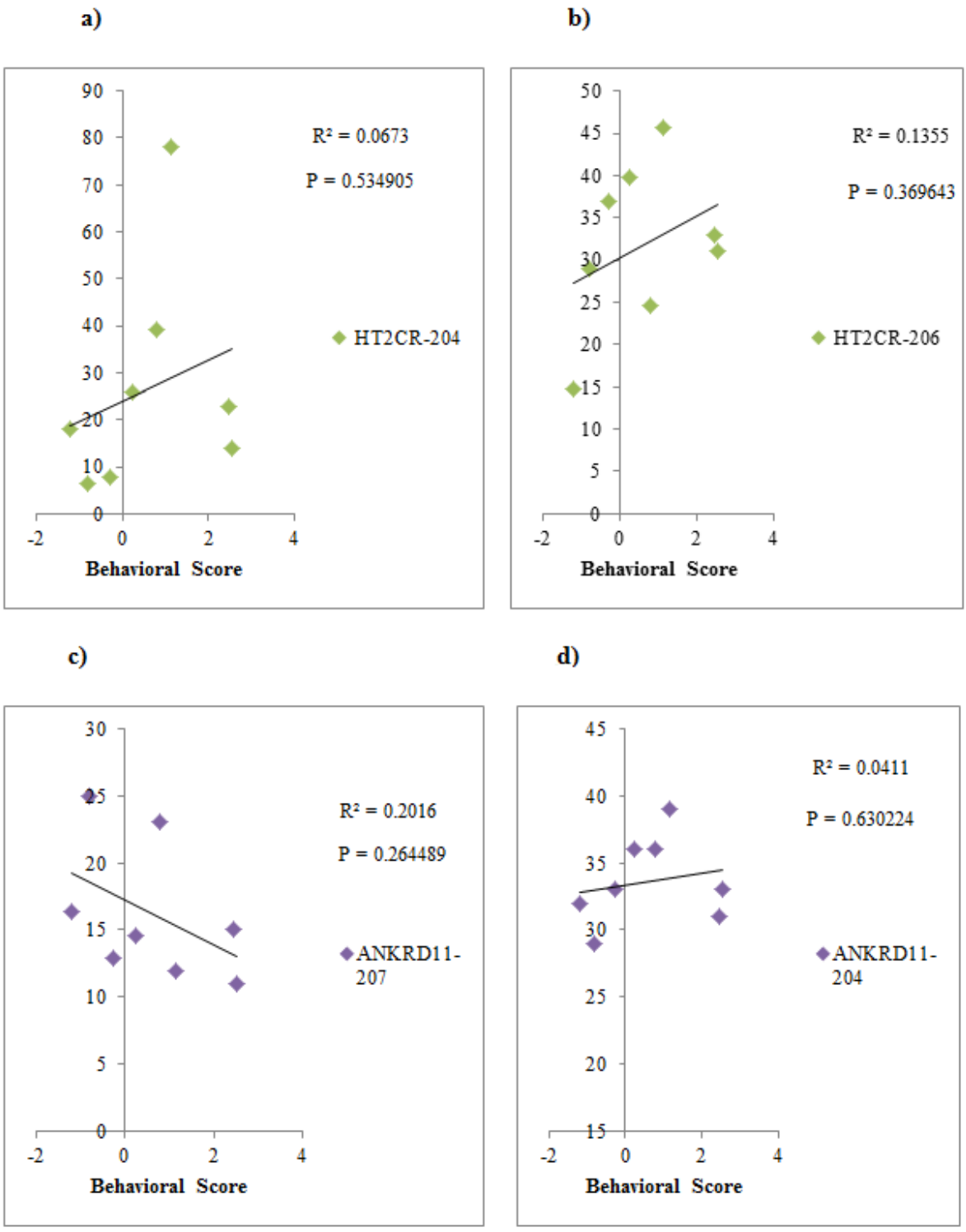


Figure 15: Correlation between SNORDs non-specific targets and behavioral score in the MC population. a) Correlation between HTR2C-204 and Behavioral Score. b) Correlation between HTR2C-206 and Behavioral Score. c) Correlation between ANKRD11-207 and Behavioral Score. d) Correlation between ANKRD11-204 and Behavioral Score. None of the correlations is significant. This expression analysis was performed on 8 individuals by using ddPCR. The y-axis shows the absolute concertation which has been calculated by ddPCR

after normalization (using β-catenin expression as housekeeping gene) for each individual as expression value. Regression analysis has been done by using lm function in R.

Correlation between SNORD115 and 116 copy numbers

Above I found quite similar correlation patterns for both SNORD families (Fig. 9 and 10). This suggests that also SNORD115 and SNORD116 copy numbers should correlate with each other, although they are located at two separate loci and they are two different gene families. By directly testing their correlation, I found that their copy numbers are indeed highly correlated in the two mouse populations (Fig. 16).

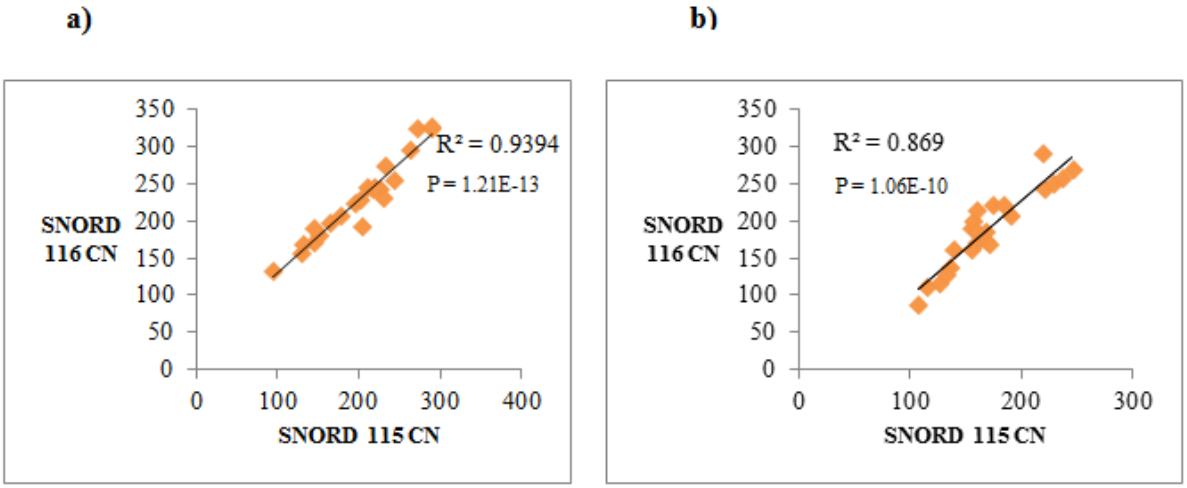


Figure 16: Correlation between SNORD115 and SNORD116 copy numbers. a) CB Mouse Population b) MC Mouse Population. Regression analysis has been done by using lm function in R.

To get more insights into whether this correlation is a general effect, a correlation analysis on other mouse populations, subspecies and species was conducted. I used an Iranian house mouse population (*Mus musculus domesticus*) which is separated from the German and French populations by about 10,000 years. Further, I used samples from the related subspecies *Mus musculus musculus* and *Mus musculus castaneus* which are separated since 0.3–0.5 million years, as well as from the species *Mus spretus*, *Mus spicilegus* and *Mus mattheyi.* These mouse populations, subspecies and species altogether span 6.6 million years of divergence (Neme et al. 2016).

10 random individuals (all male) from each population were chosen. DNA was isolated from small part of the ears and the ddPCR was run and analyzed by same primer pairs as described above (see the methods section for more details).

As Figure 17 shows, a similar trend is observed in most of the samples but the correlation is somewhat more variable among the samples (Fig. 17). The Iranian house mouse population, as well as *Mus spretus* and *Mus mattheyi* did not show significant correlations, but the trend is still visible (Fig. 17a, d, f). In contrast, the data from *Mus musculus musculus* and *Mus spicilegus* showed very strong correlations (Fig. 17b, e) and *Mus musculus castaneus* showed a moderate correlation (Fig. 17c). Interestingly, if data from individuals of all samples are pooled together, one can see a very high correlation between SNORD115 and SNORD116 copy numbers (Fig. 18). These observations suggest that there must be some mechanisms that balances the copy numbers in the long term, although the mechanisms at the level of the individuals remain unclear.

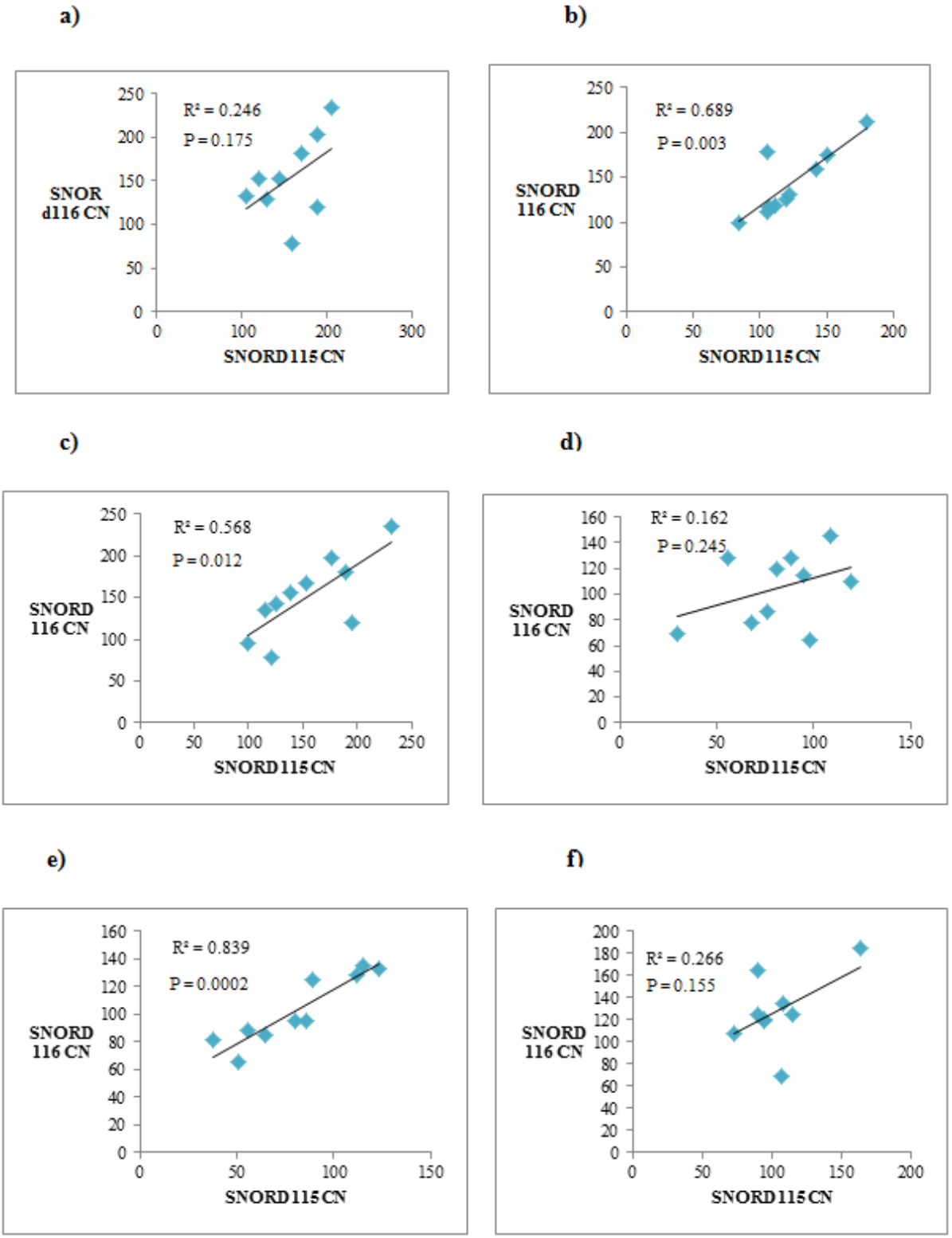


Figure 17: Correlation between SNORD115 and 116 copy numbers across different mouse populations, subspecies and species. a) Iranian house mouse population (*Mus musculus domesticus*). b) *Mus musculus musculus*. c) *Mus musculus castaneus*. d) *Mus spretus*. e) *Mus spicilegus*. f) *Mus mattheyi*. Regression analysis has been done by using lm function in R.

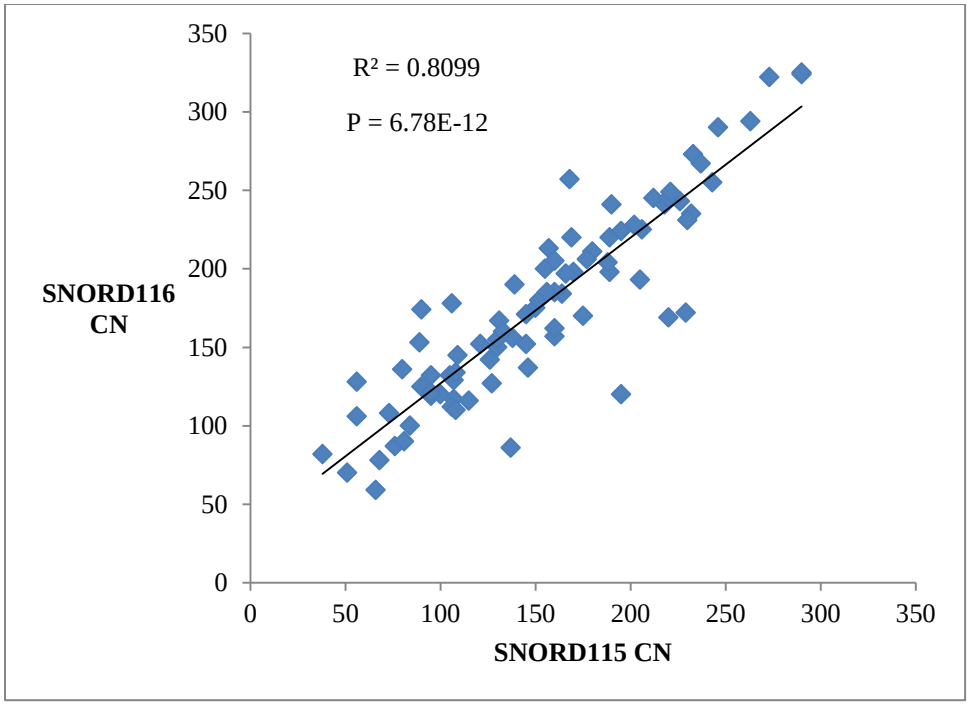


Figure 18: Correlation between SNORD115 and 116 copy numbers from pooled data of different mouse populations, subspecies and species. Regression analysis has been done by using lm function in R.

Inheritance of SNORD copy numbers

As I discussed in the general introduction, personality relationships between parent and offspring could vary from high to no similarity (Bell 1968; Patterson 1982; Sameroff 1983; Magnusson 1990; Caspi & Shiner 2006; South SC et al. 2008). Although there are several studies, consisting primarily of twin and adoption studies in humans, which towards a genetic component in individual's personality (Plomin et al. 2011), there is as yet no molecular mechanism which could explain why personality of offspring within a family can show such a high variation compared to their parents.

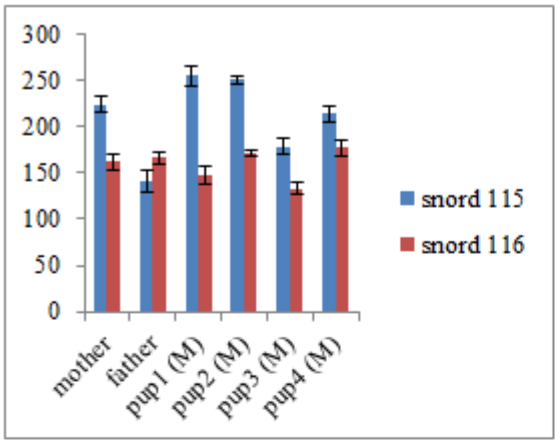
Given my finding of a high correlation between SNORD copy number and personality, I was interested in investigating the mode of inheritance of the SNORD copy number from parents to offspring. To this end, I set up 3 different types of mate pairs each from the MC population, (I): fathers with low and mothers with high copy number, (II): fathers with high and mothers with low copy number and (III): fathers and mothers with similar copy numbers.

The results from all the groups suggest a complex inheritance pattern for SNORD115 and 116 copy numbers (Fig. 19). There are cases where the offspring shows a similar copy number range as the parents (Fig. 19b, e, f) and cases where some offspring shows larger

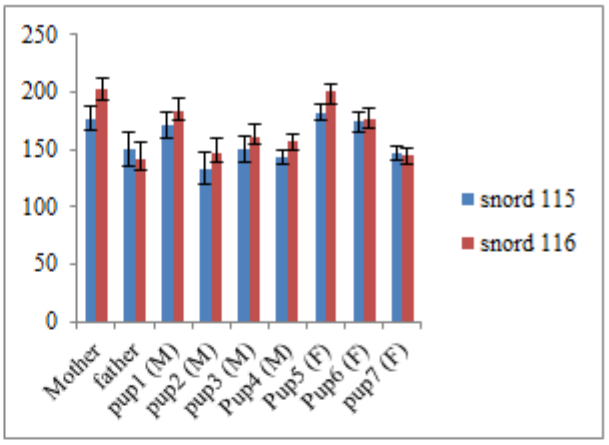
(Fig. 19a, c, d, g, h, i) or smaller CN (Fig. 19a, i) than either parent. Given that each parent should have only two alleles, one would expect that the offspring should show at most four different genotypes, i.e. four classes of allele numbers. However, even when taking the limitations of the exact measurements of copy numbers into account, there are appear to be more than four different allele class numbers among the offspring of the larger families. This would suggest that new allele versions are created by unequal cross-over in every generation. In any case, the experiments show clearly that there is a large variation among the offspring, which may be different from the variation of the parents.

To ensure that the variation of copy number with personality still holds in this experiment, I did behavioral tests on the offspring of a subset of these families. Figure 20 shows that the correlation between copy number and behavioral score holds for these as well. Hence, one can conclude that the behavioral variation that one sees among the offspring of families could be explained by the complex inheritance patterns of the copy numbers. Note that given that only the paternal alleles are expressed, while the average of both alleles is measured from the genomic DNA, one expects an even higher correlation with copy numbers.

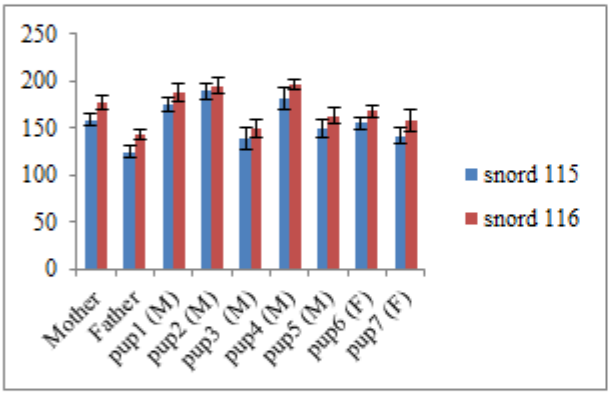
a)


300
250
200
150
100
50
0
mother
father
pup1 (M)
pup2 (M)
pup3 (M)
pup4 (M)
snord 115
snord 116

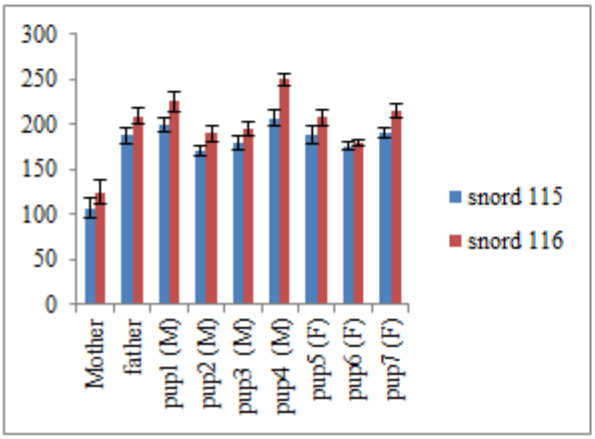
b)


250
200
150
100
50
0
Mother
father
pup1 (M)
pup2 (M)
pup3 (M)
Pup4 (M)
Pup5 (F)
Pup6 (F)
pup7 (F)
snord 115
snord 116

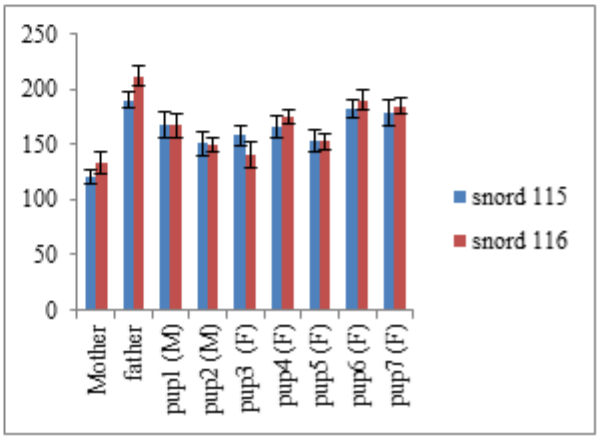
c)


250
200
150
100
50
0
Mother
Father
pup1 (M)
pup2 (M)
pup3 (M)
pup4 (M)
pup5 (M)
pup6 (F)
pup7 (F)
snord 115
snord 116

d)


300
250
200
150
100
50
0
Mother
father
pup1 (M)
pup2 (M)
pup3 (M)
pup4 (M)
pup5 (F)
pup6 (F)
pup7 (F)
snord 115
snord 116

e)


250
200
150
100
50
0
Mother
father
pup1 (M)
pup2 (M)
pup3 (F)
pup4 (F)
pup5 (F)
pup6 (F)
pup7 (F)
snord 115
snord 116

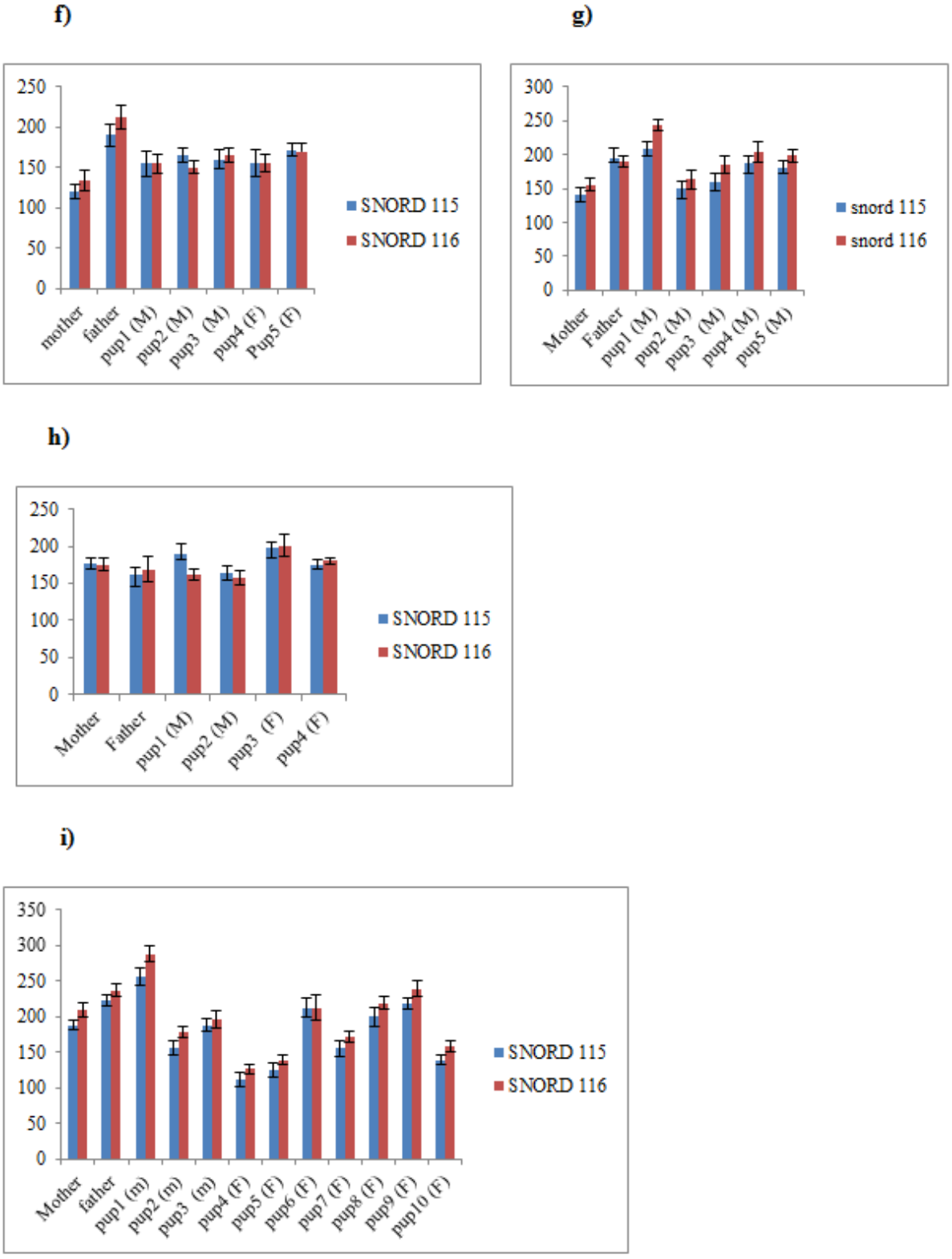
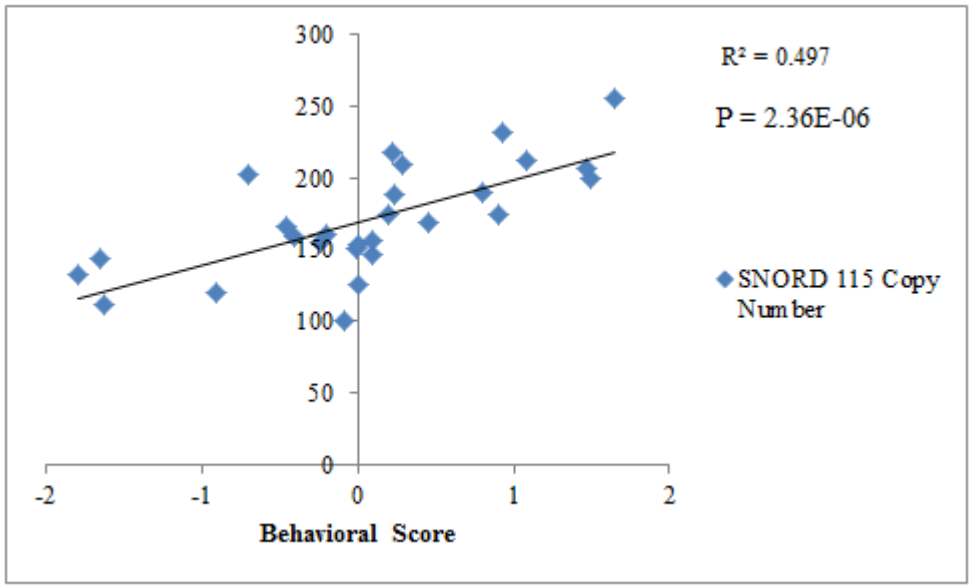


Figure 19: SNORD copy number inheritance patterns. Group I (a-c): father with low and mother with high copy number. Group II (d-g): father with high and mother with low copy number. Group III (h-i): father and mother with similar copy number. Error bar represents the variance measures that I had obtained from triplicate measurements for each individual.

a)



b)

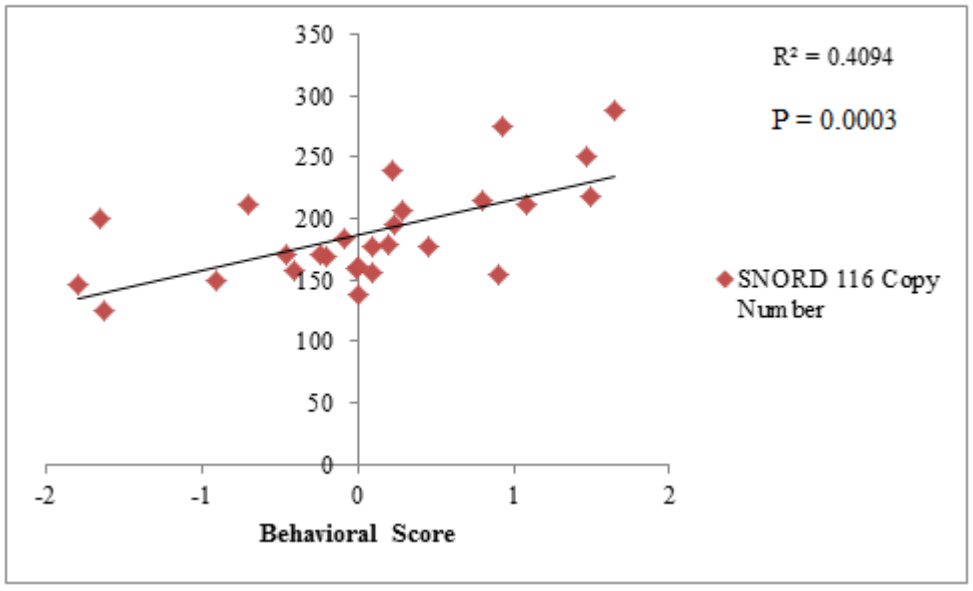


Figure 20: Correlations between SNORD CNs and Behavioral Score in the MC Population. a) Correlation between SNORD115 CN and behavioral Score. b) Correlation between SNORD116 CN and behavioral score. Regression analysis has been done by using lm function in R.

SNORD copy number differences across tissues

Generation of new copy number variants through unequal cross-over could occur in the germ line either during mitotic cycles of germ stem cells or subsequent meiosis, but may also occur post-zygotically through mitotic recombination either in early stages of embryo development (Lupski 2010; Boone 2011; Sun et al. 2013) or throughout the whole development. To check whether this SNORD CNV occurs as post-zygotic event in early stages of embryogenesis, SNORD copy number analysis across different tissues from a single animal was performed. If the variation is confined to early embryogenesis, one could expect to find mosaics of copy number between organs of the different germ layers. But as Figure 20 shows, there is no significant difference between SNORD copy numbers of different tissues. Hence, this result is compatible with either a high rate of change throughout development, thus that the cells in the tissues become on average similar to each other, or that changes occur only in the germline or during meiosis. A further explanation is that mutation are generally rare, i.e. usually not seen in a given individual.

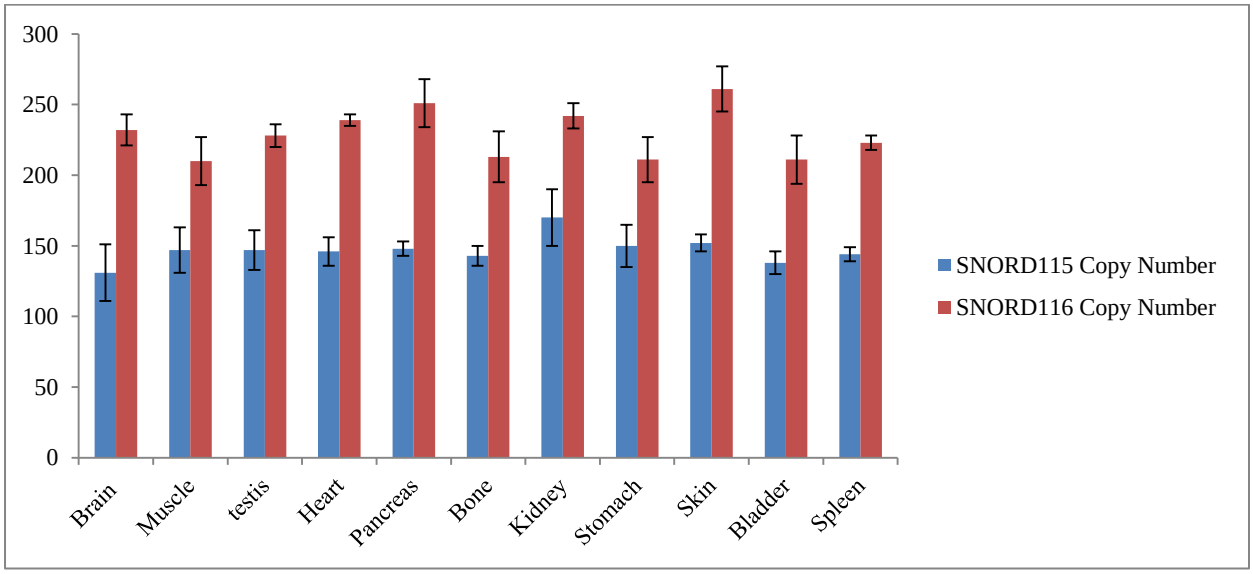


Figure 20: SNORDs CN between different tissues. Error bar represents the variance measures that I had obtained from triplicate measurements for each tissue**.** A one-way analysis of variance test was used to compare the average SNORDs copy number among tissues and $p \leq 0.05$ was considered statistically significant.

P-value for SNORD115 Copy Number = 0.87

P-value for SNORD116 Copy Number = 0.14

Variation of personality in inbred mice

It has been shown that individuals from the inbred C57BL/6 strain, which are nominally genetically identical, showed different personality and anxiety behaviors under controlled and standardized environmental conditions (Jakovcevski et al. 2008; Lewejohann et al. 2011). While the genotyping of SNP loci has shown that there are genetic differences between C57BL/6J and C57BL/6N sub-strains at 11 loci, the genetic homogeneity within a single C57BL/6 sub-strain is extremely high (Mekada et al. 2008; Zurita et al. 2011). Hence, in the above mentioned studies both genetic and environment were under control. A convincing explanation for the phenotype variations within C57BL/6 sub-strains is therefore lacking and it was suggested that an unknown component, such as early learning, may control personality (Lewejohann et al. 2011).

These inbred mouse strains are employed in many fields such as behavioral biology, neuroscience and also as models for human diseases. Therefore this behavioral variation within inbred mouse strains has raised a concern among researchers in these fields because it could generate a noise component in their studies that could lead to wrong conclusion and false discoveries (Belknap et al.1993; Owen et al. 1997; Bothe et al. 2005; Hefner et al. 2008; Lewejohann et al. 2011).

Given that the family experiments above suggested that new copy number variation could arise quickly, I was interested whether even inbred mice show SNORD copy number variation and corresponding behavioral variation. So to check SNORD copy number variation within an inbred mouse strain, I genotyped 60 individuals of C57BL/6J mice from both genders (30 female and 30 male). Interestingly, results from genotyping revealed a huge SNORD CNV between C57BL/6J mice in both females (Fig. 21) and males (Fig. 22). Note that given that the tested SNORDs are on an autosome, one would not expect an average difference between sexes and this is also what I found (Fig. 23). Subsequently I added behavioral tests to 40 of these typed individuals to check their personality. The results showed a significant correlation between individuals SNORD copy number and their personality in males (Fig. 24a-b) and females (Fig. 24c-d), suggesting that the behavior of both sexes is influenced in a similar way by the SNORDs. Hence, these data can explain why even inbred mice showed different personalities under controlled and standardized environmental conditions.

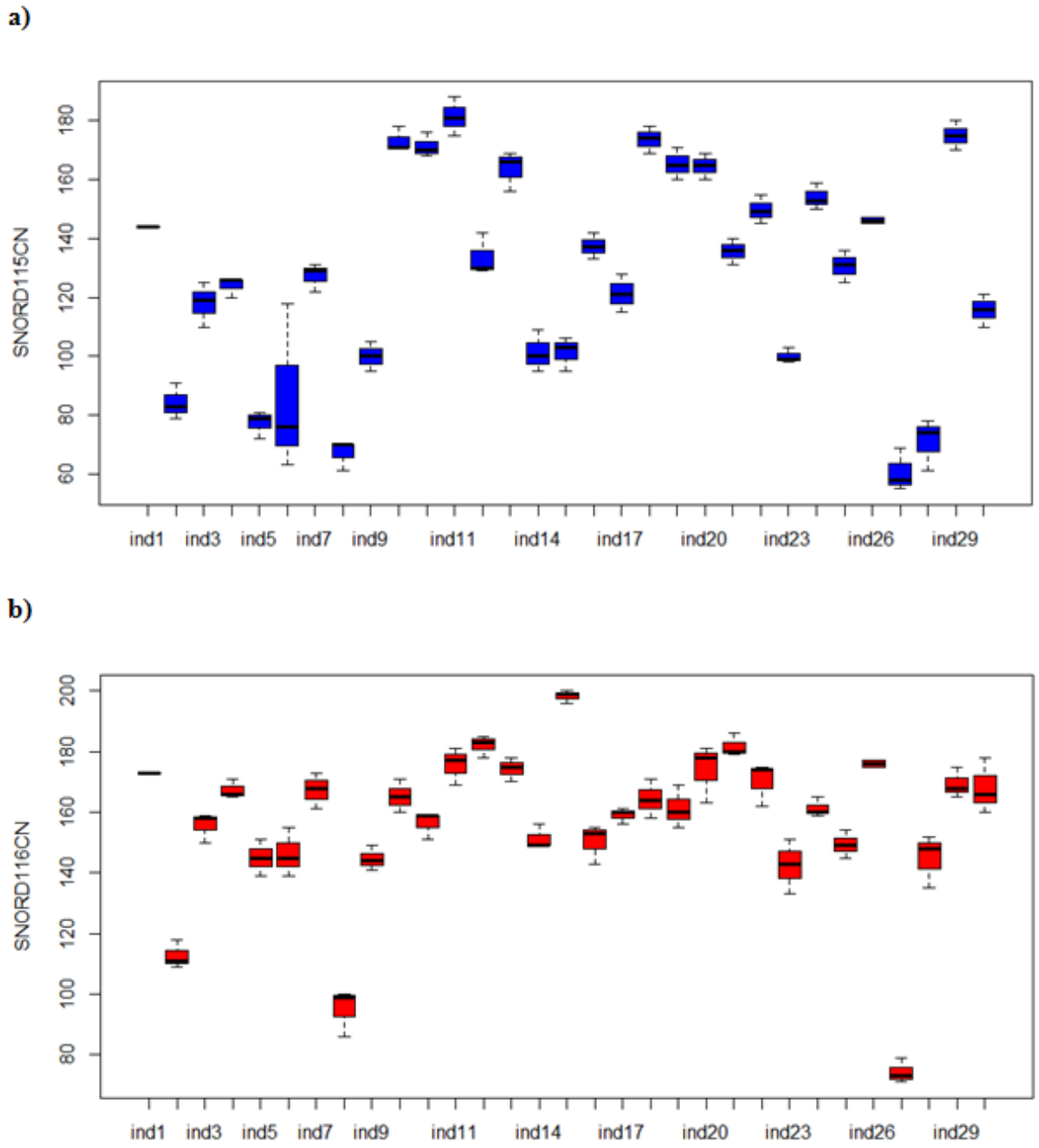
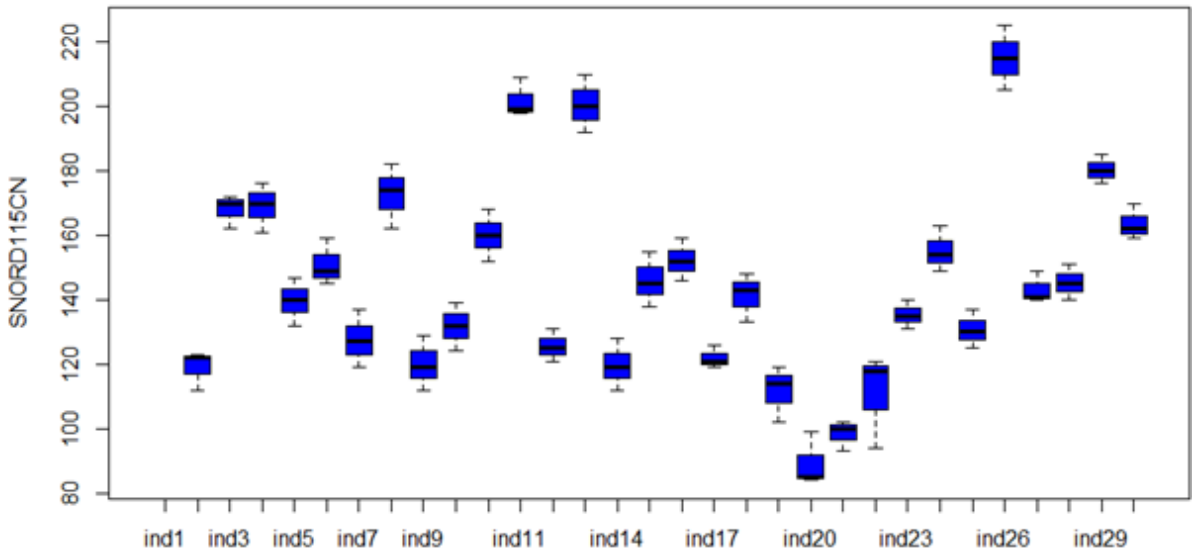


Figure 21: SNORDs CN in female C57BL/6J mice: a) SNORD115 Copy Number variation. b) SNORD116 Copy Number variation. 30 individuals were used for this analysis, each of which is depicted along the X-axis. Error bar represents the variance measures that I had obtained from triplicate measurements for each individual.

a)



b)

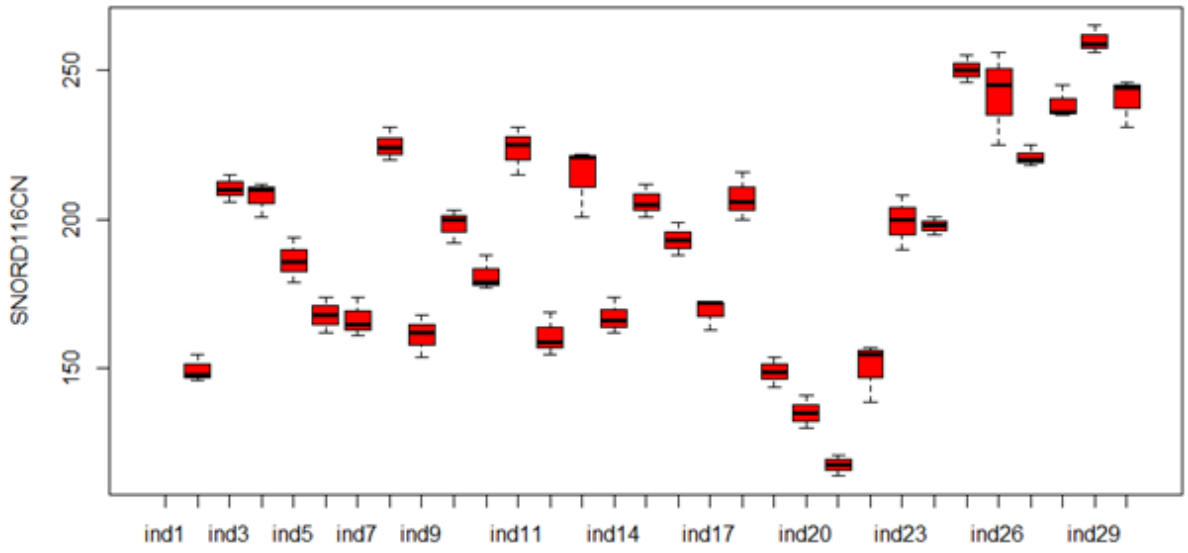


Figure 22: SNORDs CN in male C57BL/6J mice: a) SNORD115 Copy Number variation. b) SNORD116 Copy Number variation. 30 individuals were used for this analysis, each of which is depicted along the X-axis. Error bar represents the variance measures that I had obtained from triplicate measurements for each individual.

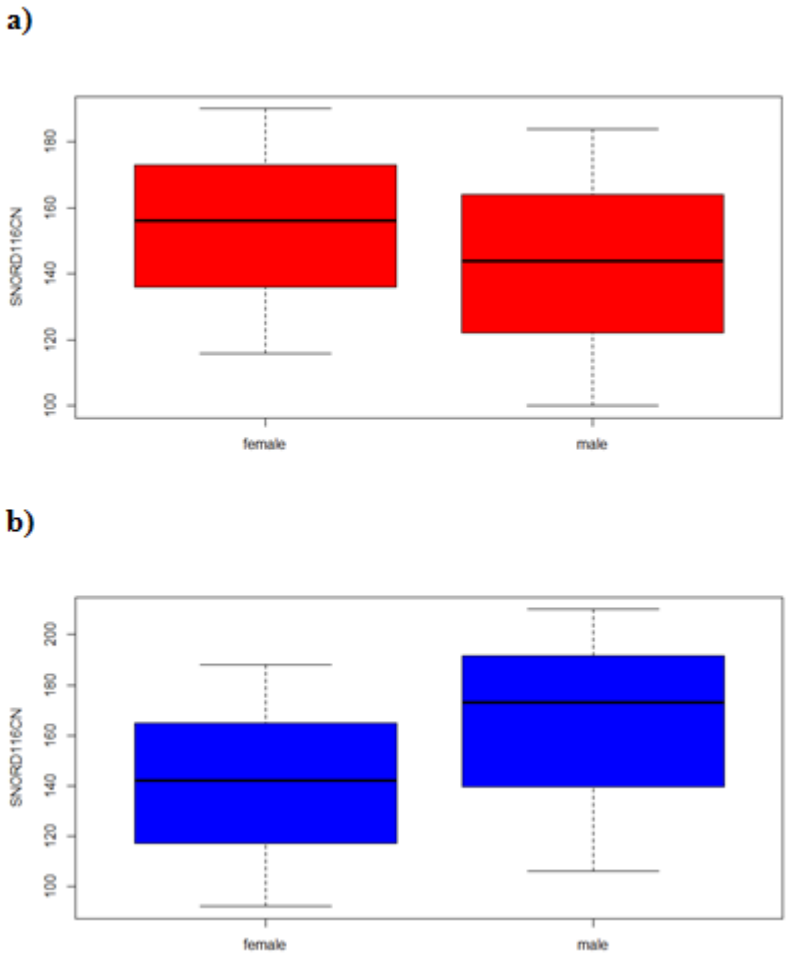


Figure 23: Comparison of SNORDs CNV in female and male C57BL/6J mice: a) SNORD115 copy number. b) SNORD116 copy number. Error bar represents the variability of SNORDs CN within each mouse population. t-test was used to determine whether there is a significant difference between the means of SNORDs CN between female and male. No significant differences were detected on average. P-value for SNORD115 Copy Number = 0.35 and for SNORD116 = 0.066

SNORD copy number influence on the brain transcriptome

To investigate the genetic network which is regulated by SNORD115/116, RNAsequencing analysis was performed on five individuals each with low and high SNORD copy number. These individuals were selected from the C57BL/6J mice in the previous part. RNA from their brain was extracted and after library preparation, libraries were subjected to llumina HiSeq sequencing (see Methods). Reads were mapped to the mm10 mouse reference genome (Mouse Genome Sequencing Consortium et al. 2002) by using Hisat2 (Kim et al. 2015) and bam files from this part were used for the RNAseq analysis.

The first analysis focused on *Ht2c* as target of SNORD115. As Figure 25 a shows, there is a high positive correlation between SNORD115 copy number and total read count from the target exon in HT2CR-201. As expected, there was no significant correlation between SNORD115 copy number and HT2CR-204 and HT2CR-206 as non-target transcripts (Fig. 25b, c). Comparable results were observed for the SNORD116 target Ankrd11 (Fig. 26). The target exon of Ankrd11 showed significant correlation to SNORD116 copy number (Fig. 26a), however there was no significant correlation between SNORD116 copy number and ANKRD11-204 and ANKRD11-207 as non-target transcripts (Fig. 26b, c). Hence, the RNAseq results confirm my previous observations from ddPCR (Fig. 7, 12 and 13).

As discussed above, *Ankrd11* is a chromatin regulator which could regulate expression of many genes by binding chromatin modifying enzymes like histone deacetylases (HDACs) (Zhang et al. 2004, 2007a, 2007b; Li et al. 2008; Neilsen et al. 2008). A study on mice which carried a mutant Ankrd11 gene, revealed more than 700 genes which are regulated by Ankrd11 (Gallagher et al. 2015). However I could recover only 635 of them, due to wrong ID information in the Gallagher et al. (2015) study. These 635 genes are listed in Appendix 1. To find which of these 635 genes are regulated by variation in the ANKRD11-202 transcript abundance, regression analysis was run on total reads from the target exon of ANKRD11-202 and total reads from all of these 635 genes. Interestingly, over 100 of these 635 genes showed a significant correlation (positive or negative) to ANKRD11-202 expression level (Appendix 2). Around 70 percent of them are down-regulated by ANKRD11-202 (Appendix 2 - Table1) and 30 percent show a positive correlation, which means they are up-regulated by ANKRD11-202 (Appendix 2 - Table2).

To gain insights about the biological role of these 100 genes, gene ontology (GO) and KEGG pathway enrichment analysis were performed using DAVID online tools (see Methods). Results from this part classified these 100 genes to five different functional categories: (I): metabolic pathway (II): proliferation and differentiation (III): anxiety (IV): intellectual ability formation (cognitive performance) and (V): osteogenesis. In addition, results from this part highlighted the Gabrg1 gene as a gene which showed the highest correlation to ANKRD11-202 expression (Table 2). Gamma-aminobutyric acid receptor subunit gamma-1 (Gabrg1) is a protein which encodes a subunit of the GABA A receptor (Hevers et al. 1998) and plays a crucial role in anxiety regulation in both human and mouse (Lydiard 2003; Tasan et al. 2011; Nuss 2015). These results suggests that SNORD116 copy number variation has a link to

regulate anxiety via the GABA A receptor, but may also interfere with the metabolic pathway, intellectual ability and also osteogenesis by targeting ANKRD11-202.

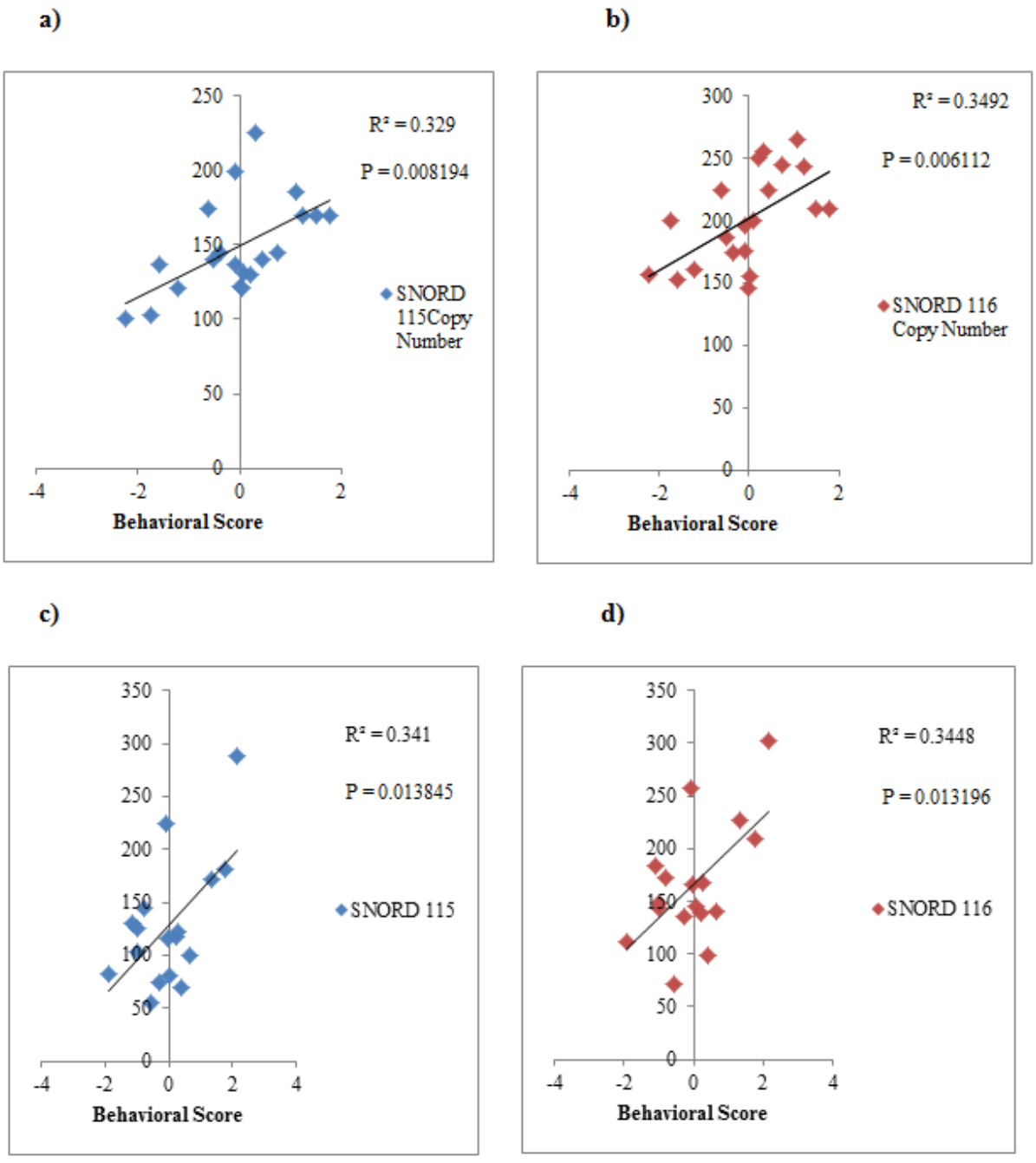
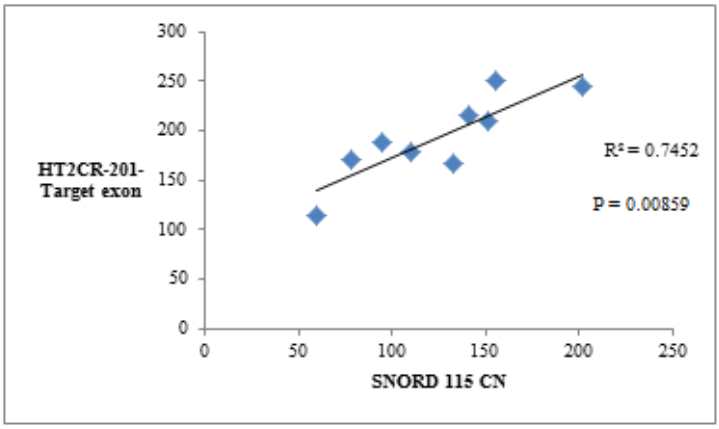
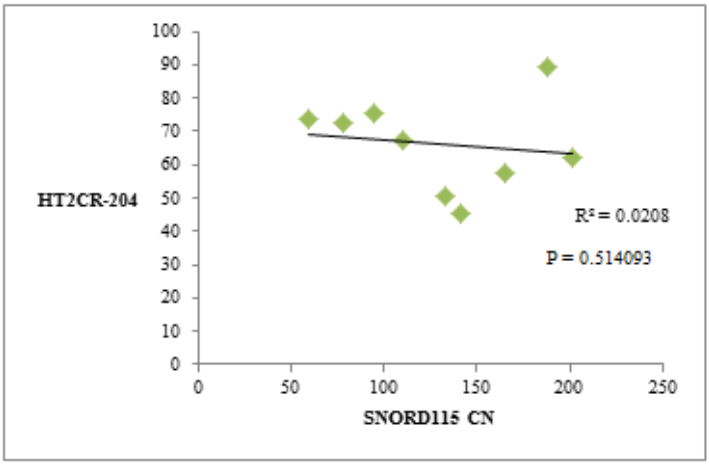


Figure 24: Correlation between SNORDs CN and Behavioral Score in C57BL/6J mice. (a-b): male mice. (c-d): female mice. (e-f): pool data (both female and male). Regression analysis has been done by using lm function in R.

a)



b)



c)

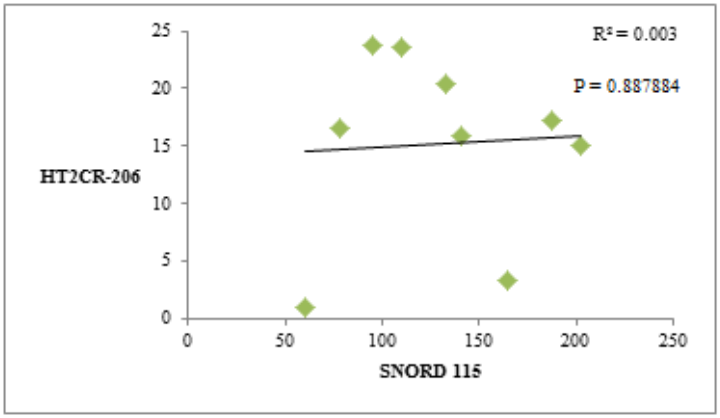
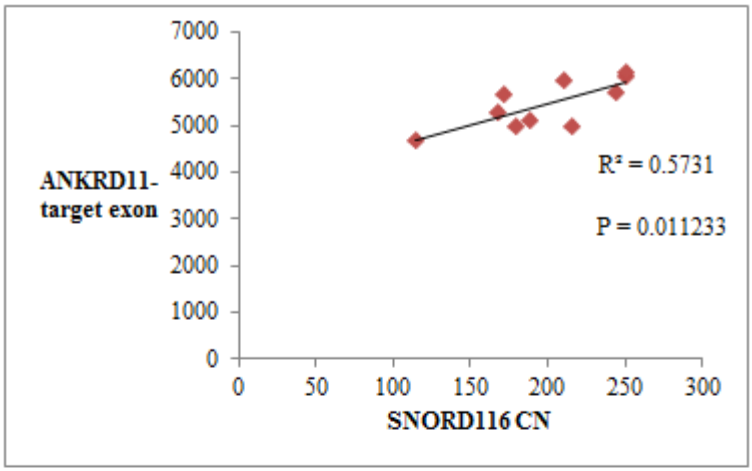
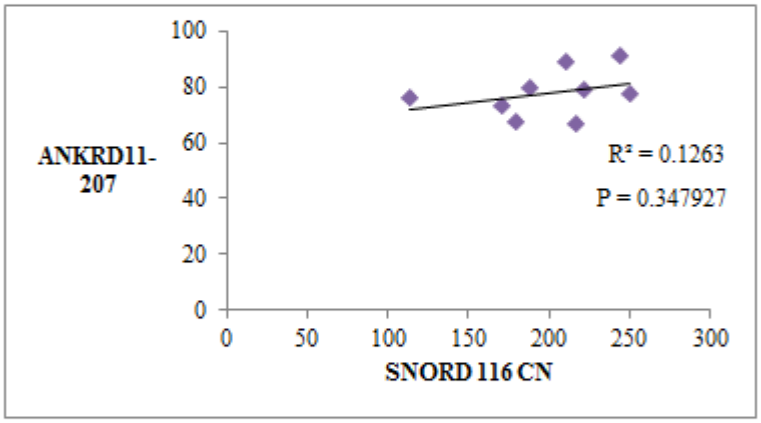


Figure 25: Correlation between SNORD115 copy number and its specific and non-specific targets. a) HTR2C-201. b) HTR2C- 204. c) HTR2C-206. Y-axis represents Transcripts Per Million (TPM) (see method section) which was used as index of gene expression. Regression analysis has been done by using lm function in R.

a)



b)



c)

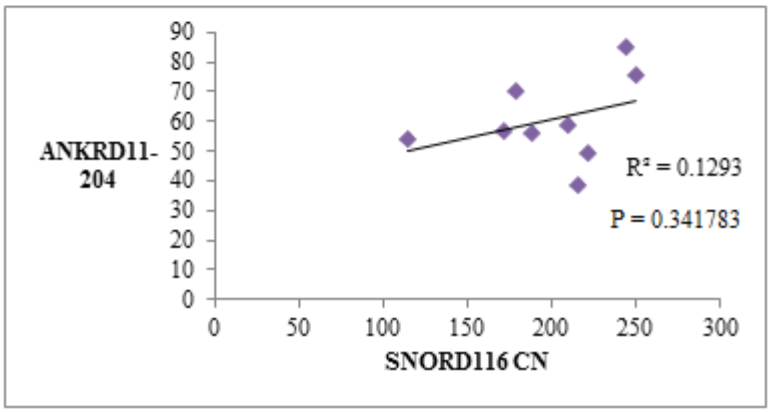


Figure 26: Correlation between SNORD116 CN and its specific and non-specific targets. a) Ankrd11 target exon. b) ANKRD11- 207. c) ANKRD11-204. Y-axis represents Transcripts Per Million (TPM) (see method section) which was used as index of gene expression. Regression analysis has been done by using lm function in R.

Table 2: List of genes with the most significant correlations to Ankrd11-202 expression

Gene Ensemble ID	R-squared	Gene Name
ENSMUSG00000001260	0.89	gamma-aminobutyric acid (GABA) A receptor, subunit gamma 1(Gabrg1)
ENSMUSG00000030067	0.68	Fork head box P1(FoxP1)
ENSMUSG00000021486	0.68	PRELI domain containing 1(Prelid1)
ENSMUSG00000051256	0.68	jagunal homolog 1(Jagn1)
ENSMUSG00000034269	0.65	SET domain containing 5(Setd5)
ENSMUSG00000000171	0.57	succinate dehydrogenase complex, subunit D, integral membrane protein(Sdhd)
ENSMUSG00000011114	0.57	transforming growth factor beta regulated gene 1(Tbrg1)
ENSMUSG00000039163	0.57	COX assembly mitochondrial protein 1(Cmc1)
ENSMUSG00000025340	0.55	Ras negative regulator Rabex-5
ENSMUSG00000044340	0.55	PH domain and leucine rich repeat protein phosphatase 1(Phlpp1)
ENSMUSG00000023961	0.55	ectonucleotide pyrophosphatase/phosphodiesterase 4(Enpp4)
ENSMUSG00000026575	0.55	NME/NM23 family member 7
ENSMUSG00000032300	0.53	D9Ertd278e
ENSMUSG00000030271	0.53	8-oxoguanine DNA-glycosylase 1(Ogg1)
ENSMUSG00000024829	0.52	mitochondrial ribosomal protein L21
ENSMUSG00000033938	0.52	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 7(Ndufb7)
ENSMUSG00000034384	0.51	BarH-like 2 (Drosophila)(Barhl2)
ENSMUSG00000048503	0.51	transmembrane protein 136(Tmem136)
ENSMUSG00000037710	0.51	CDGSH iron sulfur domain 1(Cisd1)
ENSMUSG00000022200	0.51	golgi phosphoprotein 3(Golph3)
ENSMUSG00000027509	0.51	ribonucleic acid export 1(Rae1)

SNORD116 copy number and craniofacial features

It has been reported that mutations in *Ankrd11* causes the KBG syndrome. "KBG" represents the surname initials of the first families diagnosed with the disorder. Medical studies on the KBG syndrome have shown that this syndrome is associated with a splice-site variant of *Ankrd11* (Tekin et al. 2004; Sirmaci et al. 2011; Low et al. 2017). It seems skipping exon IX, X and XI, especially exon X is the most common mutation in the KBG syndrome (Low KJ, et al. 2017). This syndrome is a rare genetic disorder characterized by intellectual disability, autism spectrum disorder, and craniofacial abnormalities (Ka et al. 2017).

Children with KBG syndrome may show abnormalities in their craniofacial features and also general skull shape. These abnormalities may include widely spaced eyes; wild eyebrows; bow-shaped lips; and/or a triangularly-shaped face. Microcephaly has been reported in some children as another abnormality in children with KBG syndrome (Ockeloen et al. 2014 and 2015; Morel Swols et al. 2017).

Interestingly a mouse study also showed that mice with a mutation in the Ankrd11 gene exhibit craniofacial abnormalities: shortened snouts, wider skulls, and deformed nasal bones. This mutation also significantly reduced bone mineral density (Barbaric et al. 2008) (Fig. 27).

These studies, together with the results from the RNASeq analysis described above, suggest that SNORD116 variation may also regulate craniofacial features. So, accordingly, I asked whether SNORD116 copy number variation could cause any changes in mouse skull and mandible and whether there is any possible link between personality and craniofacial feature by SNORD116 mediator.

To initially test this, three Bl6 mice with high SNORD116 copy number and three with low copy number were selected. Personality of these six mice was measured as described before and their heads were scanned using a computer tomography. Differences in skull shape were analyzed by land marking using the TINA tool (Schunke et al. 2012). 36 three-dimensional landmarks were positioned in the skull. Then principal components analysis was run to define shape features (see Methods).

As figure 28a indicates, there may indeed be an effect of SNORD116 copy number variations on skull shape features, as individuals could be tentatively grouped according to their SNORD116 copy number in the PCA. Furthermore, the results from correlation analysis of

SNORD116 copy number and skull score (PC1) showed a positive trend, which suggests a strong link between SNORD116 copy number and skull shape (Fig. 28b).

For Mandible, 13 three-dimensional landmarks were positioned in each hemi-mandible and principal components analysis was run to define mandible shape scores (see Methods). SNORD116 copy number showed a link to one mandible shape score (PC2) and individual's mandible feature can be tentatively grouped according to their SNORD116 copy number (Fig. 29a). Correlation analysis from this part also showed a positive trend (Fig. 29b).

Given the small number of individuals involved in this initial analysis, the results have still to be taken with caution. However, in combination with the KBG effects in humans, they suggest an intriguing unexpected link between personality scores and craniofacial shape.

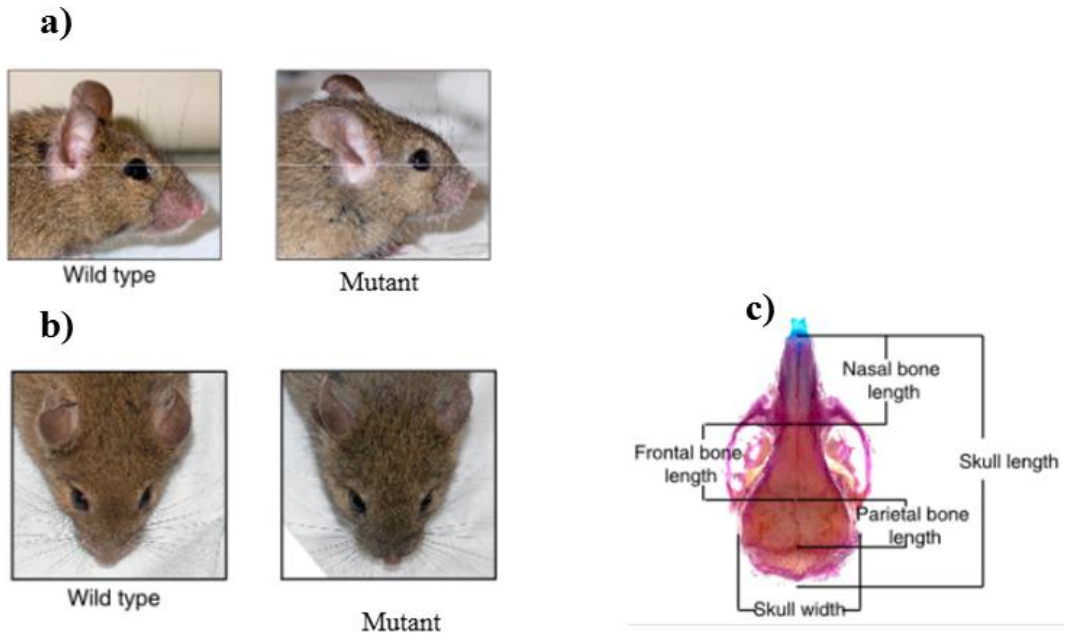


Figure 27: Craniofacial phenotype of Ankrd11 mutant mice. The mice have a shorter (lateral view, a) and wider (dorsal view, b) face appearance compared with wild-type littermates. c): dorsal view of a mouse skull stained with Alizarin red S and Alcian blue, showing distances measured on dissected skulls (Figure reproduced from Barbaric I, et al. 2008).

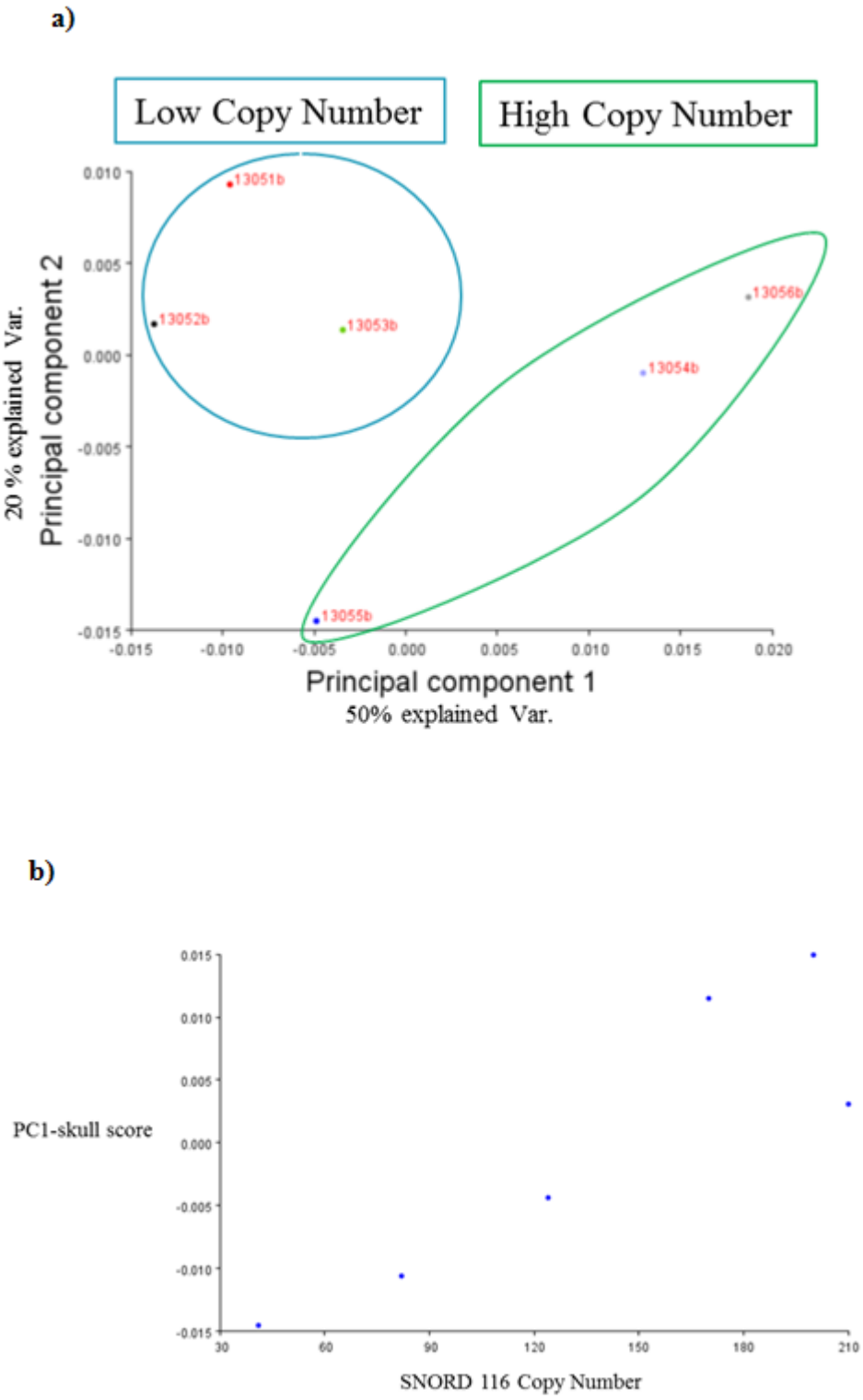


Figure 28: SNORD116 and skull shape. a) Principal components analysis was computed from 43 three-dimensional landmarks on skull of each individual. Individuals with low copy numbers are grouped by blue color and mice with high copy number by green color circles. b) Correlation analysis of SNORD116 copy number and skull score (PC1).

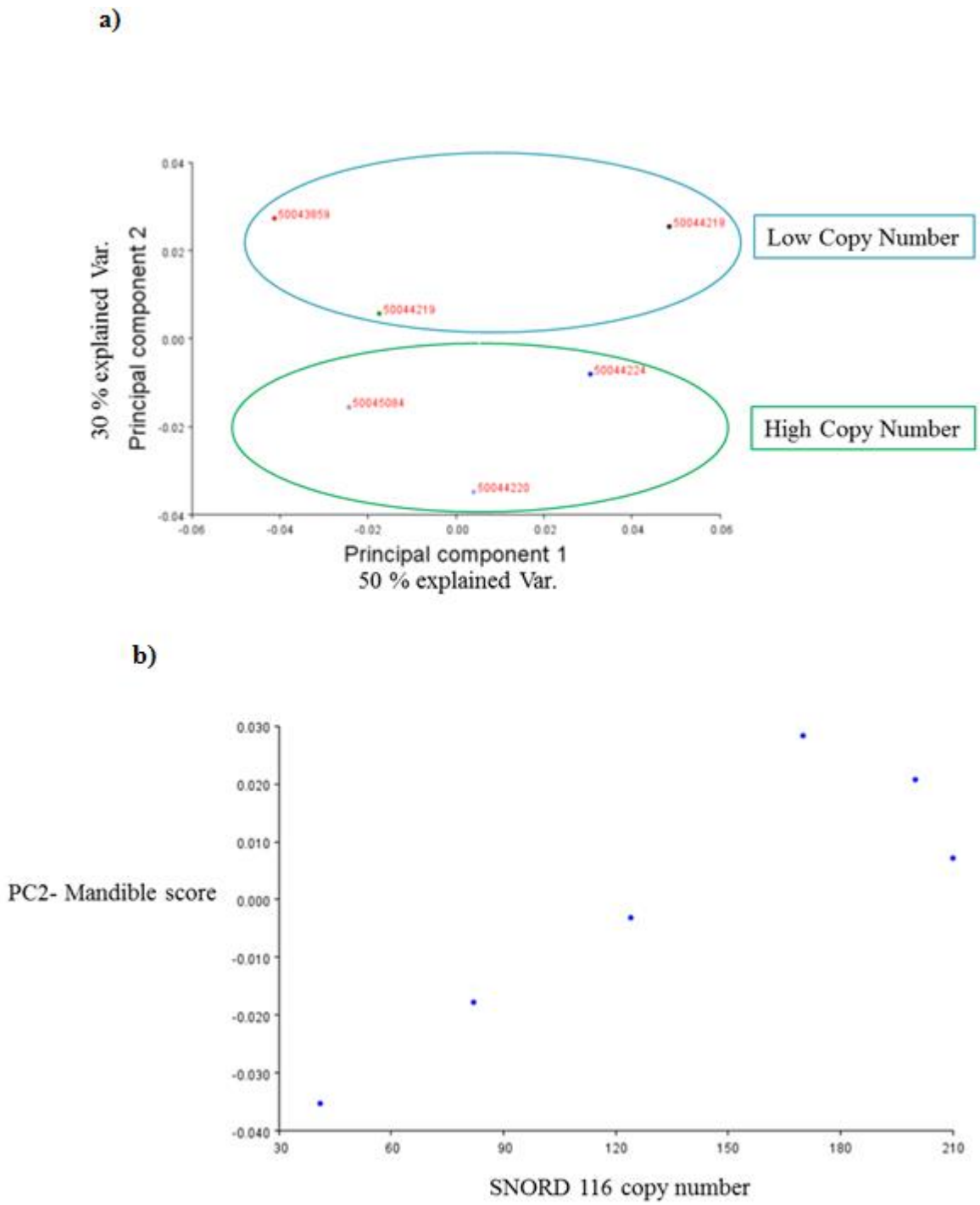


Figure 29: SNORD116 and mandible shape. a) Principal components analysis was computed from 13 three-dimensional landmarks on mandible of each individual. Individuals with low copy numbers are grouped by blue color and mice with high copy number by green color circles. b) Correlation analysis of SNORD116 copy number and skull score (PC1).

Differential gene expression through SNORD CNV

So far my results showed how SNORDs CNV changes the expression of their target genes and then subsequently the downstream molecular network, however we still don't know what other genes are differentially expressed due to the SNORD copy number variation. Above I discussed two direct targets of SNORD115 and 116 and investigated the pathway which is mediated by these two target genes (*Ht2c* and *Ankrd11*). To get a more general insight into the SNORD dependent genetic network, I did further RNAseq analysis on 10 individuals (five individuals each with low and high SNORD copy number).

For that, bam files which were generated by Hisat2 (Kim et al. 2015) were used as input for Htseq to count the aligned reads to mouse genomic feature annotations (Anders et al. 2014). Then differential expression analysis was done by using the DESeq2 package (Love et al. 2014) in the R environment. Genes with an adjusted P-value (Padj) < 0.05 were considered as differentially expressed.

The log2fold change against the mean normalized counts was plotted and is shown in Figure 30. Red dots represent those genes that are significantly differentially expressed at 10% false discovery rate (FDR). These significant genes were filtered for further analysis. 24 genes showed around two-folds change in their expression (Table 3). Further analysis from GO analysis classified these genes into three functional groups (I): proliferation and differentiation (II): intellectual ability and (III): metabolic pathway. These results revealed more genes which are regulated by SNORD115/116, but how exactly these genes are regulated by SNORD115/116 needs more exploration. This data also once again suggest that SNORD115/116 may play a key role not only in anxiety but also in intellectual ability and metabolic pathways.

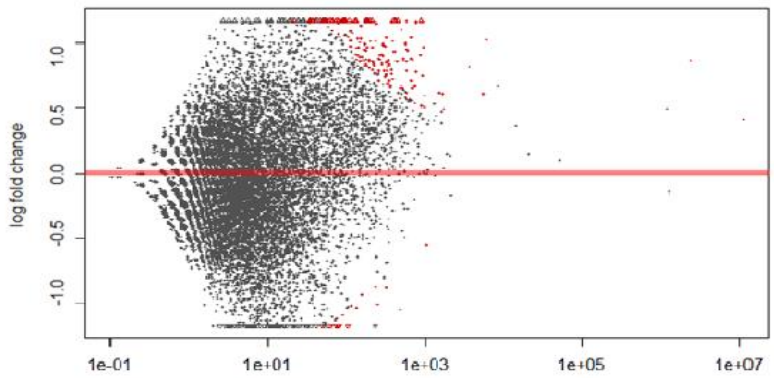


Figure 30: Plot of log2fold change against the mean normalized counts. Red colors are those genes that are significant at 10% false discovery rate (FDR).

Table 3: List of differentially expressed genes in response to SNORD115/116 CN differences. GO analysis classified these genes into three functional groups (I): proliferation and differentiation (II): intellectual ability and (III): metabolic pathway.

Gene ID	Gene description	GO Category
ENSMUSG00000089902	Sjogren syndrome antigen B	I
ENSMUSG00000062604	serine/arginine-rich protein specific kinase 2	I
ENSMUSG00000031257	NADPH oxidase 1	I
ENSMUSG00000039067	mov34	I
ENSMUSG00000036502	TMEM255A	III
ENSMUSG00000085334	zinc finger, MYM domain containing 1	I
ENSMUSG00000036168	coiled-coil domain containing 38	I
ENSMUSG00000006014	proteoglycan 4	III
ENSMUSG00000019971	centrosomal protein 290	II
ENSMUSG00000090546	CDR1	II
ENSMUSG00000018474	chromodomain helicase DNA binding protein 3	II
ENSMUSG00000026893	grancalcin	II
ENSMUSG00000037852	carboxypeptidase E	III
ENSMUSG00000087050	PHD finger protein 12	I and II
ENSMUSG00000063200	Nucleolar protein 7	III
ENSMUSG00000030935	Acyl-coenzyme A synthetase ACSM3	III
ENSMUSG00000036564	N-myc downstream regulated gene 4	I
ENSMUSG00000048330	RIC3 acetylcholine receptor chaperone	II
ENSMUSG00000052525	speedy/RINGO cell cycle regulator family, A	I
ENSMUSG00000030498	growth arrest specific 2	I
ENSMUSG00000008976	GA repeat binding protein, alpha	I
ENSMUSG00000041355	signal sequence receptor, beta	III

Discussion

SNROD115/ 116 CNV and personality

In this chapter, I showed that there is SNORD115 and 116 copy number variation not only between mouse species, subspecies and populations, but also within populations across individuals, as well between individuals of an inbred strain (Fig. 4, 17 and 18). Then I showed a correlation between SNORD CNV and expression level of predicted target transcripts, as well as a correlation with individual's personality (Fig. 9 and 10).

SNORD115 includes complementarity sequences to the alternatively spliced exon Vb of the serotonin receptor 5-HT2C receptor. Vb encodes the second intracellular loop of the receptor, which is crucial for G protein binding. Skipping of exon Vb causes a frame shift, resulting in a receptor that is truncated after the third transmembrane domain (Wang et al. 2000). Here I confirm the previous finding of Kishore et al. 2006, which showed a high positive correlation between SNORD115 expression and exon Vb usage (Fig. 7).

The 5-HT2C receptor is one of the several receptors for serotonin. Activation of this receptor by serotonin significantly regulates mood, anxiety, feeding and reproductive behavior in mice (Alex et al. 2005; Kimura et al. 2009). Interestingly several studies in humans also showed that 5-HT2C receptor plays a crucial role in several behavioral functions, including anxiety, mood, sleep cycles, aggression, appetite and learning (Woolley & Shaw 1954; Chagraoui et al. 2016). 5-HT2C receptor dysregulation or deletion causes several neuropsychiatric disorders such as schizophrenia, depression, and obesity (Hoyer et al. 2002; Fakhoury 2015). In this chapter my results also suggest a strong link between SNORD115 copy number, usage of exon Vb of 5-HT2C receptor and mouse anxiety-like personality (Fig. 9 and 10a, c, e).

Furthermore my study provides biological evidences about details of SNORD116's function and highlights Ankrd11 as SNORD116's target (Fig. 7 b, d). RNAseq analysis revealed that SNORD116 CNV changes usage of exon X of the Ankrd11-202 transcript and this correlates with changes of the expression of over 100 genes (Appendix1). One can argue that there may be several other genes involved in this molecular mechanism regulating by SNORD115/6 and variation of those genes change the expression of these 100 genes and subsequently mouse personality. However, since the results from the behavioral study and RNAseq data analysis on the C57BL/6 inbred strain with extremely high genetic homogeneity still showed high correlation between SNORD116 copy number and usage of exon X of the ANKRD11-

202 transcript, expression of these 100 genes and also mouse personality, the probability of other genes involvement in this correlation is low (Fig. 24, 26 and Table 2).

RNAseq data analysis also suggested that ANKRD11-202 may regulate expression of the GABA A receptor and by this way could significantly change mouse anxiety-like behavior (Table 2; Fig. 9 and 10b, d, f). Altered or abnormal GABA A receptor has been reported already in PWS patients (Lucignani et al. 2004). In addition, high plasma GABA levels have been reported in PWS individuals compared with control subjects (Ebert et al. 1997). However, so far there was no report which explains the causality of these observations and my research here provides the possible link how and why PWS patients show problems in their GABA level or its receptor.

So all these data together can explain how SNORD CNVs modify serotonin receptor 5-HT2C and GABA A receptors and by this way may significantly regulate individual's anxiety-like personality. Since the copy number for each individual is characteristic and probably stable across whole life, it would keep personality traits dependent on this pathway quite stable across time and environmental context.

Correlation of personality and cognitive ability

Studies on human and non-human animals such as Chimpanzee, Guinea pig and dog suggest that personality could be related to individual's cognitive ability (Guenther et al. 2013; Shimabukuro et al. 2015; Curtis et al. 2015; Wettstein et al. 2017). Phenotypic correlations suggest a shared genetics among personality traits and intellectual ability, however so far there is no molecular mechanism which could correlate these traits together. Here I showed SNORD116 CNV changes abundances of the ANKRD11-202 which in itself is expected to be involved in the changes of expression of more than 100 of its target gene. Further analysis of changes in the overall transcriptome revealed several genes which already have been reported in regulation of cognitive ability. So SNORD116 CNV not only changes individual's personality but also could change their cognitive profile and by this way these two traits would connect to each other.

In addition, a new study recently has shown that SNORD116 deletion in mice impaired their cognitive function such as memory, learning and curiosity compared to wild type through unknown mechanism (Adhikari et al. 2018). Here my data compatible with the mechanism

underlying this observation and propose how SNORD116 can indeed regulate mouse cognitive performance.

Correlation of personality and metabolism

Personality traits are relatively stable over time and consistent across different contexts. To have such a stable behavior, individuals need stable energy sources as well; otherwise they won't be able to have repeatable behavior. Hence a link between personality and metabolism has been suggested and also been shown by several researchers (Houston 2010; Terracciano et al. 2013; Careau et al. 2015). However, the molecular mechanism which keeps these two traits together has not been found so far. Here my RNAseq analysis showed how SNORD CNV correlates with the expression of several genes which play key roles in metabolic pathways such as *Cisd1* and *M6pr* (see more genes in Appendix2, table 2 and 3).

Mannose 6-phosphate receptor (MP6R) is a transmembrane glycoprotein which binds to newly synthesized lysosomal enzymes in the trans-Golgi network and transports them to lysosomes via endosomes (Hoflack & Kornfeld 1985; Varki & Kornfeld. 2017). MP6R can bind to variety of ligands such as IGF-II (insulin-like growth factor type II) (A & Nadimpalli SK. 2018). Mice lacking the MP6R show abnormal growth and usually die shortly after birth, because they were unable to regulate the levels of free IGF-II (Sohar et al. 1998).

CDGSH iron sulfur domain 1 (CISD1), also known as mitoNEET, is an outer mitochondrial membrane protein which has high expression in white adipose tissue. Overexpression of *Cisd1* in mice results in massive obesity. In parallel, reduction in *Cisd1* expression cause less weight gain on a high-fat diet (Kusminski et al. 2012).

So SNORD CNVs co-regulate both personality and rate of metabolism at the same time and would keep them connected to each other.

Correlation of personality traits and vocalization

Relationships among personality traits are also of interest. Most studies observe some degree of phenotypic correlation between different aspects of personality traits. In human neuroticism (index of anxiety) was inversely correlated with extraversion (Power& Pluess 2015; Lo et al. 2017). As table 2 shows the *Foxp1* gene is one of the genes which are

regulated significantly by ANKRD11-202. *FoxP1* dysregulation leads to cognitive dysfunction, including intellectual disability and autism spectrum disorder, together with language impairment (Bacon et al. 2012). *FoxP1* expression controls vocalization in birds and also humans (Teramitsu et al. 2004; Chen et al. 2013). A study on birds suggests that *FoxP1* together with *Foxp2* regulate significantly the amount of singing and vocalization (Chen et al. 2013). Therefore when we have SNORD116 CNV among individuals, this would not only regulate their mood and anxiety but might also regulate their vocalization and by this way could make a connection between anxiety, speech and extraversion status.

Personality and craniofacial features

A possible relationship between personality and facial features was initially suggested from the public rather than scientists. Many people believe that they can realize individual's personality only based on their facial features (Squier et al. 1981). Later several studies found a link between personality and facial characteristics in humans (Squier et al. 1981; Kramer et al. 2010), however so far no one can explain the causality of these observations. Here my data propose a mechanism which can connect personality and craniofacial features. I show how regulation of ANKRD11-202 by SNORD116 can change expression of several genes which are important in osteogenesis and by this way make a possible connection between personality and facial feature. I also provide initial behavioral and morphological evidences which suggest a link between craniofacial features and personality in mice (Fig. 28 and 29).

SNOD115/ 116 and the Prader-Willi syndrome

PWS patients show several abnormalities in their cognitive behaviors such as social communication, speech, anxiety, intellectual ability, decision making and obesity (Jauregi et al. 2007, Woodcock et al. 2009). These patients also showed abnormal craniofacial features. Usually facial problems are characterized as high, narrow forehead, thin upper lip, downturned mouth, and almond-shaped eyes. Despite much research on this syndrome, the molecular mechanism that causes these abnormalities either in cognitive behavior or in morphological aspects is poorly known. This study revealed many genes which may be involved in intellectual ability, anxiety, speech, metabolic pathway and osteogenesis. These data could open new windows in medical research to have better idea of the causality of these symptoms in Prader-Willi syndrome.

Inheritance of SNROD115 and 116

My data from the family study suggest that copy number variation may change fast for SNORD115 and 116, possibly in every generation. While this is not yet proven, it would imply that copy numbers are not inherited in a Mendelian fashion, because of a high new mutation rate. A non-Mendelian inheritance can explain why individuals even within one family can show a different degree of similarity to their parents from high to no similarity (Fig. 20). A very high mutation rate would also solve the puzzle why inbred lab mice which are supposed to be genetically very similar or even identical show different personalities under controlled environmental conditions. Here I show there is SNORD CNV within C57BL/6j mice which correlate with different personalities among individuals (Fig. 21, 22, 23 and 24). Therefore measuring of SNORD copy number could be a good way to exclude the natural phenotypic variation from raw data in inbred mice studies and by this way control the noise to avoid false conclusions and discovery (Fig. 31).

Determining the mutation rate for SNORD CNVs is clearly a major task for the future. At present I can only say that it must be much higher than the point mutation rate, given that inbred mice are polymorphic for CN. In view of the fact that previous studies have measured a certain degree of heritability for personality traits, it would also seem that not all alleles are changed at every generations, i.e. some inheritance is possible, while a subset of the offspring may carry new length variants. Unfortunately, the length of the haplotypes precludes currently an unequivocal length determination of haplotypes, i.e. also new mutations cannot be directly traced, given that the ddPCR gives always only a composite measure of both haplotypes.

Material and Methods

Mice sample

In this study, both wild and lab strain mice were used. Wild mice were offspring of mice that originated from wild populations sampled in the Massif Central region of France (MC) and the Cologne/Bonn region of Germany (CB) in 2004 and 2005 and then held under outbreeding conditions at the Max-Planck-Institute for Evolutionary Biology in Plön (see Harr et al. 2016 for details). The C57BL/6J inbred strain was purchased at the age of 3 weeks from Jackson Laboratory.

Animals were kept according to FELASA (Federation of European Laboratory Animal Science Association) guidelines, with the permit from the Veterinäramt Kreis Plön: 1401-144/PLÖ-004697. The respective animal welfare officer at the University of Kiel was informed about the sacrifice of the animals for this study.

Mouse keeping

Mice were usually kept in type III cages (Bioscape, Germany), and were weaned at the age of 3 weeks. Males were housed together with brothers or in individual cages. Females were housed in sister groups to a maximum of 5 mice per cage. Enrichment, including: wood wool, toilet paper, egg cartons and a spinning wheel (Plexx, Netherland), was provided in each cage. Mice were fed standard diet 1324 (Altromin, Germany) and provided water ad libitum. Housing prior to experiments was approximately 20–24°C, 50–65% humidity and maintained on a 12:12 light-dark schedule with lights on at 7 am.

Mouse dissection

All dissections were done by following standardized protocols and personal instructions. Prepared tissues were immediately frozen and kept at −70 ° until DNA/RNA preparation.

DNA extraction

DNA extraction was performed according to a standard salt extraction protocol. Briefly, samples were lysed by using HOM buffer (80 mM EDTA, 100 mM Tris and 0.5 % SDS) with Proteinase K (0.2 mg/mL) for 16 hours in Thermomixer (Eppendorf, Germany) at 55°C

and 500 rmp. 500 µl Sodium chloride (4.5 M) was added to each sample and was incubated on ice for 10 minutes. Then chloroform was added, mixed and spun for 10 minutes at 10,000 rpm. The upper aqueous phase was separated, mixed well with Isopropanol (0.7 volume) and spun for 10 minutes at 13,000 rpm. The pellet was washed with Ethanol (70 %), air dried and dissolved in TE-buffer (10 mM Tris, 1 mM EDTA).

DNA concentration was measured on the Nano Drop 3300 Fluorospectrometer using Quant-iT dsDNA BR Assay kit (Invitrogen) reagent.

mRNA extraction and cDNA synthesis

RNA extraction was done by using Trizol reagent. 1mL Trizol per 40mg tissue was added to each sample. Then the samples were lysed by Tissue lyser II (QIAGEN, Germany) at 30 Hertz for 5 minutes. Homogenized samples were incubated at room temperature for 5 minutes. 200µl chloroform (per 1 mL TRIzol) was added to each sample, shook vigorously by hand 15 seconds, followed by 3 minutes incubation at room temperature and spun at 12,000 g for 15 minutes at 4 °C. The aqueous phase was transferred to a new tube and 0.5 volumes Isopropanol was added, incubated at room temperature for 10 minutes and spun at 12,000 g at 4 °C (RNA forms a gel like precipitate). The supernatant was removed and the pellet was washed with 75% EtOH (made with DEPC-H_2O). Samples were mixed by hand several times and then spun at 7,500g for 5 minutes at 4 °C. The supernatant was removed and the pellet dried shortly at room temperature, dissolved in 200µl RNAse free water and stored at -20 °C for overnight. An equal volume of LiCL (5M) was added to the crude RNA extract, mixed by hand and incubated for one hour at -20 °C. Samples were spun at 16,000 g for 30 minutes. The supernatant was removed; samples were washed twice with EtOH 70% and spun at 10,000 at 4 °C. The pellet was dried at room temperature, dissolved in RNAse free water and kept in -70 °C.

The quality of the RNA samples were measured with Bio-Analyzer chips and samples with RIN values below 7.5 were discarded. cDNA was synthesized using the MMLV High Performance Reverse Transcriptase kit according to the instructions of the supplier (epicenter, an Illumina company)

RNAseq analysis

To investigate the genetic network which is regulated by SNORD115/116, RNA-sequencing analysis was performed on five individuals each with low and high SNORD copy number. These individuals were selected from the C57BL/6J mice that have been phenotyped for personality trait. The sequencing of the samples was performed using a polyA tail purification step, followed by cDNA synthesis, Illumina library preparation by using the Truseq standard RNA HT kit from Illumina. The libraries passing quality control were subjected to sequencing on an Illumina HiSeq 2000 sequencer. Each transcriptome sample was sequenced in approximately one third of a HiSeq2000 lane.

Raw sequence reads were quality trimmed using Trimmomatic (Bolger et al. 2014). The quality trimming was performed base wise, removing bases below quality score of 20 (Q20), and keeping reads whose average quality was of at least Q60.

Afterwards reads were mapped to the mouse mm10 reference genome (Genome Sequencing Consortium 2002) by using Hisat2 (Kim et al. 2015). Htseq was used for counting reads overlapping into a specific feature (gene) (Anders et al. 2015). Differential expression analysis was performed with the DESeq2 package (Love et al. 2014) in R environment. Genes with an adjusted P-value (Padj) < 0.05 were considered as differentially expressed.

To perform the alternatively spliced isoforms analysis, BAM files from Hisat2 were used as input for SAMtools (Li et al. 2009) by using option –c for total read from each exon and option –q 60 for total read of each sample.

Gene Ontology (GO) is a tool used for gene annotation by collecting defined, structured, controlled vocabulary (Ashburner et al. 2000). KEGG is a database used to categorize associated gene sets into appropriate pathway (Kanehisa et al. 2000). DAVID (Database for Annotation, Visualization and Integrated Discovery) is a web-based tool which can provide a comprehensive set of functional annotation for numerous genes.

GO and KEGG pathway enrichment analyses were performed using DAVID online tools (Version 6.8, https://david-d.ncifcrf.gov/), with the classification stringency set to “medium” P value of <0.05

Small RNA extraction, cDNA synthesis, library preparation and sequencing

Total RNA which is enriched in small RNA was extracted by using the mirVana miRNA Isolation Kit. Quality of the RNA was measured with BioAnalyzer chips and samples with RIN values below 8 were discarded. Illumina® TruSeq® Small RNA Library Prep kit was used for small RNA cDNA synthesis. The protocol in the Illumina kit takes advantage of the common natural structure in most known small RNA molecules. Most mature small RNAs have a 5'-phosphate and a 3'-hydroxyl group. So the Illumina adapters in this kit are directly ligated to these small RNAs. Then Reverse Transcriptase with a primer for this adaptor was used to synthesize cDNA from small RNA. This step was followed by PCR amplification, and purification to generate a library product. As input for library preparation I used 1 μg of total RNA which was enriched in small RNA.

Libraries were prepared separately for different mouse populations, subspecies and species that have been used in this study. This included *Mus musculus domesticus* (German, French and Iranian populations), *Mus musculus musculus*, *Mus musculus castaneus*, *Mus spretus*, *Mus spicilegus* and *Mus mattheyi*. These mouse populations, subspecies and species altogether span 6.6 million years of divergence (Neme et al. 2016).

Ligation of adaptors to small RNA is tricky and therefore it is necessary to check the quality of libraries before running the samples on an Illumina machine. To check the quality of each library, all the libraries were pooled, cloned, and clones were sequenced by the Sanger sequencing method and then analyzed using BLAST (Altschul et al. 1990).

The size distribution of each library was analyzed by running an Agilent DNA 7500 chip on Bio-Analyzer. Quantification of libraries was performed by taking advantage of the qPCR method, which is much more accurate and reliable than results from Bio-Analyzer. To this end, I designed two sets of primers which could bind to the 3' and 5' adapter sequences and by this way I measured the concentration of those fragments which have both adapters and could be sequenced by Illumina. Control template and the pool libraries for quantification were diluted to pM range and then qPCR was run by using Fast SYBR Green Master mix kit. Concentration of each library was calculated based on a standard curve generated from control template dilutions.

Based on the results from quality and quantity analysis, libraries were pooled and then sequenced on an Illumina machine using the Illumina Mid Output Kit.

Read mapping

Reads were mapped to the mouse mm10 genome reference (Genome Sequencing Consortium 2002) based on the best match option by NextGenMap (Sedlazeck et al. 2013). Then the PWS region on Chromosome 7 was extracted from the BAM file which had been generated by NextGenMap and by using read names, this region was extracted from raw data in Fastq format.

Since for further analysis I needed the Bowtie (Langmead et al. 2012) output, I mapped once again the extracted region from the raw data to mm10 by Bowtie and put the best option to map reads based on best match. The outputs were sorted uniquely by SAMtools (Li et al. 2009) and basic commands in Linux.

snoRNA analysis

For snoRNA analysis I used the snoseeker package. snoSeeker is an advanced computational package for screening of guide and orphan snoRNA genes which was introduced by Yang et al. in 2006 and then improved to a new version which is named snoseekerNGS (Yang et al. 2010; Zheng et al. 2016) as the most common tool for snoRNA discovery from next generation sequencing data. snoseekerNGS takes the output of Bowtie (Langmead 2010) as input and by using CDseeker and ACAseeker screens efficiently both guide and orphan snoRNA genes in mammalian genomes. Since both SNORD115/116 families are CD box snoRNA, CDseeker ran on final output from the previous part to find and extract SNORD115/116. Then the output was exported to Geneious 9.0.5 (Kearse et al. 2012) for alignment analysis. SNORD115/116 sequences from different mouse populations were aligned by using pairwise alignment option in Geneious 9.0.5. Global alignment with free end gaps was chosen as alignment type and cost matrix was set according to 93% of similarity. Accordingly, I found the regions in SNORD115/116 which have been conserved between mouse populations. I used these conserved regions to design the primers which have been used in droplet digital PCR procedure.

Droplet digital PCR

Digital PCR is a method enabling absolute quantification of DNA targets without the need to construct a calibration curve as used commonly in qPCR (Zhao et al. 2016). The principle of

digital PCR was first introduced in the 1990s (Sykes et al. 1992) and the recent development of droplet digital PCR (ddPCR) machines has led to wide use in research and clinical applications (Zhao et al. 2016). Hence, in this study this method was used for absolute quantification of SNORDs (SNORD copy number calculation) and also to evaluate gene expression differences.

ddPCR consists of four parts:

1- Finding a suitable reference gene

To calculate gene copy number and also normalize gene expression level, having a reference gene is essential. β-catenin, a gene expressed in most cells, was used to normalize the data from mRNA expression. However, since I have used a special kit which makes cDNA preferentially from small RNA, I could not use any standard housekeeping gene as reference gene to normalize the SNORD115 and 116 expressions and instead I have selected a small RNA with high expression in the brain. SNORD66 was used as reference gene which has high expression in heart, liver, brain, kidney, and testis and has been used in many previous studies as reference gene for small RNA analysis (Liao 2010; Lusardi et al. 2012; Kumar et al. 2014; Kulkarni et al. 2014; Lusardi et al. 2014). For copy number calculations reference genes could be any single copy gene. SNORD 66, which is located in an intron of the eukaryotic translation initiation factor 4, gamma 1 (Eif4g1), was therefore used as reference gene for copy number calculations as well.

2-Primer designing

Primers were designed with 50-70 bp amplicon length, GC content <50% and with very low potential for primer dimer structure. To design the primer for SNORD115 and 116, output from snoseekerNGS was used to find parts of these SNORDs which are conserved between mouse populations (see previous section for more information). The table below lists the primer sequences which were used in this study:

Gene name	Forward sequence 5´→3´	Reverse sequence 5´→3´
SNORD 66	GTGTCTGGGCCACTGAGAC	TTCCTCAGGTCCTCAATCCCA
SNORD 115	GGGCCTCAGCGTAATCCTAT	ACCCAATGTCATGAAGAAAGGTG
SNORD 116	ACCTCAGTTCCGATGAGAGTG	TTCCCAGTCAAACATTCCTTGG
β-catenin	GGAAAAGAGCCTCAGGGCAT	CTGCCTGACGGCCAGG
HT2CR-201	ATTCGCGGACTAAGGCCATC	GAAGTTCGGGTCATTGAGCAC
HT2CR-204	CAAAATGGCACCCTGACCTG	GCCCTGGGTTCAATATCTGTTAC
HT2CR-206	ACAGAGTTGCTTGTTTGGTGTG	TCATGCCTGACAGTGGCATAG
ANKRD11-202	TATTGCCATCGACGGAGCTG	ACCTCCTGTTAGGCAAAGGC
ANKRD11_207	GGGAAGGAGAGGAGCAGAAAC	GGGAGGATCTGTCAGTTGCTATG
ANKRD11_204	CAGGGTTCTCCTTGAGCAAGAC	TGAAATCGAGGCAGCTGGTG

3-DNA digestion and ddPCR

The QX100™ Droplet Digital™ PCR System (Bio-Rad, Hercules, CA, USA) was used in this study according to the manufacturer's instructions. Briefly, fluorescent PCR reactions for each sample were prepared in a 23 µL volume containing 12µL 2 X EvaGreen supermixes, 200nM of each forward and reverse primers, 1ng DNA or cDNA and water.

For tandem copy separation, sample viscosity reduction and improved template accessibility, DNA digestion with a restriction enzyme has been recommended. To this end, EcoRI was used for SNORD115 and BamHI for SNORD116 (Thermo Fisher Scientific). These enzymes cut only once in the repeated units. 5 units of restriction enzymes were added to each sample. Then the samples were kept 20 minutes at room temperature for complete digestion.

Droplets were generated using a Droplet Generator (DG) with an 8-channel DG8 cartridge and cartridge holder with 70 µL of DG oil/well, 20 µL of fluorescent PCR reaction mixture and a DG8 gasket. The prepared droplets were transferred to corresponding wells of a 96-well PCR plate (Eppendorf, Germany).

The PCR plate was subsequently heat-sealed with pierceable foil using a PX1™ PCR plate sealer (Bio-Rad) and then amplified in a LifeEco thermal cycler (Bioer, China). The thermocycling protocol was: initial denaturation at 95°C for 5 min, then 40 cycles of denaturation at 95°C for 30 s, annealing at 60°C for 45 s and, finally, incubation first at 4°C for 5 min and then at 90°C for 5 min (Fig. 32). After cycling, the 96-well plate was fixed into a plate holder and placed into the Droplet Reader (Bio-Rad). Droplets of each sample were analyzed sequentially and fluorescent signals of each droplet were measured individually by a detector.

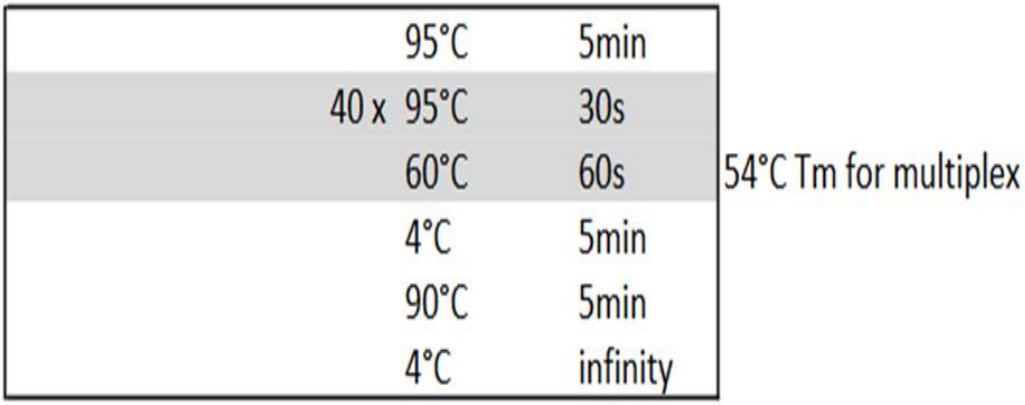


Figure 32: ddPCR cycling protocol. See the text for more information.

4-Copy Number Calculation

To have more precise calculation of SNORDs copy number, having good final output from ddPCR procedure is essential. Figure 33 shows several examples where ddPCR had failed to produce a good output. The plots in figure 33 are one-dimensional scatterplots of event number (droplets) versus fluorescence amplitude which were generated by the droplet reader machine automatically as final output of ddPCR. Blue are positive and grey are negative droplets. The final output of ddPCR is not acceptable if intermediate droplets appear between the major positive and negative bands (Fig. 33a), or where positive and negative bands are too close to each other (Fig. 33b), or the negative band is very weak (Fig. 33c), or there are only a few positive droplets (Fig. 33d). These outputs are not acceptable and in these cases the whole experiment must be repeated once again, otherwise the calculation won't be precise and reliable. For each sample and for each gene, ddPCR was ran separately and replicated three times. If the results from these three replicates for each sample were not close to each other, the experiment was repeated again.

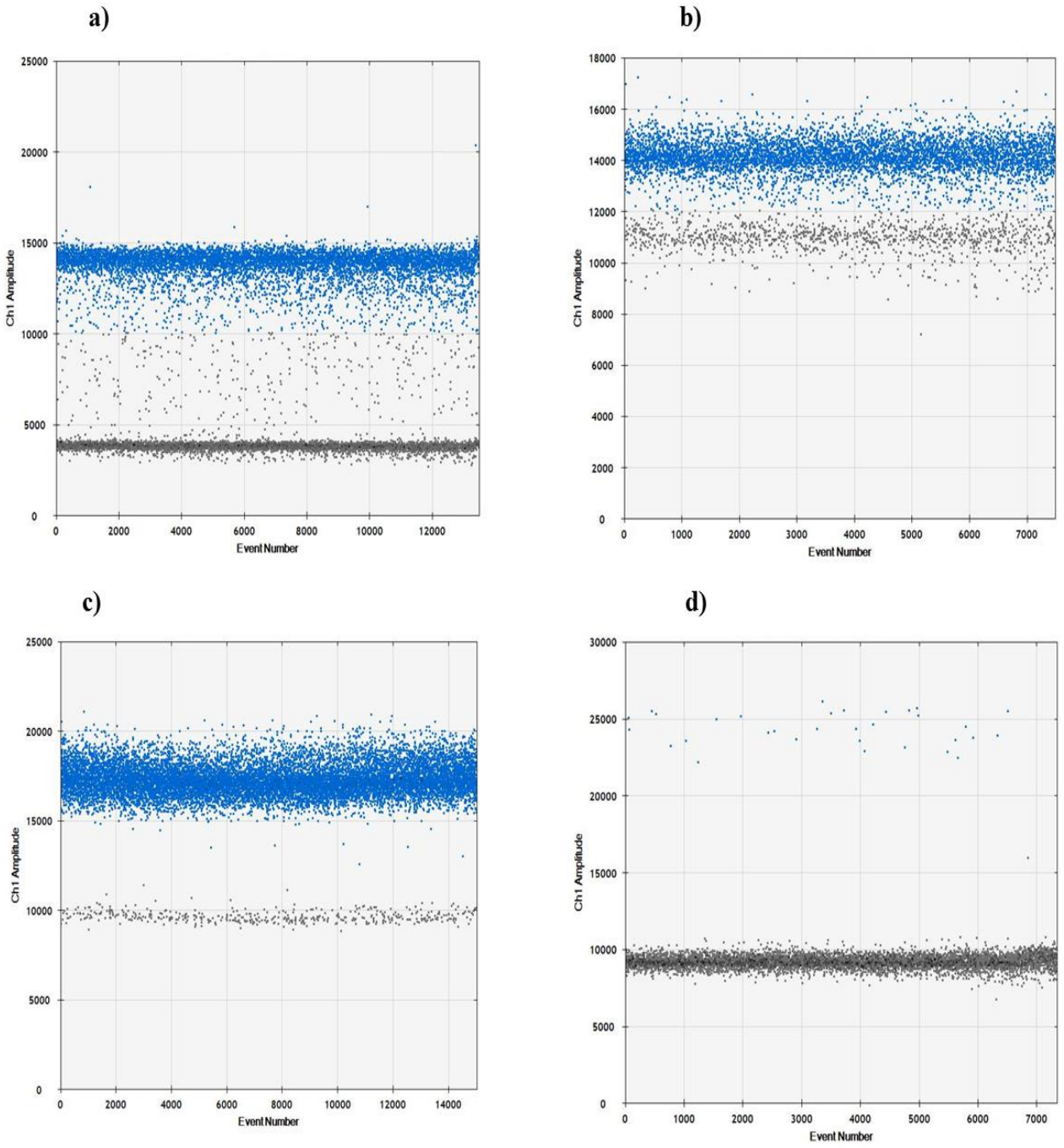


Figure 33: Several examples of data output from ddPCR uniplex assays where ddPCR had failed. a) Intermediate droplets appear between the major positive and negative bands b) close distance between positive and negative bands c) negative band is very weak d) lack of enough positive droplet.

Figure 34 shows an ideal assay with a clear separation of positive (blue) and negative (grey) droplets for target genes (either SNORD115 or SNORD116) (Fig. 34a) and reference gene (SNORD66) (Fig. 34b).

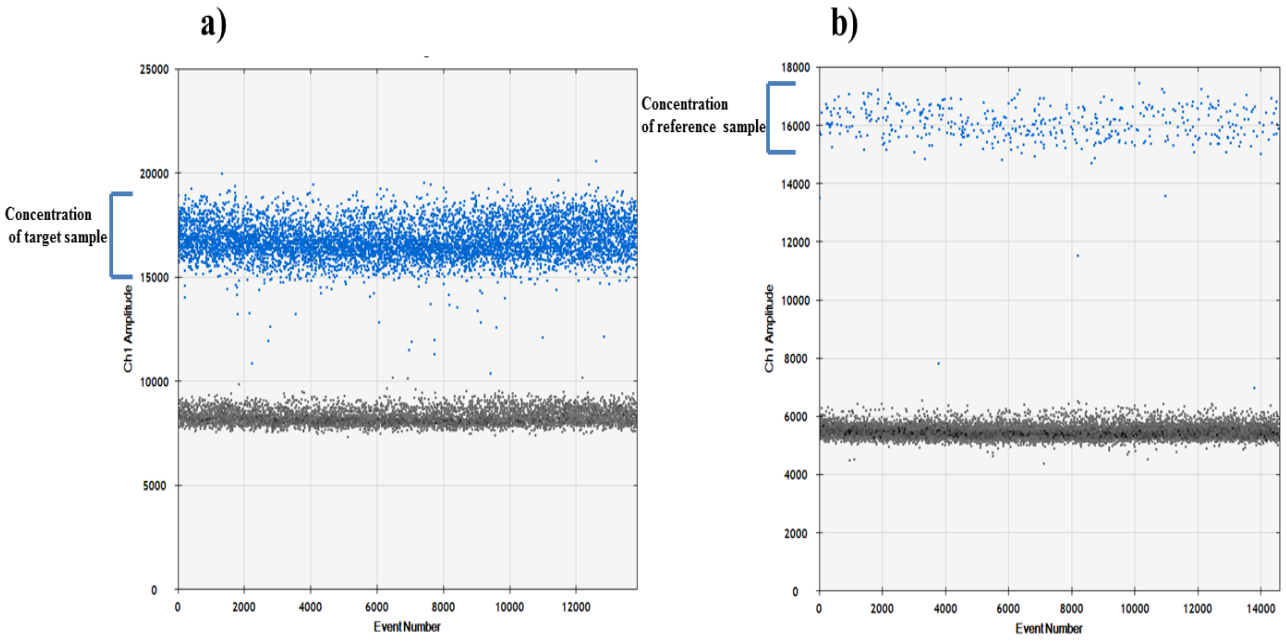


Figure 34: Ideal examples of data output from ddPCR uniplex assays. One-dimensional scatterplots of event number (droplets) vs. fluorescence amplitude; blue are positive and grey are negative droplets. a) target sample b) reference sample.

Copy Number was calculated as target divided by reference:

$CN=A/B \times Nb$

A= concentration of target sample

B= concentration of reference sample

Nb= number of copies of references species genome (here is 2)

Mouse Personality

Behavioral tests were performed on wild mice at the age of 24 weeks and on C57BL/6J inbred strain at the age of 16 weeks.

Behavioral Tests

Elevated Plus Maze

For the wild mice in this study I used a modified elevated plus maze, with clear plexiglas arms and a lid. This modification with the lid was necessary as wild mice are able to easily escape from the otherwise commonly used setup. Mice were placed in the center of an elevated plus maze (EPM) (arms are 50x50 with a 10 centimeters neutral area in the center where the arms crossed arms). Two of the arms were made of clear plexiglas and two were made of grey PVC. The floor was made of white PVC. Behavior of each mouse was monitored for 5 min (Holmes et al. 2000). During this experiment, the time spent in the dark and light arms were measured, as well as the speed and distance travelled. At the end of the experiment the setup was cleaned with 30% Ethanol (Fig. 35).

Open Field

Mice were placed in a 60x60 cm apparatus (Fig. 35) and allowed to explore it for 5 min (Wilson et al. 1994; Reale et al. 2007; Yuen et al. 2015). The speed of the mouse, the distance travelled and time spent within 10 centimeters off the wall vs in the central area were measured (Fig. 35).

Dark/Light Box

The focal mouse was placed in a test apparatus containing a small dark shelter with two exits. During the first five minutes, the time until the mouse pokes its nose out of the shelter and the first time the tail is visible was recorded. At five minutes, a set of keys was dropped next to the test apparatus, and the second part of the experiment began. The time it took for the mouse to first look out and when the entire mouse was visible was measured. If mice did not come out at all, the time was set to be 600 seconds. This test was adapted from tests in lab

mice and common voles (Young et al. 1991; Herde et al. 2013). At the end of the experiment the setup was cleaned with 30% Ethanol (Fig. 35).

Since personality is defined as a consistent behavioral trait, we need only those measurements which were consistent over the course of the experiment. Hence each behavioral test that I described above was repeated every 4 weeks for two times.

These behavioral tests were filmed by using a TSE camera (TSE system, Germany). To score the videos from each test, all the videos were transferred to Videomot2 system (TSE system, Germany). With the VideoMot system mice in nearly every arena can be tracked. Mice were detected by the software in 3 points (head/center/tail base tracking) and then the software automatically generates the numerical data of the time that each mouse spent at zones of interest. Table 3 shows component measurements of each test.

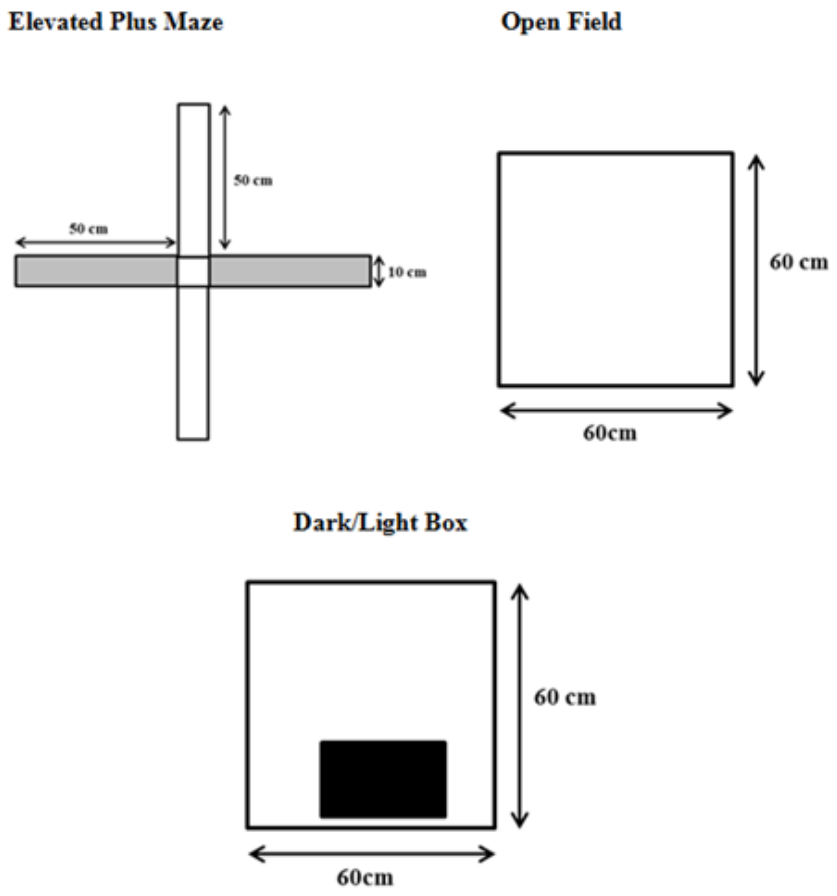


Figure 35: Schematic view of Behavioral Tests which were used in this study

Table3. The behavioral tests and their component measurements

Dark/Light Box Test	Elevated Plus Maze	Open Field Test
Nose poke out of shelter	Time in the dark arm	Speed and Distance
Mouse leaves Shelter	Speed and Distance	Time at the wall
Nose poke out of shelter post startle		Distance travelled at the wall
Mouse leaves Shelter post startle		

Statistical Analysis

Statistics were carried out using R 3.3.3 and R 3.3.2 . All the measurements from behavioral tests were not normally distributed and therefore non-parametric statistics were used.

As the repeatability of a behavior is a key component for the identification of personality trait, all single measurements assessed in the behavioral tests were subjected to repeatability analysis. Repeatability was calculated using "rptR" package (Nakagawa et al. 2010). For non-parametric measurements, "rpt.glmm" was used, with a poisson distribution and SI scores used a proportional model. Table 4 shows the repeatable measurements in each behavioral test.

To determine whether individual behavioral measurements are correlated, a Spearman correlation matrix was made. P-values were corrected using the Holm method. Behaviors were clustered using the protocol from Herde et al. 2013. An hierarchical cluster function was used from the R package "cluster" (Maechler et al. 2017), specifically "agnes", to determine the relationship between the measurements. All measurements were clustered using Manhattan clustering with complete linkage (Gyuris et al. 2011; Tremmel et al. 2013; Herde et al. 2013). The resulting dendrogram is shown in the in Figure 36. The dendrogram shows all measured variables which were repeatable. The height where the variables are joined shows the distance between two clusters (e.g. the shorter the distance the more similar variables).

Principle component analysis was made using the "psych"package (Revelle. 2016), in R functions "factor.pa" and "principle" to assign behavioral scores for anxiety and curiosity

(based on the variables which clustered in the dendrogram) (Fig. 36). In my study the first principle component for anxiety part was used as behavioral score.

The whole statistical analysis was validated in the thesis of Rebecca Krebs (Krebs 2018) and the original behavioral data is added as an appendix to the thesis (Appendix 4).

Table4. The behavioral tests and their component measurements which were repeatable

Dark/Light Box Test	Elevated Plus Maze	Open Field Test
Mouse leaves Shelter	Time in the dark arm	Time at the wall
Nose poke out of shelter post startle		
Mouse leaves Shelter post startle		

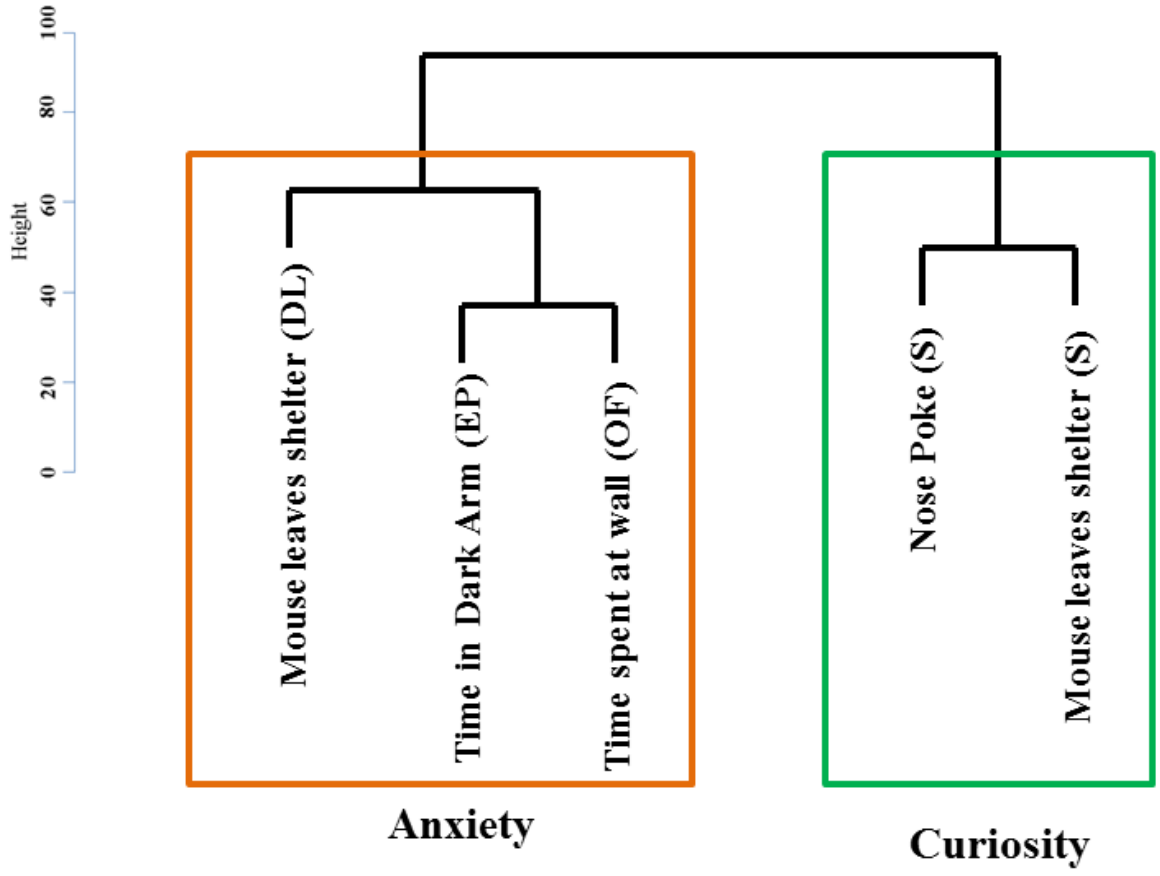


Figure 36: Hierarchical clustering of repeatable measurements. Height of branches indicates similarity between measurements. Measurements could be pooled into two main categories, later classified as curiosity, anxiety based on the types of measurements in the clusters. Dark/Light box (DL), Startle Test (S) Elevated Plus Maze (EP) and Open Field Test (OF**).**

Shape phenotyping

Morphological analysis was performed as described previously (Pallares et al. 2015). Briefly, Mouse heads were scanned using a computer tomograph (micro-CT—vivaCT 40; Scanco, Bruettisellen, Switzerland) at a resolution of 48 cross-sections per millimeter. Using the TINA landmarking tool (Schunke , et al. 2012), 36 three-dimensional landmarks were positioned in the skull, and 14 in each hemi mandible (Fig. 37, Table 5). The semi-automatic landmark annotation extension implemented in the TINA landmarking tool was used to reduce digitation error and accelerate the phenotyping process (Bromiley et al. 2014). The raw 3D landmark coordinates obtained in TINA tool were exported to (MorphoJ Klingenberg CP. 2011) for further morphometric analyses.

The symmetric component of the mandible and skull were obtained following (Klingenberg et al. 2002). In short, for mandible a full generalized Procrustes analysis (GPA) was performed with the land-mark configurations of the right and left hemi mandibles. The GPA eliminates the variation due to size, location, and orientation of the specimens, and generates a new dataset that only contains shape variation. For each individual, we recorded an average of the right and left resulting configurations, which represents the symmetric component of shape variation. For skull, a mirror image of the landmark configuration of each individual was generated, and a full GPA was performed with the original and mirror configurations. Again, the resulting con-figurations were averaged to obtain the symmetric component of shape variation. The new landmark coordinates generated by the GPA are called “Procrustes coordinates”.

To define shape features, I computed in MorphoJ principal components (PCs) from the n x 3k covariance matrix of Procrustes coordinates, where n is the number of samples and k is the number of landmarks; 3k represents the number of Procrustes coordinates. PC loadings computed in this analysis define as morphological score in this study.

All morphological measurements and analysis have been done by Dr. Peter Refki who is working as postdocs in our research group in Ploen. The method description above has been adopted from our previous paper (Pallares et al. 2015).

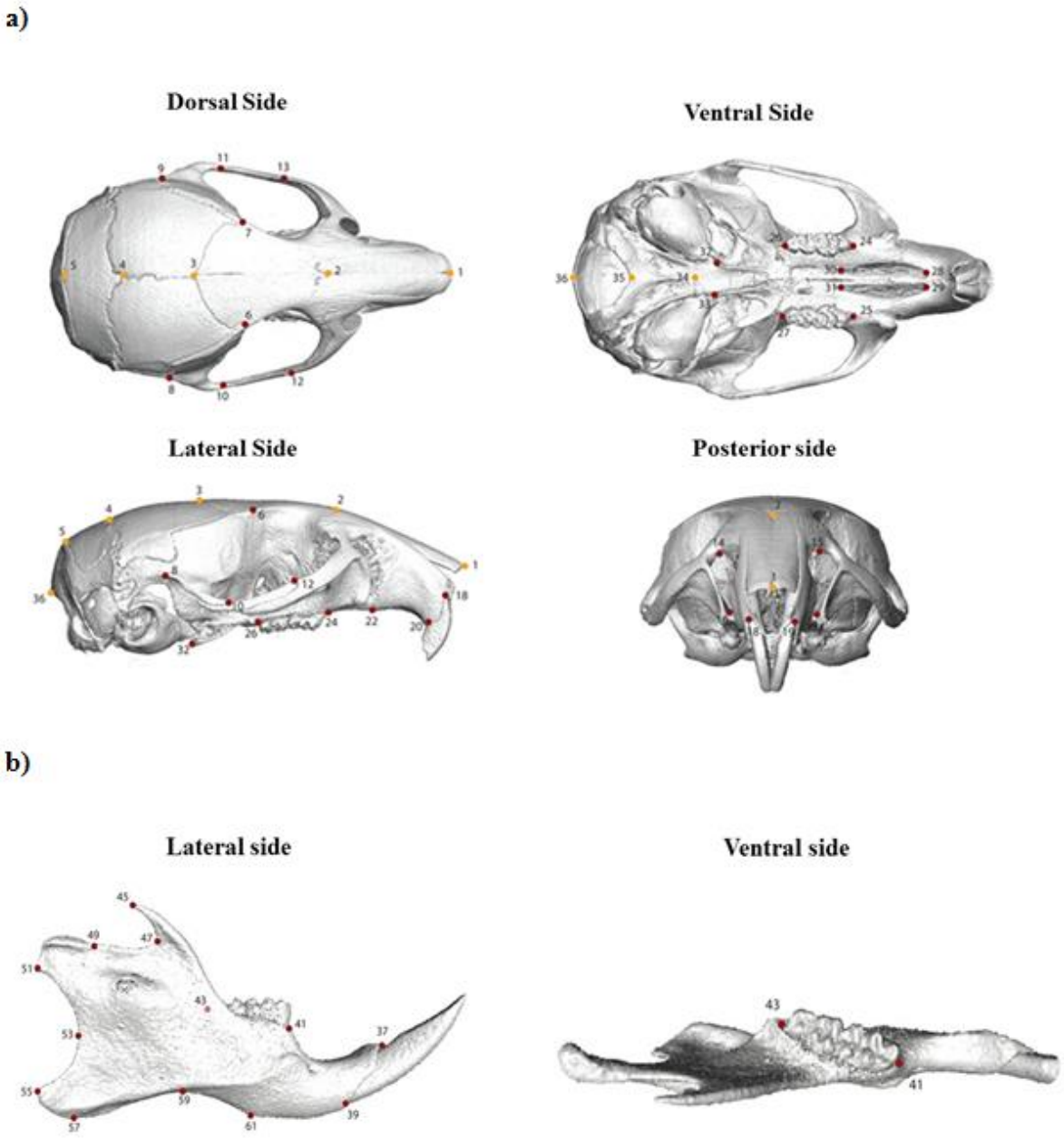


Figure 37: Three-dimensional landmarks positions in skull and mandible. a) Skull b) Mandible. This figure was provided by Dr. Peter Refki.

Table 4: list of Landmark positions in Skull and Mandible

	SKULL LANDMARKS
1	Nasal bones most anterior intersection
2	Nasal bones most posterior intersection
3	Frontal bones most posterior intersection
4	Parietal bones most posterior intersection
5	Interparietal middle posterior point
6	R Frontal-Parietal most anterior intersection
7	L Frontal-Parietal most anterior intersection
8	R Squa/zigosqua posterior contact
9	L Squa/zigosqua posterior contact
10	R Zigosqua/Jugal most anterior juntion
11	L Zigosqua/Jugal most anterior juntion
12	R Jugal/Zigomaxila most anterior juntion
13	L Jugal/Zigomaxila most anterior juntion
14	R Infraorb foramen upper edge
15	L Infraorb foramen upper edge
16	R Infraorb foramen lower edge
17	L Infraorb foramen lower edge
18	R Up incisor alveo most superior edge
19	L Up incisor alveo most superior edge
20	R Up incisor alveo most inferior edge
21	L Up incisor alveo most inferior edge
22	R Premax/maxila ventral juntion
23	L Premax/maxila ventral juntion
24	R Up first molar alveo most anter point
25	L Up first molar alveo most anter point
26	R Up last molar alveo most post point
27	L Up last molar alveo most post point
28	R Ant palatine foramen most anterior point
29	L Ant palatine foramen most anterior point
30	R Ant palatine foramen most posterior point
31	R Ant palatine foramen most posterior point
32	R Tip pterigoid process,most posterior point
33	L Tip pterigoid process, most posterior point
34	Occ/basisphenoid middle point intersection
35	Foramen magnum most anterior point, Basion
36	Foramen magnum most posterior point, Bregma

	MANDIBLE LANDMARK
37	R Low incisor alveolus anterior most point
38	L L inc alve anterior most point
39	R L incisor alveolus posterior most point
40	L L inc alve posterior point
41	R L first molar alveolus anterior most point
42	L L first molar alveolus anterior most point
43	R L last molar alveolus posterior most point
44	L L last molar alveolus posterior most point
45	R Tip of coronoid process (posterior)
46	L Tip of coronoid process (posterior)
47	R Posterior end of coronoid process
48	L Posterior end of coronoid process
49	L Posterior end of coronoid process
50	R Condyle articular surface most anterior point
51	L Condyle articular surface most anterior point
52	R Condyle most posterior point
53	L Condyle most posterior point
54	R Most concave point condyle/angular process
55	L Most concave point condyle/angular process
56	R Tip of angular processs (post)
57	L Tip of angular process (post)
58	R Most inferior point of angular process
59	L Most inferior point of angular process
60	R Ascending ramus ventral most concave point
61	R Ascending ramus ventral most concave point
62	L Ascending ramus ventral most concave point
63	R Alveolar region most inferior point
64	L Alveolar region most inferior point

Chapter 2

Functional analysis of the paternally expressed gene *Peg13*

Introduction

The study of Lorenc et al. (2015) found the *Peg13* transcript within the *Trappc9* gene region on Chromosome 15 as a second region which is paternally imprinted and highly differentiated between CB and MC house mouse populations. Hence, they proposed a potential role of *Peg13* in influencing paternal mate choice preference. However, a functional role of this region during mouse evolution and their paternal mate choice preference remains unknown. Therefore in this chapter, I investigated the biological role of *Peg13* in mouse behavior, based on analyzing the phenotype of a mouse carrying a knockout of the gene.

Peg13 (also known as paternally expressed 13) is a single exon gene located in the intron 16 of *Trappc9* on mouse Chromosome 15 (Fig. 38) (Smith et al. 2003; Ruf et al. 2007). The *Trappc9* gene is maternally expressed and encodes for the trafficking protein particle complex subunit 9. It may function in neuronal cell differentiation and plays a role in vesicular transport from endoplasmic reticulum to Golgi. This gene has been implicated in humans with intellectual disability disorders and *Peg13* may be connected with its regulation (Court et al. 2014).

Peg13 has a CpG island (CGI) over the promoter region and this CGI shows germ line–derived maternal methylation (gDMR) (Suzuki et al. 2004). It is consistent with the paternal expression of *Peg13*, so this CGI is thought to be essential for *Peg13* imprinting. Evolutionary analysis by using the evolutionary conserved regions in the corresponding intron of *Trappc9* gene on orthologous regions in human, mouse, dog, elephant, and opossum genomes revealed that this CGI is conserved in the expected position between human and mouse, but not in dog, elephant, and opossum, suggesting that the CGI is not conserved in Metatheria, Afrotheria, and Laurasiatheria. It seems that the genomic region corresponding to *Peg13* was inserted into the genome of the Euarchontoglires ancestor (Fig. 39), (Suzuki et al. 2004). It remains unclear whether the inserted DNA itself was CpG rich or whether CpG sequences were accumulated after the insertion event. However, in both cases, the insertion of DNA was potentially the trigger for the emergence of a novel CGI forming the gDMR in this locus (Suzuki et al. 2004).

Although this CGI is conserved between human and mouse, the *Peg13* sequence is not highly conserved between these two species, suggesting rapid sequence evolution of *Peg13* transcripts (Suzuki et al. 2004; Court et al. 2014). This is also how our group (Lorenc et al.

2015) initially has detected *Peg13* as a fast evolving gene between CB and MC house mouse populations.

Peg13 is highly expressed in brain, but its biological role is unknown. For geneticists, the targeted deletion of a gene in animal model like mouse provides an important means to determine the biological role of a gene (Hall et al. 2009). So to investigate the role of *Peg13*, here I conducted a functional study on *Peg13* in mouse, based on a transgenic mouse model.

Figure 38: Schematic view of *Peg13* mouse gene organization. *Peg13* is a single exon gene located in the intron 16 of *Trappc9* gene on mouse Chromosome 15. Red transcripts are maternally expressed, blue are paternally expressed.

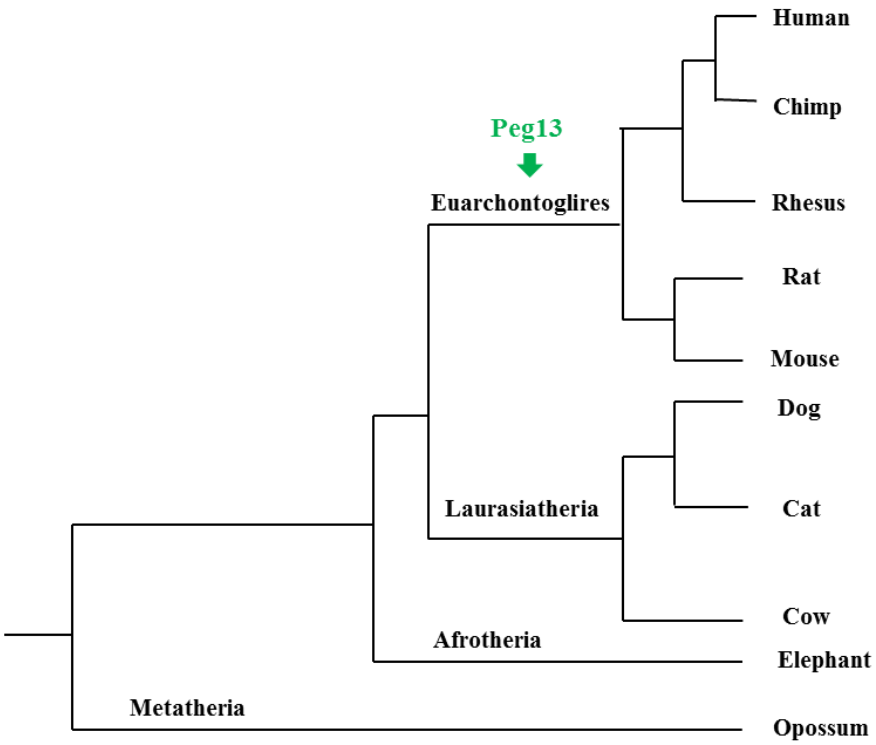


Figure 39: Timing of the CGI emergence in the *Peg13* locus during mammalian evolution. The arrowhead represents the acquisition of the CGI according to Suzuki et al. (2004).

Results

Peg13 and *Trappc9* Expression

Peg13 has two transcripts (Fig. 40). The short transcript is fully contained within the long transcript. The difference is only at the 5´-end, but they share the same polyadenylation site.

To get a general overview of *Peg13* expression across different tissues in mouse, RNAseq data of the ENCODE mouse dataset from brain, bladder, kidney, lung, liver, placenta, testis, thymus, stomach, small intestine, large intestine, duodenum, adrenal glands, adipose, heart, colon, spleen, gonadal fat pad (G.F.P), mammary glands and ovary was downloaded. Raw RNAseq data from each tissue were mapped to the mouse reference genome (mm10) (Genome Sequencing Consortium 2002) by Hisat2 (Kim et al. 2015). The mapped data were used for gene expression analysis (see Methods).

Figure 41 shows a heat map diagram of *Peg13* and *Trappc9* expression across different tissues. Both genes show highest expression in brain and *Peg13* has an approximately 5-fold higher expression level than *Trappc9*. The other tissues show low to moderate expression of these two genes (Fig. 41). Interestingly, correlation analysis of *Peg13* and *Trappc9* expression across different tissues showed that expression of Peg13 and its surrounding gene, *Trappc9*, is highly correlated (Fig. 42b). *Kcnk9* is another imprinted gene in this cluster which also has maternal expression (Fig. 42a). Further analysis on *Kcnk9* showed that this gene does not have a broad expression and is mainly expressed in brain and adrenal glands. To check whether *Kcnk9* is co-regulated with *Peg13*, a correlation analysis on expression of these two genes in hindbrain, midbrain, forebrain, frontal cortex and cerebellum was run. *Kcnk9* showed high correlation with *Peg13* expression (Fig. 42c).

Chrac1 is another gene which is located near *Peg13* (Fig. 42a), however according to the atlas of genomic imprinting in mouse and the results of Lorenc et al. (2015), this gene is not imprinted . This gene did not show any correlation to Peg13 expression across different tissues (Fig. 42d).

Figure 40: *Peg13* isoforms. See text for more information. *Peg13* has two transcripts. The short transcript is fully contained within the long transcript. The difference is only at the 5´-end, but they share the same polyadenylation site. The figure was generated and provided by genOway.

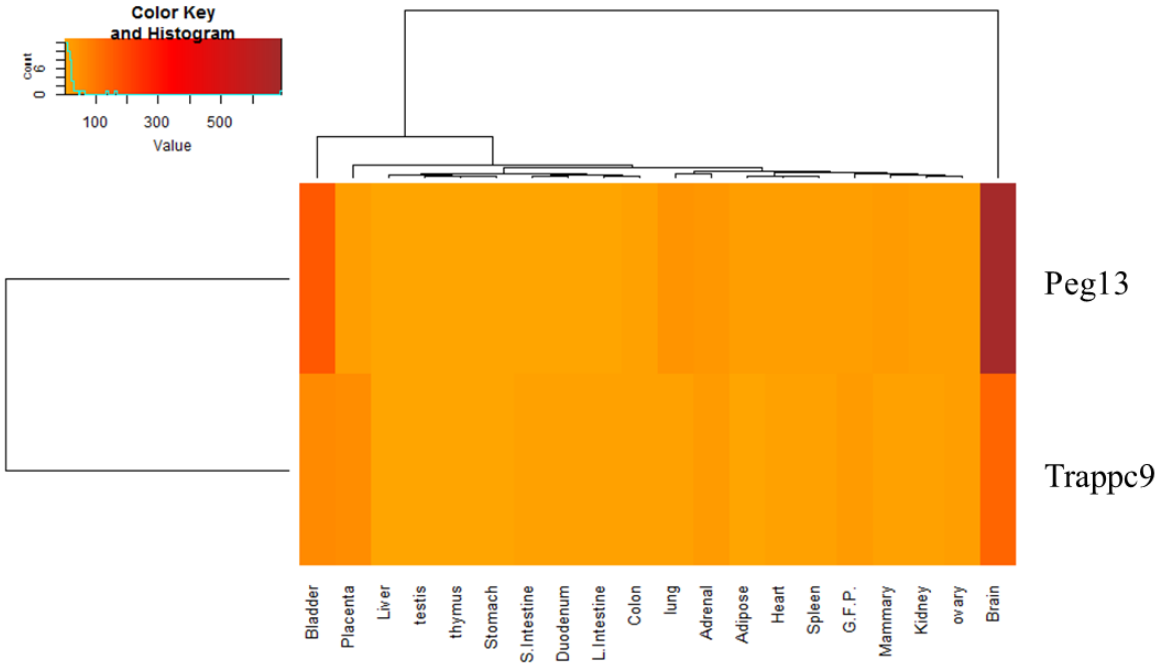


Figure 41: Heat map diagram of *Peg13* and *Trappc9* expression across different tissues. The length of each branch in the dendrogram shows the degree of similarity between tissues based on the *Peg13* and *Trappc9* expression. The color bar on the left side demonstrates gene expression level. Transcripts Per Million (TPM) was used as the value for gene expression.

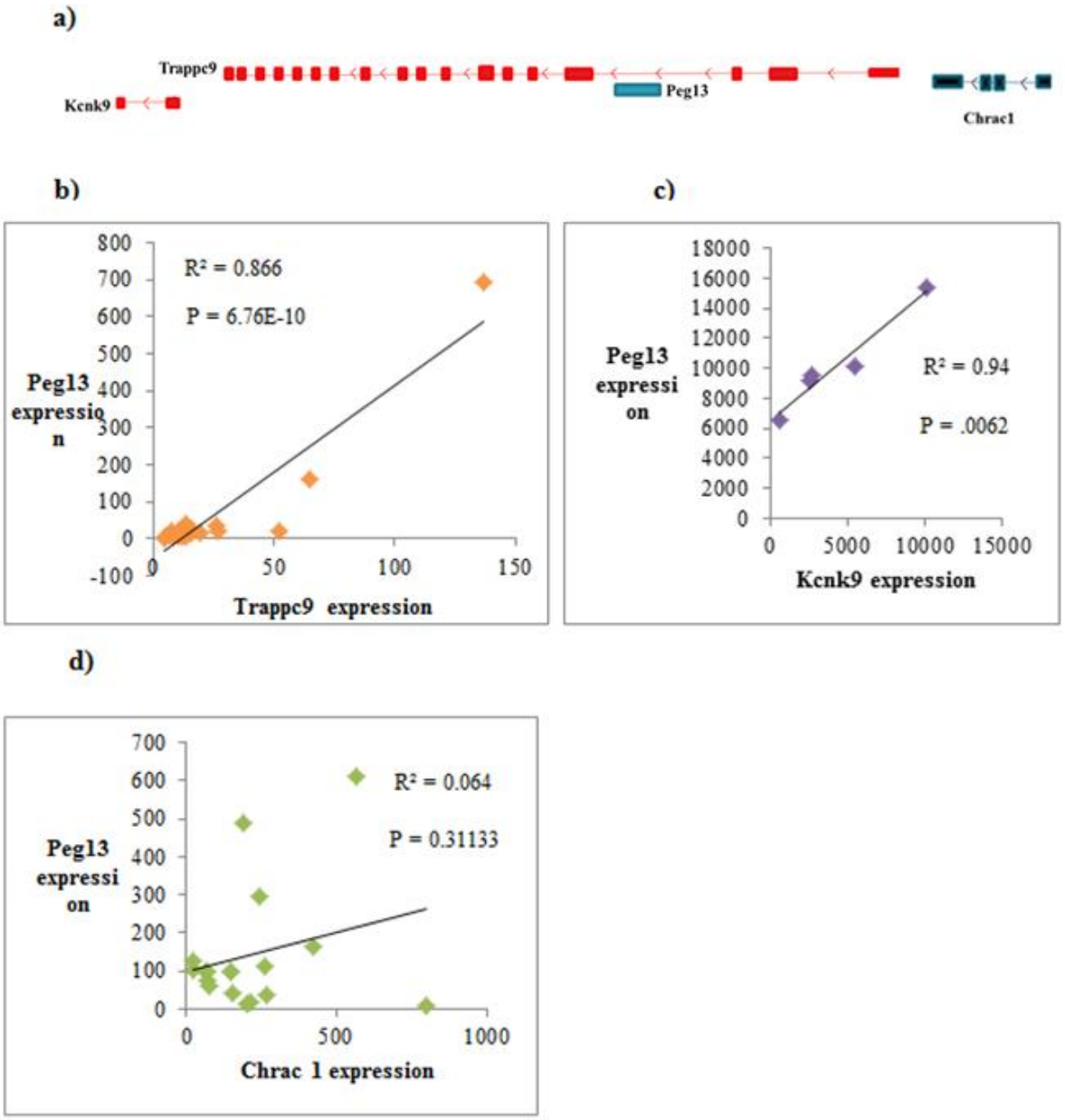


Figure 42: Expression correlation analysis of the genes around the *Peg13* imprinted domain. a) Schematic representation of the *Peg13* imprinted domain. Red transcripts are maternally expressed, blue are paternally expressed and black are expressed from both parental alleles. b) Correlation analysis of *Peg13* and *Trappc9* expression across different tissues. Dots are calculated TPM of these two genes from each tissue. Brain, bladder, kidney, lung, liver, placenta, testis, thymus, stomach, small intestine, large intestine, duodenum, adrenal glands, adipose, heart, colon, spleen, gonadal fat fad (G.F.P), mammary glands and ovary were used for this analysis. c) Correlation analysis of *Peg13* and *Kcnk9* expression in hindbrain, midbrain, forebrain, frontal cortex and cerebellum d) Correlation analysis of *Peg13* and *Trappc9* expression across different tissues same as part b.

Peg13 expression in the brain

Results from the last part showed that *Peg13* has the highest expression in brain tissue (Fig. 41). There are three major divisions of the brain with each division having specific functions. The major divisions of the brain are the forebrain, midbrain and hindbrain (Fig. 43). RNAseq analysis (from Encode) from different parts of the mouse brain showed that both *Peg13* and *Trappc9* have higher expression in hindbrain and midbrain than forebrain (Fig. 44).

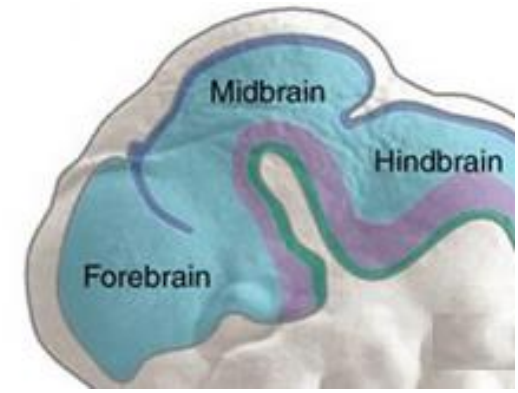


Figure 43: Schematic view of three major divisions of mouse brain. This figure was adopted from Sanes et al. (2011).

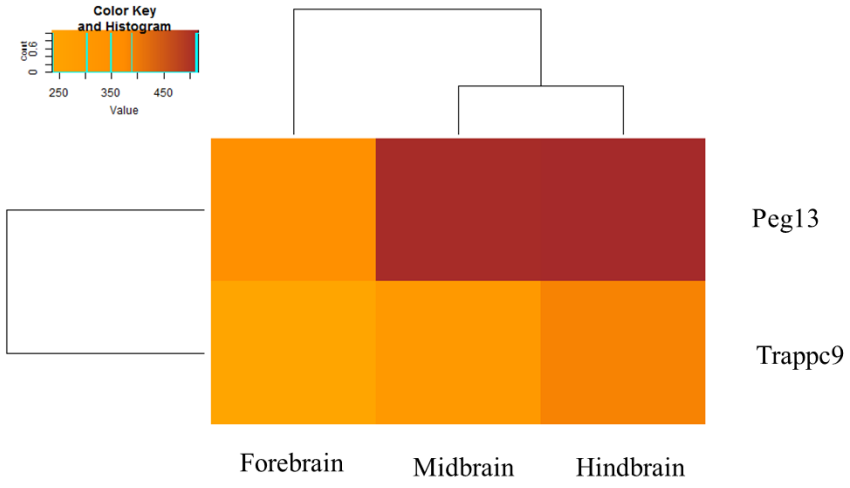


Figure 44: Heat map diagram of *Peg13* and *Trappc9* expression between three major divisions of the mouse brain. The length of each branch in the dendrogram shows the degree of similarity between three major divisions of mouse brain based on the *Peg13* and *Trappc9* expression. The color bar on the left side demonstrates gene expression level. Transcripts Per Million (TPM) was used as the value for gene expression.

Peg13 expression during developmental stages

Brain is a complex organ which has many different parts to act as the control center of the body. Gene expression is one of the main molecular processes regulating the differentiation, development, and functioning of cells and tissues (Naumova et al. 2013). The transcriptome of the brain like other organs also changes during development to form and determine functional specialization of brain regions (Liscovitch et al. 2013).

Hence to investigate role of *Peg13* in brain development, RNAseq data available in our group from 4 different stages of embryo development were analyzed. *Peg13* has the highest expression at day 12.5 of embryogenesis. Expression of *Trappc9* increases a bit at day 12.5, but compared to *Peg13* is around 5-fold less (Fig. 45).

According to the mouse brain development timeline (Finlay et al. 1995), day 12.5 represents a peak of neurogenesis for ventral posterior nucleus and ventrobasal nucleus of the thalamus, nucleus of lateral olfactory tract, cortical layer VI and preoptic nucleus. So results from this part suggest that *Peg13* may have an important role in the formation and differentiation of these parts in the brain.

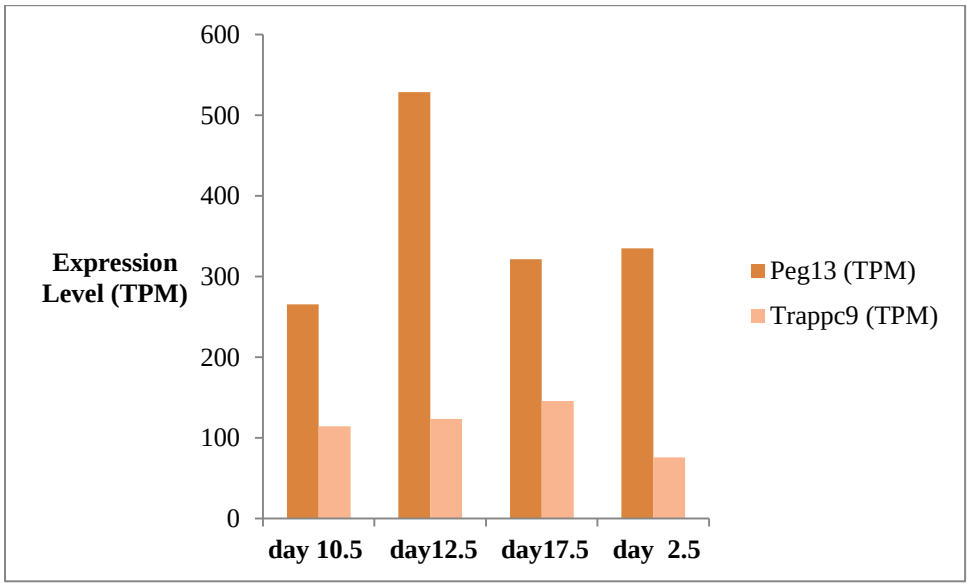


Figure 45: *Peg13* expression during mouse brain development. Day 10.5, 12.5 and 17.5 (E- day) from embryonic stages and 2.5 day from postnatal stage (P- day) were selected for this analysis. Transcripts Per Million (TPM) was used as the value for gene expression.

A possible PEG13 protein

Sequence analysis on the long transcript of PEG13 suggested that this transcript has a potentially translated short open reading frame (ORF). The ORF is initiated by the first ATG and has a good Kozak consensus sequence (Fig. 46). The same reading frame can also be found in rat. The short transcript does not appear to have credible ORFs.

Ribosome profiling is a deep-sequencing-based method which facilitates the detailed measurement of translation globally and in vivo (Ingolia et al. 2009). This approach works basically according to the fact that a translating ribosome strongly protects about 30 nucleotides of a mRNA from nuclease activity (Wolin et al. 1998 and Steitz 1969). Sequencing of these ribosome-protected fragments, termed ribosome footprints, could provide a relatively precise record of the position of translation activity and protein synthesis in the genome. Measuring the density of protected fragments on a given transcript provides a proxy for the rate of protein synthesis.

To assess whether the PEG13 long transcript is truly translated, ribosome profiling data analysis was performed. To this end, published ribosome profiling data (Castañeda et al. 2014; Gonzalez et al. 2014) from brain, testis, heart and skeletal muscle tissues were analyzed. Results from this part revealed that *Trappc9* has a high rate of protein translation in brain and testis and a moderate rate in muscle and heart (Fig. 47). Interestingly, the short ORF in the long transcript of *Peg13* in brain and especially in hippocampus is translated and could synthesize a short protein. However there was no evidence for *Peg13* translation in testis, heart and skeletal muscle (Fig. 47). These results suggest that *Peg13* is not a simple noncoding RNA as it had been described (Smith et al. 2003; Suzuki et al. 2004; Ruf et al. 2007) and it can synthesize a short peptide in mouse brain.

.

Figure 46: Kozak consensus sequence and short ORF of the long PEG13 transcript. The purple arrow designates the Kozak consensus sequence and the orange one shows the ORF which is initiated by ATG.

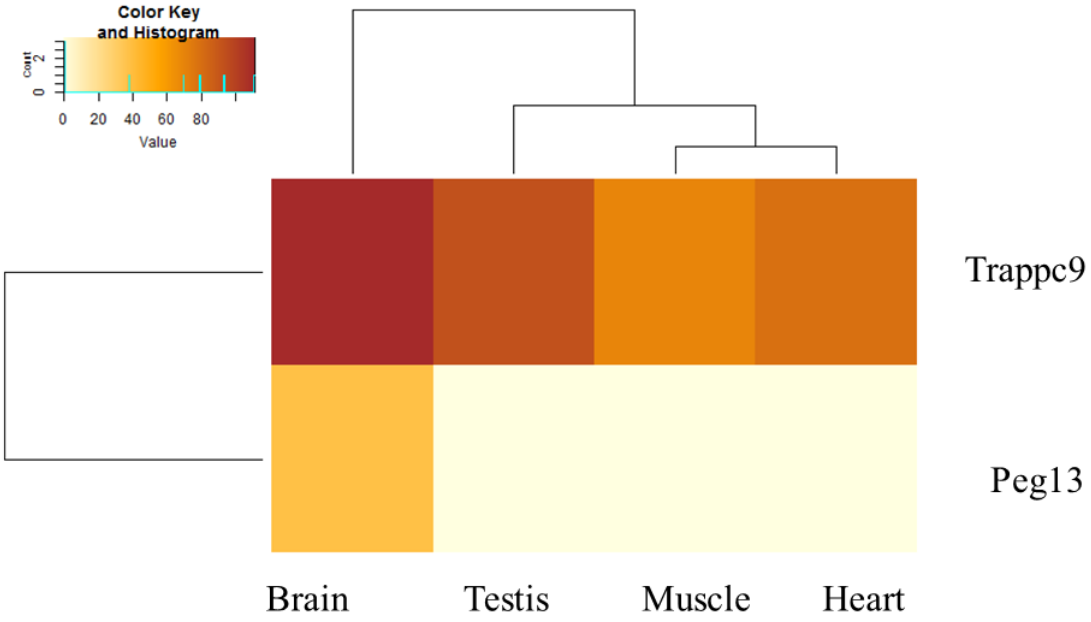


Figure 47: Heat map diagram of PEG13 and TRAPPC9 protein synthesis across different tissues. The length of each branch in the dendrogram shows the degree of similarity between different tissues based on PEG13 and TRAPPC9 protein level. The color bar on the left side demonstrates protein level (see Method section for protein level calculation).

Generation of Knock-out mice

The generation of knock-out mice was initiated in cooperation with the company genOway. They analyzed the gene region and found that the genomic regions across *Peg13* is rich in DNaseI-protected regions, a CpG island, CTCF binding sites and several transcription factor binding regions (Fig. 48). Modifications in these regions could result in dysregulation of both *Peg13* and the co-regulated genes *Trappc9* and *Kcnk9* (see above). This could imply that the deletion of the whole *Peg13* gene would potentially result in a triple knock-out: *Peg13*, *Trappc9* and *Kcnk9* genes.

Consequently, it was decided to apply a double strategy which consists of a constitutive deletion of the *Peg13* 3'-region and the option to delete the whole gene. The respective 3' region is devoid of predicted regulatory elements. The optional deletion of the whole gene was planned to be mediated by Cre recombinase.

Cre recombinase is a 38 kDa protein from the bacteriophage P1 that mediates intramolecular and intermolecular site-specific recombination between two loxP sites (locus of X-over of P1) (Bouabe et al. 2013). The loxP sequence is 34 bp long and consists of two 13 bp inverted repeats which is separated by an 8 bp non-palindromic (asymmetric) sequence. Two loxP sequences in same orientation mediate excision of the intervening DNA between the sites by Cre recombinase (Bouabe et al. 2013).

So in this study the whole *Peg13* 5'-region was flanked by loxP sites enabling its possible subsequent whole-body or tissue-specific deletion by Cre recombinase (Fig. 49) (see Appendix 5 for more details).

Figure 48: Bioinformatics assessment of the presence of regulatory elements within the targeted region. The targeted region is depicted as the *Peg13* locus, the DNAse-I protected region is indicated as blue box. Transcription Factor Binding Sites (TFBS) are indicated as green boxes. The CpG Island and DNA methylation region are indicated in red and black, respectively. The figure was generated and provided by genOway.

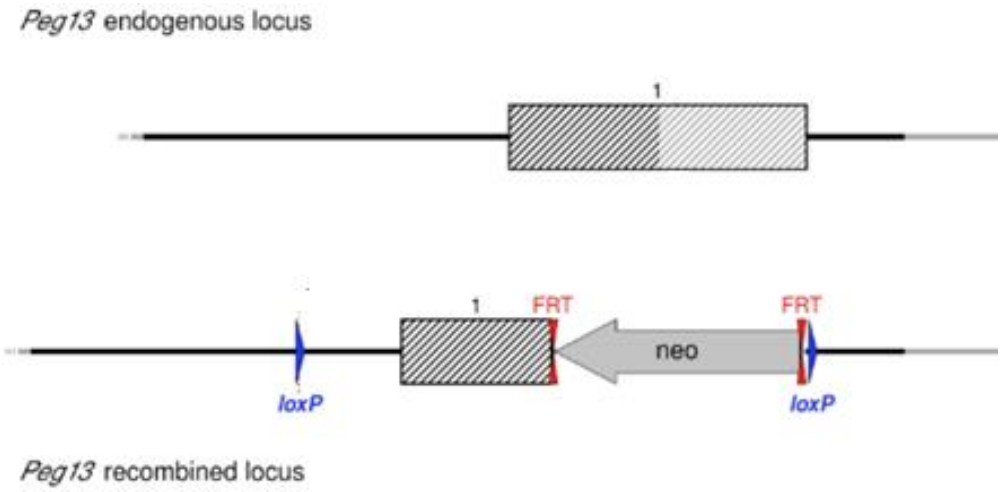


Figure 49: Schematic view of *Peg13* wild type and recombinant alleles. The recombinant allele consists of the *Peg13* 3'-deletion and the *Peg13* 5'-region which is flanked by loxP sites enabling its subsequent whole-body or tissue-specific deletion by Cre recombinase. LoxP sites are represented by blue triangles and FRT sites by double red triangles. Neo: neomycin positive selection cassette. Boxes represent exons and solid lines represent intronic sequences. The figure was generated and provided by genOway.

***Peg13* 3'-knockout mouse**

Mice carrying the deletion of the *Peg13* 3'-region were generated by genOway and 3 heterozygous mice (2 female and 1 male) were obtained from them. As a first step, I bred the mice to generate wild type and homozygous mouse individuals. These breeding experiments allowed already some insights. As Table 6 shows, pairs where both male and female carried the homozygous *Peg13* 3'-deletion had very low mating success. Of 33 pairs only twelve females became pregnant and only four of these could have and keep their offspring. Eight females ate their pups right after birth. So overall, from 33 pairs, we have only twelve percent rate of success in homozygous/homozygous pairs. The corresponding wild type pairs, representing the background in which the manipulation was done, did not have any problem in breeding (Table 6). Of the heterozygous/heterozygous pairs about half were successful in their mating and offspring production (Table 6). Note that it was not possible for the heterozygous animals to infer whether they received the deleted allele from the father or the mother. If they had received it from the father, it would be silenced, due to paternal expression only and this might explain why only half of the pairs were successful. Hence, these initial observations suggest that the *Peg13* gene may not be essential (i.e. homozygous animas are viable), but that it has a potential role in mating and breeding behavior.

Table 6: rate of success in *Peg13* 3'-knockout compared to heterozygous and homozygous mice.

Female Genotype	Male Genotype	Total Pairs	Total Pregnancy	# pairs who ate their pups	Rate of success
+/+	+/+	15	15	0	100%
+/-	+/-	25	19	6	52%
-/-	-/-	33	12	8	12%

Behavioral tests on *Peg13* 3'-region knockout mice

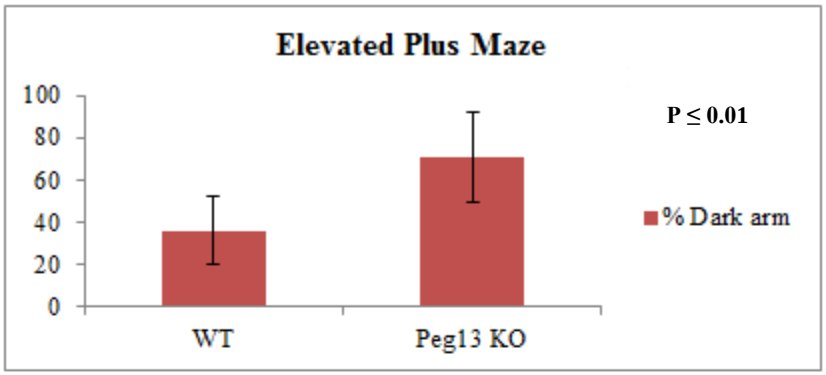
Given the possibility that behavior is impaired in *Peg13* 3'-region knockout (3'-KO) mice, I did behavioral anxiety tests on 40 homozygous and wild type mice each. *Peg13* 3'-KO mice spent more time (around two-fold) in the dark arm of the elevated plus maze compare to wild type mice (Fig. 50a). Results from the open field test also showed that mutant mice stayed next to the wall more than wild type mice (Fig. 50b). During the whole dark light box experiment, *Peg13* 3'-KO mice stayed in the dark box and they did not come out from the box (Fig. 50c). All of these observations suggest that even the partial deletion of *Peg13* might increase the anxiety type behavior in mice.

However, as pointed out in the first chapter of the thesis, the anxiety test could be influenced by SNORD115/116 copy numbers, even in inbred strains. So it is essential to check SNORD copy number as a co-variable for each mouse.

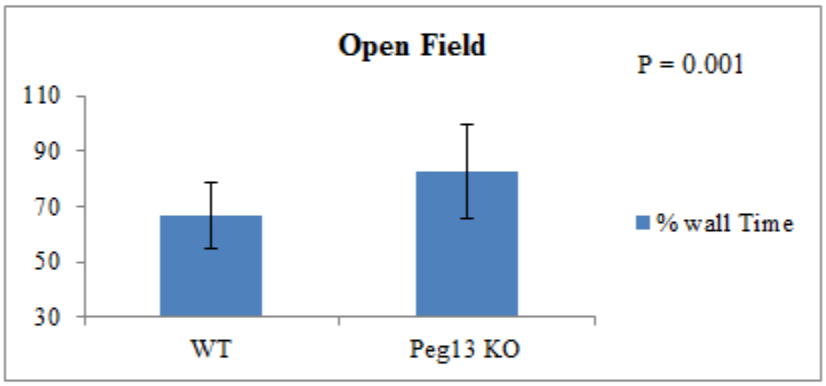
SNORD copy number was determined for all mice in the experiment, as described in the first chapter. This showed that indeed there is a significant difference between *Peg13* 3'-KO and wild type mice, with the former having on average a higher copy number (Fig. 51). Higher numbers of these SNORDs increase anxiety behavior (chapter 1). Hence, all the scores from the elevated plus maze, open field and dark light box tests had to be normalized accordingly. To this end, the strength of the association between SNORDs copy number and measurement from each behavioral test was calculated by Pearson's correlation. In fact, the value of correlation for each behavioral test explained the effect of SNORDs copy number on the measurement of that behavioral test. Therefore the raw data from behavioral tests were adjusted according to the correlation value of each test to SNORDs copy number.

Figure 52 shows the results after normalization. The results from the elevated plus maze do not change much, while the open field still shows a significant difference, but with a higher p-value (Fig. 52a, b). The results from the dark light box become non-significant after this normalization (Fig. 52c). Hence, although these results suggest that *Peg13* 3'-KO has still an effect on anxiety-like behavior, it is less strong than the first test would have suggested. This proves that the determination of SNORD115/116 copy numbers needs to be taken into account in behavioral tests as co-variable.

a)



b)



c)

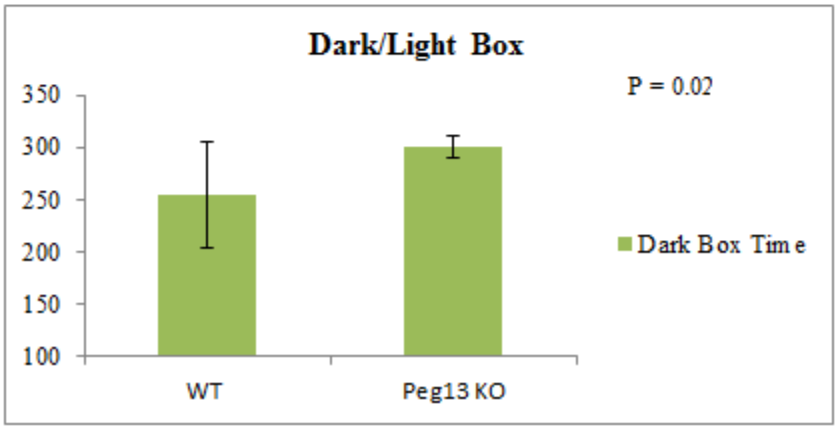


Figure 50: Results of the behavioral tests on *Peg13* 3'-KO mice. a) Elevated plus maze, b) Open field and c) Dark light box. Error bar represents the variability of measurements within each group .t-test was used to determine whether there is a significant difference between Wild Type (WT) and Knockout (KO) mouse.

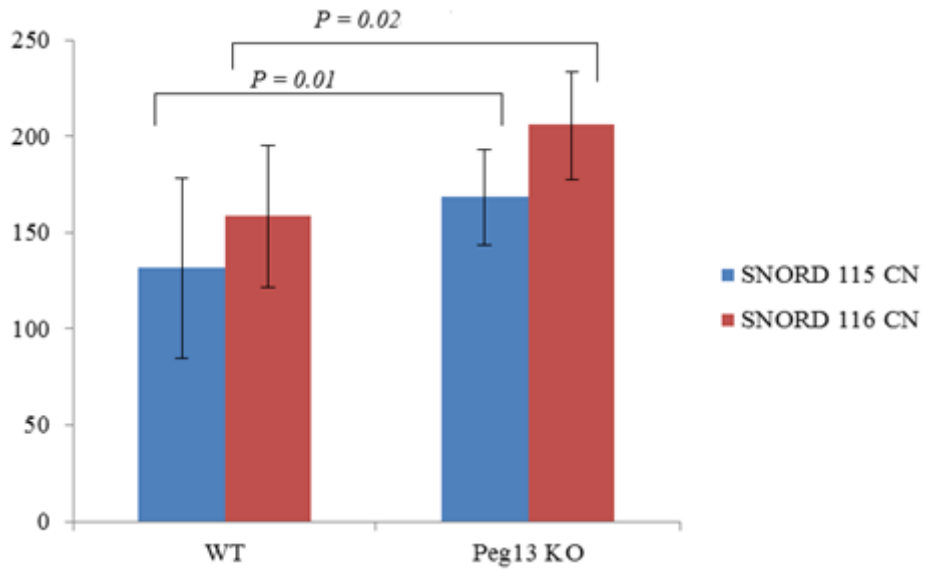
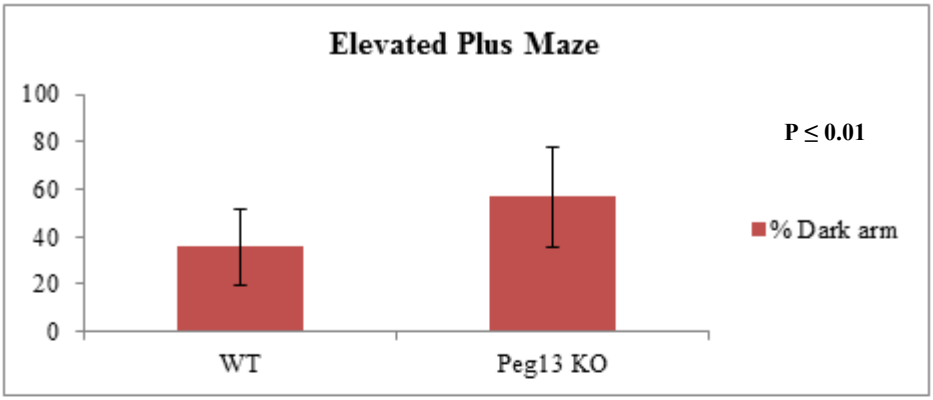
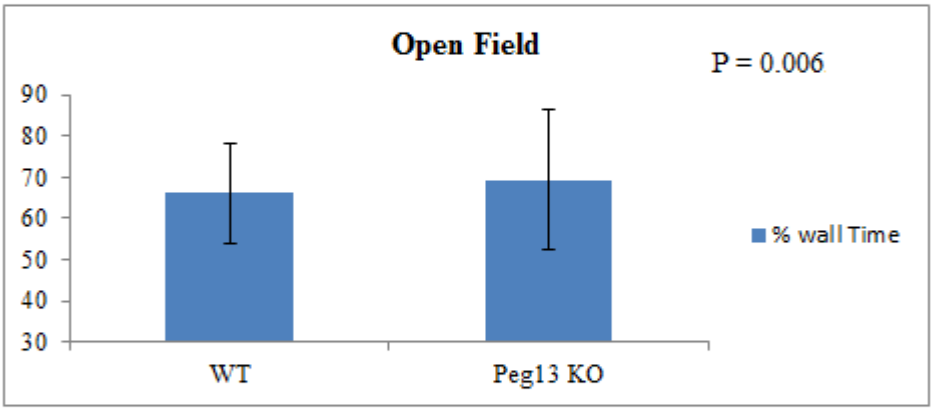


Figure 51: SNORDs Copy Number Variation between *Peg13* KO and wild type (WT) mice. Error bar represents the variability of the means of SNORDs CN within each group .t-test was used to determine whether there is a significant difference between WT and KO.

a)



b)



c)

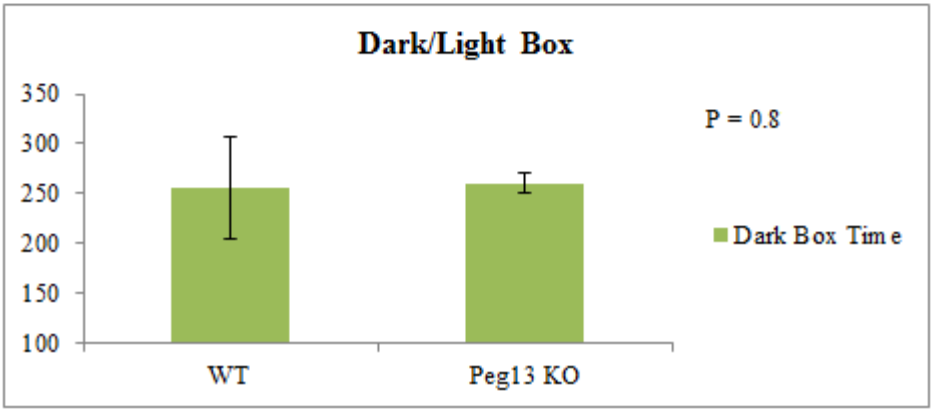
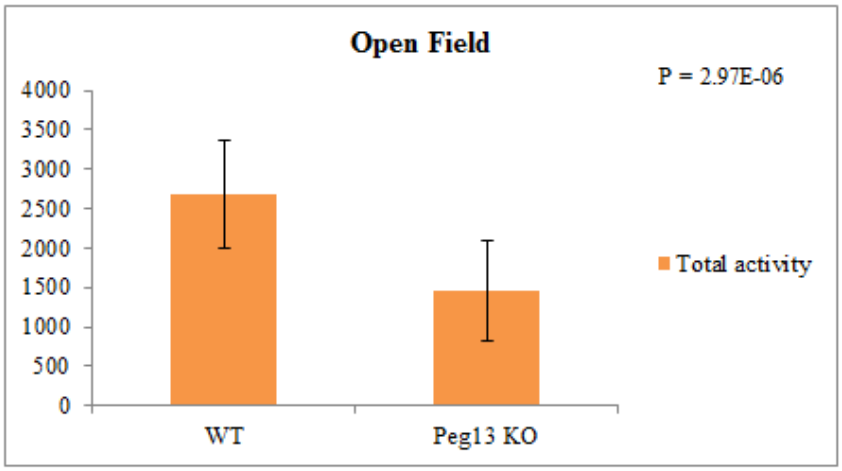


Figure 52: Results of the behavioral tests on *Peg13* 3'-KO mice after correction for SNORD115/116 copy number. a) Elevated plus maze, b) Open field and c) Dark light box. Error bar represents the variability of measurements within each group .t-test was used to determine whether there is a significant difference between Wild Type (WT) and Knockout (KO) mouse .

The open field test is also an experiment which is used to assay general locomotor activity of mice and rat. To this end, the total movement distance of mice during the five minute test is measured and used as an activity index. As Figure 53a shows, *Peg13* 3'-KO mice traveled less distance in the open field apparatus compared to wild type mice. To assess general curiosity of *Peg13* 3'-KO mice, the novel object test was performed. Results from this test revealed that these mice are significantly less curios than wild type mice (Fig. 53b).

a)



b)

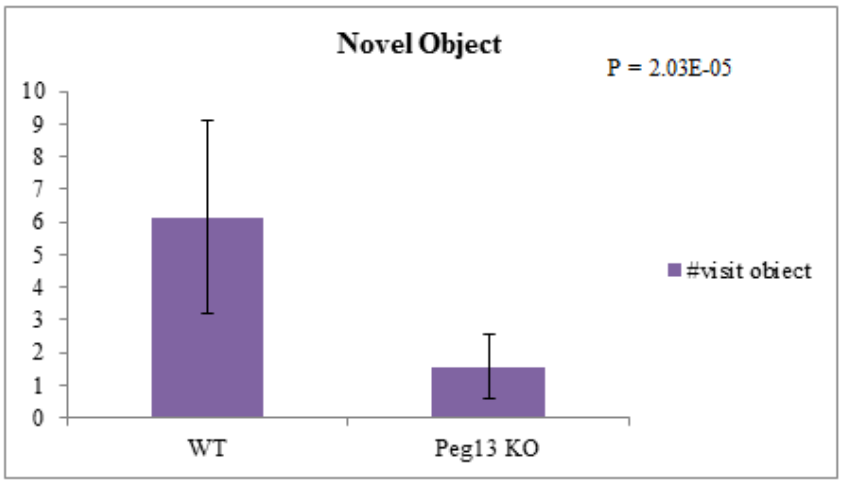


Figure 53: Results of the behavioral test on *Peg13* 3'-KO mice to assess their activity and curiosity. a) Open Field test for activity b) Novel Object test for curiosity. Error bar represents the variability of measurements within each group .t-test was used to determine whether there is a significant difference between Wild Type (WT) and Knockout (KO) mouse.

RNAseq analysis on *Peg13* 3'- region knockout mice

Measuring transcription changes through RNAseq is a suitable approach to unravel the gene network associated with a gene when knocked out (Wang et al 2009). To this end, 10 *Peg13* 3'-KO and 10 wild type mice at the age of 16 weeks were selected. RNA from their brain was extracted and after library preparation these were subjected to Illumina sequencing. Reads were mapped to the mouse mm10 reference genome (Genome Sequencing Consortium 2002) by using Hisat2 as described in the methods section. Differential expression analysis was done using the DESeq2 package in the R environment. Genes with an adjusted P-value (Padj) < 0.05 were considered as significantly differentially expressed.

The log2fold change against the mean normalized counts was plotted and is shown in Fig 54. Red dots represent those genes that are significantly differentially expressed at a 10% false discovery rate (FDR). These significant genes were filtered for further analysis. 479 genes showed a significant change in their expression by *Peg13* 3'- region deletion (Appendix 3).

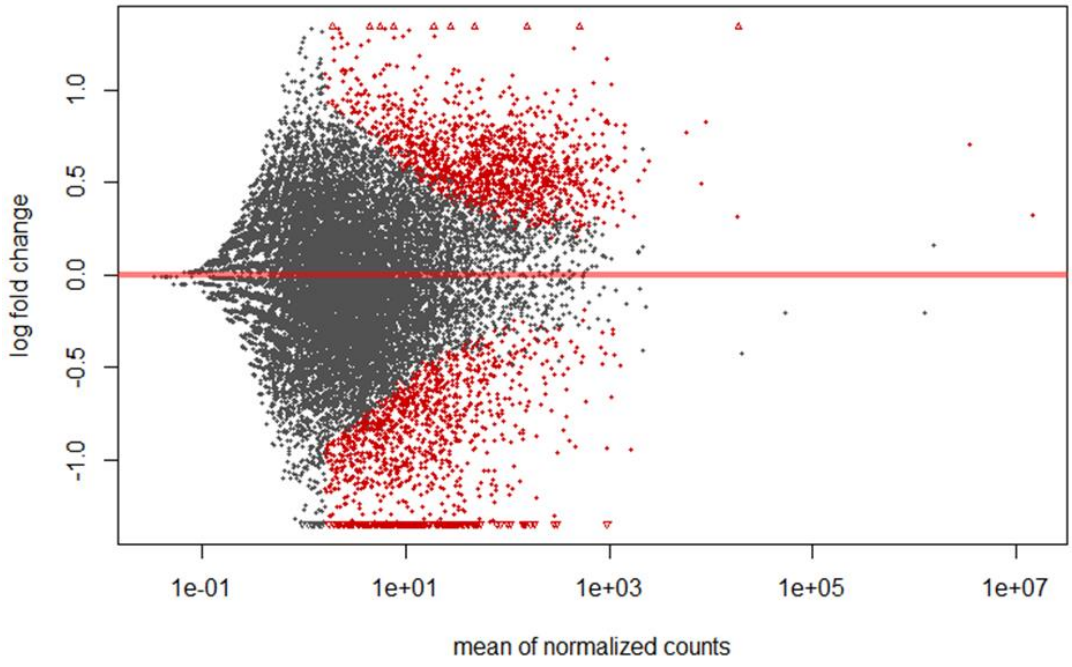


Figure 54: Plot of log2fold change against the mean normalized counts. Red dots represent those genes that are significant at 10% false discovery rate (FDR).

GO analysis

To gain insight and have better interpretation of the differentially expressed genes, GO and KEGG pathway enrichment analyses were performed using DAVID online tools (see methods in chapter one). The results from the GO analysis revealed an enrichment of genes which are important in GTPase activity, metabolic pathway, synapse formation, immunity and inflammatory response, proliferation, differentiation and developmental pathways (Wnt/beta-catenin signaling pathway) and also genes which have been reported for intellectual disability diseases. This suggests that *Peg13* 3'-KO has indeed a major impact on regulatory networks in the brain.

Discussion

In this chapter, I started to conduct a functional study on *Peg13* to investigate a possible biological role of this gene in mouse behavior. I summarized the information about expression of *Peg13* and genes which are located in its imprinted cluster, *Kcnk9* and *Trappc9*, in different tissues and also during different developmental stages of brain formation. Further, by analyzing *Peg13* 3'-KO mice, I investigated the functional role of this gene in mouse behavior. During the time of my thesis I was not able to analyze phenotypic effects of the complete knockout of *Peg13*, but already the partial knockout yielded interesting insights.

Expression of *Peg13, Kcnk9* and *Trappc9* are co-regulated

The results from the correlation analysis suggest that expression of *Peg13, Kcnk9* and *Trappc9* are co-regulated (Fig. 42). Imprinted genes are often co-regulated (Verona et al. 2003) and tend to show coordinated expression. This may reflect an evolutionary history where natural selection has built co-inheritance of interacting alleles (Hartl et al. 1997; Wolf 2013) to make co-adapted associations among beneficially interactors.

The potential regulatory regions across the *Peg13* region in the mouse genome includes a DNaseI-protected region, a CpG island, CTCF binding sites and several transcription factor binding regions. It seems this 5'-part of *Peg13* constitutes the imprinting control (IC) region which may regulate the expression of this whole imprinted cluster (Fig. 48).

Peg13 is not a simple non-coding RNA

Peg13 has been referred to as long non-coding RNA in several studies (Smith et al. 2003; Suzuki et al. 2004; Ruf et al. 2007; Court et al. 2014). Long non-coding RNAs are classified as such in databases because, according to a number of criteria such as lack of a long ORF, the absence of amino acid sequence conservation, and the lack of known protein domains, they are unlikely to encode functional proteins. These criteria include the lack of a long ORF, the absence of amino acid sequence conservation, and the lack of known protein domains (Harrow et al. 2012). In many cases, the transcripts containing ORFs will be classified as non-coding, if the ORF is not well conserved across different species or if the ORF is very small. Here I showed that *Peg13* has two transcripts and the short transcript is fully contained within the long transcript (Fig. 40). The difference is only at the 5´-end, but they share the

same polyadenylation site. Both short and long transcripts are expressed in all tissues, but the short one has an around 10-fold higher level of expression than the long one. Interestingly, sequence analysis on the long transcript of *Peg13* suggested that this transcript has a potentially translated short open reading frame (ORF). The ORF is initiated by the first ATG and has a good Kozak consensus sequence (Fig. 46). Then my further analysis by using ribosomal profiling data confirmed that indeed the long transcript synthesizes a short peptide in the brain (Fig. 47) which so far has not been described. This short ORF is conserved in mouse and rat, but not in humans. In fact long non-coding RNAs may act as a repository for the synthesis of new peptides, new function and contribute to fast speciation (Ruiz-Orera et al. 2014).

Behavior of *Peg13* 3'-region knockout mice

Peg13 3'-KO mice showed problems in reproduction with either failing mating or becoming pregnant, or when they were pregnant ate their offspring after birth (Table 6).

Peg13 expression is increased significantly at day 12.5 of mouse embryogenesis (Fig. 45). According to the mouse brain development timeline, day 12.5 is the peak of neurogenesis for the preoptic nucleus (Finlay et al. 1995). In parallel brain transcriptome analysis in this study suggests that partial deletion of the *Peg13* gene could significantly change expression of many genes which are involved in proliferation, differentiation and brain developmental pathways. These two results may suggest a role of the *Peg13* gene in preoptic area development.

The preoptic area (also called POA) is part of the anterior hypothalamus. Male rats with lesions in POA showed a complete loss of copulatory behaviors (Paredes 2003). Conversely, electrical stimulation of this area triggers male copulatory behavior, as measured by decreases in the latency to ejaculate (Paredes 2003)

Studies using female Syrian hamsters have shown that the POA is important for sexual odor preference. While control females will investigate male odors more than female odors, those with bilateral lesions to the POA showed no difference in odor preference (Graham et al, 2013).

The POA has been implicated in parental care in both males and females. In rats, oxytocin and vasopressin are associated with maintaining maternal care by local release in the POA

(Bosch et al, 2010). Problem in the POA disrupt the maternal behavior, nest-building and pup retrieval (Numan 1988). Also in fathers, studies have shown that when they receive ultrasonic or pheromone cues from their mates, their c-fos expression in the POA further increased, suggesting that rat paternal behavior is mediated through the POA (Zhong et al. 2014).

So these data altogether suggest that *Peg13* 3'-KO mice may have abnormalities in the preoptic area and that this impairs their sexual and parental behavior, however further experiments, including detailed histological analyses will be required to elucidate this more deeply.

Anxiety behavior of *Peg13* 3'-KO mice

Peg13 3'-KO mice showed significantly higher levels of anxiety and lower activity and curiosity compared to wild type mice (Fig. 52, 53). RNAseq analysis on brain of *Peg13* 3'-KO mice revealed several genes which are important for synapse formation, GTPase activity, immunity and inflammatory response and also metabolic pathway. This suggests that *Peg13* 3'-KO mice may have problems in these pathways.

The nervous system consists of billions of neurons. The capacity of a neuron to innervate and function within a network is mediated via the synapses. So synapse formation is essential for brain function (Colón-Ramos 2009). Defects in synapse formation leads to several brain disorders and causes abnormal behavior (Aincy et al. 2017, Fujita et al. 2017).

GTPases are a large family of hydrolase enzymes that can bind and hydrolyze guanosine triphosphate (GTP). This hydrolyzation is important for signal transduction at the intracellular domain of transmembrane receptors (Cansado et al. 2018). Several studies suggest a strong link between GTPase activity and mouse behavior such as anxiety and intellectual ability (Das et al. 2017, Carlson et al. 2016, Shahani et al. 2017, Musilli et al. 2013).

Depression and anxiety are a common symptom observed in patients with inflammatory disease. Increasing lines of evidence indicates that immune-inflammatory responses are involved in mood and behavior regulation (Haj-Mirzaian et al 2017).

Energy metabolism is the process of generating energy (ATP) from nutrients. This process is essential for normal behavior and locomotor activity. Therefore any disruption in metabolic networks alters locomotor activity and subsequently other behavior (Chapnik et al. 2013).

However, the RNAseq results need to be taken with caution at present, since they may also be influenced by the difference in SNORD copy number between the knockout and wild type controls. Future experiments will need to be designed such that this interfering effect can be excluded, for example by matching the copy numbers of the wild type controls.

Conclusion

The current data suggest that *Peg13* plays an important role in brain development and mouse behavior, since even a partial deletion of this gene could change expression of approximately 500 genes and significantly change mouse behavior such as anxiety, activity and curiosity. Given that the gene has its highest expression in day 12.5, I propose that *Peg13* may have a specific role in preoptic area differentiation with possible consequences in mate choice and breeding behavior.

Material and Methods

mRNA extraction

RNA extraction was done by using Trizol reagent. 1mL Trizol per 40mg tissue (here is brain) was added to each sample. Then the samples were lysed by Tissue lyser II (QIAGEN, Germany) at 30 Hertz for 5 minutes. Homogenized samples were incubated at room temperature for 5 minutes. 200µl chloroform (per 1 mL TRIzol) was added to each sample, shook vigorously by hand 15 seconds, followed by 3 minutes incubation at room temperature and spun at 12,000 g for 15 minutes at 4 °C. The aqueous phase was transferred to a new tube and 0.5 volumes Isopropanol was added, incubated at room temperature for 10 minutes and spun at 12,000 g at 4 °C (RNA forms a gel like precipitate). The supernatant was removed and the pellet was washed with 75% EtOH (made with DEPC-H2O). Samples were mixed by hand several times and then spun at 7,500g for 5 minutes at 4 °C. The supernatant was removed and the pellet dried shortly at room temperature, dissolved in 200µl RNAse free water and stored at -20 °C for overnight. An equal volume of LiCL (5M) was added to the crude RNA extract, mixed by hand and incubated for one hour at -20 °C. Samples were spun at 16,000 g for 30 minutes. The supernatant was removed; samples were washed twice with EtOH 70% and spun at 10,000 at 4 °C. The pellet was dried at room temperature, dissolved in RNAse free water and kept in -70 °C.

RNA-seq analysis between wild type and knockout mice

To investigate the genetic network which is regulated by *Peg13*, RNA-sequencing analysis was performed on 10 Peg13 3'-KO and 10 wild type mice at the age of 16 weeks. The sequencing of the samples was performed using a polyA tail purification step, followed by cDNA synthesis, Illumina library preparation by using the Truseq standard RNA HT kit from Illumina. The libraries passing quality control were subjected to sequencing on an Illumina HiSeq 2000 sequencer. Each transcriptome sample was sequenced in approximately one third of a HiSeq2000 lane.

Raw sequence reads were quality trimmed using Trimmomatic (Bolger et al. 2014). The quality trimming was performed base wise, removing bases below quality score of 20 (Q20), and keeping reads whose average quality was of at least Q60.

Afterwards reads were mapped to the mouse mm10 reference genome (Genome Sequencing Consortium 2002) by using Hisat2 (Kim et al. 2015). Htseq was used for counting reads

overlapping into a specific feature (gene) (Anders et al. 2015). Differential expression analysis was performed with the DESeq2 package (Love et al. 2014) in R environment. Genes with an adjusted P-value (Padj) < 0.05 were considered as differentially expressed.

Gene Ontology (GO) is a tool used for gene annotation by collecting defined, structured, controlled vocabulary (Ashburner et al. 2000). KEGG is a database used to categorize associated gene sets into appropriate pathway (Kanehisa et al. 2000). DAVID (Database for Annotation, Visualization and Integrated Discovery) is a web-based tool which can provide a comprehensive set of functional annotation for numerous genes.

GO and KEGG pathway enrichment analyses were performed using DAVID online tools (Version 6.8, https://david-d.ncifcrf.gov/), with the classification stringency set to "medium" P value of <0.05

RNAseq analysis across different tissues

RNAseq data were downloaded from ENCODE mouse dataset to do transcriptome analysis across different tissues. These tissues include brain, duodenum, adrenal glands, adipose, heart, colon, spleen, bladder, kidney, lung, liver, placenta, testis, thymus, stomach, small intestine, large intestine, gonadal fat pad (G.F.P), mammary glands and ovary. Raw RNAseq data from each tissue were mapped to the mouse reference genome (mm10) (Genome Sequencing Consortium 2002) by Hisat2 (Kim et al. 2015). Mapped data were used for gene expression analysis. Transcripts Per Million (TPM) was calculated and used as index of gene expression. TPM was calculated as follows:

First, the read counts were divided by the length of each gene in kilobases. This value is called reads per kilobase (RPK). Then, all the RPK values were counted up in a sample and were divided by 1,000,000 (this is the "per million" scaling factor). Finally, the RPK values were divided by the "per million" scaling factor. This gives the TPM. Graphs for this part were drawn by the gplots package using the R environment.

Ribosomal profiling data analysis

Raw ribosome profiling data from brain, testis, heart and skeletal muscle tissues were downloaded from published data in Gene Expression Omnibus (GEO) (Gonzalez et al. 2014

and Castañeda et al. 2014). Raw sequencing data from each tissue were mapped to the mouse reference genome (mm10) (Genome Sequencing Consortium 2002) by bowtei2 (Langmead et al. 2012). Read counts were divided by the length of the protein of each gene in kilobases. RPK values were calculated as described above.

Generation of knock-out mice by genOway

The knockout mice were generated by the company genOway (www.genoway.com), after discussing the goals and the procedures with us. The detailed experimental description is added as an appendix to the thesis (Appendix 5).

Novel Object Test

In this chapter, mouse curiosity was assessed by using Novel Object test. This test was performed in the same test apparatus as the open field test, but this time a LEGO toy was placed in the center as a novel object. The time of the novel object was explored, and then the time spent sniffing the new object was measured over a period of 5 minutes (Birke et al. 1983; Verbeek et al. 1994; Yuen et al. 2015). At the end of the experiment the setup was cleaned with 30% Ethanol.

Other procedures

Mouse handling and Open Field, Elevated Plus Maze and Dark/Light Box tests were done according to the descriptions in chapter I.

General Discussion

The general background underlying this PhD thesis

Mate choice is one of most important decisions in an individual's life (Wedell & Ritchie 2004). Choice of the sexual partner could be at random, where there would be an equal chance for any individuals to mate with any other individuals within a population (Hedrick 2017). Mate choice comes into play when the choice of partner is nonrandom, i.e. where mate selection occurs according to the attractiveness of an individual's phenotype (Ryan et al. 1990; Vincent & Sadler 1995; Jones & Ratterman 2009; Xu et al. 2016; Fernald 2017). Non-random mate choice implies that inheritance does not occur strictly to Hardy-Weinberg rules, but that some alleles may be favored. Hence reproductive strategy and mate choice can have a direct influence on the allele distribution and the evolutionary change in populations (Vines & Schluter 2006; Xue et al. 2014).

There is a long standing interest in exploring mate choice patterns, especially assortative mating patterns, since these are of relevance for speciation mechanisms, including adaptive and sympatric speciation (Dieckmann et al. 2004). In an experiment to trace the evolution of assortative mating patterns previous work at our institute had shown that mate choice decisions between two allopatric populations of the Western house mouse, one from Germany (CB) and one from France (MC) appear to be influenced by paternal cues (Montero et al. 2013). In the follow up study, our group could identify two imprinted clusters of paternally expressed genes which are genetically highly differentiated between these two Western house mouse populations and which could therefore constitute a genetic basis for these fast evolving paternal cues (Lorenc et al. 2014). The first region is on Chromosome 7 and is orthologous to the region involved in the Prader-Willi Syndrome (PWS) in humans. The second is on Chromosome 15 and codes for a non-coding RNA (*Peg13*) which was also found to be associated with behavioral phenotypes in humans. However, the functional role of these paternally expressed genes in mouse mate choice preference remained unclear. Therefore this thesis was conducted to investigate how these two imprinted regions could be associated to mate choice decisions in mice and how they could influence their paternal mate choice preference.

The PWS region may be associated to the paternal mate choice preference in Western house mouse by regulating mouse personality

The PWS locus on Chr7 expresses two snoRNA gene families which are organized in large, tandemLy repeated clusters known as SNORD115 and SNORD116. The first chapter was focused on these two snoRNA families to investigate their potential role in mouse behavior. Results from this chapter showed that there is SNORD115 and 116 copy number variation (CNV) among mouse individuals that change the expression of over hundred genes in the brain, partly by directly regulating splice variants of two key genes. Gene ontology analysis showed these genes may play important role in the regulation of anxiety, mood, intellectual ability and metabolic pathways with possible link to personality traits. This inference is supported by the results from behavioral personality tests (Open Field Test, Dark/Light Box and Elevated Plus Maze). Most interestingly I found that there is a high correlation between individual's personality and their snoRNA copy numbers (Fig. 9 and 10).

Evolution of personality

Personality is a partially heritable trait which has been found so far in variety of species and from recent studies, it has been estimated that it may exist across the whole animal kingdom (Ogden 2012). The heritability component of personality provides a fascinating question for evolutionary biologists, namely how genetic variation could be maintained in a population where selection actively operates on a trait to keep beneficial and more adaptable genotypes (Verweij et al. 2012). So far several theories have been developed to explain the evolutionary basis of personality heritability while maintaining genetic variation that causes different personalities within a population. Below I discuss these theories briefly.

1-Selective Neutrality

Tooby & Cosmides (1990) have proposed a highly influential idea on the evolutionary genetics of personality traits. They highlighted selective neutrality (Kimura1983) as an evolutionary mechanism which could keep genetic variance in a population. In their view, genetic variation which causes different personalities does not influence individuals' fitness and therefore can freely change in frequency with not being affected by selection. Hence, in their perspective, personality is a neutral trait with no general impact on fitness (Tooby & Cosmides 1990). This theory had been accepted by many scientists in this field until

researchers found some evidence which showed that personality is not a neutral trait and can have indeed an important role in animal mate choice and reproductive behavior (Dingemanse et al. 2005).

2-Mutation-selection balance

The second explanation for the maintenance of variation in personality traits is mutation-selection balance (Lande 1975; Zhang and Hill 2005). In this theory, it is assumed that there is an optimal personality level in a population, maintained by selection acting against deviation from the optimal personality and thus eliminating alleles which do not fit to this optimal level. This effect should reduce genetic variation within a population. However, it is assumed that in parallel, new mutations arise in the population. Mutations with deleterious effect which cause strong deviation from the optimum will quickly be removed by selection. Mutations with weak effect which are less visible to the selection may persist in a population for many generations, but unlikely to be common within population due to the selection against them (Eyre-Walker &Keightley 2007; Eyre-Walker 2010). They also assume that in this case, personality must be affected by large number of loci with small effect to have a large mutational target size. Therefore since too many new mutations with small effect on the trait are continually reintroduced, selection is unable to act against them and delete these new mutations at a sufficiently high rate (Houle 1998).

My results from the first chapter do not really fit to this theory due to two reasons. First, because in this theory they predict a large number of loci with a small effect for personality traits, however my data showed that SNORD115/116 (i.e. two closely linked loci) could actually explain a large amount of variation in mouse personality traits (Fig. 9, 10, 20). Especially my results from C57BL/6 inbred strain with extremely high genetic homogeneity, strongly suggest that it is unlikely that a large number of loci are involved in the genetic variation underlying the personality traits that were measured here (Fig. 24).

In addition, from their view, selection acts actively against extreme phenotypes to remove highly deleterious mutations and that's why they are not common in a population. However, the normal distribution of SNORD115/116 copy number in both MC (Fig. 55) and C57BL/6 mice (Fig. 56) does not support this. A normal distribution suggests that extreme genotypes and phenotypes reflect simply a random distribution. Lab mice have not experienced natural environment and therefore selection cannot actively act on them to remove the extreme

phenotype or genotype but still we see a normal distribution of SNORD115/116 copy numbers among them (Fig. 56). Therefore, the mutation-selection balance theory cannot explain the genetic variation underlying personality trait in mouse.

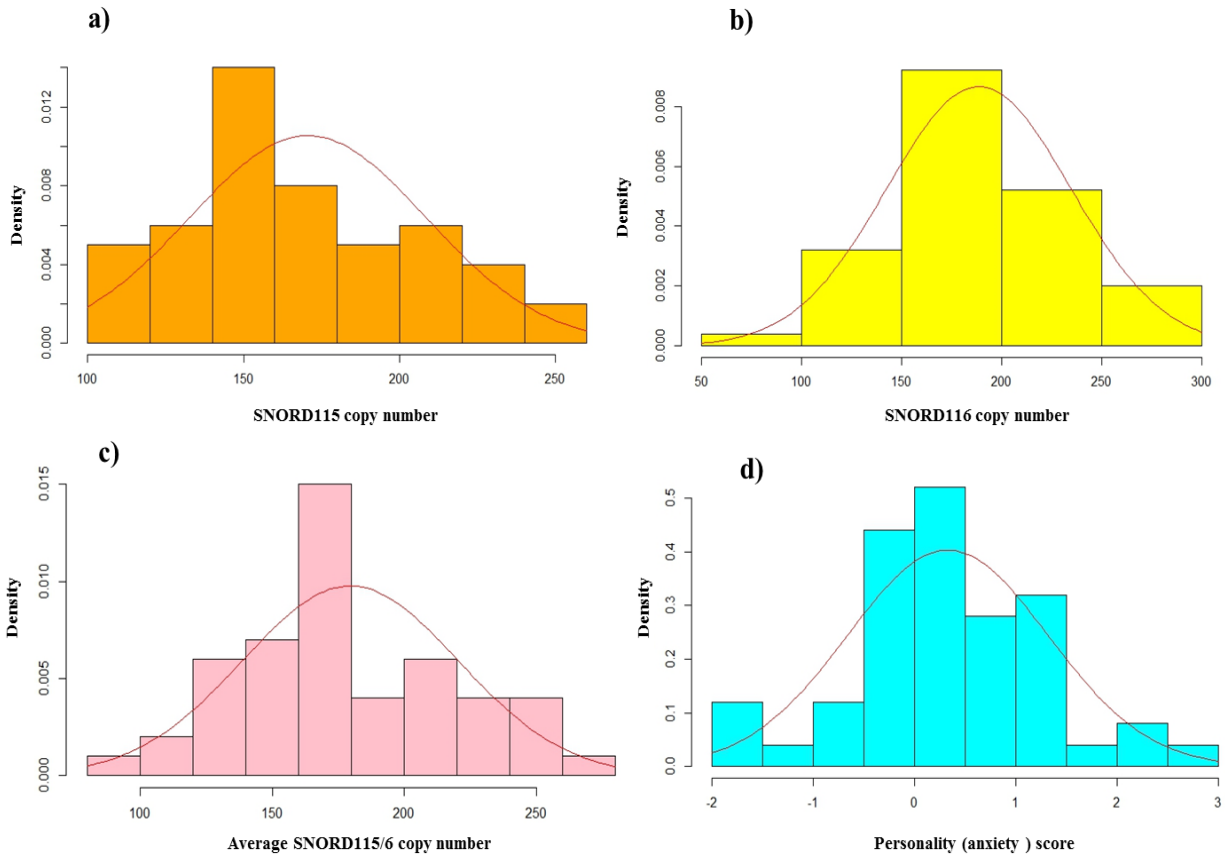


Fig. 54: Distribution of SNORD copy number and personality in MC mice. In this analysis, 50 individuals were involved. These individuals have been used in the first chapter for both genetics and behavioral analysis. Shapiro–Wilk (SW) (Shapiro & Wilk 1965) and Anderson Darling (AD) tests were used to assess the normal distribution by using the "nortest" package in R environment (Gross 2006). If the p-value in these tests is less than or equal to the significance level (0.05), then the null hypothesis is rejected and we can conclude that the data would not have a normal distribution. If the p-value is larger than the significance level, then the null hypothesis won't be rejected.

a) SNORD115 copy number. *p value from SW = 0.34 and p value from AD = 0.24*

b) SNORD116 copy number. *p value from SW = 0.62 and p value from AD = 0.45*

c) Average SNORD115/116 copy number. *p value from SW = 0.30 and p value AD = 0.16*

d) Personality (anxiety) score. *p value from SW = 0.49 and p value from AD = 0.51*

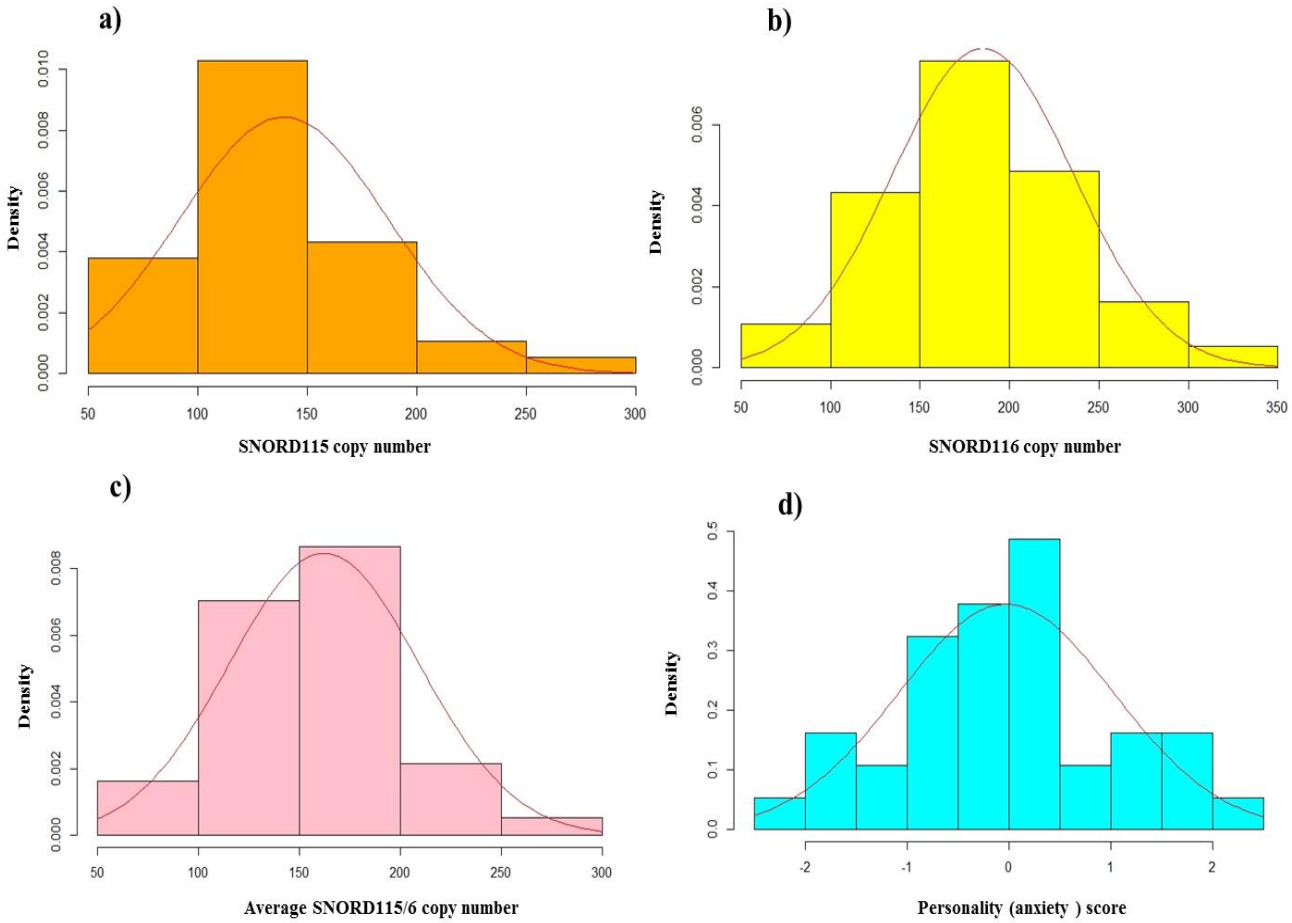


Fig. 55: Distribution of SNORD copy number and personality in C57BL/6J inbred mice. In this analysis, 40 individuals were involved. These individuals have been used in the first chapter for both genetics and behavioral analysis. Shapiro–Wilk (SW) (Shapiro & Wilk 1965) and Anderson Darling (AD) tests were used to assess the normal distribution by using "nortest" package in R environment (Gross 2006). If the p-value in these tests is less than or equal to the significance level (0.05), then the null hypothesis is rejected and we can conclude that the data does not have a normal distribution. If the p-value is larger than the significance level, then the null hypothesis won't be rejected.

a) SNORD115 copy number. *p value from SW = 0.08 and p value from AD = 0.12*

b) SNORD116 copy number. *p value from SW = 0.62 and p value from AD = 0.45*

c) Average SNORD115/6 copy number. *p value from SW = 0.88 and p value from AD = 0.60*

d) Personality (anxiety) score. *p value from SW = 0.68 and p value from AD = 0.51*

3-Balancing Selection

Both the Selective Neutrality and Mutation-Selection Balance theory explain the genetic variation of personality traits based on that these variations are invisible for selection to delete them, either due to neutrality of personality or to large mutational target size. In the third theory, Balancing Selection, selection itself maintains this genetic variation. Balancing selection can occur, when different degrees of trait dimensions are favored equally by selection due to environmental heterogeneity. In this case, effect of personality on individual's fitness in a population differs across space or time and therefore selection maintains the polymorphisms (Penke et al. 2007). Dingemanse et al. (2004) showed for birds that selection pressure fluctuated due to environmental conditions (food and space availability). It has been noted that genetic variation in human personality traits is also most likely maintained by balancing selection as a result of spatial or temporal fluctuations in selection pressure in different social or physical environments that human have had during their evolutionary history (Penke et al 2007).

My results fit to this theory only partially. The idea of environmental heterogeneity can be applicable here, since in this situation having different strategies to deal with different environmental conditions work much better than a single strategy. Different personalities within a population promote the average fitness of whole population where we have changes in environmental condition. However the reason and mechanism which maintain the genetic variation of personality is not balancing selection, but recurrent mutations. Data from the family study (Fig. 19, 20) and also inbred mice (Fig. 21, 22, 24) in the first chapter suggest a very fast mutational rate for these loci (SNORD115/116) which imply mutation itself is sufficient and robust to create and keep the genetic variation underlying personality. However, this raises the question why such a polymorphic system has evolved and has been maintained at least in mammals. It would suggest that it is generally favorable for a population to maintain variance in personality traits, but there is a general consensus in evolutionary biology that population level selection (i.e. group selection) is difficult to explain, because the interests of some individuals ("cheaters") have short term advantages. While balancing selection as such is not a group selection problem, it would appear that the rather many interactions in the SNORD115/116 system are too complex to be maintained by balancing selection alone.

Personality and imprinting

SNORD115/116 which regulate significantly mouse personality, are imprinted genes. So an intriguing question is why such an important trait should be regulated by only one of the parental alleles. I would like to address and discuss this question from two different prospective.

1-Different recombination rate between male and female

It has been shown that most, if not all, imprinted regions in the human genome have extremely high recombination rates with possible differences between males and females, especially in the vicinity of the 15q11-q13 region (Paldi et al .1995; Robinson & Lalande 1995; Lercher & Hurst 2003; Sandovici et al. 2006). Paldi et al. (1995) reported a higher recombination rate for males than females in 15q11-13. However, later some other studies failed to validate their findings and instead found the opposite result, i.e. higher recombination rate in females compared with males (Robinson & Lalande 1995; Lercher & Hurst 2003; Sandovici et al. 2006).

15q11-q13 is the location of PWS region, including SNROD 115/116 in the human genome. From my knowledge, there is no study on mouse and other mammals on this region to investigate a sex-specific recombination rate in this region and therefore I cannot generalize this finding to all other mammalian genomes. However, these observations suggest a possible difference in the recombination rate for SNORD115/116 in the eutherian genome between females and males. In this situation, egg and sperm will have different information and different recombination outcomes for SNORD115/116 copy numbers and subsequently selection may act on them to provide fitter alleles to the offspring compared with the other parents.

2-Co-adapted gene expression

Wolf (2013) proposed a new idea for the evolution of genomic imprinting. He suggests that when one locus is evolved as imprinted gene for any reason, selection favors imprinting of its interacting partners at different loci to match its pattern of imprinting. Therefore imprinting of some genes is not directly favored; instead they may evolve only as a co-evolutionary

response. This could also explain why many imprinted genes appear as clusters in the eutherian genome (Wolf 2013).

So the second explanation that I would like to propose to explain the imprinted status of SNORD115/116 is a co-adapted gene expression model. SNORD115/116 may evolve as imprinted genes mostly because of their location in the genome and the imprinted state of their expression may be coordinated by selection only to make a beneficial association and better interaction with other genes in this cluster.

Although I tried to propose two different explanations for the causality of imprinted manner of SNROD 115/116, it still remains a quite open and fascinating topic for more studies and ideas.

***Peg13* could be involved in sexual behavior in mouse**

In the second chapter of this PhD thesis, I conducted a functional study on *Peg13* in mice by using *Peg13* KO mice. The results from this part suggest that *Peg13* may play an important role in the development of the preoptic area in the brain and subsequently mouse sexual behavior. The role of the preoptic area in sexual orientation has been well described in the literature for both, rat (Paredes 2003; Bosch et al, 2010; Graham et al, 2013; Zhong et al. 2014) and humans (Hofman & Swaab 1989; LeVay 1991; Swaab 1995; Dominguez & Hull 2005; Poeppl 2016). *Peg13* KO mice also showed significant difference in their activity, anxiety, curiosity and decision making ability that I have discussed in the second chapter. Here I would like to discuss additional aspects of these results.

It has been a fascinating question in evolutionary biology that why individuals within a population shows different sexual preference and orientation. Homosexual and asexual individuals decrease the rate of reproduction, but still persist in a population over evolutionary time. There are several views to explain the causality of having different sexual preference which has been reviewed by Ryabko & Reznikova (2015). One of the earliest, but still famous theories is kin selection (Hamilton 1994), which suggests an indirect adaptive effect to have a different sexual orientation. From this view, homosexuals can indirectly increase reproductive success in theirs families, by providing some obvious or non-obvious benefit for their relatives. Although this theory is one of the oldest theories, still there is poor support for that.

Here in the second chapter of this thesis, I provided data which suggest *Peg13* may play an important role not only in sexual behavior (probably by influencing the formation of the preoptic area), but also regulate significantly other mouse behaviors such as anxiety, activity and also curiosity possibly by a role in synapse formation, GTPase activity, immunity and inflammatory response and also metabolic pathway (see discussion in the second chapter for more details). These data altogether could suggest a multifunctional role of *Peg13,* which can be involved in different molecular and behavior processes. So one allele may have a non-adaptive function in a sexual trait, but can have an adaptive role in another trait and that's why it still persists in a population. Therefore genes which result in homosexuality and asexuality may also be involved in other pathways and traits which are otherwise beneficial and adaptive and that's why they may persist in a population. I think *Peg13* could be good example of such multifunctional gene which plays important role not only in sexual behavior but also other behavioral phenotypes.

Conclusion

This PhD thesis has dealt with the genetic mechanisms which may be involved in the regulation of behavioral mating preferences in mice. Two previously identified imprinted loci were studied to identify possible molecular mechanisms that can lead to the fast divergence of mating preferences.

The first locus is known as the Prader-Willi Syndrom (PWS) region in humans and is known to be involved in behavioral and intellectual ability traits. My focus was on clusters of regulatory RNAs in the PWS region, the SNORD115/116 genes. I found that the copy number variation in these genes correlates with behavioral measures, collectively termed "personality". Molecular analysis showed that this occurs by influencing the expression of more than 130 genes, which are themselves known to be involved in a variety of pathways. The fast evolution through copy number variation of the SNORD115/116 genes suggests a mechanism for divergence of mating preferences and population differentiation. My data can also nicely explain the coevolution of personality and other traits such as metabolism, cognitive ability, vocalization, craniofacial feature. This finding is very unexpected, since it suggests the existence of a major control locus for basic behavioral traits, with an inbuilt mechanism to ensure variance in behavioral strategies within a population. This finding could open new windows in several different fields such as zoology, behavioral biology, molecular biology, genetics, neurobiology, evolutionary biology and also medicine.

The second locus has been known as *Peg13*, but its molecular function was unknown. This thesis provided new insights into its function through the analysis of a knockout mouse model. The results suggest that this gene may also regulate paternal mate choice preferences in mice, possibly through a role in preoptic area differentiation and correspondent influence on sexual affinity behavior.

Reference

A AK, Nadimpalli SK. Takifugu rubripes cation independent mannose 6-phosphate receptor: Cloning, expression and functional characterization of the IGF-II binding domain. Int J Biol Macromol. 2018 Jul 1;113:59-65.

Adhikari A, et al. Cognitive deficits in the Snord116 deletion mouse model for Prader-Willi syndrome. Neurobiol Learn Mem. 2018 May 23; pii: S1074-7427(18)30119-9.

Aincy M, et al. Synaptic dysfunction in amygdala in intellectual disorder models. Prog Neuropsychopharmacol Biol Psychiatry. 2017 Aug; pii: S0278-5846(17)30393-7.

Alex KD, et al. Modulation of dopamine release by striatal 5-HT2C receptors. Synapse. March 2005; 55 (4): 242–51.

Altschul SF, et al. Basic local alignment search tool. J Mol Biol. 1990 Oct 5; 215(3):403-10.

Anders S, et al. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015 Jan; 15; 31(2):166-9.

Ashburner M, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 2000; 25:25–9.

Babak T, et al. Genetic conflict reflected in tissue-specific maps of genomic imprinting in human and mouse. Nat Genet. 2015 May; 47(5):544-9.

Badyaev AV. Stress-induced variation in evolution: from behavioral plasticity to genetic assimilation. Proc R Soc Lond B. 2005; 272:877–886

Bando Y, et al. Dysfunction of KCNK potassium channels impairs neuronal migration in the developing mouse cerebral cortex. Cereb Cortex. 2014; 24(4):1017–29.

Barlow DP, et al. Genomic Imprinting in Mammals. Cold Spring Harb Perspect Biol. 2014 Feb; 6(2): a018382.

Bates D, et al. "Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software. 2015; 67(1), 1-48.

Bazeley PS, et al. snoTARGET shows that human orphan snoRNA targets locate close to alternative splice junctions. Gene. 2008 Jan 31; 408(1-2):172-9.

Belknap J K, et al. Voluntary consumption of ethanol in 15 inbred mouse strains. Psychopharmacology (Berl). 1993; 112, 503–510

Bell AC, et al. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. Nature. May 2000; 405 (6785): 482–5.

Bell RQ. A reinterpretation of the direction of effects in studies of socialization. Psychological Review. 1968; 75:81–85

Benus RF, et al. Heritable variation for aggression as a reflection of individual coping strategies. Experientia. 1991; 47(10): 1008-1019.

Bergthorsson U, et al. Gene Copy-Number Changes in Evolution. (May 2016) In: eLS. John Wiley & Sons Ltd, Chichester. http://www.els.net

Bierbach D, et al. Personality affects mate choice: bolder males show stronger audience effects under high competition. Behavioral Ecology, Volume 26, Issue 5, 1 September 2015, Pages 1314–1325.

Birke LIA, et al. Some issues and problems in the study of animal exploration. van Nostrand Reinhold (UK) Co. Ltd, Wokingham, Berkshire, England. (1983).

Bolger AM, et al. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014 Aug 1; 30(15):2114-20.

Bonduriansky R. The genetic architecture of sexual dimorphism: the potential roles of genomic imprinting and condition. Oxford University Press: 2007 Oxford, UK; 176–184.

Boone PM, et al. Genomic medicine and neurological disease. Hum. Genet. 2011;130:103–121.

Bortolin-Cavaillé ML, et al. The SNORD115 (H/MBII-52) and SNORD116 (H/MBII-85) gene clusters at the imprinted Prader-Willi locus generate canonical box C/D snoRNAs. Nucleic Acids Res. 2012 Aug; 40(14):6800-7.

Bosch OJ, et al. Maternal behaviour is associated with vasopressin release in the medial preoptic area and bed nucleus of the stria terminalis in the rat. Journal of Neuroendocrinology. 2010; 22 (5): 420–429.

Bothe GW, et al. Behavioral differences among fourteen inbred mouse strains commonly used as disease models. Comp. Med. 2005; 55, 326–334.

Bouabe H, et al. Gene targeting in mice: a review. Methods Mol Biol. 2013; 1064:315-36.

Branchi I, et al. Ultrasonic vocalizations by infant laboratory mice: A preliminary spectrographic characterization under different conditions. Dev Psychobiol. 1998; 33:249–256.

Cansado J. To finish things well: cysteine methylation ensures selective GTPase membrane localization and signalling. Curr Genet. 2018 Apr;64(2):341-344.

Careau V, Garland T Jr. Energetics and behavior: many paths to understanding. Trends Ecol Evol. 2015 Jul;30(7):365-6.

Carlson GC, et al. Dexras1 a unique ras-GTPase interacts with NMDA receptor activity and provides a novel dissociation between anxiety, working memory and sensory gating. Neuroscience, 2016 May 13; 322:408-15.

Caspi A, et al. Personality development across the life course. In: Eisenberg N, Damon W, Lerner RM, editors. Handbook of child psychology: Vol. 3, Social, emotional, and personality development. 6th ed. John Wiley & Sons Inc.; Hoboken, NJ: 2006. pp. 300–365.

Castañeda J, et al. Reduced pachytene piRNAs and translation underlie spermiogenic arrest in Maelstrom mutant mice. EMBO J 2014 Sep 17; 33(18):1999-2019.

Castelo-Branco G, et al. Neural stem cell differentiation is dictated by distinct actions of nuclear receptor corepressors and histone deacetylases. Stem Cell Reports. 2014 Sep 9; 3(3):502-15.

Cavaillé J. Box C/D small nucleolar RNA genes and the Prader-Willi syndrome: a complex interplay. Wiley Interdiscip Rev RNA. 2017 Jul; 8(4).

Chapnik N, et al. A superactive leptin antagonist alters metabolism and locomotion in high-leptin mice. J Endocrinol. 2013 Apr 29; 217(3):283-90.

Chen Q, et al. Expression analysis of the speech-related genes FoxP1 and FoxP2 and their relation to singing behavior in two songbird species. J Exp Biol. 2013 Oct 1; 216(Pt 19):3682-92.

Cho J, et al. Multiple repressive mechanisms in the hippocampus during memory formation. Science 2015 Oct 2; 350(6256):82-7.

Clegg, MT, et al. (1995) Science and the Endangered Species Act (Natl. Acad. Press, Washington, DC)

Colas D, et al. Sleep disturbances in Ube3a maternal-deficient mice modeling Angelman syndrome. Neurobiol Dis. 2005; 20(2):471–78.

Court F, et al. The PEG13-DMR and brain-specific enhancers dictate imprinted expression within the 8q24 intellectual disability risk locus. Epigenetics Chromatin. 2014 Mar 25; 7(1):5.

Crespi BJ. Causes of assortative mating in arthropods. Animal Behaviour. 1989;38:980–1000.

Curley JP, et al. Coadaptation in mother and infant regulated by a paternally expressed imprinted gene. Proc R Soc B. 2004; 271(1545):1303–9.

Curley JP. Is there a genomically imprinted social brain? BioEssays. 2011; 33(9):662–68.

Das A, et al. Affecting long-term fear memory formation through optical control of Rac1 GTPase and PAK activity in lateral amygdala. Sci Rep. 2017 Oct 24; 7(1):13930.

Day T, et al. Intralocus sexual conflict can drive the evolution of genomic imprinting. Genetics. 2004; 167:1537–1546.

Dessauer CW. Shining a light on GPCR complexes. J Biol Chem. 2017 Aug 25; 292(34):14290-14291.

Diaz-Muñoz MD, et al. The RNA-binding protein HuR is essential for the B cell antibody response. Nat Immunol 2015 Apr; 16(4):415-25.

Dieckmann U, et al. Adaptive Speciation. Cambridge University Press. (2004).

Ding F1, et al. SnoRNA Snord116 (Pwcr1/MBII-85) deletion causes growth deficiency and hyperphagia in mice. PLoS One. 2008 Mar 5; 3(3):e1709.

Dingemanse NJ, et al. Repeatability and heritability of exploratory behaviour in great tits from the wild. Animal Behaviour. 2002; 64: 929-938.

Dingemanse NJ, et al. Natural selection and animal personality. Behaviour. 2005; 142, 1159-1184.

Dominguez JM, Hull EM. Dopamine, the medial preoptic area, and male sexual behavior. Physiol Behav. Physiol Behav. 2005 Oct 15; 86(3):356-68.

Duckworth RA. The role of behavior in evolution: a search for mechanism. Evol Ecol. 2009; 23: 513.

Dupuis-Sandoval F, et al. The emerging landscape of small nucleolar RNAs in cell biology. Wiley Interdiscip Rev RNA. 2015 Jul-Aug;6(4):381-97.

Ehrlich PR, et al. Differentiation of populations. Science. 1969; 165, 1228–1232.

Guo H, et al. Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature. 2010 Aug 12;466(7308):835-40.

Eyre-Walker A & Keightley PD. The distribution of fitness effects of new mutations. Nat Rev Genet. 2007; 8:610–618

Eyre-Walker A. Genetic architecture of a complex trait and its implications for fitness and genome-wide association studies. Proceedings of the National Academy of Sciences of the United States of America. 2010; 107:1752–1756

Falaleeva M & Stamm S. Processing of snoRNAs as a new source of regulatory non-coding RNAs: snoRNA fragments form a new class of functional RNAs. Bioessays. 2013 Jan; 35(1):46-54

Felsenstein J. Skepticism towards Santa Rosalia, or why are there so few kinds of animals? Evolution. 1981; 35:124–138

Fernald RD. Cognitive skills and the evolution of social systems. J Exp Biol. 2017 Jan 1;220

Ferrón SR, et al. Differential genomic imprinting regulates paracrine and autocrine roles of IGF2 in mouse adult neurogenesis. Nat Commun. 2015; 6:8265.

Finlay BL, et al. Linked regularities in the development and evolution of mammalian brains. Science. 1995 Jun 16; 268(5217):1578-84.

Flavell RB. Inactivation of gene expression in plants as a consequence of specific sequence duplication. Proc Natl Acad Sci U S A. 1994 Apr 26; 91(9):3490-6.

Fritz J. et al. NextGenMap: fast and accurate read mapping in highly polymorphic genomes. Bioinformatics, Vol. 29, No. 21. 1 November 2013, pp. 2790-2791,

Fujita Y, et al. Decreased cohesin in the brain leads to defective synapse development and anxiety-related behavior. J Exp Med. 2017 May 1; 214(5):1431-1452.

Fuller RC, et al. Sensory bias as an explanation for the evolution of mate preferences. Am Nat. 2005 Oct; 166(4):437-46.

Gallagher D, et al. Ankrd11 is a chromatin regulator involved in autism that is essential for neural development. Dev Cell. 2015 Jan 12; 32(1):31-42.

Garfield AS, et al. Distinct physiological and behavioural functions for parental alleles of imprinted Grb10. Nature. 2011; 469(7331):534–38.

Gonzalez C, et al. Ribosome profiling reveals a cell-type-specific translational landscape in brain tumors. J Neurosci 2014 Aug 13; 34(33):10924-36.

Gonzalez C, et al. Ribosome profiling reveals a cell-type-specific translational landscape in brain tumors. J Neurosci 2014 Aug 13; 34(33):10924-36.

Gosling S. D. "From mice to men: What can we learn about personality from animal research?" Psychological Bulletin. 2001; 127(1): 45-86.

Gosling SD. Personality in non-human animals. Social and Personality Psychology Compass. 2008; 2(2):985-1001.

Gotter AL, et al. TASK-3 as a potential antide-pressant target. Brain Res. 2011; 1416:69–79.

Graham MD, et al. "Infusions of ascorbic acid into the medial preoptic area facilitate appetitive sexual behavior in the female rat". Physiology and Behavior. 2013; 122 (1): 140–146.

Grau-Bové X, et al. A genomic survey of HECT ubiquitin ligases in eukaryotes reveals independent expansions of the HECT system in several lineages. Genome Biol. Evol. 2013; 5, 833–847.

Gregg C, et al. High-resolution analysis of parent-of-origin allelic expression in the mouse brain. Science. 2010; 329(5992):643–48.

Gross J. nortest: Tests for normality. R package (2015)

Gyuris E. et al. Individual behaviour in firebugs (Pyrrhocoris apterus). Proc R Soc B. 2011; 278 (1705): 628-633.

Haj-Mirzaian A, et al. Anxiety- and Depressive-Like Behaviors are Associated with Altered Hippocampal Energy and Inflammatory Status in a Mouse Model of Crohn's Disease. Neuroscience. 2017 Dec 16; 366:124-137.

Hall B, et al. Overview: generation of gene knockout mice. Curr Protoc Cell Biol. 2009 Sep; Chapter 19: Unit 19.12 19.12.1-17.

Hamilton WD. The genetical evolution of social behaviour. II. J. Theor. Biol. (1964) 7, 17–52.

Hammerschmidt K, et al. Female mice respond to male ultrasonic ‘songs’ with approach behaviour. Biol Lett, 2009; 5: 589–592.

Hardouin EA, et al. Eurasian house mouse (Mus musculus L.) differentiation at microsatellite loci identifies the Iranian plateau as a phylogeographic hotspot. BMC Evol Biol. 2015 Feb 25;15:26.

Härmälä SK, et al. Copy Number Variation Analysis by Droplet Digital PCR. Methods Mol Biol. 2017; 1654:135-149.

Harris JR. The nurture assumption: Why children turn out the way they do. Free Press; New York: 1998.

Harris JR. Where is the child's environment? A group socialization theory of development. Psychological Review. 1995; 102:458–489.

Harrow J, et al. GENCODE: the reference human genome annotation for The ENCODE Project. Genome Research. 2012; 22:1760–1774.

Hartl DL, et al. Principles of Population Genetics. (Sinauer, Sunderland, MA). (1997)

Hedrick PW. Assortative Mating and Linkage Disequilibrium. G3 (Bethesda). 2017 Jan 5; 7(1):55-62.

Hefner K, et al. Impaired fear extinction learning and cortico-amygdala circuit abnormalities in a common genetic mouse strain. J. Neurosci. 2008; 28, 8074–8085

Herde A, et al. Consistency in boldness, activity and exploration at different stages of life. BMC Ecol. 2013 Dec 7; 13:49.

Hertel J, et al. SnoReport: computational identification of snoRNAs with unknown targets. Bioinformatics. 2008 Jan 15; 24(2):158-64.

Hevers W, et al. The diversity of GABAA receptors. Pharmacological and electrophysiological properties of GABAA channel subtypes. Mol. Neurobiol. August 1998; 18 (1): 35–86.

Hoflack B, Kornfeld S. Lysosomal enzyme binding to mouse P388D1 macrophage membranes lacking the 215-kDa mannose 6-phosphate receptor: evidence for the existence of a second mannose 6-phosphate receptor. Proc Natl Acad Sci U S A. 1985 Jul;82(13):4428-32.

Hofman MA & Swaab DF. The sexually dimorphic nucleus of the preoptic area in the human brain: a comparative morphometric study. J Anat. 1989 Jun; 164:55-72.

Hohoff C. Anxiety in mice and men: a comparison. J Neural Transm (Vienna). 2009 Jun; 116(6):679-87

Holmes A, et al. Behavioral profile of wild mice in the elevated plusmaze test for anxiety. Physiology & Behavior. 2000; 71(5): 509-516.

Holy TE. et al. Ultrasonic Songs of Male Mice. PLoS Biol. 2005 3: e386.

Hooper PL, Miller GF. Mutual mate choice can drive costly signaling even under perfect monogamy. Adaptive Behavior. 2008; 16:53–70.

Hopwood CJ, et al. Genetic and environmental influences on personality trait stability and growth during the transition to adulthood: a three-wave longitudinal study. J Pers Soc Psychol. 2011 Mar; 100(3):545-56.

Houle D. How should we explain variation in the genetic variance of traits? Genetica. 1998; 102–103:241–253.

Houston AI. Evolutionary models of metabolism, behaviour and personality. Philos Trans R Soc Lond B Biol Sci. 2010 Dec 27; 365(1560):3969-75.

Huey RB, et al. Behavioral drive versus behavioral inertia in evolution: a null model approach. Am Nat. 2003; 161:357–366.

Tooby J, Cosmides L.On the universality of human nature and the uniqueness of the individual: the role of genetics and adaptation. Journal of Personality J Pers. 1990 Mar; 58(1):17-67.

Ingolia NT. Ribosome profiling: new views of translation, from single codons to genome scale. Nat Rev Genet. 2014 Mar; 15(3):205-13.

Ingolia NT, et al. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science. 2009; 324, 218–223.

Jauregi J, et al. A neuropsychological assessment of frontal cognitive functions in Prader-Willi syndrome. J Intellect Disabil Res. 2007 May; 51(Pt 5):350-65.

Jiang Y, et al. Assortative mating in animals. Am Nat. 2013 Jun; 181(6):E125-38.

Jiang YH, et al. Altered ultrasonic vocalization and impaired learning and memory in Angelman syndrome mouse model with a large maternal deletion from Ube3a to Gabrb3. PLoS One. 2010 Aug 20; 5(8):e12278.

Johannesson K, et al. Incipient reproductive isolation between two sympatric morphs of the intertidal snail Littorina saxatilis. Evolution. 1995 Dec; 49(6):1180-1190.

Jones AG, Ratterman NL. Mate choice and sexual selection: what have we learned since Darwin? .Proc Natl Acad Sci U S A. 2009 Jun 16; 106 Suppl 1:10001-8.

Ka M, et al. ANKRD11 associated with intellectual disability and autism regulates dendrite differentiation via the BDNF/TrkB signaling pathway. Neurobiol Dis. 2017 Dec 21; 111:138-152.

Kakar N, et al. A homozygous splice site mutation in TRAPPC9 causes intellectual disability and microcephaly. Eur J Med Genet. 2012; 55: 727-731.

Kanehisa M, et al. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 2000; 28:27–30.

Keverne EB, et al. Genomic imprinting and the differential roles of parental genomes in brain development. Brain Res Dev Brain Res. 1996; 92(1):91–100.

Kim D, et al. HISAT: a fast spliced aligner with low memory requirements. Nat Methods. 2015 Apr; 12(4):357-60.

Kim D, et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 2013 Apr 25; 14(4):R36.

Kimura A, et al. Overexpression of 5-HT2C receptors in forebrain leads to elevated anxiety and hypoactivity. Eur J Neurosci. 2009 Jul; 30(2):299-306.

Kimura M. The neutral theory of molecular evolution. Cambridge, U.K.: Cambridge (1983).

Kishore S, et al. The snoRNA HBII-52 regulates alternative splicing of the serotonin receptor 2C. Science. 2006 Jan 13; 311(5758):230-2.

Klingenberg CP, et al. Shape analysis of symmetric structures: quantifying variation among individuals and asymmetry. Evolution. 2002; 56: 1909–1920.

Kondrashov AS, Shpak M. on the origin of species by means of assortative mating. Proc Biol Sci. 1998 Dec 7; 265(1412):2273-8.

Koolhaas, J. M. et al. Coping styles in animals: current status in behavior and stress physiology. Neuroscience and Bio behavioral. 1999; Reviews 23(7): 925-935.

Kotov R, et al. big personality traits to anxiety, depressive, and substance use disorders: a meta-analysis. Psychol Bull. 2010 Sep; 136(5):768-821.

Krebs, Rebecca. The influence of personality on territory use and fitness in Western house mice (Mus musculus domesticus) (Doctoral dissertation). Kiel, Germany: Christian-Albrechts Universität zu Kiel. (2018).

Krieger J, et al. Olfactory reception in invertebrates. Science. 1999 Oct 22; 286(5440):720-3.

Kulkarni SR, et al. Caloric restriction-mediated induction of lipid metabolism gene expression in liver is enhanced by Keap1-knockdown. Pharm Res. 2013 Sep; 30(9):2221-31.

Kumar M, et al. Integrated analysis of microRNAs and their disease related targets in the brain of mice infected with West Nile virus. Virology. 2014 Mar; 452-453:143-51.

Kusminski CM, et al. MitoNEET-driven alterations in adipocyte mitochondrial activity reveals a crucial adaptive process that preserves insulin sensitivity in obesity. Nat Med. 2012 Oct;18(10):1539-49.

Lande R. The maintenance of genetic variability by mutation in a polygenic character with linked loci. Genetical Research. 1975; 26:221–235.

Langmead B, et al. Fast gapped-read alignment with Bowtie2. Nature Methods. 2012; 9:357-359.

Lefebvre L, et al. Abnormal maternal behaviour and growth retardation associated with loss of the imprinted gene Mest. Nat Genet. 1998; 20(2):163–69.

Lercher MJ, Hurst LD. Imprinted chromosomal regions of the human genome have unusually high recombination rates. Genetics. 2003 Nov; 165(3):1629-32.

LeVay S. A difference in hypothalamic structure between heterosexual and homosexual men. Science. 1991 Aug 30; 253(5023):1034-7.

Lewejohann L, et al. Personality in laboratory mice used for biomedical research: a way of understanding variability? Dev Psychobiol. 2011 Sep; 53(6):624-30.

Lewontin R, Kirk D, Crow J. Selective mating, assortative mating, and inbreeding: definitions and implications. Eugen Q. 1968 Jun; 15(2):141-3.

Li CW, et al. Ankyrin repeats-containing cofactors interact with ADA3 and modulate its co-activator function. Biochem J. 2008 Jul 15; 413(2):349-57.

Li H, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009 Aug 15; 25(16):2078-9.

Li H, et al. A Bcl-xL–Drp1 complex regulates synaptic vesicle membrane dynamics during endocytosis. Nature. 2013; 15(7):773–85.

Liao J, et al. Small nucleolar RNA signatures as biomarkers for non-small-cell lung cancer. Mol Cancer. 2010 Jul 27; 9:198.

Lilja T, et al. Like a rolling histone: epigenetic regulation of neural stem cells and brain development by factors controlling histone acetylation and methylation. Biochim. Biophys. Biochim Biophys Acta. 2013 Feb; 1830(2):2354-60.

Linden AM, et al. TASK-3 knockout mice exhibit exaggerated nocturnal activity, impairments in cognitive functions, and reduced sensitivity to inhalation anesthetics. J Pharmacol Exp Ther. 2007;323(3):924–34.

Liscovitch N, et al. Specialization of gene expression during mouse brain development. PLoS Comput Biol. 2013; 9(9):e1003185.

Lo MT, et al. Genome-wide analyses for personality traits identify six genomic loci and show correlations with psychiatric disorders. Nat Genet. 2017 Jan; 49(1):152-156.

Lo-Castro A, et al. Neurobehavioral phenotype observed in KBG syndrome caused by ANKRD11 mutations. Am J Med Genet B Neuropsychiatr Genet. 2013 Jan; 162B(1):17-23.

Lo-Castro A, et al. Neurobehavioral phenotype observed in KBG syndrome caused by ANKRD11 mutations. Am J Med Genet B Neuropsychiatr Genet. 2013 Jan;162B(1):17-23
Lorenc A, et al. Genetic differentiation of hypothalamus parentally biased transcripts in populations of the house mouse implicate the Prader-Willi syndrome imprinted region as a possible source of behavioral divergence. Mol Biol Evol. 2014; 31: 3240-3249.

Losos JB, et al. Predator-induced behaviour shifts and natural selection in field-experimental lizard populations. Nature. 2004; 432:505–508.

Love MI, et al. Moderated estimation of fold change and dispersion for RNAseq data with DESeq2. Genome Biol. 2014; 15(12):550.

Lui L, et al. Small nucleolar RNAs and RNA-guided post-transcriptional modification. Essays Biochem. 2013; 54:53-77.

Lupski JR. New mutations and intellectual function. Nat. Genet. 2010; 42:1036–1038.

Lusardi TA, et al. Effect of (S)-3,5-DHPG on microRNA expression in mouse brain. Exp Neurol. 2012 Jun; 235(2):497-507.

Lusardi TA, et al. MicroRNA responses to focal cerebral ischemia in male and female mouse brain. Front Mol Neurosci. 2014 Feb 11; 7:11.

Lydiard RB. The role of GABA in anxiety disorders. J Clin Psychiatry. 2003; 64 Suppl 3:21-7

Lynch M, and Walsh B. Genetics and analysis of quantitative traits. Sinauer, Sunderland, MA. (1998).

Magnusson D. Personality development from an interactional perspective. In: Pervin LA, editor. Handbook of personality: Theory and measurement. Guilford Press; New York: 1990. pp. 193–222.

Marangi G, et al. TRAPPC9-related autosomal recessive intellectual disability: report of a new mutation and clinical phenotype. Eur J Hum Genet. 2013; 21: 229-232.

Marín, I. Animal HECT ubiquitin ligases: evolution and functional implications. BMC Evol Biol. 2010 Feb 22; 10:56.

Marshall CR, et al. Structural variation of Chromosomes in autism spectrum disorder. Am J Hum Genet. 2008 Feb; 82(2):477-88.

Mayr E. Animal species and evolution. Harvard University Press, Cambridge. (1963).

McGlothlin JW, et al. Interacting phenotypes and the evolutionary process. III. Social evolution. Evolution. 2010; 64:2558–2574.

Mead LS, Arnold SJ. Quantitative genetic models of sexual selection. Trends Ecol Evol. 2004 May; 19(5):264-71.

Mir A, et al. Identification of mutations in TRAPPC9, which encodes the NIK- and IKK-beta-binding protein, in nonsyndromic autosomal-recessive mental retardation. Am J Hum Genet. 2009; 85: 909-915.

Mochida GH, et al. A truncating mutation of TRAPPC9 is associated with autosomal-recessive intellectual disability and postnatal microcephaly. Am J Hum Genet. 2009; 85: 897-902.

Montero I, et al. Paternal imprinting of mating preferences between natural populations of house mice (Mus musculus domesticus). Mol Ecol. 2013 May; 22(9):254962.

Moore AJ, et al. Interacting phenotypes and the evolutionary process.1. Direct and indirect genetic effects of social interactions. Evolution. 1997; 51:1352–1362.

Morel Swols D, et al. KBG syndrome. Orphanet J Rare Dis. 2017 Dec 19; 12(1):183.

Mouse Genome Sequencing C, et al. Initial sequencing and comparative analysis of the mouse genome. Nature. 2002; 420:520-562.

Musilli M, et al. Behavioral effects of Rho GTPase modulation in a model of Alzheimer's disease. Behav Brain Res. 2013 Jan 15; 237:223-9.

Musolf K, et al. Ultrasonic courtship vocalizations in wild house mice, Mus musculus musculus. Anim Behav. 2010; 79: 757–764.

Nakagawa S, et al. Repeatability for Gaussian and non-Gaussian data: a practical guide for biologists. Biological Reviews .2010; 85: 935-956.

Nakatani J, et al. Abnormal behavior in a Chromosome-engineered mouse model for human 15q11–13 duplication seen in autism. Cell. 2009 Jun 26; 137(7):1235-46.

Naumova OY, et al. Gene expression in the human brain: the current state of the study of specificity and spatiotemporal dynamics. Child Dev. 2013 Jan-Feb;84(1):76-88.

Neilsen PM. et al. Identification of ANKRD11 as a p53 coactivator. J Cell Sci. 2008 Nov 1;121(Pt 21):3541-52.

Neme R, et al. Fast turnover of genome transcription across evolutionary time exposes entire non-coding DNA to de novo gene emergence. Elife. 2016 Feb 2; 5:e09977.

Nicholls RD, et al. Genetic imprinting suggested by maternal heterodisomy in nondeletion Prader-Willi syndrome. Nature. 1989; 342(6247):281–85.

Numan M. Neural basis of maternal behavior in the rat. Psychoneuroendocrinology. Psychoneuroendocrinology. 1988;13(1-2):47-62.

Nuss P. Anxiety disorders and GABA neurotransmission: a disturbance of modulation. Neuropsychiatr Dis Treat. 2015 Jan 17; 11:165-75.

Nyby J. Ultrasonic vocalizations during sex behavior of male house mice (Mus musculus): A description. Behav Neural Biol.1983; 39: 128–134.

Ockeloen CW, et al. Further delineation of the KBG syndrome phenotype caused by ANKRD11 aberrations. Eur J Hum Genet. 2015 Sep; 23(9):1176-85.

Ogden LV. Do Animals Have Personality? The importance of individual differences. BioScience, Volume 62, Issue 6, 1 June 2012, Pages 533–537.

Owen EH, et al. Assessment of learning by the Morris water task and fear conditioning in inbred mouse strains and F1 hybrids: implications of genetic background for single gene mutations and quantitative trait loci analyses. Neuroscience.1997; 80: 1087–1099.

Ozer DJ, Benet-Martinez V. Personality and the prediction of consequential outcomes. Annu Rev Psychol. 2006; 57:401-21.

Pallares LF, et al. Mapping of Craniofacial Traits in Outbred Mice Identifies Major Developmental Genes Involved in Shape Determination. PLoS Genet. 2015 Nov 2; 11(11):e1005607.

Panksepp JB, et al. Affiliative Behavior, Ultrasonic Communication and Social Reward Are Influenced by Genetic Variation in Adolescent Mice. PLoS One. 2007 Apr 4; 2(4):e351

Paredes RG. Medial preoptic area/anterior hypothalamus and sexual motivation. Scandinavian Journal of Psychology. 2003; 44 (3): 203–212.

Patten MM, et al. The evolution of genomic imprinting: theories, predictions and empirical tests. Heredity (Edinb). 2014 Aug; 113(2):119-28.

Patterson GR. Coercive family process. Castalia; Eugene, OR: 1982.

Penke L, et al. The evolutionary genetics of personality. European Journal of Personality. 2007; 21:549–587.

Perez JD, et al. New Perspectives on Genomic Imprinting, an Essential and Multifaceted Mode of Epigenetic Control in the Developing and Adult Brain. Annu Rev Neurosci. 2016 Jul 8; 39:347-84.

Piaget J. Behavior and evolution. Pantheon Books, New York. (1978).

Pinheiro J, et al. nlme: Linear and Nonlinear Mixed Effects Models. R package version (2016) 3.1-128.

Plotkin HC. Behavior and evolution. In: Plotkin HC (ed) The role of behavior in evolution. The MIT Press, Cambridge. (1988).

Poeppl TB, et al. A neural circuit encoding sexual preference in humans. Neurosci Biobehav Rev. 2016 Sep; 68:530-536.

R foundation for statistical computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org

Reale D. What do we really know about selection on personality? European Journal of Personality. 2007; 21 (5): 618-620.

Redden DT, Allison DB. The effect of assortative mating upon genetic association studies: spurious associations and population substructure in the absence of admixture. Behav Genet. 2006 Sep; 36(5):678-86.

Revelle W. psych: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA, (2016). http://CRAN.R-project.org/package=psych Version = 1.5.8

Robinson WP, Lalande M. Sex-specific meiotic recombination in the prader-willi/angelman syndrome imprinted region. Hum. Mol. Genet. 1995; 4: 801–806.

Rozengurt E, et al. Taste receptor signaling in the mammalian gut. Curr Opin Pharmacol. 2007 Dec; 7(6):557-62.

Rudenko A, Tsai LH. Epigenetic modifications in the nervous system and their impact upon cognitive impairments. Neuropharmacology. 2014 May; 80:70-82.

Ruf N, et al. Sequence-based bioinformatics prediction and QUASEP identify genomic imprinting of the KCNK9 potassium channel gene in mouse and human. Hum Mol Genet. 2007 Nov 1;16(21):2591-9.

Ryabko D, Reznikova Z. On the evolutionary origins of differences in sexual preferences. Front Psychol. 2015 May 21;6:628.

Ryan MJ, et al. Sexual selection for sensory exploitation in the frog Physalaemus pustulosus. Nature. 1990 Jan 4; 343(6253):66-7.

Sameroff AJ. Developmental systems: Contexts and evolution. In: Kessen W, editor. Handbook of child psychology, 4th ed. Vol. 1. History, theory, and methods. Wiley; New York: 1983. pp. 237–294.

Sanchez-Roige S, et al. The genetics of human personality. Genes Brain Behav. 2018 Mar; 17(3):e12439.

Sanes et al. Development of the Nervous System. Academic Press .(2011).

Sandovici I, et al. Human imprinted chromosomal regions are historical hot-spots of recombination. PLoS Genet. 2006 Jul; 2(7):e101. Epub 2006 May 22.

Sato M. Early Origin and Evolution of the Angelman Syndrome Ubiquitin Ligase Gene. Front Cell Neurosci. 2017 Mar 7; 11:62.

Schmidt-Edelkraut U, et al. Zac1 regulates cell cycle arrest in neuronal progenitors via Tcf4. Mol Cell Biol. 2014; 34(6):1020–30

Schunke A, et al. TINA manual landmarking tool: software for the precise digitization of 3D landmarks. Front Zool. 2012 Apr 5;9(1):6.

Shahani N, et al. Forebrain depletion of Rheb GTPase elicits spatial memory deficits in mice. Neurobiol Aging. 2017 Feb; 50:134-143.

Shapiro SS, Wilk MB. "An analysis of variance test for normality (complete samples)". Biometrika. 1965; 52 (3–4): 591–611.

Shine R, et al. Pick on someone your own size: ontogenetic shifts in mate choice by male garter snakes result in size-assortative mating. Animal Behaviour. 2001; 61:1133–1141.

Sih A, et al. Behavioral syndromes: An integrative overview. Quarterly Review of Biology. 2004; 79(3): 241-277.

Sinervo B, et al. Correlational selection and the evolution of genomic architecture. Heredity. 2002; 89:329–338.

Sirmaci A, et al. Mutations in ANKRD11 cause KBG syndrome, characterized by intellectual disability, skeletal malformations, and macrodontia. American Journal of Human Genetics .2011; 89, 289–294.

Sirmaci M, et al. Mutations in ANKRD11 cause KBG syndrome, characterized by intellectual disability, skeletal malformations, and macrodontia. Am J Hum Genet. 2011 Aug 12;89(2):289-94.

Smith JM. Theories of sexual selection Trends. Ecol Evol. 1991 May; 6(5):146-51.

Smith RJ, et al. Identification of novel imprinted genes in a genome-wide screen for maternal methylation. Genome Res. 2003; 13(4):558-69.

So D, et al. Behavioral drive or behavioral inhibition in evolution: subspecific diversification in holarctic passerines. Evolution. 2005; 59:2669–2677.

Sohar I, et al. Mouse mutants lacking the cation-independent mannose 6-phosphate/insulin-like growth factor II receptor are impaired in lysosomal enzyme transport: comparison of cation-independent and cation-dependent mannose 6-phosphate receptor-deficient mice. Biochem J. 1998 Mar 1;330 (Pt 2):903-8.

South SC, et al. Adolescent personality moderates genetic and environmental influences on relationships with parents. J Pers Soc Psychol. 2008 May; 94(5):899-912.

Spahn L, et al. An ICE pattern crystallizes. Nat Genet. 2003; 35(1):11–12.

Squier RW, et al. The relationship between facial structure and personality characteristics. Br J Soc Psychol. 1981 Sep; 20(Pt 3):151-60.

Squier RW, et al. The relationship between facial structure and personality characteristics. Br J Soc Psychol. 1981 Sep; 20(Pt 3):151-60.

Steitz JA. Polypeptide chain initiation: nucleotide sequences of the three ribosomal binding sites in bacteriophage R17 RNA. Nature. 1969 Dec 6;224(5223):957-64.

Störchel PH, et al. A large-scale functional screen identifies Nova1 and Ncoa3 as regulators of neuronal miRNA function. EMBO J. 2015; 34:2237–54.

Sun Z, et al. Replicative mechanisms of CNV formation preferentially occurs as intrachromosomal events: evidence from Potocki-Lupski duplication syndrome. Hum Mol Genet. 2013 Feb 15;2 2(4):749-56.

Suzuki S, et al. The evolution of mammalian genomic imprinting was accompanied by the acquisition of novel CpG islands. Genome Biol Evol. 2011; 3:1276-83.

Swaab DF, et al. Brain research, gender and sexual orientation. J Homosex. 1995; 28(3-4):283-301.

Sykes PJ, et al. Quantitation of targets for the polymerase chain reaction by use of limiting dilution. Biotechniques. 1992; 13(3): 444–449.

Takumi T. A humanoid mouse model of autism. Brain Dev. 2010; 32(9), 753–758.

Tamada K, et al. Decreased exploratory activity in a mouse model of 15q duplication syndrome: implications for disturbance of serotonin signaling. PLOS ONE. 2010; 5(12):e15126.

Tasan RO, et al. Altered GABA transmission in a mouse model of increased trait anxiety. Neuroscience. 2011 Jun 2; 183:71-80.

Tekin M, et al. The KBG syndrome: Confirmation of autosomal dominant inheritance and further delineation of the phenotype. Am J Med Genet A. 2004 Oct 15;130A(3):284-7.

Templeton AR. Evolution (Lawrence, Kans.). 1996; 50, 909–915.

Terracciano A, et al. Personality, metabolic rate and aerobic capacity. PLoS One. 2013; 8(1):e54746.

Teyssier A. Partners' personality types and mate preferences: predation risk matters. Behavioral Ecology, Volume 25, Issue 4, 1 July 2014, Pages 723–733.

Thompson ER. Development and Validation of an International English Big-Five Mini-Markers. Personality and Individual Differences. October 2008; 45 (6): 542–548.

Tooby J, Cosmides L. On the universality of human nature and the uniqueness of the individual: the role of genetics and adaptation. J Pers. 1990 Mar; 58(1):17-67.

Tremmel M, et al. Insect personality depends on environmental conditions. Behavioural Ecology. 2013; 24 (2): 386-392.

Tsai TF, et al. Paternal deletion from Snrpn to Ube3a in the mouse causes hypotonia, growth retardation and partial lethality and provides evidence for a gene contributing to Prader-Willi syndrome. Hum Mol Genet. 1999 Aug; 8(8):1357-64.

Urraca N, et al. The interstitial duplication 15q11.2-q13 syndrome includes autism, mild facial anomalies and a characteristic EEG signature. Autism Res. 2013 Aug;6(4):268-79.

van Oers K, and Mueller J. Evolutionary genomics of animal personality. Philos Trans R Soc Lond B Biol Sci. 2011; 365(1560): 3991–4000.

van Oortmerssen GA. Artificial selection for short and long attack latencies in wild Mus musculus domesticus. Behav Genet. 1981 Mar; 11(2):115-26.

Varki A, Kornfeld S. P-Type Lectins. Essentials of Glycobiology. 3rd edition. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2015-2017. Chapter 33.

Veenema, A. H. et al. Genetic selection for coping style predicts stressor susceptibility. J Neuroendocrinol. 2003 Mar; 15(3):256-67.

Verbeek MEM, et al. Consistent individual differences in early exploratory -behaviour of male great tits. Animal Behaviour. 1994; 48(5): 1113-1121.

Verona RI, et al. Genomic imprinting: intricacies of epigenetic regulation in clusters. Annu Rev Cell Dev Biol. 2003; 19:237–259.

Verweij KJ, et al. 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. 2012 Oct; 66(10):3238-51.

Vincent ACJ, Sadler LM. Faithful pairbonds in Wild seahorses, Hippocampus whitei. Anim. Behav. 1995; 50, 1557-1569.

Vines TH, Schluter D. Strong assortative mating between allopatric sticklebacks as a by-product of adaptation to different environments. Proc Biol Sci. 2006 Apr 22; 273(1589):911-6.

von Merten S, et al. A role for ultrasonic vocalisation in social communication and divergence of natural populations of the house mouse (Mus musculus domesticus). PLoS One. 2014 May 9; 9(5):e97244.

Waddington CH. Evolutionary adaptation. In: Sax S (ed) The evolution of life. University of Chicago Press. (1960).

Wang Q, et al. Altered G protein-coupling functions of RNA editing isoform and splicing variant serotonin2C receptors. J Neurochem. 2000 Mar; 74(3):1290-300.

Wang Z, et al. RNAseq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009 Jan; 10(1):57-63.

Washburn CP, et al. Serotonergic raphe neurons express TASK channel transcripts and a TASK-like pH- and halothane-sensitive K+conductance. J Neurosci. 2002; 22(4):1256–65.

Wedell N. and Ritchie MG. Male age, mating status and nuptial gift quality in a bushcricket. Animal Behaviour. 2004; 67: 1059-1065.

Wilkins JF, et al. What good is genomic imprinting: the function of parent-specific gene expression.Nat. Rev. Genet. 2003; 4(5):359–68.

Wilkins JF. Genomic imprinting and conflict-induced decanalization. Evolution. 2011; 65:537–553.

Wilkinson LS, et al. Genomic imprinting effects on brain development and function. Nat Rev Neurosci. 2007; 8(11):832–43.

Wilson DS, et al. Shyness and boldness in humans and other animals. Trends Ecol Evol. 1994; 19:442-446.

Wolf JB, et al. A maternal-offspring coadaptation theory for the evolution of genomic imprinting. PLoS Biol. 2006; 4:2238–2243.

Wolf JB, et al. Selective abortion and the evolution of genomic imprinting. J Evol Biol. 2009; 22:2519–2523.

Wolf JB, et al. The coadaptation of parental and offspring characters. Evolution. 1998; 52:299–308.

Wolf JB, et al. What are maternal effects (and what are they not) Philos Trans R Soc Lond B Biol Sci. 2009; 364:1107–1115.

Wolf JB. Evolution of genomic imprinting as a coordinator of coadapted gene expression. Proc Natl Acad Sci U S A. 2013 Mar 26; 110(13):5085-90.

Wolin SL, Walter P. Ribosome pausing and stacking during translation of a eukaryotic mRNA. EMBO J. 1988 Nov;7(11):3559-69.

Woodcock KA, et al. Task-switching deficits and repetitive behaviour in genetic neurodevelopmental disorders: data from children with Prader-Willi syndrome Chromosome 15 q11–q13 deletion and boys with Fragile X syndrome. Cogn Neuropsychol. 2009; 26(2):172–94.

Wright, S. Variability Within and Among Natural Populations (University of Chicago Press, London), Vol. 4. (1978).

Wyles JS, et al. Birds, behavior, and anatomical evolution. Proc Natl Acad Sci U S A. 1983 Jul; 80(14):4394-7.

Xue HJ, et al. Assortative mating between two sympatric closely-related specialists: inferred from molecular phylogenetic analysis and behavioral data. Sci Rep. 2014 Jun 25; 4:5436.

Yang JH, et al. deepBase: a database for deeply annotating and mining deep sequencing data. Nucleic Acids Res 38(Database issue): D123–D130. 2010.

Young R, et al. A fully automated light/dark apparatus useful for comparing anxiolytic agents. Pharmacol Biochem Behav. 1991 Dec; 40(4):739-43.

Yuen CH, et al. Personality does not constrain social and behavioural flexibility in African striped mice. Behavioral Ecology and Sociobiology. 2015; 69(8): 1237-1249.

Zhang A, et al. Characterization of transcriptional regulatory domains of ankyrin repeat cofactor. Biochem Biophys Res Commun. 2007a Jul 13; 358(4):1034-40.

Zhang A, et al. Subcellular localization of ankyrin repeats cofactor-1 regulates its corepressor activity. J Cell Biochem. 2007b Aug 1; 101(5):1301-15.

Zhang PL. et al. Identification of a novel family of ankyrin repeats containing cofactors for p160 nuclear receptor coactivators. J Biol Chem. 2004 Aug 6; 279(32):33799-805.

Zhang XS, Hill WG. Genetic variability under mutation selection balance. Trends in Ecology & Evolution. 2005; 20:468–470.

Zhang YJ, et al. Rapid birth-and-death evolution of imprinted snoRNAs in the Prader-Willi syndrome locus: implications for neural development in Euarchontoglires. PLoS One. 2014 Jun 19; 9(6):e100329.

Zhao Y, et al. Comparison of Droplet Digital PCR and Quantitative PCR Assays for Quantitative Detection of Xanthomonas citri Subsp. citri. PLoS One. 2016 Jul 18; 11(7):e0159004.

Zheng L-L, et al. deepBase v2.0: identification, expression, evolution and function of small RNAs, LncRNAs and circular RNAs from deep-sequencing data. Nucleic Acids Res (2016) 44(D1): D196–D202.

Zhong J, et al. c-Fos expression in the paternal mouse brain induced by communicative interaction with maternal mates. Molecular Brain. 2014; 7 (1): 1–23.

Zurita E, et al. Genetic polymorphisms among C57BL/6 mouse inbred strains. Transgenic Res. 2011 Jun; 20(3):481-9.

Acknowledgements

Four years ago, when I decided to study evolutionary biology for my PhD, I had no background in this field and it was really tricky to convince group leaders in evolutionary biology to give me a PhD position in their groups. So foremost I would like thank my supervisor Diethard Tautz as the first person in evolutionary biology that believed my motivation and interest in this field and trusted my ability and knowledge to give me this opportunity to start my PhD here at MPI for evolutionary biology. He has made very friendly atmosphere in his group which makes even a very young scientist like me able to not only express freely her own ideas but also encouraged me to develop and work on it. I highly thank him for his incredible support, the freedom and the independence he gave me during these three years which made my PhD time very enjoyable.

I highly thank Rebecca Krebs who introduced me to behavioral biology and help me to understand this field better. I thank her for her valuable work during her PhD and establishment of a nice system to work on mouse personality. The work presented here would not have been possible to achieve in three years if her work would not have been available.

I thank Peter Refki who introduced me to morphological traits especially craniofacial feature and skull development. I appreciate him for being one of the collaborators in my PhD project who made me able to make a nice connection between behavioral genetics and morphological trait.

I am grateful to Chen Xie for his valuable advice with transcriptome and ribosomal profiling data analysis.

I thank Stefan Dennenmoser and Cemalettin Bekpen who trained me for gene copy number analysis from both conceptual and methodological aspects.

I appreciate Guy Reeves for his valuable advice and great discussions that we had together especially in the early stage of my PhD to establish a new system to study small RNAs.

I am grateful to Derek Caetano-Anolles and Luisa Pallares for providing me RNAseq data from different mouse brain development timelines.

I thank Rafik Neme who taught me genome and transcriptome assembly in the early stage of my PhD.

I thank Sven Künzel for the sequencing and Cornelia Burghardt, Ellen McConnell, Nicole Thomsen, Heike Buhtz, Barbara Kleinhenz, Elke Blohm-Sievers and Elke Bustorf for their great lab-related support.

I am grateful to Christine Pfeifle who helped and advised me to organize the work with mice. I thank also the mouse team, in particular Susanne Holz, Camilo Medina, Anastasia Vock, Anika Jonas for taking care of my mice and Heike Harre for her help with behavioral tests and brain dissection.

I thank Sabine Meier and Britta Baron who helped me to arrange my business trips during my PhD time.

I had many up and down during my PhD time. I would like thank my friends Zahra Khomarbaghi, Elzbieta Iwaszkiewicz, Jatin Arora, Neva Skrabar, Wioleta Rasmus and Nazanin Alavinejad for all the mental support in both bad and good moments.

I want to say special thanks to my husband, Mojtaba Moeini, for his unconditional support and love. I thank him for being the best company I could have wished in my life who even quitted his job in our country to accompany me in Germany for my PhD.

I am also grateful to my family in Iran for their support and love that I always felt even from far geographical distance.

Contributions to the thesis

Chapter one

I designed the study, genotyped mice, generated small RNA library, performed transcriptome and gene ontology analysis, recorded and scored behavioral videos and also performed statistical analysis for mouse personality. Rebecca Krebs-Wheaton helped me for running the behavioral tests and the entire mouse team took care of my mice during my PhD study. Elke Blohm-Sievers helped scanning the mice and all morphological measurements and analysis have been done by Peter Refki.

Chapter two

I designed the study, performed transcriptome analysis, ran different behavioral tests, and conducted gene expression correlation analysis. Derek Caetano-Anolles and Luisa Pallares provided RNAseq data from different mouse brain development timelines and I analyzed the data. The knockout mice were generated by the company genOway (www.genoway.com), after discussing the goals and the procedures with me and my supervisor Diethard Tautz. The entire mouse team took care of my mice during my PhD study.

Declaration

I hereby declare that:

i. Apart from my supervisor's guidance, the content and design of this dissertation is the product of my own work. The co-authours contributions to specific paragraphs are listed in the thesis outline section;

ii. This thesis has not already been submitted either partially or wholly as part of doctoral degree to another examination body, and no other materials are published or submitted for publication than indicated in the thesis;

iii. The preparation of the thesis has been subjected to the Rules of Good Scientific Practice of the German Research Foundation.

Plön, July 17th, 2018

Maryam Keshavarz

Curriculum Vitae

Personal Information

Name: Maryam Keshavarz

Gender and Civil status: Female & married

Place of Birth: Kazeroon, Iran

Date of Birth: November 26, 1988

Citizenship: Iranian

Education

PhD- Evolutionary Biology/ Behavioral Genetics

Max Planck for Evolutionary Biology (since December 2014)

M.Sc. –Animal Cell Developmental Biology

University of Tehran (2011-2013)

B.Sc. –General Biology

University of Shiraz (2007-2011)

High School Diploma in Sciences

Fatemieh High School and Alzahra Pre-University Centre, Shiraz, Iran, (2002-2007)