On the trail of anxiety – Analysis of copy number variants as a factor influencing anxiety-related behavior in mice

Dissertation

der Fakultät für Biologie

der Ludwig-Maximilians-Universität München

Angefertigt am Max-Planck-Institut für Psychiatrie, München

vorgelegt von

Julia Brenndörfer

München, 31. Januar 2013

Erstgutachter: Prof. Dr. Rainer Landgraf

Zweitgutachter: Prof. Dr. John Parsch

Eingereicht am: 31. Januar 2013

Mündliche Prüfung am: 18. Juni 2013

On the trail of anxiety – Analysis of copy number variants as a factor influencing anxiety-related behavior in mice

Dissertation

der Fakultät für Biologie

der Ludwig-Maximilians-Universität München

Angefertigt am Max-Planck-Institut für Psychiatrie, München

vorgelegt von

Julia Brenndörfer

München, 31. Januar 2013

Für meine Eltern

Phantasie ist wichtiger als Wissen,
denn Wissen ist begrenzt.

Albert Einstein

Table of contents

Table of contents ..... **I**

List of figures ..... **V**

List of tables ..... **VII**

List of abbreviations ..... **IX**

I Abstract **1**

II Introduction **3**

1 The definition of anxiety 3

2 The comorbidity of anxiety and depression 4

3 Can human anxiety be modeled in animals? 5

4 The HAB/LAB mouse model 8

5 Definition and forming mechanisms of CNVs and SVs 9

6 How CNVs mediate their effect on gene expression 13

7 The impact of CNVs on human diseases and anxiety phenotypes 15

8 Aim of the study 16

III Materials and Methods **19**

1 Equipment, chemicals and remarks 19

2 Housing and testing conditions of animals 20

2.1 Elevated plus-maze (EPM) test 20

2.2 Open field (OF) test 21

2.3 Forced swim test (FST) 22

2.4 Stress reactivity test (SRT) 23

2.5 Tail suspension test (TST) 24

3 Analysis of CNVs in the HAB/LAB mouse model 25

3.1 Detection of CNVs by genome-wide oligonucleotide array-based comparative genomic hybridization (aCGH) 25

3.1.1 DNA extraction 25

3.1.2 Processing of the samples in aCGH 25

3.1.3 Interpretation and further analysis of aCGH data 26
3.2 Detection of CNVs by the Jax Mouse Diversity Genotyping Array (Jax MDGA) 27
3.2.1 Performance of the assay 27
3.2.2 Genotyping using the MouseDivGeno R package 27
3.3 Detection of CNVs and structural variants (SVs) by next-generation sequencing (NGS) 28
3.3.1 DNA extraction 28
3.3.2 Performance of sequencing process 29
3.3.3 Data handling to detect CNVs and structural variants (SVs) 30
3.3.3.1 Quality control and alignment 30
3.3.3.2 CNV detection 31
3.3.3.3 SV detection 32
3.4 Influence of CNVs on gene expression 34
3.4.1 Detection of differentially expressed genes in HAB/LAB mice by gene expression microarray analysis 34
3.4.2 Confirmation of candidate genes by quantitative real-time PCR (qPCR) 35
3.4.2.1 Selection of candidate genes and primer design 35
3.4.2.2 RNA extraction 36
3.4.2.3 cDNA synthesis 37
3.4.2.4 Performance of qPCR 38
3.4.3 Correlation of CNV data with expression data 39
3.5 Influence of a single CNV comprising the glyoxalase 1 (*Glo1*) locus on anxiety-related behavior 41
3.5.1 Breeding protocol 41
3.5.2 Genotyping of the animals 42
3.5.3 Behavioral testing 44
3.5.4 Statistical analysis 44
4 Analysis of CNVs in CD-1 mice 45
4.1 Testing of CD-1 mice in different behavioral experiments 45
4.2 Detection of CNVs in CD-1 mice 46
4.2.1 JAX Mouse Diversity Genotyping Array (JaxMDGA) 46

4.2.1.1 Performance of the array *46*

4.2.1.2 Computational analysis of Jax MDGA data *46*

4.2.2 Multiplex Ligation-dependent Probe Amplification (MLPA) 49

4.2.2.1 Designing of synthetic MLPA probes *49*

4.2.2.2 Performance of MLPA *56*

4.2.2.3 Performance of capillary electrophoresis *57*

4.3 Association of CNVs with behavior 58

4.4 Cluster analysis of interesting genes using DAVID Bioinformatics Database 58

IV Results 59

1 Position and impact of CNVs in HAB/LAB mice 59

1.1 Position of CNVs detected by aCGH 59

1.2 Position of CNVs detected by JaxMDGA 62

1.3 Position of CNVs and structural variants (SVs) detected by NGS 67

1.4 Comparison of CNVs detected by aCGH, JaxMDGA and NGS 70

1.5 Expression versus CNV data 77

1.5.1 Differentially expressed genes detected by expression microarray 77

1.5.2 Confirmation of differentially expressed genes by qPCR 79

1.5.3 Correlation of differentially expressed genes with CNV data 80

1.6 Influence of the *Glo1* CNV on behavior 82

2 Position and impact of CNVs in CD-1 mice 84

2.1 CNVs examined by MLPA 84

2.2 Behavior of CD-1 mice screened for CNVs by JaxMDGA 86

2.3 Position of CNVs detected in CD-1 mice by JaxMDGA 86

2.4 CNVs associated with behavior 92

3 Candidate genes of anxiety-related behavior 96

3.1 Candidate genes in HAB/LAB mice 96

3.2 Candidate genes in CD-1 mice 99

3.3 Identified clusters in the candidate genes 101

V Discussion 105

1 The detection of CNVs 105

1.1 CNVs in the HAB/LAB mouse model 105
1.2 CNVs in CD-1 mice 109
2 The contribution of CNVs to phenotypic variation 111
2.1 The influence of CNVs on expression levels in HAB/LAB mice 112
2.2 A breeding approach to examine the impact of a single CNV comprising the *Glo1* locus on anxiety-related behavior 114
2.3 The effect of CNVs on anxiety-related behavior in CD-1 mice 115
3 Candidate genes of anxiety-related behavior 117
3.1 Candidate genes in the HAB/LAB mouse model 117
3.2 Candidate genes in CD-1 mice 122
4 Conclusion 124

VI Perspectives 127

VII References 129

VIII Supplementary data 149

Acknowledgements/ Danksagung 153

Curriculum Vitae 155

Declaration/ Erklärung 157

List of figures

Figure II-1: Development of anxiety-related behavior in the HAB/LAB mouse model over 45 generations 9
Figure II-2: Mechanism of NAHR 11
Figure II-3: Mechanism of NHEJ 12
Figure II-4: Mechanism of FoSTeS 13
Figure II-5: Gene dosage alterations and their influence on gene expression 14
Figure III-1: Schematic drawing of EPM test. A mouse is sitting on one of the open arms 21
Figure III-2: Bird's eye view of the open field 22
Figure III-3: Drawing of FST 23
Figure III-4: Mouse in SRT 24
Figure III-5: Experimental set-up of TST 24
Figure III-6: Principle of bin counts 32
Figure III-7: Principle of SV detection 33
Figure III-8: Positions of genotyping primers 43
Figure III-9: Test battery of CD-1 panel 45
Figure III-10: Situation of potential information loss based on applied algorithm 47
Figure III-11: Breakpoint definition for CD-1 CNVs 48
Figure IV-1: Frequency distribution of CNVs detected by aCGH in HAB *vs.* LAB mice, in size classes of 0.5 Mbp 60
Figure IV-2: Detailed frequency distribution of CNVs < 0.5 Mbp, detected by aCGH in HAB *vs.* LAB mice 60
Figure IV-3: Chromosomal distribution of CNVs detected by aCGH in HAB *vs.* LAB mice 61
Figure IV-4: Frequency distribution of CNVs detected by JaxMDGA in HAB *vs.* LAB mice 63
Figure IV-5: Detailed frequency distribution of CNVs < 50 kbp, detected by JaxMDGA in HAB *vs.* LAB mice 64
Figure IV-6: Frequency distribution of CNVs < 1 kbp, by JaxMDGA detected in HAB *vs.* LAB mice 64
Figure IV-7: Chromosomal distribution of CNVs detected by JaxMDGA in HAB *vs.* LAB mice 65
Figure IV-8: Frequency distribution of CNVs detected by NGS in HAB *vs.* LAB mice, in size classes of 50 kbp 67

Figure IV-9: Detailed frequency distribution of CNVs < 50 kbp, detected by NGS in HAB *vs.* LAB mice 68

Figure IV-10: Detailed frequency distribution of CNVs < 5 kpb, detected by NGS in HAB *vs.* LAB mice 68

Figure IV-11: Frequency distribution of CNVs detected by NGS in HAB *vs.* LAB mice over all chromosomes 69

Figure IV-12: Genomic positions of CNVs on chromosomes 1 – 4, detected in HAB *vs.* LAB mice 72

Figure IV-13: Genomic positions of CNVs on chromosomes 5 – 8, detected in HAB *vs.* LAB mice 73

Figure IV-14: Genomic positions of CNVs on chromosomes 9 – 12, detected in HAB *vs.* LAB mice 74

Figure IV-15: Genomic positions of CNVs on chromosomes 13 – 16, detected in HAB *vs.* LAB mice 75

Figure IV-16: Genomic positions of CNVs on chromosomes 18, 19, X & Y, detected in HAB *vs.* LAB mice 76

Figure IV-17: Genomic position of CNVs on chromosome 17, detected in HAB *vs.* LAB mice 77

Figure IV-18: Picture of a gel showing both HAB- and LAB-specific genotypes 82

Figure IV-19: Results of behavioral testing in animals of the 6th *Glo1* breeding generation 83

Figure IV-20: Behavior of 64 CD-1 mice screened for CNVs by JaxMDGA 86

Figure IV-21: Genomic positions of CNVs detected in 64 CD-1 mice by JaxMDGA 87

Figure IV-22: Frequency distribution of CNVs detected in 64 CD-1 mice by JaxMDGA 88

Figure IV-23: Detailed frequency distribution of CNVs < 50 kbp detected in CD-1 mice 89

Figure IV-24: Frequency distribution of CNVs < 1 kbp detected in 64 CD-1 mice by JaxMDGA 89

Figure IV-25: Chromosomal distribution of CNVs detected in 64 CD-1 mice 90

Figure IV-26: Association of copy number with behavior 95

List of tables

Table III-1: List of oligonucleotides used as primers for qPCR 35
Table III-2: Thermocycler program for cDNA synthesis..... 38
Table III-3: Conditions of qPCR run..... 39
Table III-4: Similarity coefficients of Glo1 breeding approach 42
Table III-5: Primers for targeting CNV breakpoint of *Glo1* locus..... 43
Table III-6: PCR conditions of Glo1 genotyping. 44
Table III-7: Sequence information of MLPA probes 51
Table III-8: Genomic position of MLPA probes..... 54
Table IV-1: Detailed information on aCGH-detected CNVs per chromosome 61
Table IV-2: Detailed information on CNVs detected by JaxMDGA in HAB *vs.* LAB mice..... 65
Table IV-3: Detailed information on CNVs detected by NGS in HAB *vs.* LAB mice..... 69
Table IV-4: Comparison of CNVs detected in HAB *vs.* LAB mice by three different methods..... 71
Table IV-5: Expression differences of genes analyzed later on in qPCR, detected by expression microarray 78
Table IV-6: Expression differences of genes successfully tested in qPCR 79
Table IV-7: Count of genes that are differentially expressed between HAB and LAB mice and overlap with CNVs..... 81
Table IV-8: Result of the correlation of CNVs with expression 81
Table IV-9: Comparison of the CNV status of 33 different loci determined by MLPA with the CNV status detected by aCGH, JaxMDGA and NGS in a HAB sample..... 84
Table IV-10: Detailed information on CNVs detected in 64 CD-1 mice by JaxMDGA 90
Table IV-11: List of CNVs detected in 64 CD-1 mice that were significantly associated with the time animals spent on the open arm of the EPM 92
Table IV-12: Protein coding genes in CNVs detected by aCGH, JaxMDGA and NGS in HAB/LAB mice 96
Table IV-13: Protein coding genes in CNVs detected in HAB/LAB and CD-1 mice 99
Table IV-14: Protein coding genes in CNVs detected by JaxMDGA in 64 CD-1 mice 100
Table IV-15: Identified clusters in 15 candidate genes of anxiety-related behavior in HAB/LAB mice 102
Table IV-16: Identified clusters in 53 candidate genes of anxiety-related behavior in CD-1 mice 103

List of abbreviations

BLA	basolateral amygdala
bp	basepair(s)
CeA	central amygdala
Cg	cingulate cortex
cM	centimorgan
Cort	corticosterone
DNA	desoxyribonucleic acid
EPM	elevated plus-maze, elevated plus-maze
FoSTeS	fork stalling and template switching
FST	forced swim test
HAB	high anxiety-related behavior
Ig	immunoglobulin
IGPs	invariant genomic probes
IQR	interquartile range
kbp	kilobasepairs
LAB	low anxiety-related behavior
LCRs	low-copy repeats
LHS	left hybridizing sequence
LMP	long mate-paired library
LPO	left probe oligonucleotide
Mbp	megabasepairs
MGI	mouse genome informatics
MHC	major histocompatibility complex, major histocompatibility complex
MLPA	multiplex ligation-dependant probe amplification
NAHR	nonallelic homologues recombination
NCBI	National Center for Biotechnology Information
NGS	next-generation sequencing
NHEJ	nonhomologous end-joining
nt	nucleotide(s)

OF	open field
PCR	polymerase chain reaction, polymerase chain reaction
PVC	polyvinylchloride
PVN	hypothalamic paraventricular nucleus
QC	quality control
qPCR	quantitative real-time PCR, quantitative real-time PCR
RHS	right hybridizing sequence
RNA	ribonucleic acid
RPO	right probe oligonucleotide
SD	segmental duplication
SNP	single nucleotide polymorphism
SRT	stress reactivity test
SV	structural variant
TBE	tris - boric acid - EDTA
TST	tail suspension test
UCEs	ultraconserved elements
UCSC	University of California Santa Cruz

I Abstract

With the advances in genome-wide screening arrays and sequencing techniques scientists were enabled to examine genetic variations and their effects on behavioral phenotypes. While single nucleotide polymorphisms (SNPs) are the most widely studied form of genomic variations to date, another type of variants has become increasingly important in recent research, the copy number variants (CNVs). These large segments of DNA that can comprise up to several megabasepairs and differ in copy number with respect to a reference genome have been associated with several disorders and behavioral phenotypes before.

This study investigated the influence of CNVs on anxiety-related behavior. The detection of these variants turned out to be a major challenge since all methods available are biased by limitations of the design of the approach and the subsequent computational analyses. Therefore, three different techniques (next-generation sequencing and two distinct whole-genome genotyping arrays) were employed to identify CNVs in a CD-1-derived mouse model consisting of two mouse strains showing high (HAB) and low (LAB) anxiety-related behavior, respectively. By comparing CNVs in HAB *vs.* LAB mice with expression data of four distinct brain regions of high relevance to the limbic system (central and basolateral amygdala, cingulate cortex and the hypothalamic paraventricular nucleus), it was shown that CNVs can influence the expression of protein coding genes by the alteration of the genes' copy number *per se*. Therefore, the genes mapping into regions where CNVs were detected in HAB *vs.* LAB mice (by all three detection methods) were suggested to be possible effectors of anxiety-related behavior. Amongst these candidate genes those were considered to be the most interesting ones that were additionally found to map into regions of CNVs associated with anxiety-related behavior in CD-1 mice. CNVs in these mice were detected by means of a whole-genome genotyping array and subsequent processing of the raw data with a novel computational approach that was adapted from existing analysis methods.

Furthermore, to test the effect of a specific CNV on anxiety-related behavior *in vivo*, a breeding approach was used to generate animals with a full genetic background of HAB

mice except for one LAB-derived locus harboring a CNV that included the *Glo1* gene. No direct effect on the phenotype could be observed, however, the respective CNV might be involved in the manipulation of anxiety-related behavior taken into account the interaction with other factors.

Taken together, this study provides not only a comprehensive catalogue of CNVs in HAB/LAB mice but also the evidence that these variants can influence anxiety-related behavior. Furthermore, it gives a first insight into the functionality of CNVs with respect to anxiety-related behavior. Therefore, this thesis provides a profound basis for multiple advanced studies.

II Introduction

1 The definition of anxiety

When talking about anxiety, it is inevitable to address the term of fear since both are often conflated in the literature (Sylvers *et al.*, 2011). The relationship between fear and anxiety has been described controversially. Beck and Emery (2005) assessed fear as the cognitive response to threat whereas anxiety is the emotional response to that fear, while Sylvers and colleagues (Sylvers *et al.*, 2011) postulated that anxiety and fear encompass distinct emotional states and are largely unrelated. However, there is consensus in the definition of fear being an emotion that triggers adaptation mechanisms in response to the exposure to a distinct threat (Belzung and Philippot, 2007; Sylvers *et al.*, 2011), while anxiety is considered to be a future-oriented (emotional) state that does not require the presence of an acute stressor (Barlow, 2000; Belzung and Philippot, 2007; Sylvers *et al.*, 2011). Both anxiety and fear could be differentiated in either a state emotion that affects adaptations to specific situations, or a trait emotion that refers to affective characteristics of an individual across time and situation (Sylvers *et al.*, 2011). Hence, state anxiety is a type of anxiety that occurs at a particular moment in time when an individual approaches an ambiguous and uncertain threat. In contrast, trait anxiety is a persistent feature of an organism leading to a persistent hypervigilance for potential threats (Clement *et al.*, 2007; Sylvers *et al.*, 2011). However, since in most publications on anxiety the authors refrained from discriminating these two subtypes of anxiety, this distinction was also not made in this thesis. It could be kept in mind, though, that behavioral tests like the elevated plus-maze and open field test are considered to assess state anxiety, while trait anxiety would reflect a consistent characteristic (Belzung and Griebel, 2001; Clement *et al.*, 2007).

The complex phenomena of anxiety and fear are accompanied by specific responses including a behavioral (e.g., flight), physiological (e.g., increase in hart rate) and expressive (e.g., vocalization) component (Belzung and Philippot, 2007; Gross and Hen, 2004). Each of these components evolved to protect the individual from danger. However, in a pathological form of anxiety, they can severely lower the quality of life. In

humans six major anxiety disorders have been classified: generalized anxiety disorder, social and simple phobia, panic disorder, post-traumatic stress disorder and obsessive-compulsive disorder (American Psychiatric Association, 1994; Gross and Hen, 2004). The 1-year and lifetime prevalence rate of anxiety disorders in adults have been estimated to be higher than 10% and 16%, respectively, though these numbers vary greatly between published reports (Sommers, 2006). This huge number of affected people turns the elucidation of molecular and genetic causes of anxiety into an important field of current clinical and basic research.

2 The comorbidity of anxiety and depression

The relationship between anxiety and depression has been in the focus of interest since the 1980s (Mineka *et al.*, 1998). Although anxiety disorders have been classified distinct from mood disorders, like the major depressive disorder, there is evidence of an existing complex comorbidity (American Psychiatric Association, 1994; Hettema *et al.*, 2003; Kessler *et al.*, 2005; Mineka *et al.*, 1998; Stein and Heimberg, 2004; Yerevanian *et al.*, 2001). This comorbidity was shown on a diagnostic as well as molecular level. Thus, for example, it was reported that 40% of patients suffering from anxiety disorders also were afflicted with depression and, the other way round, that 67% of patients affected by depression had experienced any anxiety disorder (Hranov, 2007). Furthermore, a dysfunctional hypothalamic-pituitary-adrenal (HPA) axis, a system, which is involved in stress response, was connected with both types of disease (Cowen, 2009; Faravelli *et al.*, 2012; Hettema *et al.*, 2006; Kallen *et al.*, 2008). The involvement of the stress hormone system in anxiety and depression was further shown in clinical studies, which found elevated serum glucocorticoid concentrations in depressive patients, and animal studies, which revealed a correlation of increased corticosterone levels and measures of risk assessment (Finn *et al.*, 2003; Krishnan and Nestler, 2008). Other studies have focused on the genetic level of comorbidity (e.g., Lawford *et al.*, 2006; Molina *et al.*, 2011; van Veen *et al.*, 2012), leading to the assumption that the underlying genetics of anxiety and depression are nearly identical (Kendler and Prescott, 2006).

It therefore seems that both anxiety and depression may be triggered by a complex system of different molecular mechanisms, of which some are unique to a respective type of disorder and others are shared. Although the relationship between anxiety and depression is not yet fully understood, the comorbidity cannot be denied. On this account, even though the focus of this thesis was on the analysis of anxiety, tests for depression-like behavior and stress-reactivity (stress-reactivity test and tail suspension test) were performed to provide the basis for future examinations on this topic.

3 Can human anxiety be modeled in animals?

Believing in the theory of evolution, each complex phenotype is supposed to have its roots in more simple behavioral and molecular patterns. The investigation of a behavioral phenotype requires not only to question the underlying molecular characteristics, but also to consider how and why these behavioral and molecular patterns could have evolved. Especially for the study of complex traits like emotions (e.g., fear and anxiety) the knowledge of the traits' origin is important to comprehend the complex interactions of molecular processes involved. It has been postulated that a steadily changing environment could have led to the evolution of emotions to coordinate the activity of multiple body systems to meet the new challenges (Andrews *et al.*, 2002; Andrews and Thomson, 2009; Tooby and Cosmides, 1990). This coordination is likely to have evolved in a step-wise manner following the construction of new systems. Therefore, a gain in complexity of emotional processes and their related logistical systems should be apparent over evolutionary time (Andrews and Thomson, 2009; Belzung and Philippot, 2007). Indeed, such a trend was observed in a study investigating human anxiety from a phylogenetic perspective (Belzung and Philippot, 2007). For example, triggered by fear, biogenic amines like adrenaline, noradrenaline and dopamine have been shown to be released in humans. With some variations, these amines have been described in all vertebrates, and some of them were also found to be increased under stressful situations in invertebrates (Belzung and Philippot, 2007; Lacoste *et al.*, 2001; Weiger, 1997). Furthermore, stress hormones linked to human anxiety like the corticotropic-releasing hormone (CRH) are highly conserved across species. CRH has been found in vertebrates (e.g., mammals, birds and amphibians) but also as CRH-like molecules in some

invertebrates, for example annelids, insects and mollusks (Belzung and Philippot, 2007; Deussing and Wurst, 2005; Merali *et al.*, 2004; Ottaviani and Franceschi, 1996). There is further evidence that most animals share expressive components of fear and anxiety. Vocalizations to threat can be observed all across the phylum, from insects (e.g., crickets emit ultrasonic signals specific to the danger they are exposed to) to higher mammals (e.g., vervet monkeys have alarm calls specific to different predators) (Belzung and Philippot, 2007; Seyfarth *et al.*, 1980; Wyttenbach *et al.*, 1996). Moreover, chemical cues are known in this context. Social hymenoptera for instance use pheromones as alarm calls, thereby triggering fight or flight reactions in conspecifics (Maschwitz, 1966). The excretion of odors when experiencing fear or anxiety has also been shown in rodents, primates, and humans (Belzung and Philippot, 2007; Chen *et al.*, 2006; Haegler *et al.*, 2010; Kiyokawa *et al.*, 2006).

Besides, there are components of anxiety that seem to be unique to higher mammals and humans. The most important is certainly the subjective feeling state, thus the awareness of the emotion, which might appear at different levels (Belzung and Philippot, 2007; Lane *et al.*, 1998). It has been controversially discussed if animals, and primates in particular, are able to mentally experience past events or to imagine how a future might look like or what future experiences might feel like (Roberts and Feeney, 2009; Sajithlal *et al.*, 2002; Suddendorf and Corballis, 2007; Wheeler *et al.*, 1997). Such ability would imply a kind of self-consciousness, also called autonoetic consciousness (Chiong, 2011; Tulving, 1985). It is subject of discussion whether it exists in non-humans, and which level is required to experience anxiety. While the debate on this topic is still going on, it has be shown that human anxiety can be discriminated from anxiety of lower animals by the brain regions engaged (insofar as these even exist in animals). Thus, the involvement of the hypothalamus, hippocampus and amygdala in emotional response has been proven for many mammals and animals even lower in the phylogenetic tree. In contrast, there seem to be aspects of anxiety that are unique to great apes and humans since they engage the insular and cingulate cortex in these species solely (Belzung and Philippot, 2007).

In conclusion, human anxiety comprises facets that could be observed already in insects, but also others that seem to be unique to humans and maybe hominids (Belzung and Philippot, 2007). Therefore, it is likely that critical aspects of anxiety are regulated in a

similar way in humans and animals. To understand the complexity of molecular interactions triggering and triggered by anxiety in humans, it is inevitable to reveal the distinct systems involved in these complex patterns. These systems can be best examined in animals showing a rudimentary form of the respective behavior and molecular mechanisms. Additionally, specific genomic and physiological manipulations are required to investigate the underlying mechanisms, which are not feasible in humans. Hence, although not every aspect of anxiety can be modeled in animals, model organisms are an important and indispensable tool in science and in anxiety research in particular.

The use of the mouse as model organism in neuropsychiatric and other preclinical research has rapidly escalated in the last decades (Cryan and Holmes, 2005). There are different reasons for this trend. First, due to their high reproduction rate and their relatively small size, mice are easy to host and breed in a quick and cost effective manner. Furthermore, and even more important, there is evidence of a large genetic homology between mice and men. About 99% mouse orthologs were reported to be found in human genes and *vice versa* (Tecott, 2003). Therefore, genes influencing anxiety-related behavior in mice could be regarded as excellent candidate genes for human anxiety disorders. Besides the genetic there is also an extensive neuroanatomical homology between mice and men and, finally, mice show a similar response to specific drugs that are used to treat human disease (Blanchard *et al.*, 2003; Hovatta and Barlow, 2008; Tecott, 2003).

To date, several mouse models of anxiety and depression have been developed based on genetic engineering or inbreeding approaches (Cryan and Holmes, 2005; El Yacoubi and Vaugeois, 2007). A genetically homogeneous population resulting from sufficient inbreeding is beneficial for the investigation of genetic influences on specific traits since it reduces genomic variability. The individuals of that population are genetically almost identical and thus highly suitable for long-term investigations (Tecott, 2003). Following a breeding strategy of selecting mice showing extremes for a specific trait like anxiety, one can enhance the genetic features associated with the respective trait. Thereby, the mice's phenotype is shifted bidirectionally from the mean behavior of the original population, finally resulting in two distinct mouse lines (Krömer *et al.*, 2005). These two lines provide a powerful tool to investigate genetic underpinnings of the respective trait. As described

below, such a selective and bidirectional breeding approach was applied to generate the mouse model that was used in the context of this study.

4 The HAB/LAB mouse model

Following bidirectional selective breeding protocols, which were successfully applied in rats before (Landgraf *et al.*, 2007; Landgraf and Wigger, 2002; Liebsch *et al.*, 1998), a mouse model of two lines reflecting extremes in anxiety-related behavior was generated. The crucial breeding parameter, for which the animals were tested, was the time they spent on the open arms of the elevated plus-maze (EPM). The procedure of this test, a standard test to determine anxiety-related behavior in rodents (Pellow *et al.*, 1985), is described in chapter III-2.1. Depending on their performance on the EPM, the individuals of a founder population of more than 250 animals from over 25 litters of outbred Swiss CD-1 mice were used to form the basis for either the HAB (high anxiety-related behavior) or LAB (low anxiety-related behavior) mouse line. For nine generations outbreeding across families but within behavioral restrictions was performed. Afterwards, a strict inbreeding protocol (brother-sister mating) was followed (Krömer *et al.*, 2005). As illustrated in Figure II-1 the two mouse lines show a stable anxiety-related phenotype over generations of inbreeding, with HAB mice spending about 10% of the test time on the open arm of the EPM while LAB mice stay there for about 50% of the time. 23 generations after establishing the HAB/LAB mouse model the breeding of a control line (NAB; normal anxiety-related behavior) was initiated. The individuals of this line show an intermediate anxiety-related behavior, which is similar to the mean behavior shown by CD-1 mice, as depicted in Figure II-1.

To this day more than 40 generations of HAB/LAB mice have served as a basis for multiple studies with genetic, proteomic, cognitive, developmental and brain physiological background (e.g., Czibere *et al.*, 2011; Ditzen *et al.*, 2009; Hambsch *et al.*, 2010; Ionescu *et al.*, 2012; Krömer *et al.*, 2005; Sah *et al.*, 2012; Sotnikov *et al.*, 2011; Yen *et al.*, 2012). In the field of genetics several experiments have been performed to test the effect of, for instance, single nucleotide polymorphisms (SNPs) (e.g., Kessler *et al.*, 2007) or epigenetic factors (studies are still ongoing) in the HAB/LAB mouse model. So far, however, no

studies have been performed to investigate the influence of another type of genetic variance, the so called copy number variants (CNVs). This is at the central focus of the present study.

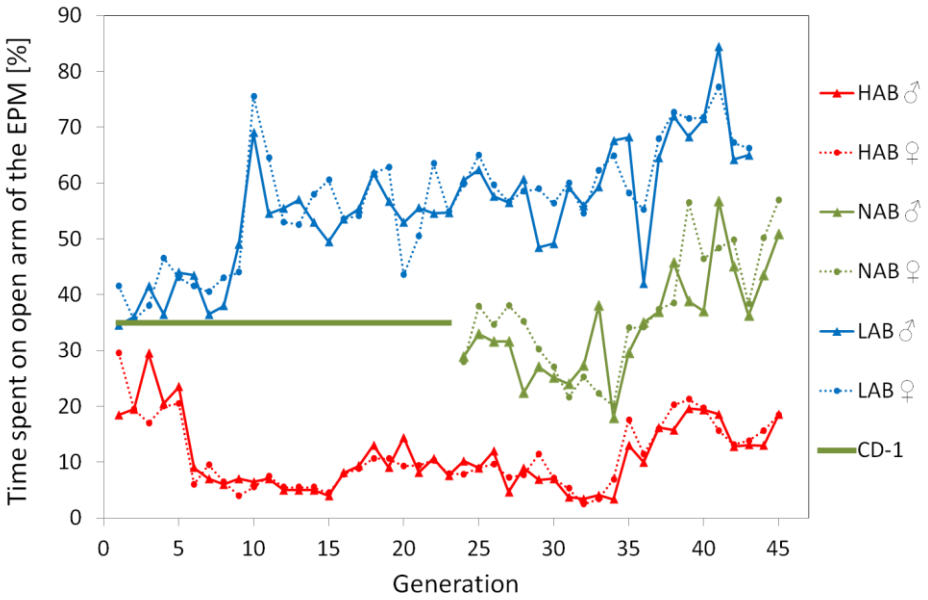

Figure II-1: Development of anxiety-related behavior in the HAB/LAB mouse model over 45 generations. The crucial parameter for the breeding was the time the animals spent on the open arm of the EPM, which is plotted on the y-axis. The x-axis shows the respective generation. The behavior of HAB animals is indicated by red lines, of LAB by blue lines, of NAB by green lines and of CD-1 mice by a thick green line. Dashed lines represent females, solid males.

5 Definition and forming mechanisms of CNVs and SVs

Lupski and colleagues (Lupski *et al.*, 1991) were the first to report that large genomic rearrangements and gene dosage effects, and not small alterations of the coding DNA, were causing an autosomal dominant neurodegenerative disorder (Canales and Walz, 2011). Since 1991, not least thanks to the development of advanced genotyping and sequencing techniques, many more examples of diseases caused by these so called copy number variants (CNVs) have been found. However, the understanding of the extent and influence of this type of genetic variations is still in its infancy (Canales and Walz, 2011).

Based on a widely accepted definition, CNVs are considered as large regions of duplicated or deleted DNA regions that vary in copy number with respect to a reference genome, and range in size from 1 kilobasepair (bp) to several megabasepairs (Kim *et al.*, 2008).

However, taken into account recent findings, it seems like the definition of the lower bound needs to be eased as it relies on an operational definition and approved detection methods enable the revelation of even smaller variants. For example, based on the application of new techniques, CNVs were found with a lower bound of merely 50 bp (Arlt *et al.*, 2012).

Reading the literature, caution needs to be exercised with respect to a term often mentioned in the context of CNVs, the so called segmental duplications (SDs). SDs, also called low-copy repeats (LCRs) were defined as duplicated regions, not necessarily adjacent to each other, larger than 1 kbp with over 90% sequence identity among the duplicates (Bailey *et al.*, 2004; Gu *et al.*, 2008; Kim *et al.*, 2008). In some publications no clear distinction between the two terms was drawn and the reader has to take care not to be confused. Furthermore, especially for CNVs detected by array-based methods, the sequence similarity between the copies of a specific region cannot be determined and thus duplications that should strictly be termed SDs are specified as CNVs. To prevent confusion, in the present study no classification of CNVs and SDs was established and the definition of detected CNVs did not comprise any lower limit. However, another term was used in this study, namely structural variations (SVs), which are genomic rearrangements that could be of complex nature like inversions or be simple insertions or deletions (Malhotra and Sebat, 2012).

Three major mechanisms, as outlined below, have been proposed to form CNVs and SVs: nonallelic homologues recombination (NAHR), nonhomologous end-joining (NHEJ), and fork stalling and template switching (FoSTeS) (Gu *et al.*, 2008).

Most of the recurrent CNVs, i.e. CNVs of the same genomic sequence that have been observed in different individuals, are caused by NAHR (Stankiewicz and Lupski, 2002). This mechanism is mediated by LCRs (also called SDs) at non-allelic loci during meiosis and, at lower frequency, during mitosis (Lam and Jeffreys, 2006; Malhotra and Sebat, 2012; Turner *et al.*, 2008). Due to the high sequence similarity between the LCRs, misalignments can occur, not only on an interchromosomal but also an inter- and intrachromatid level (Figure II-2), leading to subsequent crossing over and thus rearrangement of the genomic region. The intrachromatid rearrangements can only lead to deletions (Figure II-2, right),

while the others could induce both deletions and duplications (Figure II-2, left and middle). Therefore, at least theoretically, a heightened occurrence of deletions would be expected (Gu *et al.*, 2008; Stankiewicz and Lupski, 2002; Turner *et al.*, 2008). It has been shown that for NAHR to take place specific "hotspots" are required, that is segments of a minimal length inside the LCRs sharing identity or at least extremely high similarity between the aligned LCRs. Small changes in the sequence of these "hotspots" could lead to a reduced recombination frequency. Furthermore, there is evidence that the distance between two LCRs is influencing the efficiency of NAHR (Rubnitz and Subramani, 1984; Sharp *et al.*, 2005; Waldman and Liskay, 1988).

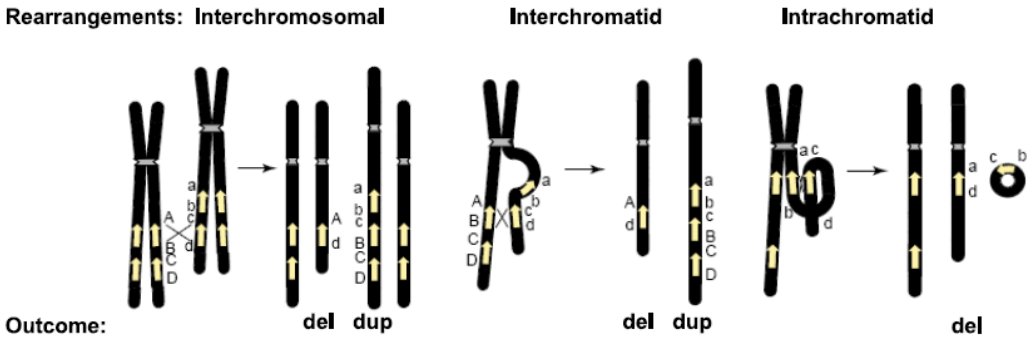

Figure II-2: Mechanism of NAHR (adapted from Gu *et al.*, 2008). Interchromosomal (left) and interchromatid (middle) misalignment could result in the formation of deletions and duplications. The misalignment on an intrachromatid (right) level leads to deletions solely.

The mechanism of NHEJ is known to be one of the two major mechanisms to repair double-strand breaks (Gu *et al.*, 2008; Lieber *et al.*, 2003). NHEJ is mediated by the activity of different enzymes that first detect the broken DNA ends and then modify the ends to make them compatible for the final ligation step (Weterings and van Gent, 2004). Thereby, small deletions or insertions could easily be formed (Figure II-3). If combined with a preceding step at which one of the broken ends invades and copies from the sister chromatid, NHEJ could also be used to explain the synthesis of duplications, thus CNVs (Woodward *et al.*, 2005). Although NHEJ-mediated CNVs often fall within repetitive elements such as long terminal repeats (LTRs), long interspersed nuclear elements (LINEs), mammalian interspersed repeats (MIRs), and short interspersed repeat elements (SINEs), sequences of extended homology are not necessarily required (Gu *et al.*, 2008; Malhotra and Sebat, 2012; Zhang *et al.*, 2009).

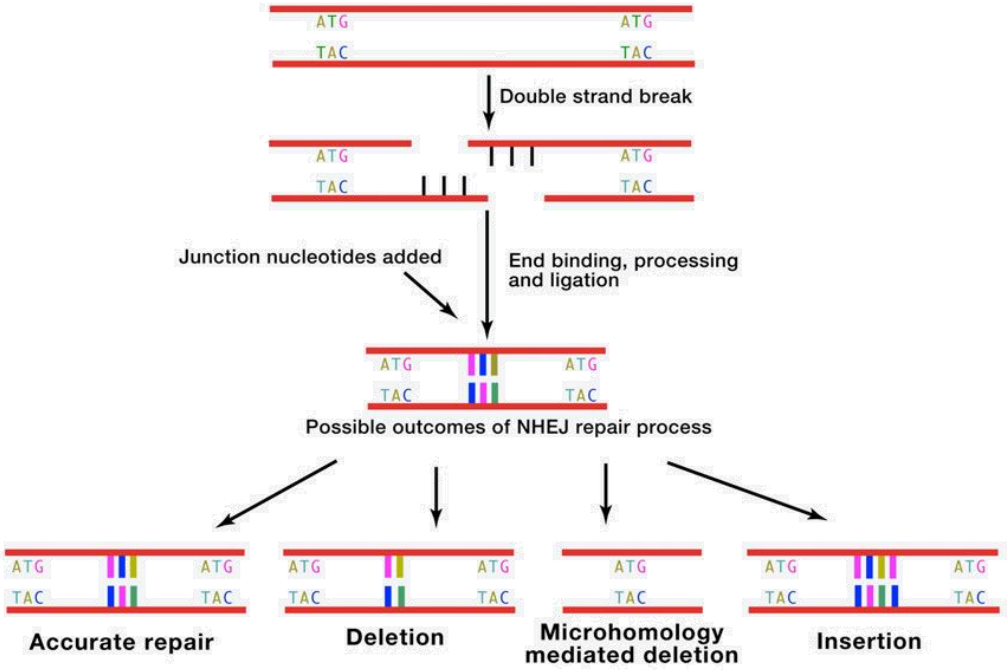

Figure II-3: Mechanism of NHEJ (adapted from Malhotra and Sebat, 2012). The classical mechanism could contribute to the formation of small SVs by errors occurring during process of strand break repair. If one of the broken ends invades and copies from the sister chromatid (not depicted here) duplications, thus CNVs, can be formed as well.

Especially large genomic tandem duplications of several megabases might be explained best by the mechanism of FoSTeS (Lee *et al.*, 2007), which was further generalized to the microhomology-mediated break-induced replication (MMBIR) model (Hastings *et al.*, 2009). According to this model errors during the DNA replication process, provoked by stalling of the DNA replication fork and further annealing and restart of the DNA synthesis at another replication site, lead to the formation of complex genomic rearrangements (Gu *et al.*, 2008; Zhang *et al.*, 2009). A simplified illustration of this mechanism is depicted in Figure II-4.

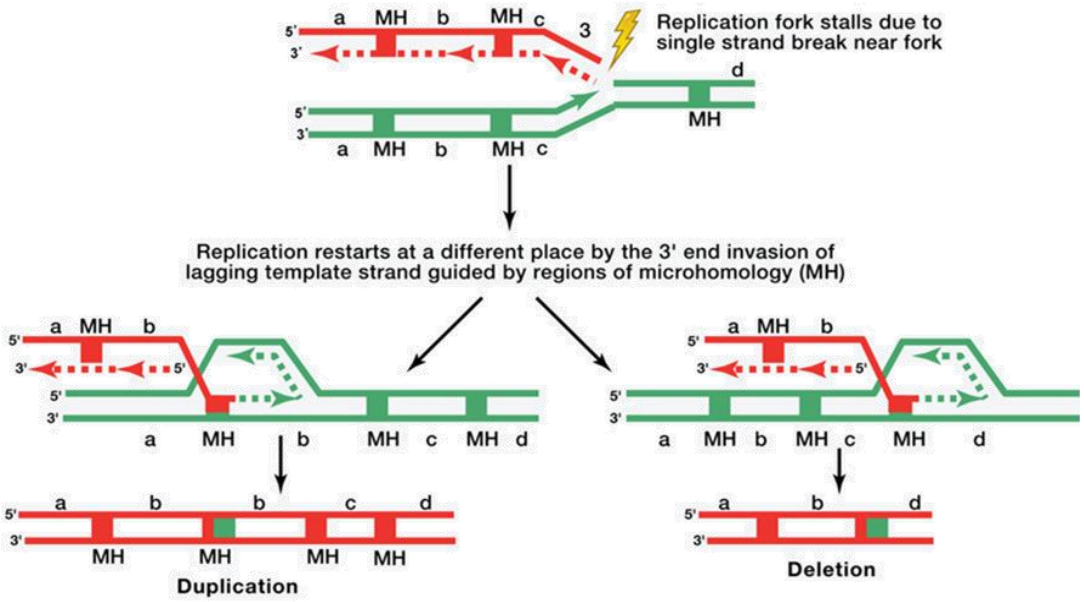

Figure II-4: Mechanism of FoSTeS (adapted from Malhotra and Sebat, 2012). Here a simple model of FoSTeS is depicted resulting in either a duplication or deletion. MH = regions of microhomology.

The extent to which the different mechanisms contribute to the formation of SVs and CNVs is not yet clear however a recent study (Mills *et al.*, 2011) on that topic suggested NAHR to generate most of the large deletions or duplications and FoSTeS/MMBIR to be responsible for the occurrence of most tandem duplications (Malhotra and Sebat, 2012).

6 How CNVs mediate their effect on gene expression

CNVs are supposed to affect phenotypes by the alteration of expression. Genes and putative regulatory elements found in regions of CNVs have been shown to play a significant role in gene expression in mice and humans (Gamazon *et al.*, 2011; Orozco *et al.*, 2009). A weak positive correlation between the CNV status of genes and their relative expression levels has been found (Henrichsen *et al.*, 2009b). There are several explanations why an enhancement of gene dosage does not necessarily result in a significant increase of expression products (Figure II-5A+B) or *vice versa*, as might be expected (Henrichsen *et al.*, 2009a). First, an increase of the gene product could activate a negative feedback loop that prevents further expression of the respective gene (Figure II-5C). Second, if the enhancer of the gene is not included in the copied locus, the

influence of the single enhancer might be too weak to trigger a huge increase in expression (Figure II-5D). Further, if the copies of the gene map to different chromosomes and thus chromatin environments, they might be regulated differentially (Figure II-5E). An initial level of gene products proportional to the copy number might induce the expression of a repressor, which reduces or abolishes further expression of the gene (Figure II-5F). Finally, according to another hypothesis, steric hinderance resulting from the gene duplication impairs the access of transcription factors (Sexton *et al.*, 2007) (Figure II-5G).

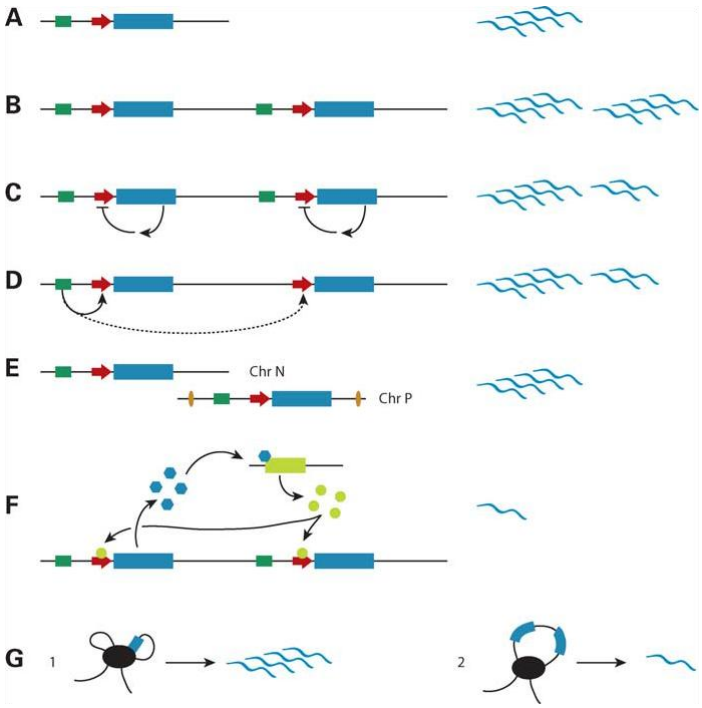

Figure II-5: Gene dosage alterations and their influence on gene expression (adapted from Henrichsen *et al.*, 2009a). (A) Relative amount of transcription products (blue wiggly lines) of a single gene copy (blue box) is mediated by the gene promoter (red arrow) and its enhancer (green arrow). **(B)** A complete tandem duplication of the locus. **(C)** Negative feedback loop activated by complete tandem duplication of the locus. **(D)** Tandem duplication excluding the enhancer region. **(E)** Complete non-tandem duplication on another chromosome, where a different chromatin context, e.g., insulators (yellow ellipses) modifies expression levels. **(F)** Complete tandem duplication results in increase of the gene product (blue hexagons), which induces a repressor (light green box), the product of which blocks the expression of the gene. **(G)** The access of transcription factors is physically impaired as a result of tandem duplication (2).

For genes located in the vicinity of CNVs, alterations of expression have been shown as well (Chaignat *et al.*, 2011; Henrichsen *et al.*, 2009b). Here, the effects might be mediated

by enhancers, repressors or other transcription regulators that are copied and thus increase the transcription of their original target gene or other genes at greater distance (Chaignat *et al.*, 2011; Ricard *et al.*, 2010). Another mechanism conceivable in this context is a physical dissociation of the transcription unit from its *cis*-acting regulators or modifications of the chromatin structure, leading to silencing of a gene outside the copied region (Reymond *et al.*, 2007).

Furthermore, CNVs could act through effects of transvection. These effects enable the communication between homologues chromosomes and were first described in Drosophila (Lewis, 1954; Wu and Morris, 1999) and later also in humans (Liu *et al.*, 2008) and mice (Rassoulzadegan *et al.*, 2002; Sandhu *et al.*, 2009). CNVs might interrupt that communication, which was shown to occur on allelic and nonallelic level (Duncan, 2002; Kennison and Southworth, 2002; Sandhu *et al.*, 2009), by disrupting or fully deleting a coding sequence or a regulatory element.

7 The impact of CNVs on human diseases and anxiety phenotypes

The authors of a study published some years ago concluded that common CNVs are unlikely to have a major role in the genetic basis of human diseases (Wellcome Trust Case Control Consortium, 2010). This conclusion was drawn from a genome wide association study (GWAS). However, the use of GWAS in the context of complex diseases and CNVs has been criticized in a more recent review (Lee *et al.*, 2012). In contrast, current findings indicate that it would be premature to neglect CNVs as contributors to complex phenotypes (Gamazon *et al.*, 2011) and, indeed, many diseases have been linked to CNVs based on GWAS but also other studies (Girirajan *et al.*, 2011; Klopocki and Mundlos, 2011; Malhotra and Sebat, 2012).

Besides common disorders like asthma (Brasch-Andersen *et al.*, 2004), metabolic diseases like type 2 diabetes and obesity (Bochukova *et al.*, 2010; Jeon *et al.*, 2010), cancer (Shlien and Malkin, 2009), and other diseases, CNVs have been reported to affect the disease susceptibility of neurological disorders like Parkinson's disease (Singleton *et al.*, 2003), Alzheimer's disease (Rovelet-Lecrux *et al.*, 2006), Autism (Levy *et al.*, 2011), Schizophrenia (Vrijenhoek *et al.*, 2008), bipolar disorders (Lachman *et al.*, 2007; Malhotra *et al.*, 2011)

and anxiety disorders. Not many of the studies published to date focused on the correlation between anxiety (pathological or non-pathological) and CNVs. In its pathological form, anxiety was shown to be associated with CNVs in a study of panic disorder in a Japanese population (Kawamura *et al.*, 2011). Non-pathological anxiety with respect to CNV-related changes in behavior has been addressed by animal studies. As an example, an increased maternal separation-induced anxiety was shown in mouse pups harboring an extra copy of a specific (Gtf2i) gene (Mervis *et al.*, 2012). Furthermore, a positive relationship between the duplication of the genomic region harboring the glyoxalase1 (*Glo1*) locus and the *Glo1* gene expression was reported in the HAB/LAB and other mouse lines (Hambsch *et al.*, 2010; Williams *et al.*, 2009). The *Glo1* gene has been shown to be related to anxiety, however published studies disagree about whether *Glo1* expression is negatively or positively correlated with anxiety-related behavior (Ditzen *et al.*, 2006; Hambsch *et al.*, 2010; Hovatta *et al.*, 2005; Krömer *et al.*, 2005; Landgraf *et al.*, 2007; Thornalley, 2006).

In conclusion, the general impact of CNVs (as well as SVs) on anxiety and other phenotypes is not yet clear. However, there is evidence from many distinct studies that CNVs are indeed involved in phenotypic expressions and, considering the mechanisms by which they might act on gene expression (see chapter 6) and their large extend over the genome, their contribution might be important (Feuk *et al.*, 2006). A recent review claimed that the conduction of several large-scale correlation studies should be performed, not only with single CNVs but also a combination of other genetic and environmental factors, before a general conclusion on the implication of CNVs on health could be drawn (Almal and Padh, 2012).

8 Aim of the study

The main aim of this study was to generate the first comprehensive catalogue of CNVs in the HAB/LAB mouse model and thereby lay the foundation for future investigations on the complex interactions of multiple genetic factors (e.g., SNPs, epigenetic factors or other SVs) influencing anxiety-related behavior.

Although there is evidence from previous studies that CNVs contribute to the phenotypic expression of complex behavior, the impact of the variants on anxiety-related behavior had to be proven for the HAB/LAB mouse model, which was a second goal of this thesis.

Finally, if an effect of CNVs on the HAB/LAB phenotype could be shown, then a first insight into CNV-based candidate genes potentially influencing the phenotypic characteristics of anxiety, not only in HAB/LAB but also CD-1 mice, should be delivered.

III Materials and Methods

1 Equipment, chemicals and remarks

Unless otherwise specified, all chemicals mentioned were purchased at Sigma-Aldrich (Taufkirchen, Germany).

The agarose gels for gel electrophoreses were prepared using TBE buffer and UltraPure agarose (Invitrogen, LifeTechnologies, Darmstadt, Germany) in the respectively required concentration, mixed with 0.5 µl/ml ethidiumbromide (Roth, Karlsruhe, Germany). The TBE buffer was prepared at 5-fold concentration with 54 g Tris (Roth), 27.5 g boric acid and 20 ml of 0.5 M EDTA (pH 8.0) in a total volume of one liter distilled and autoclaved water. After the electrophoresis performance, the signals were visualized under UV light using the GelDoc 2000 UV-Transilluminator and the Quantity One software (BioRad, Munich, Germany).

Volumes of up to 1 ml have been transferred with pipettes model Pipetman Classic™ P2, P20, P100, P200 and P1000, respectively (Gilson, Middleton, WI, USA). Larger volumes were handled with serological pipettes. Those, as well as the normal and filtered pipette tips, were supplied by Sarstedt (Nürmbrecht, Germany).

All centrifugation steps were done in a Hermle Z216 MK centrifuge (Hermle Labortechnik, Wehingen, Germany), shaking and heating steps in the PHMT SC-20 Thermo-shaker (Grant Instruments, Cambridge, UK).

DNA and RNA concentrations were measured on an Implen NanoPhotometer (Implen, Munich, Germany). Depending on the amount of DNA/RNA in the respective sample, lidfactor 10 or 50 was chosen. The ratio of absorption at 260 nm and at 280 nm, which was calculated by the photometer automatically, served as a parameter for quality control. Values of around 1.8 for DNA and around 2.0 for RNA were considered to show satisfactory quality.

A thermocycler of type Primus 96 advanced (PeqLab, Erlangen, Germany) was used for the reverse transcription of RNA into cDNA.

All information on the genomic position of genes and probes mentioned in this thesis refer to UCSC genome browser version mm9 (Kent *et al.*, 2002) and NCBI build m37, respectively.

Here, non-human proteins' abbreviations were written lower-case with a capital as first letter. The same applies to non-human genes except that they were italicized. Human proteins were written in upper case as well as human proteins, while the latter were additionally italicized.

If not indicated otherwise, all computational analyses were done using R software (R Development Core Team, 2010).

2 Housing and testing conditions of animals

All animals used for data generation of this thesis were housed in the animal facility of the Max Planck Institute of Psychiatry under standard conditions, thus, a temperature of 23 ± 2 °C, a relative air humidity of 60 ± 5 % and a 12/12-hour light-dark cycle with beginning of the light phase at 8 a.m. Animals of the same sex were kept in groups of up to four animals per type II standard cage with nesting and bedding material, having access to food pellets (Altromin GmbH, Lage) and tap water *ad libitum*.

All behavioral experiments were carried out with the sanction of the local authorities in accordance with the German law and the Council of the European Communities' "directive of 24 november 1986 on the approximation of laws, regulations and administrative provisions of the Member States regarding the protection of animals used for experimental and other scientific purposes (86/609/EEC)". All tests were conducted between 9 a.m. and 1 p.m. under standard housing conditions regarding temperature and humidity.

2.1 Elevated plus-maze (EPM) test

The elevated plus-maze, which consists of an elevated plus-shaped platform with two open and two closed arms, was developed as a test for anxiety-related behavior in rats

(Pellow *et al.*, 1985). Nowadays it is used as a commonly accepted standard trial for rats and mice, since animals were shown to spend more time in the aversive environment of the open arm after treatment with clinical effective anxiolytics (Hogg, 1996; Lister, 1987; Pellow *et al.*, 1985). The EPM used here (Figure III-1) was made of polyvinylchloride (PVC), elevated by four 40 cm-long legs and consisted of two open arms (30 x 5 cm), two closed arms (30 x 5 x 15 cm), and an inner zone (5 x 5 cm) in between. To reinforce the protective nature of the closed arms they were with 10 lx less brightly illuminated than the open arms with 300 lx.

In each trial one mouse was placed into the inner zone facing one of the closed arms. The animal's behavior was recorded for 5 min and analyzed afterwards using the tracking software Any-maze version 4.72 (Stoelting, West Lane, IL, USA). Before the next trial started, the apparatus was cleaned with 70% ethanol and soap water.

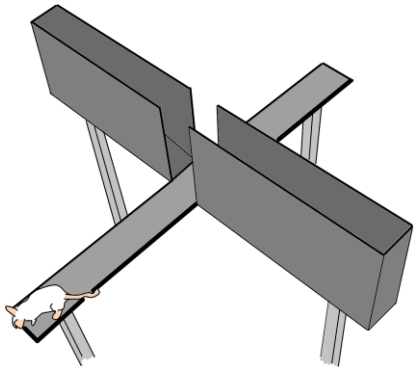

Figure III-1: Schematic drawing of EPM test. A mouse is sitting on one of the open arms. Closed arms are surrounded by walls. Illumination is brighter on the open arms.

2.2 Open field (OF) test

The open field test could not only be used to test for anxiety-related behavior, but also as a test for locomotion and exploratory behavior (Prut and Belzung, 2003; Walsh and Cummins, 1976). Independent of the approach, there are always two parts to be discriminated, an inner zone with brighter illumination and a darker outer zone. The brighter the illumination the more aversive appears the zone to the animals. Here, the OF was applied to test rather for locomotion than anxiety-related behavior, hence the light

intensity was adjusted to differ only about 15 lx between the central part and the periphery of the OF. The apparatus itself consisted of a round PVC wall of 40 cm height framing a field of 60 cm in diameter (Figure III-2).

Similar to the EPM test procedure, the animals were placed in the inner zone and tested for 5 min. The tracking and analysis was done using Any-maze software. At the end of each trial, the apparatus was washed with 70% ethanol and soap water.

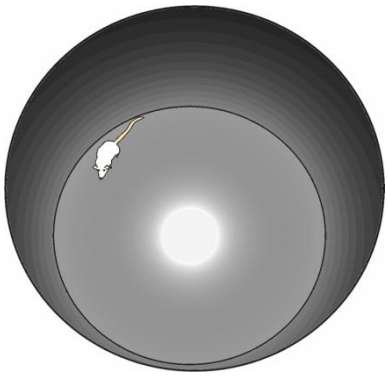

Figure III-2: Bird's eye view of the open field. A mouse is placed into a field, surrounded by cylindrical wall. The middle of field is brighter illuminated (in current study around 15 lx brighter) than the outside. Here the field diameter was around 60 cm, the height of walls about 40 cm.

2.3 Forced swim test (FST)

Originally, the FST was designed to test animals for depression-like behavior (Porsolt *et al.*, 1977, 1978). Even if the FST is still used as the most common test for screening antidepressants, considering the discussion about the assumption of psychological states that probably not exist in mice, it might be more precise to talk about an alternative active and passive reactivity to stress instead of depression-like behavior (El Yacoubi and Vaugeois, 2007; Holmes, 2003).

The animal's behavior was recorded with a video camera for 5 min. At the beginning of the test, the mouse was placed into a 2 l glass beaker filled with 1.75 l of 23 °C warm tap water (Figure III-3). Before putting it back into its home cage the mouse was dried with a towel. The data were analyzed later on using a custom Eventlog program, differentiating between four kinds of behavior: freezing, floating, struggling and swimming. The latter

was defined as movements of all four paws underneath the water surface, whereas struggling meant extensive paddling and breaking through the surface with at least the two fore limbs. The animal was evaluated as floating if it did not show any movements except very slight balancing movements. In contrast, when no movements at all could be observed and the mouse showed some kind of rigor, the behavior was referred to as freezing.

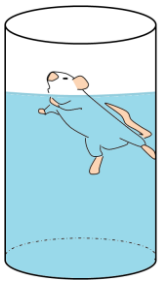

Figure III-3: Drawing of FST. A mouse is placed into a 2 l glass beaker filled with warm tap water (blue).

2.4 Stress reactivity test (SRT)

The stress reactivity of mice was tested via the increase of corticosterone (Cort) in the blood after suffering restraint stress (Touma *et al.*, 2008).

For that purpose an initial blood sample was collected from the tail vessel directly after taking the animal out of its home cage, using Micro haematocrit tubes that then were sealed with a Haematocrit sealing compound (both from Brand, Wertheim, Germany). Afterwards, the mouse was placed into a 50 ml Falcon tube (Sarstedt, Nürmbrecht, Germany) for 15 min. The ventilation was assured by a hole at the tip of the tube (Figure III-4). Before the animal could recover in its home cage a reaction blood sample was collected as described above. The approximately 50 µl blood of each sample were centrifuged for 5 min at 14,800 g to separate cellular components and plasma. The latter was collected and stored at -20°C until the Cort concentration was measured by radioimmunoassay. Finally, the Cort increase was defined by comparing Cort concentrations of the initial and reaction blood sample.

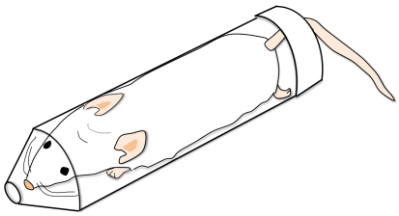

Figure III-4: Mouse in SRT. The falcon tube has a hole in the tip to ensure air flow and another one to lead the tail to the outside of the tube.

2.5 Tail suspension test (TST)

In the TST the mouse was suspended from a metal frame by fixing its tail tip with a tape. Typically in TST, mice immediately behaved in escape-oriented manner, followed by periods of immobility and movement. Hence the immobility time could be reduced by a variety of antidepressants the TST is commonly used to test for depression-like behavior (Cryan and Holmes, 2005; Steru *et al.*, 1987; Steru *et al.*, 1985). Again, referring to the description of the FST in chapter 2.3, one might rather talk about an alternative coping style in adapting to stress.

Here, an apparatus was used enabling to test four animals simultaneously (Figure III-5). After fixing all four mice, they were recorded with a video camera for 6 min. The video tapes were analyzed using a custom Eventlog program. Two different kinds of behavior were distinguished: mobility, when the animal showed movements with the head, the limbs or the head and immobility, when the animal was not moving at all.

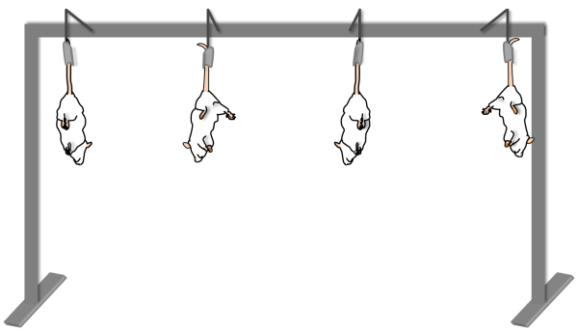

Figure III-5: Experimental set-up of TST. Four mice in a row are fixed on metal hooks with glue tape. There are no dividing walls between the mice.

3 Analysis of CNVs in the HAB/LAB mouse model

3.1 Detection of CNVs by genome-wide oligonucleotide array-based comparative genomic hybridization (aCGH)

3.1.1 DNA extraction

Genomic DNA was extracted from tail tips of two 16-week-old HAB and LAB mouse males, respectively (both generation 35), as well as from brain tissue of only one pair of these mice, using the NucleoSpin® Tissue kit (Macherey Nagel, Düren, Germany). Following the manufacturer's instruction, about 0.5 cm tail or 25 mg brain tissue were incubated and gently shaken over night at 56 °C in 180 µl buffer T1 and 10 µl proteinase K. After addition of 200 µl lysis buffer B3 and an incubation time of 10 min at 70 °C DNA was precipitated by adding 210 µl 99% pure ethanol. The sample was transferred to a spin column provided in the kit. During a 1-minute-centrifugation step at 11,000 g the DNA was bound to the silica membrane, where it was washed with 500 µl buffer BW first and 600 µl buffer B5 thereafter, each time followed by the centrifugation step described above. The DNA was dried by repeating the centrifugation twice, discarding the flow-through in between. Finally, DNA was eluted in 100 µl elution buffer BE, pre-warmed at 70 °C.

The DNA concentration was measured on an Implen NanoPhotometer and quality control was performed by gel electrophoresis. Therefore, diluted DNA (150 ng in 12 µl) combined with an appropriate amount of 6x DNA loading dye (Fermentas, St. Leon-Rot, Germany) was loaded onto a 1% agarose gel (see chapter 1 for details). The ZipRuler^{TM} Express DNA ladder 2 (Fermentas) was added as DNA standard ladder. The gel ran for 50 min at 90 V.

3.1.2 Processing of the samples in aCGH

The extracted DNA of the samples was screened for CNVs on the Mouse CGH 3x720 K Whole-Genome Tiling Array. Therefore, the DNA was sent to Roche NimbleGen (Madison, WI, USA) to access NimbleGen's full CGH microarray service. Besides the manufacture of the array, the service involved the complete process of hybridization and data analyses. Briefly described, the DNA was labeled with fluorescent dyes Cy3 and Cy5, respectively

and co-hybridized to the array that includes about 720,000 probes of length 50-75mer (based on NCBI build m37, UCSC genome browser mm9), resulting in a median probe spacing of 3,537 bp. A microarray scanner was used to scan the fluorescent signal intensities, which were then extracted from the scanned images using NimbleGen NimbleScan software. After correction and normalization steps, the segMNT algorithm was applied to the data to define genomic segments that differ in signal intensity between HAB and LAB samples. Both the final results and the Roche NimbleGen SignalMap software to review those were provided on a DVD.

3.1.3 Interpretation and further analysis of aCGH data

In the provided analyzed data, differences in probe signal with higher signal intensity in HAB were indicated by a positive $\log_2$ ratio, those with higher signal intensity in LAB by a negative one. In a first step, plots showing the $\log_2$ ratio for all probes were reviewed visually. As a consequence, the background noise was estimated to be in the range of -0.5 to +0.5 (values of single probes). In a second step, a visual assessment of the mean $\log_2$ ratios of segments determined by the segMNT algorithm, defined only segments with a $\log_2$ ratio < -0.09 and > +0.09 (values of segments) to be potential CNVs.

Since only CNVs consistent in different HAB/LAB individuals and tissues were of interest for further studies, the data of the three arrays had to be combined. Therefore, files of all three arrays that list the $\log_2$ ratio of all probes and the corresponding genomic position were combined using Microsoft Excel. Before the resulting tables were merged with data of files that list the predicted segments, all probes with a $\log_2$ ratio between -0.5 and +0.5 (values of single probes) were discarded. Next, only those probes were considered that show a mean $\log_2$ ratio < -0.09 or > +0.09 (values of segments) for the corresponding segment in all three arrays, with no contradictions in algebraic sign allowed. In a final step, breakpoints of CNVs were defined after careful consideration of segment data in the remaining probe table.

3.2 Detection of CNVs by the Jax Mouse Diversity Genotyping Array (Jax MDGA)

3.2.1 Performance of the assay

The Jax Mouse Diversity Genotyping Array, applied here to screen for CNVs, was designed by The Jackson Laboratory (Bar Harbor, ME, USA). It is a high-density array containing not only 623,124 probes to detect SNPs but also 916,269 invariant genomic probes (IGPs), capturing 93.4% of the exons defined in Ensembl (version 49) by three probes each, tiling 238 of 481 ultraconserved elements (UCEs) identified between human and mouse, and covering 950 100-kbp intervals with fewer than 10 of the previously described probes, most of which were known to contain segmental duplications (Yang *et al.*, 2009).

Tail tips of one HAB and one LAB mouse male were sent to The Jackson Laboratory for accessing the basic service including the extraction of high-molecular weight DNA, the sample preparation, the array hybridizations, and the supply of raw data. Furthermore, the manufacturer provided the R package “MouseDivGeno” (Yang, 2010) for subsequent analysis, together with its description, the MouseDivGeno Vignette (version 1.0.0).

3.2.2 Genotyping using the MouseDivGeno R package

A function called “simpleCNV”, which was comprised in the “MouseDivGeno” R package (Yang, 2010), was applied to reveal CNVs of the HAB/LAB samples in the Jax MDGA data, following the instructions of the MouseDivGeno Vignette. This function integrates normalized intensities from SNP probes and exons (IGPs) and uses the HiddenMarkov “HMM” R package to infer the most likely CNV states. The function always compares a reference with one sample. That is why both had to be defined, with the LAB sample serving as reference. In the output of the function all segments were listed that showed significantly more intensity, thus a copy number gain, or significantly less intensity, thus a copy number loss, in HAB *vs.* LAB animal.

3.3 Detection of CNVs and structural variants (SVs) by next-generation sequencing (NGS)

3.3.1 DNA extraction

First of all, six HAB and six LAB males, 29 weeks of age each, were carefully chosen from generation 40 and 41. As first selection criterion, animals had to show distinct line-specific anxiety behavior on the EPM. Besides, they should belong to different generations and be the less related as possible, as the samples were supposed to be pooled for the sequencing process. Thus, CNVs that were specific to single individuals and those that remained not stable over generations could be recognized in the analysis of sequencing data and excluded from the final results.

Here, the DNeasy blood & tissue kit (Qiagen, Düsseldorf, Germany) was used for DNA extraction. Directly after taking, the tail tips were stored at -80 °C. About one month later they were thawed to extract the DNA, following the manufacturer's instructions. Hence, 25 mg of tissue were gently shaken while incubated in 180 µl buffer ATL and 20 µl proteinase K for nearly 3 hours at 56 °C. After centrifugation for 3 min at 6,000 g, the supernatant was transferred to a new tube, mixed with 4 µl RNAse A and incubated another 2 min at room temperature. Next, 400 µl of pre-mixed 1:1-solution of buffer AL and 100% pure ethanol were added and all was loaded on the provided column. During a 1-minute-centrifugation step at 6,000 g, the precipitated DNA was bound to the membrane of the column, where it could be washed with 500 µl buffer AW1 first, and 500 µl buffer AW2 thereafter. In order to dry the DNA, the column was centrifuged at 20,000 g for 3 min. As a last step, DNA was eluted in 200 µl buffer AE, centrifuged a last time at 6,000 g for 1 min.

Finally, the concentration of DNA was measured on a NanoPhotometer and the quality of DNA was checked by loading 150 ng onto a 0.7%-agarose gel that run for 90 min at 80 V. Using ZipRuler Express DNA Ladder 2 as a standard marker, the DNA of all samples could be determined to be of high-molecular weight, with a length of about 20,000 bp. Until its use, the DNA was stored at 4 °C.

3.3.2 Performance of sequencing process

The next-generation sequencing (NGS) was done in the facilities of Max Planck Institute of Psychiatry in Munich. The data were generated in two runs on the SOLiD™ 4 System (Applied Biosystems, Foster City, CA, USA).

As previously mentioned, the DNA of six HAB and LAB males each was pooled after extraction. In advance, a quality check was performed by measuring the integrity, purity and concentration of all 12 samples. For the final pools 1 µg DNA of each animal was used to end up with a total amount of 6 µg HAB DNA and LAB DNA, respectively.

With a 2 x 60 bp long mate-paired library (LMP library) the appropriate library type was chosen to detect CNVs and SVs. For its preparation, the DNA was sheared using the Covaris S2 System (Covaris, Woburn, MA, USA) to generate fragments of length 2,000 bp. All further steps followed the Applied Biosystems' Mate-Paired Library Preparation guide (part # 4460958 Rev. A, revision date march 2011). Briefly, the DNA fragments were end-repaired using End Polishing E1 and E2 enzymes before MP adaptors were ligated to, using T4 DNA ligase. The adaptors were missing a 5' phosphate at the non-joining end, resulting in a nick on each strand when the DNA was circularized after purification with the Agencourt AMPure® XP Reagent. The circularization occurred through intramolecular hybridization at low concentrations, after removal of the blocking oligonucleotides that protected the 3' overhangs of the MP Adaptors from self annealing, via heat denaturation. Before the DNA was purified using the Agencourt AMPure® XP Reagent again, uncircularized DNA was eliminated by Plasmid-Safe™ DNAse-treatment. Next, the nick was translated into the genomic DNA region using *E.coli* DNA polymerase I and the DNA was purified using the SOLiD™ Library Micro Column Purification Kit. Directly afterwards, 60 bp mate-paired tags were cleaved from the circularized template using first T7 exonuclease, that recognized the nicks and digested the unligated strand away from the tags to create a gap in the sequence that could then be recognized more easily by the S1 nuclease. Another Agencourt AMPure® XP Reagent-based purification step was followed by the adding of a dA tail to the S1-nuclease-treated DNA by A-Tailing Enzyme II that should increase the efficiency of the proximate ligation to P1-T and P2-T Adaptors using T4 DNA ligase. Before and after the ligation, the library was purified from side

products by binding streptaviding beads (Dynabeads® MyOneTM Streptavidin C1) to the biotin-labeled MP adaptors in the library molecules, and subsequent washing of the bead-DNA complex. After a trial amplification to determine the optimal cycling number, resulting in 12 cycles for the HAB sample and 15 cycles for the LAB sample, the library was nick-translated and amplified using Library PCR Primers 1 and 2 with the Platinum® PCR Amplification Mix. The LMP library preparation was finished with the purification of DNA with the SOLiD™ Library Micro Column Purification Kit. The quality of library was checked on the Agilent 2100 Bioanalyzer (Agilent Technologies, Böblingen, Germany) followed by purification with Ampure beads to remove primer peaks and another quality check on the bioanalyzer. The quantification was done by qPCR using the SOLiD Library TaqMan Quantification kit (Cat. # A12127) following the manufacturer's protocol.

Before the LMP libraries could be sequenced they required to be clonally amplified on SOLiD™ P1 DNA Beads by emulsion PCR (ePCR). The bead preparation was conducted according to the manufacturer's protocol. For both LMP libraries a 2 x E80 bead preparation scale was chosen, using an input of 1 pM library template each. The ePCR was followed by enrichment of the templated beads, which then were loaded onto slides. Each E80 preparation was loaded on one full slide. The sequencing was carried out in two runs with one HAB and one LAB library at a time.

3.3.3 Data handling to detect CNVs and structural variants (SVs)

Here, the following part is written in a straightforward manner. A more detailed description of data handling can be found in the supplementary data (report S1).

3.3.3.1 Quality control and alignment

The first step after data generation included the quality control (QC) of single reads. A mean quality value (QV) above 10.0 was considered to indicate sufficient quality. For the further analysis it should be kept in mind that a mate-paired library was used, as described above. Hence, every single read (60 bp in length) could be associated with its mate, another single read, which is known to appear at a distance of 2,000 bp in the

original genome. If only one mate of the pair passed the QC, both reads were not considered for the paired-end analysis of CNVs and SVs.

In a second step all reads were aligned against a reference genome (based on UCSC genome browser mm9). Therefore, two different alignment programs, bowtie (version 0.12.7) and BWA aligner (version 0.5.7), were applied according to the following scheme. First, only read pairs that survived QC were aligned with bowtie. Second, those pairs that could not be aligned with bowtie were aligned with BWA. Finally, BWA was also applied on single reads that passed QC but whose mate did not. All resulted alignments (single reads) were post-processed in order to remove those reads that could be mapped at more than just one single position in the genome.

3.3.3.2 CNV detection

In order to detect CNVs, a "depth of coverage" analysis (DOC analysis), comparing the coverage between the HAB and LAB mouse lines, was performed using a custom Python program. In detail, that means that the whole reference genome was split into bins of 200 bp length and all reads mapping into a single bin were counted (Figure III-6A). Theoretically, the paired end sequencing is not required to reveal CNVs. However, the additional pairing information encouraged the coverage estimation by counting also the bins in between the bins where the mates mapped and thus, more accurate results could be achieved (Figure III-6B). If two mates appeared to have an insert size much larger than 2,000 bp (the z-score cutoff was calculated based on the first 200,000 reads) only the two bins of direct mapping were counted (Figure III-6A).

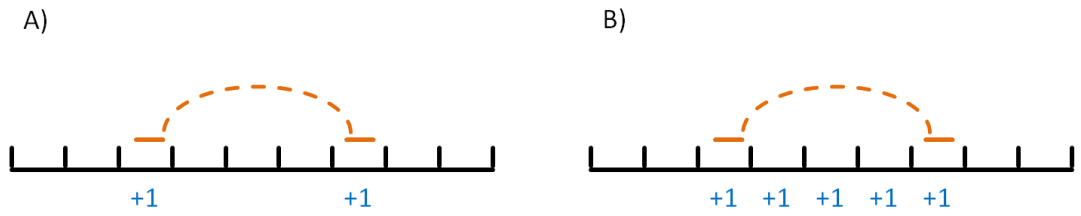

Figure III-6: Principle of bin counts. Reference genome (black horizontal lines) split into bins of 200 bp length (black vertical lines). Two single reads of a mate-pair (orange lines), appearing at a distance of 2,000 bp in the sample genome (indicated by orange dashed line), were mapped against the genome. (A) Bin count of bins where reads could be aligned was increased (blue). (B) Additional increase in bin count of bins in between led to encouraged coverage.

The subsequent statistical analysis was done bin- and chromosome-wise using Fisher's Exact test, with a 2x2 contingency table per bin, containing the count of reads in the bin and the count of reads in the remaining bins of the chromosome, each for the HAB and the LAB sample. The resulting p-value was converted into a q-value using the Bonferroni-Hochberg method. Since the test is a one-sided test, it had to be conducted twice to calculate q-values for the enrichment in HAB samples over LAB samples and vice versa. In a final step CNVs could be defined by finding "unusual" accumulations of specific q-values. That was done by requiring all –log10 of q-values within a window of 8 bins to exceed a initial finding threshold of 12. All bins right and left to those bins were still considered to be part of the CNV, as long as they exceeded an extension threshold of 10. The window size, initial finding threshold and extension threshold were defined based on carefully conducted pretests.

3.3.3.3 SV detection

With respect to the statistical model, the analyses to detect SVs did not differ to the one to define CNVs. Thus, the calculation of p-values from counts of aligned reads in 200 bp bins, the transformation of p-values into q-values, and the definition of accumulation differences between HAB and LAB samples was mainly done as described in chapter 3.3.3.2. Regarding the counting of the aligned reads in the bins, however, the SV analyses differ from each other and from the previously described one.

For detecting large deletions within one mouse line, all "regularly" aligned mate-pairs were counted, as well as those that showed a significantly increased insert size, which

might indicate a deletion. Q-values were calculated as described above, except that the window size was increased to 12, the initial finding threshold to 50 and the extension threshold was decreased to 3, based on the fact that most of the bins would not comprise any mate-pairs with abnormal length of insert size.

Based on the fact that the actual insert size was about 2,000 bp with a standard deviation of 850, in order to be significantly decreased, the aligned insert size was calculated to account for a maximum size of -400 bp, which is evidently not feasible. Thus, the analysis used for detection of large deletions couldn't be applied for detecting large insertions. Therefore, followed up on the idea that the amount of aligned single reads, derived from mate-pairs with only one mate mapping to the reference genome (Figure III-7), should show a distinct elevation up- and downstream of the insertion, this type of SVs could be calculated assuming a window size of 7 bins, an initial finding threshold of 12 and an extension threshold of 3.

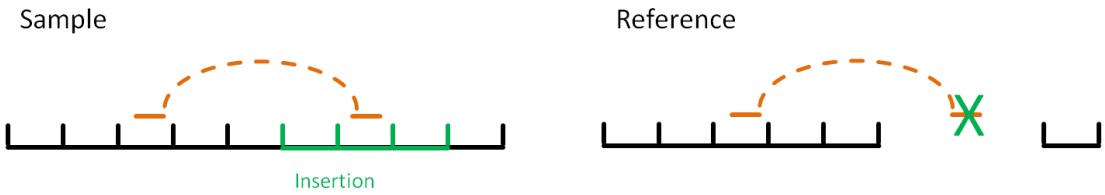

Figure III-7: Principle of SV detection. Genome (black horizontal lines) split into bins of 200 bp (black vertical lines). If sample genome (left) implied insertion (bins colored in green) compared to reference genome (left), mate-pairs with one mate not mapping (indicated by green X) should be increased near insertion breakpoint.

The analysis of inversions started with counting mates that mapped in the wrong direction. Theoretically, the single reads R3 and F3 of a mate-pair should both align to the reference genome on the forward or the reverse strand, respectively. Thus, if mates were found to map on different strands, that could be an indication for a present genomic inversion. The statistical analysis was conducted using a window size of 8 bins, an initial finding threshold of 40 and an extension threshold of 3.

Finally, the wrong orientation of R3 and F3 reads was examined, which may be caused by complex SVs. To detect these genomic regions, those mate-pairs were counted, whose

R3 read could not be aligned upstream to the F3 read as expected, but the other way round. Based on these counts and a window size of 8 bins, an initial finding threshold of 30 and an extension threshold of 3 complex SVs could be detected.

3.4 Influence of CNVs on gene expression

3.4.1 Detection of differentially expressed genes in HAB/LAB mice by gene expression microarray analysis

The raw data of genes differentially expressed between HAB and LAB mice in the basolateral (BLA) and central amygdala (CeA), the hypothalamic paraventricular nucleus (PVN), and the anterior part of cingulate cortex (Cg), respectively, were acquired by a former study on expression profile. In that study the RNA of eight HAB and eight LAB animals (generation 25) was tested separately on the MouseWG-6 v1.1 Expression BeadChip (Illumina, San Diego, CA, USA), which enabled to identify 46,000 individual gene transcripts (Czibere, 2008). Due to the fact that the former data analysis refers to information on probe position based on UCSC genome browser mm8, but the CNV data here refer to version mm9, and considering that in recent years new and better statistical approaches were developed to handle array data, a reappraisal of raw data was conducted.

A detailed description of evaluating the differential expression data can be found in the supplementary data (report S2). In summary, raw data analysis was done using different functions in R, such as “vsn” (Huber *et al.*, 2002) for data normalization. All data of probes found in all samples (46,657) were further filtered by three criteria. First, probes had to map to genes with EntrezGene-ID. Second, they had to be aligned uniquely to the genome with only two mismatches allowed. Third, only those genes with a detection p-value of less than $1x10^{-4}$ were selected. Based on these filtering conditions, the measured values of 12,171 probes were used for subsequent differential expression analysis, where the functions of the “Limma” R package (Smyth, 2005) were applied. In order to stabilize the statistical analysis (Smyth *et al.*, 2005), significantly regulated genes were ranked using an empirical Bayes method (Smyth, 2004). Finally, statistical significant

expressed genes were determined based on a contrast analysis (for more detail see supplementary report S2). As final output of this analysis a list was created, including all 12,171 probes with their respective p-value. Here, a p-value less than 0.05 indicated significant expression differences.

3.4.2 Confirmation of candidate genes by quantitative real-time PCR (qPCR)

3.4.2.1 Selection of candidate genes and primer design

The candidate genes to be analyzed by qPCR were chosen based on the results of the expression microarray reappraisal (see chapter 3.4.1). The first selection criterion was a different expression between HAB and LAB mice in at least one of the four investigated brain regions. The eligible candidates were further analyzed for functional protein association networks formed by these candidates, using the STRING online software version 9.0 (see supplementary report S3). Finally, 12 genes were accepted that fulfilled both selection criteria. Then primers for qPCR were designed using the online tool Primer-Blast provided at the NCBI homepage. A list of these primers is shown in Table III-1.

Table III-1: List of oligonucleotides used as primers for qPCR. The table is sorted by chromosome. Columns show (left to right): chromosome, gene represented by primer, primer orientation, primer sequence (direction 5' to 3'), melting temperature and size of the resulting PCR product.

Chr.	Gene symbol	Orientation	Primer sequence 5'-->3'	T_m [°C]	Prod. size [bp]
1	Rgs16	forward	GGC TCA CCA CAT CTT TGA CG	59.2	110bp
		reverse	*TGG TAG TGG CAG CTT GTA GG*	*59.39*	
2	Ghrh	forward	CTC ATC CTC ACC AGT GGC TC	59.82	115bp
		reverse	*ACA GCT GGC TCA GGA GTT TC*	*59.96*	
7	Fgfr2	forward	GGC AGT AAA TAC GGG CCT GA	59.82	84bp
		reverse	*CAG CAC TTC TGC ATT GGA GC*	*60.46*	
7	Prkcdbp	forward	CTG CAC GTC CTG CTC TTC A	60.01	85bp
		reverse	*GGT CCT CCG GAC CCA AGA*	*60.28*	

Chr.	Gene symbol	Orientation	Primer sequence 5'-->3'	T_m [°C]	Prod. size [bp]
9	Glb1	forward	GTT CTC CGG TCT TCT GAC CC	59.75	85bp
		reverse	*AGA GCA GGG GCT TCA TCT TG*	*59.74*	
11	Epn2	forward	CTA GCC TCC CAC CCT AAT GG	58.94	130bp
		reverse	*GCT CCT CTT CTC CGC TTG TC*	*60.46*	
11	Pdk2	forward	GAG ATG ACC CCG TCT CCA AC	59.82	120bp
		reverse	*GGT TGG TGC TGC CAT CAA AG*	*60.04*	
12	Rhoj	forward	ACG CCT TCC CAG AGG AAT AC	59.17	81bp
		reverse	*GCA AGT GCT GCT TGC CTC*	*59.74*	
17	Alk	forward	GGC AAG CCT GTG ATT TCC AC	59.76	128bp
		reverse	*GAG TGG ACT TTG GGT CCA GC*	*60.61*	
17	Glo1	forward	GGA TTG CCG TTC CTG ATG TC	60.04	123 bp
		reverse	*AGC CGT CAG GGT CTT GAA TG*	*58.98*	
17	Slc30a6	forward	TGG ACC CTT GGA TTT GGC TC	59.96	153bp
		reverse	*CCG AAT CCA GTC GTC CTT GA*	*59.47*	
19	Gnaq	forward	CAG GAG TGC TAC GAC AGA CG	60.18	82bp
		reverse	*CGG CTA CAC GGT CCA AGT C*	*60.45*	

3.4.2.2 RNA extraction

Four different brain regions (BLA, CeA, PVN, Cg) were of interest for the qPCR analysis. The RNA was extracted of 11 HAB and 8 LAB males (generation 46), aged between 7 and 10 weeks. The animals were killed by decapitation after being anesthetized with isoflurane. The brains were collected and directly frozen on dry ice, then stored at -20 °C until they were cut (rostral to caudal) into 200 µm thick slices by use of a cryostat (Microm HM560, Microm, Walldorf, Germany) and fixed on Superfrost microscope slides (Menzel, Braunschweig, Germany). The next step was carried out on dry-ice so that slides could not thaw completely. From the slides, the brain areas of interest were gained by micropuncture as described before by Palkovits (1973), using sample corers of 0.5 mm (BLA, CeA) and 1.0 mm (Cg, PVN) in diameter (Fine Science Tools, Heidelberg, Germany). The appropriate slides were detected by help of the mouse brain atlas

(Paxinos and Franklin, 2001). Thus, slices corresponding to the bregma reference points -1.46 to -1.82 mm were used to punch the BLA, the reference points -1.22 to -1.58 mm (CeA), -0.58 to -0.94 mm (PVN), and 1.34 to -0.22 mm (Cg), respectively, were considered for punching the other brain regions.

The RNA of each sample was extracted from the pooled punches of each brain region using the RNeasy Plus Micro Kit (Qiagen, Düsseldorf, Germany). Based on the kit's protocol the sample was vortexed for 30 s with 350 µl buffer RLT plus and 5 µl carrier RNA working solution, the latter was prepared in advance as recommended. After transferring the sample to a gDNA Eliminator spin column and a 30-second-centrifugation at 8,000 g, the flow-through was mixed with 350 µl of 70% ethanol. Now the mix was centrifuged in a RNeasy MinElute spin column for 15 s at 8,000 g, mixed with 700 µl buffer RW 1, centrifuged under same conditions, subsequently mixed with 80 µl of 80% ethanol before centrifuged first for 2 min at 8,000 g and finally for 5 min at maximum speed and open lid. Then the RNA was eluted in 20 µl RNase-free water with an incubation time of 3 min at room temperature and spun down for 1 min at maximum speed. Before the RNA was stored at -20 °C until the day of cDNA synthesis, its concentration was measured on the Implen NanoPhotometer.

3.4.2.3 cDNA synthesis

Before the cDNA synthesis was performed, the RNA samples were adjusted to a concentration of about 50 ng/µl. The synthesis was then started with the preparation of a master mix as claimed in the protocol of the used High Capacity cDNA Reverse Transcriptase Kit (Applied Biosystems, Foster City, CA, USA). Thus, 10 µl master mix consisted of 2 µl RTbuffer, 0.8 µl dNTPs (100mM), 2 µl RT random primers, 1 µl MultiScribe Reverse Transcriptase and 4.2 µl nuclease-free water. The ice-cold master mix was then merged 1:1 with the sample RNA on a 96-well plate. For cDNA synthesis the thermocycler program was adjusted to the conditions listed in Table III-2.

Table III-2: Thermocycler program for cDNA synthesis.

Temperature	Time
25 °C	10 min
37 °C	120 min
85 °C	5 min
4 °C	forever

3.4.2.4 Performance of qPCR

The qPCR was performed in a 384-well format, using the Roche LightCycler 480 (Roche Diagnostics Deutschland, Mannheim, Germany). Each sample was analyzed in duplicates using a 1:5 dilution of the original cDNA concentration (see chapter 3.4.2.3 for details on cDNA synthesis). In order to test the primer functionality, a standard curve was created for each primer pair, using 1:5, 1:25 and 1:125 dilutions of the original cDNA concentration of one HAB sample. Two housekeeper genes, beta-2 microglobulin (*B2mg*) and DNA polymerase II large subunit (*Pol2b*), were used for data normalization.

In total, seven runs were required to test all candidate genes. Directly before the PCR was performed, each of the 384 wells was loaded with 2 µl cDNA (1:5) and 8 µl master mix, the latter consisting of 1 µl forward primer (10 µM), 1 µl reverse primer (10 µM), 1 µl RNase free water and 3 µl of 2x SYBR Green MasterMix (Qiagen, Düsseldorf, Germany). The PCR ran under the conditions shown in Table III-3. Data analysis was performed using the absolute quantification fit points method, provided with the LightCycler software. The so created crossing points (Cp) were further analyzed by the ΔΔCT-method (Livak and Schmittgen, 2001) to normalize the sample data to the housekeeper genes. Each sample's fold regulation value was additionally normalized to the mean values of the HAB group, before expression differences between HAB and LAB animals were statistically analyzed by applying the Mann-Whitney U test (SPSS v16.0.1 software).

Table III-3: Conditions of qPCR run.

Temperature	Time
95 °C	5 min
-------- Start loop (40x) --------	
95 °C	10 s
60 °C	30 s
72 °C	10 s
-------- End of loop --------	
95 °C	5 s
50 °C	10 s
42°C	30 s

3.4.3 Correlation of CNV data with expression data

In a first step, the list of 12,171 microarray probes created during data analysis (see chapter 3.4.2) was adapted. Thus, all information but the ones about probe position, p-values and fold-change were removed. Afterwards, the probe positions were checked for mapping into any gene based on the Mouse Genome Informatics (MGI) database (referring to UCSC database version mm9) and the gene information was added to the list. The intention of this step was to ensure that the expression data could be compared to the CNV data without problems occurring caused by confusion about different genome databases consulted. Excluding all probes that could not be assigned to any of the chromosomes 1-19, X and Y, a table of 12,120 probes resulted. Next, a new list was generated, comprising all genes that appeared to be differentially expressed in at least one probe and one of the examined brain regions (BLA, CeA, Cg, PVN), combined with the information on how many probes mapped into these genes and how many of them showed a p-value less than 0.05. As final step, the genomic position of genes was aligned to the regions of CNVs defined by aCGH, JaxMDGA and NGS, respectively. Subsequently, an overlap count could be conducted and statistical analysis was performed.

In a first approach the hypothesis that CNVs have an influence on protein coding gene expression was tested using a two-proportion z-test. The calculation was executed in R as described below:

```
n1 = amount of genes not showing significant expression difference

n2 = amount of genes differentially expressed

cnvDiff = amount of differentially expressed genes mapping in CNVs

cnvNodiff = amount of genes not differentially expressed, mapping
            in CNVs
```

$$pro1 = \frac{cnvNodiff}{n1}$$

$$pro2 = \frac{cnvDiff}{n2}$$

$$pro3 = \frac{cnvDiff + cnvNodiff}{n1 + n2}$$

$$SE = \sqrt{pro3 \cdot (1 - pro3) \cdot \left(\frac{1}{n1} \cdot \frac{1}{n2}\right)}$$

$$z = \frac{pro1 - pro2}{SE}$$

$$p - value = 2 \cdot pnorm(-abs(z))$$

In a second approach, the question of whether the amount of gene copies (i.e. a copy number gain or loss) could influence the amount of gene expression (i.e. more or less expression) was examined. Initially, a CNV value and an expression value were assigned to every gene that was shown to be differentially expressed and that mapped into any CNV detected by aCGH, JaxMDGA or NGS, with a value of +1 for a copy number gain and more expression, respectively, and a value of -1 for the opposite conditions. In case the CNV data showed contradictory results, the gene appeared twice in the created list, once with a CNV value of +1 and once with a value of -1. A similar strategy was applied when assigning the expression values. If a gene was covered by several microarray probes and at least one of the probes showed more expression while the other(s) showed less, the gene was listed twice, once with an expression value of +1, once with a value of -1. For statistical analysis, the Pearson's product-moment correlation was applied on the table of CNV and expression values, using the "cor.test" function of "stats" R package. The resulting p-values were corrected for multiple testing by bonferroni correction, using the "p.adjust" function in the just mentioned R package. The analysis was done for each of the four tested brain regions separately.

3.5 Influence of a single CNV comprising the glyoxalase 1 (*Glo1*) locus on anxiety-related behavior

In order to test the influence of a single CNV on anxiety-related behavior, a selective breeding approach based on the HAB/LAB mouse lines was conducted. The decision to focus on the CNV comprising the *Glo1* gene was made since *Glo1* was shown before to be linked to anxiety and depression, not only in the HAB/LAB mouse model (Hambsch *et al.*, 2010; Kromer *et al.*, 2005; Landgraf *et al.*, 2007) but also in different human and animal studies (Hovatta *et al.*, 2005; Tanna *et al.*, 1989; Williams *et al.*, 2009), although there was no consensus in these studies if the relationship is a positive or negative one. Thus, in the text below, the locus is referred to as *Glo1* locus. However, it should be kept in mind that the locus not only comprised the *Glo1* gene, but also *Dnahc8* (dynein, axonemal, heavy chain 8) and parts of *Glp1r* (glucagon-like peptide 1 receptor).

3.5.1 Breeding protocol

The idea of the breeding approach was to end up with animals having the genetic background of HAB animals, with exception of the *Glo1* locus and all its genetically linked loci, and further phenotype these animals to reveal a potential influence of that locus on behavior. The first generation arose from a HAB-LAB cross-breeding. The subsequent generations were selectively bred for the *Glo1* locus by mating the individuals harboring the LAB-specific *Glo1* locus with HAB animals. A detailed breeding scheme can be found in the supplementary data (report S4).

In order to estimate the genomic similarity between HAB animals and the current generation, a similarity coefficient was evaluated. The calculation of that coefficient based on the probability of heredity of the *Glo1* locus. An animal inherits 0.5 of its genome from its father and 0.5 from its mother. Thus, as animals of the first generation (G1) had a HAB mother and a LAB father, their similarity coefficient was 0.5. An animal of the next generation (G2) would inherit half of its genome from a HAB animal and the other half from an animal of G1. Since the similarity of the G1 genome to the original HAB genome is 0.5 and half of the genome is passed down to the offspring, the theoretic similarity to the HAB genome of the G1-originated half of the G2 genome is 0.25

(= 0.5 * 0.5). The similarity coefficient of G1 is therefore 0.75, the sum of 0.5 (HAB-originated half of the G2 genome) and 0.25 (G1-originated half). The same principle was true for the subsequent generations, that is for generation 3 (G3), the similarity coefficient was a sum of 0.5 (HAB-originated) and 0.375 (G2-origianated; = 025 + ½*0.25). All similarity coefficients calculated up to the 10th generation are shown in Table III-4.The following formula could be applied to accelerate the calculation of the similarity coefficient *x* for a specific generation *n*:

$$x_n = \frac{1}{2^{n-1}} \cdot x_1 + \frac{2^{n-1} - 1}{2^{n-1}}$$

Table III-4: Similarity coefficients of Glo1 breeding approach. In the right column the similarity coefficient for the generation named in the left column is shown. The coefficient is an indicator for the resemblance with a HAB genome (with 1 as 100% resemblance).

Generation *n*	Similarity coefficient x_n
1	0.5
2	0.75
3	0.875
4	0.9375
5	0.96875
6	0.984375
7	0.9921875
8	0.99609375
9	0.998046875
10	0.999023438

3.5.2 Genotyping of the animals

The determination of the animals' genotype was based on the knowledge of different copy number states for the *Glo1* allele between HAB (one copy only) and LAB (more than one copy). Thus, a PCR analysis was performed using oligonucleotide primers targeting sequences around the CNV breakpoint. The binding positions of primers are depicted in

Figure III-8. A control primer pair was positioned in front of the potentially duplicated region to test the functionality of PCR, whereas the breakpoint primers embraced the CNV breakpoint, resulting in a PCR product of length 290 bp and 150 bp, respectively. Target primers only led to the generation of a specific product if a CNV was present, thus, a 150 bp product should only be seen in animals with the LAB locus. The primers were adapted from Williams *et al.* (2009) with their sequences shown in Table III-5.

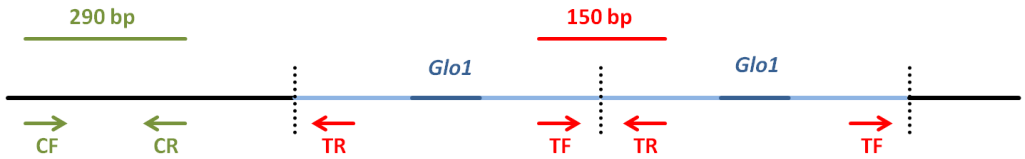

Figure III-8: Positions of genotyping primers. Black lines indicate the genomic region of chromosome 17, that embrace the CNV (blue lines). The target primers (red arrows) bind near to the CNV breakpoint, but outside of the *Glo1* sequence (dark blue lines). PCR results only in a specific product of 150 bp in length (red line) if a duplication of the locus exists, which is the case in LAB, but not HAB animals. The control primers (green arrows) generate a product of 290 bp length and were used to test functionality of PCR. *CF = forward control primer, CR = reverse control primer, TF = forward target primer, TR = reverse target primer.*

Table III-5: Primers for targeting CNV breakpoint of *Glo1* locus.

Chr.	Target	Orientation	Primer sequence (5' → 3')	T_m [°C]	Prod. size [bp]
17	Breakpoint	forward	CTC TGC CCC AGA GAA CAG TC	64.0	150
		reverse	*TGA TAG AGG CCA CAC AGC AG*	*64.1*	
17	Control	forward	CAG TCG TCG ACA GTC ATC GT	64.1	290
		reverse	*GAG CTG AAG GGA TCT GCA AC*	*63.9*	

The genotyping was done using genomic DNA isolated from tail tips. Analogues to the protocol described in chapter 3.1.1 the DNA was extracted by use of the NucleoSpin® Tissue kit (Macherey Nagel, Düren). Subsequently, after measuring of DNA concentration, the PCR reaction mix was prepared with 4 µl DNA (25 ng/µl), 1 µl of each of the four primers (10 pmol/µl), 2 µl Taq buffer including KCl (Fermentas, St. Leon Rot, Germany), 1.2 µl $MgCl_2$ (25 mM) 0.4 µl dNTP mix (10 mM each), 7.4 µl distilled water and 1 µl Taq

DNA polymerase (1 U/µl). The PCR ran according to the conditions listed in and PCR products were screened for by gel electrophoreses using a 1.5% agarose gel.

Table III-6: PCR conditions of Glo1 genotyping.

Temperature	Time
95 °C	2 min
-------- Start loop (35x) --------	
95 °C	30 s
60 °C	30 s
72 °C	1 min
-------- End of loop --------	
72 °C	5 min
8 °C	forever

3.5.3 Behavioral testing

All 94 animals of the 6th generation were tested at the age of 8 weeks on the elevated plus-maze (EPM). One day later, they were tested in the open field (OF) and after one day of recovery in the tail suspension test (TST). The final test, the forced swim test (FST), was conducted another two days later. A detailed description of the test procedures can be found in chapters 2.1, 2.2, 2.3, and 2.5.

3.5.4 Statistical analysis

Since the idea of the *Glo1* breeding approach was to reveal the influence of the CNV on phenotypes, the results of the behavioral tests were analyzed for genotype specific differences in behavior, applying the Mann-Whitney U test. Data acquired from males and females were examined separately.

4 Analysis of CNVs in CD-1 mice

4.1 Testing of CD-1 mice in different behavioral experiments

384 outbred CD-1 mouse males were phenotypically characterized in a series of five tests covering different aspects of anxiety. The aim of this project, called "CD-1 panel", was to reveal the genetic contribution to anxiety-related behavior by examine later on the mice's genotypes under different aspects. All mice were bred in the Charles River Laboratories (Sulzfeld, Germany) and delivered at the age of eight weeks in eight different batches each of 48 animals. Each batch of animals went through the identical test battery, as illustrated in Figure III-9. After a 4-day-habituation, half of the animals were tested on the elevated plus-maze (EPM) the other half in the open field (OF) and two days later *vice versa*. Another two days later all 48 animals were tested in the forced swim test (FST). Before the stress reactivity test (SRT) on day 12 and the tail suspension test (TST) on day 14 after arrival were performed, the animals could recover for four days. Subsequent to the last test all mice were anesthetized by inhalation of Isoflurane (Curamed Pharma, Karlsruhe, Germany) and then sacrificed by decapitation. For further studies, tail tips were taken and immediately frozen on dry ice before stored at -80 °C.

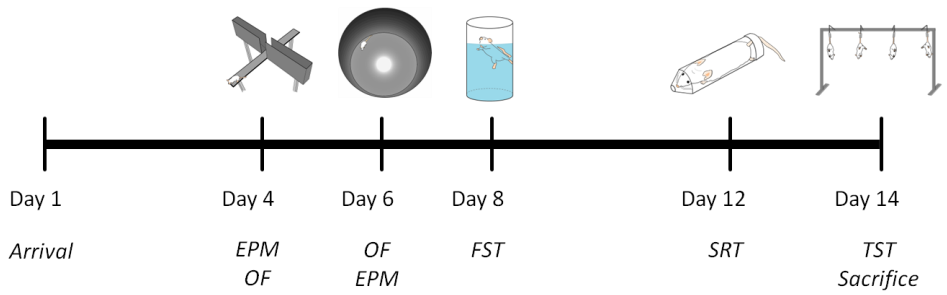

Figure III-9: Test battery of CD-1 panel. Animals that performed the EPM test on day 4 were tested in the OF on day 6 and vice versa.

4.2 Detection of CNVs in CD-1 mice

4.2.1 JAX Mouse Diversity Genotyping Array (JaxMDGA)

4.2.1.1 Performance of the array

One way to detect CNVs in the phenotyped CD-1 mice was by use of the JAX Mouse Diversity Genotyping Array. The same array type was applied before to detect CNVs in HAB/LAB animals, though this time a 32-sample-format was chosen. More details on the array can be found in chapter 3.2.1. Again, tail tips were sent on dry ice to The Jackson Laboratory, where DNA extraction and subsequent processing was performed. Finally, the company provided the raw data for further analysis.

In total 64 of the 384 phenotyped CD-1 animals (see chapter 4.1) were tested in two different approaches with 32 animals each. The mice's behavior averaged over all performed tests provided the basis for choosing the 12 most anxious, 12 least anxious and 8 mean individuals to be analyzed in the first array approach (for details see Widner-Andrä, 2011). The same selection strategy was pursued for the second approach, though not considering the animals already screened in the first approach.

4.2.1.2 Computational analysis of Jax MDGA data

The raw data provided by Jackson Laboratory were analyzed for each set of 32 animals separately. Analogues to the analysis of the HAB/LAB data, the "simple CNV" function of the "MouseDivGeno" R package was applied as described in chapter 3.2.2. That function required a reference sample to compare all other samples with. Since assigning only one sample to be the reference would lead to missing information on CNVs, the function was applied multiple times by looping over all samples to end up with comparisons of all sample pairs. The following R script section was written to loop over the samples:

```
files=list.files(pattern='.CEL',full.names=F)

for(i in 1:length(files)){
 for(j in i+1:length(files)){
  if(j+1 <=32){
   result <- paste('Result_',as.character(files[i]),'_ggRef_',
                   as.character(files[j]),sep='')

   simpleCNV(snpProbeInfo,snpInfo,snpReferenceDistribution,
      invariantProbeInfo,invariantProbesetInfo,
      invariantReferenceDistribution,
      celFiles=files[i],
      referenceCelFile=files[j],
      summaryOutputFile=result) }}}
```

The potential loss of information, if referring all samples to only one reference, is in consequence of the algorithm used by the "simple CNV" function. Briefly, the algorithm defines sets of array probes whose mean fluorescence values are considered to show significant differences between sample and reference. The first and last probes of such sets represent the breakpoints of the respective CNV. Theoretically it might happen that for a specific set of probes no significant difference could be found between samples and the reference, but if two samples would be compared to each other, the significance threshold might be passed, as exemplary depicted in Figure III-10.

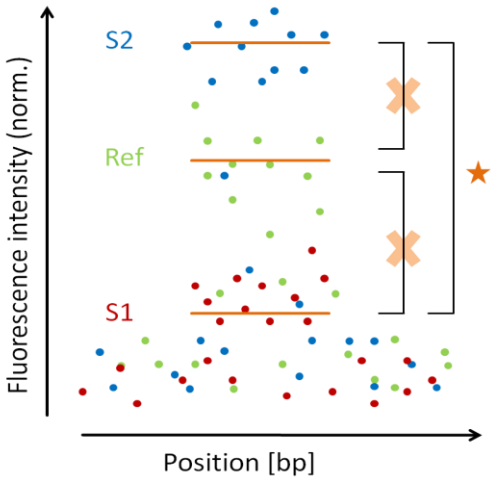

Figure III-10: Situation of potential information loss based on applied algorithm. Normalized fluorescence intensities of single probes from a reference sample (Ref, green), sample 1 (S1, red) and sample 2 (S2, blue) at a specific genome position are depicted as single dots. Comparisons of sample intensities to the reference intensities might not show significant differences, indicated by the orange crosses, and hence a potential CNV (orange lines) would be missed. A direct comparison of sample 1 with sample 2 intensities could pass the significance threshold (orange star).

The provided output of the "simple CNV" function was a list of CNVs with the mean value of the normalized fluorescence intensities of probes considered for CNV definition, and with information about chromosome, start and end point. After merging all output lists of all comparisons, duplicated elements were removed. However, there were still regions incorporated showing only slight shifts in breakpoints, thus, were very likely to represent just a single CNV. Therefore, those regions were outlined as one CNV only, with breakpoints as far away from each other as possible (Figure III-11).

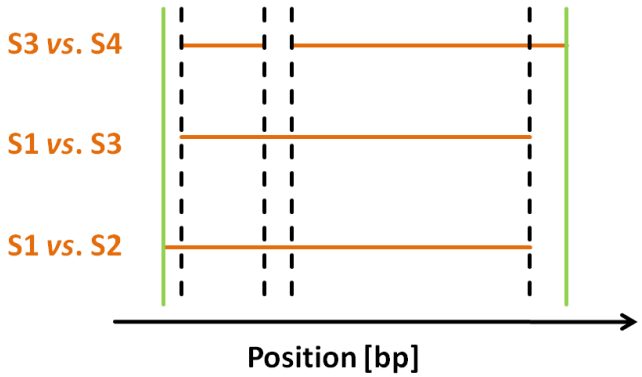

Figure III-11: Breakpoint definition for CD-1 CNVs. If results of different comparisons showed CNVs (orange lines) with huge overlap, the regions were outlined as a single CNV. Breakpoints for the newly defined CNV are indicated by green lines. Dashed lines depict former breakpoints. S1 = sample 1, S2 = sample 2, S3 = sample 3.

As a consequence of the new breakpoint definition another problem occurred. Enlarging the size of the CNV meant also to increase the amount of array probes mapping into the CNV. Therefore, the mean normalized intensities of all relevant probes had to be recalculated, since the values were required for further analyses. As mentioned before, applying the "simple CNV" function only yielded a list of final results, but not the normalized intensities of any single probe. Thus, the "simple CNV" function had to be decrypted before subsequently the R code was rewritten to get the desired information. Since that new R script would fill several pages, it can be found in the supplementary data (report S5). Therein, all parts of the original function's code that were rewritten were highlighted.

Shortly summarized, for each of the two array approaches the CNVs were revealed by comparing all 32 animals against each other and listed in two separate tables (one per approach). Each table comprised all CNVs that occurred in at least one pair of animals together with the information on chromosome, start and end point. For each animal and CNV the mean normalized intensities of the respective probe sets were calculated as the values were needed for further analysis on the correlation of CNVs and behavior.

4.2.2 Multiplex Ligation-dependent Probe Amplification (MLPA)

The Mulitplex Ligation-dependent Probe Amplification (MLPA), developed by MRC-Holland (Amsterdam, Netherlands) is a method to detect CNVs of multiple DNA sequences in a single multiplex PCR-based reaction. Fundamental for this method is that not the sample DNA is amplified during the PCR, but MLPA probes, each consisting of two oligonucleotides: the left probe oligonucleotide (LPO) and the right probe oligonucleotide (RPO). The LPO as well as the RPO consist of two parts, a primer binding sequence and a hybridizing sequence. In order to be ligated, both the LPO and the RPO have to bind to adjacent DNA target sequences via their hybridizing sequences, named the left and right hybridizing sequence (LHS and RHS), respectively. After ligation the probes are amplified exponentially by use of PCR primers specific for the forward primer binding sequence GGGTTCCCTAAGGGTTGGA, which forms the 5' end of the LPO, and the reverse primer binding sequence TCTAGATTGGATCTTGCTGGCAC, which forms the 3' end of the RPO. Here, differences in PCR product length could be obtained by designing hybridizing sequences of variable size, enabling the PCR products to be separated and quantified by capillary electrophoreses.

4.2.2.1 Designing of synthetic MLPA probes

In a first step, interesting target regions of the MLPA probes had to be defined. These regions were chosen based on preliminary data of the first JaxMDGA approach. CNVs that showed a correlation coefficient (correlation with time animals spent on the open arm) with a value belonging to the upper 20% of the frequency distribution, and whose genomic position overlapped with any CNV found by aCGH or JaxMDGA in HAB/LAB mice

were considered to be of interest. In a second step, three MLPA probes per region were designed for 11 CNVs, resulting in a total of 33 probes.

Designing synthetic MLPA probes turned out to be a big challenge and was done by strictly following the recommendations in the "Designing synthetic MLPA probes" protocol (version 12) provided by MRC-Holland. In order to find candidate probes the genomic region of interest was retrieved from the UCSC genome browser, assembly NCBI37/mm9, and subsequently submitted to the freely available online tool for probe design MAPD (Zhi, 2010). The resulted list of potential probe sequences had to be further analyzed and checked for the required probe conditions. First, all LHS and RHS were tested for uniqueness using the BLAST tool provided at the Ensembl database (Flicek *et al.*, 2011). Since probes were used in three different MLPA approaches, three sets of 11 probes with each probe having a unique total length (LPO + RPO), had to be defined. The melting temperature for each probe was determined using the RaW program, provided and recommended by MRC-Holland, and had to be ≥ 71.0 °C for LHS/RHS of length 29 - 35 nt (nucleotides), ≥ 70.0 °C for 37 – 39 nt and ≥ 68 °C for 41 - 55 nt. Between probes, a total length difference of 4 nt was required. Furthermore, the GC-content should amount to ~50 %. Finally, for probe design it was taken into account that the 5' end of the LHS did not start with an adenine.

All synthetic probes were ordered at Sigma-Aldrich (Taufkirchen, Germany), with all probes purified by HPLC and PAGE, respectively, and all RPOs 5' phosphorylated. In Table III-7 all probes used in the experiment are listed, including information on the probe characteristics. Table III-8 shows information on the total probes' genomic positions and also on the CNVs (based on final data) the probes mapped into.

Table III-7: Sequence information of MLPA probes. Here the sequences of LHS and RHS, respectively are listed as well as their size in nucleotides, their GC content in percent and their melting temperature, determined by use of RaW program. LHS and RHS together with the primer binding sequences form one MLPA probe.

Probe No.	Probe part	Sequence	Size [nt]	GC [%]	Tm [°C]
1	LHS	GGC TTT GTT AAA CTT GGC GTC TAC AGG TTG CTT TCC AGC CTA ACT AG	47	47	81.48
1	RHS	GTT TTG CCG TGG TGA GGA AAT GAG GGT TTT AGA TGG GGA AGG GAA AG	47	49	85.23
2	LHS	GAT GAT GGT TTC ACA GGT AGC AAT CAG GTG TGG CCC ATC ATA AAT CAT G	49	45	83.69
2	RHS	AGC ACC TCC ATT GTG ACC AGA AGC CCC ATA TGC GGT TTG TTC CAG GTT G	49	53	87.84
3	LHS	TAC TGA GGG GCC TGA AAC CTT ACA GCA ATA C	31	48	73.75
3	RHS	AAG AGC GAT CAC CAC CCC ATA CTA CAC ACA C	31	52	75.61
4	LHS	GTC ACA GTA ACA TTT CTC AGA AGT TAG CCC ACT GCC TTC CA	41	46	79.95
4	RHS	GCT TTC AGT GTA GGA GAA GAG ATA TTC GAC ATC AGT GTG CC	41	46	78.94
5	LHS	TTC CAA ATA GAC ACC CCT GCT TTG AAA GC	29	45	72.87
5	RHS	TCC CAC CTT AGT TGG CAG ATG GGA AGA TG	29	52	75.92
6	LHS	TGC AAA GGT AGG ATC CAT GTA TGT GCT CTG CAT	33	45	75.59
6	RHS	ACC ATA CTG CGG TTG TAC TGA CTC TCA TCT CCA	33	48	75.52
7	LHS	TTC GGT AAA GTG AGG ACT CCT TCT CAG AGC CAA AGA G	37	49	78.23
7	RHS	AAC GTT GCT AGG ACT GAA CGA TGT CAT CAA TAC CAC G	37	46	77.6
8	LHS	TCA CTT GGG CAG CTC CAT AGA ACA CAG TCC ATG CT	35	51	80.01
8	RHS	GAC GGG GGG CGG AGA TTT ATC TAA CTG GTG AAT GA	35	51	80.63
9	LHS	GGT CTA CTC TGG CTA ATT TGG TAA CAG ACC GTA GAT TGC CCT AGT	45	47	79.6
9	RHS	AGA GAG TAC CTG CAA TAA TCC ATG CCT GAG AGA GTG GGG TGA GCA	45	51	83.66
10	LHS	GTG ACT AGT CTT GCC GAG TTG CTC TTC TGC CTG TGA ATT	39	49	80.16
10	RHS	CTT GCT AAA GCA TGC CTA TGC AAC CCT TGG CTC TGC ATC	39	51	82.36
11	LHS	GGC AGA TGC AGA AAA AAG GGT AAC TAT CGG GGC TCC CTA CTT T	43	49	82.1
11	RHS	TTT CTA AAC CCC CCT TTC CTA GTC TCT TAT GCC CCA CAA GTA G	43	47	79.48

Probe No.	Probe part	Sequence	Size [nt]	GC [%]	Tm [°C]
12	LHS	TAT CCA AGG TCA CGG TTC CAG AGG CCA TCT G	31	55	79.08
12	RHS	GTA CTT TGG GTC ATT GTC CTA CCG GGC TTG T	31	52	77.22
13	LHS	TAT CAA TTA CTG CGC TTG CGT GGG GAG GTC AAG TC	35	51	79.76
13	RHS	CAC AAC AGG GCT GAT GCG TGC TCT TCC GAT AAT TT	35	49	79.86
14	LHS	GCT CTT CAC AAA TAA TCC ATT CGG CTC AGA GAC TGT CGT AGA G	43	47	80.08
14	RHS	AAG AAG TTG ATC ACG GGC CAC CTG C TC ACG TCT GCA CCG AAA	43	53	86.35
15	LHS	TTC TCC TGT CGC TCT TCA AGC ACG AGT GT	29	52	75.61
15	RHS	AAC AGA GTG ATC GCA GAC AGG TGT GCA CT	29	52	75.25
16	LHS	GTC CCC ATA CTC AGG AAG TCT ACC CAC CCA CAT GAG CTA GA	41	54	82.87
16	RHS	CTT TAA TGT TTT CAT GGG ACG CCT GGG AGG ACT GAG ATT TC	41	46	81.09
17	LHS	GAA GAG GCA ATG ATT CCT TAT AGC CCC CCC CCA ATA AAA	39	46	80.31
17	RHS	GGG AGG CCC CTA AAG TGT CCC TAA GTC ATC ACC TAT CCT	39	54	81.99
18	LHS	GAC AGA GTT AAT GGG AGA GCG TTC TTC CCG CAT AA GCA CTT CTG T	45	49	82.7
18	RHS	GGA CCG TGT AAC CAC AGA TTA CTC ATT ATG GCC AAA GAA GCC TCA	45	47	82.15
19	LHS	TGT GTA TCT AGA GGG AAT GTT GGA GAG CAG TCA	33	45	74.06
19	RHS	GCT AGA AGC GCC AAG TAG GAA AAT CAG TCA GGT	33	48	76.13
20	LHS	GTT GCC TCT TGG TAC AGG GTT GGG AAA GGG AGC AGA TAT GTC TAA GA	47	49	83.42
20	RHS	ACT GAT GAT CCA TCC TTC TCA GAC TTT ACT GGG TCC AAC TGA TAT CC	47	45	80.47
21	LHS	TCT TGA GCA GTA GAG GCC CTA ATA TGA GAT GGA CAC T	37	46	75.83
21	RHS	AAA TAG AAT CAG GAC ATT CAG GGG ATT CAG GGG CTG T	37	46	78.5
22	LHS	GTG GGA GTC CTC TGG TTA CCA AGT CAA CCA AGC ACT ACT CAG AGA TCT A	49	49	83.01
22	RHS	GAG CTG GCT GAC ACT ACA TGA CAC CCA AAA GCA GGA CAA TGT TAT ATG T	49	45	82.22
23	LHS	TGC CTA GAG TTT ATG CTA TAG CCT AGG CTG CCT AAC C	37	49	75.7
23	RHS	CCA CCA GCA ACC ACT CAT GGC AGT CTT TCT GTT TTA G	37	49	79.98

Probe No.	Probe part	Sequence	Size [nt]	GC [%]	Tm [°C]
24	LHS	GCA TGT GTG TGA GGG ATA AGA GTA GTG ACT GGC ATC ATT GG	41	49	81.29
24	RHS	GCA CTT TCC CCC TTT TCA GAT GGA TCT GGA CAG AGC AGA AG	41	51	83.43
25	LHS	GTT GAA CAT TGA GGT CCA GAT TTG GCA CCG TCT CCA GCA GAG T	43	51	84.71
25	RHS	TTG GGG GTT CAG CTT TCT GGA CTA CAT TTT CTC CTC TTG GCA C	43	49	82.65
26	LHS	TGT GAG GGT TGT TAG AGC AGA CGT AGC TCC AGG AA	35	51	78.81
26	RHS	GAA CAG CCA AGT CAG CCC TTC ATG AGC TCT GAA AT	35	49	79.14
27	LHS	GAA GAG AGC AGG CAT TCG TCA GCT ACT AAC TTG GGG TAG GTA TGT TAA G	49	47	81.05
27	RHS	TGT CTG CAC CTC ATG TTA GCT TTA GGA AAA CCA GGA CTC AGT GAG CTA G	49	47	81.83
28	LHS	TCT GCA GAG AGG AAG CAT GAT ACT TAC GTA GTG	33	45	72.52
28	RHS	ATG GAG ACA CTA CGT AGA GAT GGC ACT CCC GAC	33	55	77.83
29	LHS	GTC AGG GAC ACA CAG CAG TGG CAG GCC TCC ACA CTT ACT TTT TCG	45	56	86.98
29	RHS	GTC CTG TAA CAA ACA CTC GGA AGA CAC CGT TGA AAG GAG GGA AAG	45	49	83.31
30	LHS	TAG GGC AGG ATG ACA GAA ATG CTT GTT GCT T	31	45	74.73
30	RHS	AGC AAT GCA AAG CCC GTT GTC CTC ACA AAG G	31	52	78.88
31	LHS	TGC CTA GCT CCA CCT TCC TTC CAG TTT GC	29	55	77.16
31	RHS	TTC TCC ACC CTG ACC AGA TCA TCA GGC TC	29	55	77.28
32	LHS	GTG AAA GGC TAT CTC CAC ATT CTC CTG TAC TAG CTG CAA	39	46	77.88
32	RHS	ACT TGG ACC TGG TAG GGC AAC ACA TGT ATC AGG GCT GTG	39	54	82.74
33	LHS	GAG AGC AGT GAT CAA CCA GGA CAC AAC AGC AGA GGA AGC CCA GAT CT	47	53	86.42
33	RHS	GCC TGG AGT ATG GGT GAC AAT GAC AAC ATA CTT TCT TCT ACC TTC TG	47	45	80.61

Table III-8: Genomic position of MLPA probes. The genomic position of combined LHS and RHS sequences (chromosome and basepair information) was determined by use of BLAST tool in Ensembl database. If the probe mapped into any gene sequence, the gene symbol is listed in column "in gene". The fifth column shows the p-value of the final association of the behavior animals showed on the EPM (% time on the open arm) with the CNV detected in CD-1 mice by JaxMDGA (see chapter 4.3), to which the MLPA probe is targeted to. Columns 6-8 show the copy number status (HAB *vs.* LAB) of the HAB/LAB CNVs (detected by three different methods) the MLPA probe could be mapped to.

Probe No.	Chr	Genomic Position (LHS+RHS) [bp]	in gene	p-value Asso.	Status HAB/LAB aCGH	Status HAB/LAB JaxMDGA	Status HAB/LAB NGS
1	2	157,152,804 - 157,152,897	no	0.3297	loss	loss	-
2	2	157,153,700 - 157,153,797	no	0.3297	loss	-	-
3	2	157,153,831 - 157,153,892	no	0.3297	loss	-	-
4	12	68,608,045 - 68,608,126	no	0.2125	-	loss	loss
5	12	68,608,383 - 68,608,411	no	0.2125	-	loss	loss
6	12	68,610,071 - 68,610,136	no	0.2125	-	loss	-
7	14	9,483,225 - 9,483,298	no	0.0563	-	gain	-
8	14	9,481,584 - 9,481,653	no	0.0563	-	gain	-
9	14	9,480,633 - 9,480,722	no	0.0563	-	gain	-
10	11	120,875,255 - 120,875,332	no	0.0536	-	gain	-
11	11	120,880,146 - 120,880,231	no	0.0536	-	gain	-
12	11	120,883,311 - 120,883,372	no	0.0536	-	gain	-
13	17	30,743,662 - 30,743,731	Glo1 *(intron)*	0.4246	loss	loss	loss
14	17	30,870,646 - 30,870,731	no	0.4246	loss	loss	loss
15	17	30,895,305 - 30,895,362	Dnahc8 *(exon)*	0.4246	loss	loss	loss
16	6	101,281,458 - 101,281,539	no	0.4259	-	-	-

Probe No.	Chr	Genomic Position (LHS+RHS) [bp]	in gene	p-value Asso.	Status HAB/LAB aCGH	Status HAB/LAB JaxMDGA	Status HAB/LAB NGS
17	6	101,285,391 - 101,285,436	no	0.4259	-	gain	gain
18	6	101,286,806 - 101,286,895	no	0.4259	-	gain	-
19	1	111,651,529 - 111,651,594	no	0.5699	loss	-	loss
20	1	111,662,394 - 111,662,487	no	0.5699	loss	loss	loss
21	1	111,663,531 - 111,663,604	no	0.5699	loss	loss	loss
22	1	155,368,537 - 155,368,634	no	0.3545	loss	loss	loss
23	1	155,364,562 - 155,364,635	no	0.3545	loss	loss	-
24	1	155,368,432 - 155,368,513	no	0.3545	loss	loss	loss
25	5	54,036,835 - 54,036,920	Rbpj *(exon)*	0.3798	-	gain	-
26	5	54,041,592 - 54,041,661	Rbpj *(exon)*	0.3798	-	gain	-
27	5	54,051,259 - 54,051,356	no	0.3798	-	-	-
28	17	13,267,096 - 13,267,161	no	0.0642	loss	loss	-
29	17	13,297,698 - 13,297,787	no	0.0642	loss	-	-
30	17	13,298,603 - 13,298,664	no	0.0642	loss	-	-
31	6	107,067,619 - 107,067,676	no	0.0955	-	gain	-
32	6	107,078,729 - 107,078,806	no	0.0955	-	gain	-
33	6	107,079,573 - 107,079,666	no	0.0955	-	gain	-

4.2.2.2 Performance of MLPA

After their arrival, the MLPA probes were dissolved in TE buffer (10 mM Tris/HCl, 1 mM EDTA, pH 8.0) to obtain 10 µM stock solutions for ultramers (> 60 nt) and 100 µM ones for normal oligonucleotides (≤ 60 nt). The stock solutions were further diluted to 1 µM final solutions, which were stored at -20 °C until required for further processing.

The DNA used to screen for CNVs by MLPA was extracted from tail tips of 54 CD-1 mouse males, previously tested in different behavioral tests (see chapter 4.1). Based on the percent time the animals spent on the open arm of the EPM, the 27 most anxious and 27 least anxious ones were chosen. Additionally, one HAB and one LAB sample (both aged 8 weeks, generation 46) were used as positive control and reference, respectively. The DNA extraction was done by use of the Wizard® Genomic DNA Purification Kit (Promega, Madisonm WI, USA), following the recommendations of the delivered protocol. Hence, about 0.5 cm tail tissue were digested over night at 55 °C in 17.5 µl proteinase K (20 mg/ml) and 600 µl previously prepared master mix of 3 ml EDTA (0.5 M, pH 8.0) and 12.5 ml Nuclei Lysis Solution. The next day, 3 µl RNase Solution was added and incubated for 30 min at 43 °C and subsequently 5 min at room temperature. Then the sample was mixed with 200 µl Protein Precipitation Solution and vortexed for 20 s. After 5 min of incubation on ice, a 4-minute-centrifugation step was performed at 15,000 g. The supernatant was transferred into a new tube and mixed gently with 600 µl isopropanol. This time the sample was centrifuged for 2 min at 15,000 g and the supernatant was discarded. Another centrifugation step under the same conditions was performed after the addition of 600 µl ethanol (70 %). The supernatant was discarded and the DNA pellet was dried for 15 min at room temperature before it was dissolved at 65 °C for 1 hour in 100 µl TE buffer. The DNA concentration was measured and 150 ng of each sample were checked for sufficient quality on a 1 % agarose gel. Finally, the samples were diluted with low-EDTA TE buffer (10 mM Tris/HCl, 0.5 mM EDTA, pH 8.0 – 8.5) to a final concentration of 20 ng/µl.

The MLPA procedure was done in two approaches of 96-well format using the SALSA MLPA kit PM200-A1 Mouse reference (MRC-Holland), by following the given instructions. Thus, three sets of probes were prepared, called the synthetic basic probemix 1 – 3. For

each mixture 182.4 µl TE buffer were added to 0.8 µl of each of 11 different final solutions, premixed as described above. The first synthetic basic probemix contained the final solutions of MLPA probes 1 to 11, the second of probes 12 to 22 and the third of probes 23 to 33 (see Table III-7 and Table III-8).

Just before the MLPA run was performed, 0.5 µl of the respective synthetic basic probemix, 1 µl PM200 probemix and 1.5 µl MLPA buffer (both provided in the kit) were combined for each sample tested to form the MLPA master mix. 3 µl of that mixture were added to 5 µl DNA (20 ng/µl) after a pre-incubation for 5 min at 98 °C and cooling down to 25 °C. The samples were heated to 95 °C for 1 min and then incubated over night at 60 °C. The day after, before proceeding with the protocol, a ligase master mix for 100 samples was prepared, including 2.5 ml distilled water, 300 µl Ligase buffer A, 300 µl Ligase buffer B and 100 µl Ligase-65. Then, the over-night-incubated samples were cooled down to 54 °C and 32 µl ligase master mix were added to each of it. The PCR program proceeded with 54 °C for 15 min, 98 °C for 5 min and 20 °C until the polymerase mastermix for 100 samples was prepared, consisting of 750 µl distilled water, 200 µl SALSA PCR primermix and 50 µl SALSA polymerase. Afterwards, 10 µl polymerase mastermix were added to each sample and PCR program was continued with 35 cycles of 30 s at 95°C, 30 s at 60 °C and 60 s at 72 °C. Finally, after another 20 min at 72 °C, the PCR program finished achieving 15 °C and samples were stored at 4 °C until capillary electrophoresis was conducted.

4.2.2.3 Performance of capillary electrophoresis

The samples amplified in the MLPA PCR were diluted 1:50 and 1:100 with distilled water, before they were tested in two separate capillary electrophoresis runs. 1 µl of each sample was transferred to a 96-well sequencing plate and mixed with 12 µl electrophoresis master mix. The latter was prepared just before usage, consisting of 1197.6 µl HiDi formamide and 2.4 µl GeneScan – 500 LIZ size standard (both delivered from Applied Biosystems) for samples diluted 1:50, and 1198.8 µl HiDi formamide and 1.2 µl 500 LIZ standard for samples diluted 1:100. Subsequently, the plate was heated to 95 °C for 3 min before it was stored on ice. The sequencing process was performed in the facilities of the institute of human genetics (Helmholtz Zentrum Munich, Germany) on an

Applied Biosystems Sequencer 3730. Read outs were provided as fsa-files and analyzed using the intermediate version 1.0.0.43 of the Coffalyser software, developed by MRC-Holland, strictly following the engineer's instructions (Coffa and van den Berg, 2011).

4.3 Association of CNVs with behavior

The CNVs detected by JaxMDGA in 64 phenotyped CD-1 mice (see chapter 4.2.1) were statistically analyzed for a potential association with anxiety-related behavior, using a generalized linear model, i.e. the "glm" function in R, considering the mean normalized intensities of the CNVs and the measurement values of different behavioral tests. The p-values were calculated by a likelyhood-ratio test as implemented in the "anova" function of the "stats" R package.

The just described method was applied to the combined data sets of all 64 animals. The merging was done in analogy to the CNV detection described in chapter 4.2.1.2, that is, by defining overlapping CNVs as one with a size as large as possible, and calculating the mean normalized intensities of all newly defined CNVs.

4.4 Cluster analysis of interesting genes using DAVID Bioinformatics Database

After the association of CNVs with anxiety-related behavior, the list of CNVs that could be associated with the time animals spent on the open arm of the EPM was aligned with the table of all known mouse genes based on the MGI database (NCBI build m37). A functional cluster analysis was performed for all protein coding genes mapping into any of these CNVs. This was done by submitting the list of candidate genes to the DAVID bioinformatics database (version 6.7), a database for annotation, visualization, and integrated discovery. Following the protocol of Da Huang and colleagues 2009a, a functional annotation table was generated. Clusters with an enrichment score larger than 1.3 were considered to show sets of genes having significant biological impact.

The same functional annotation analysis was performed for all HAB/LAB candidate genes.

IV Results

1 Position and impact of CNVs in HAB/LAB mice

1.1 Position of CNVs detected by aCGH

The comparison of the three different CGH arrays, performed with DNA extracted from tail tissue and brain of two different HAB and LAB mouse males, respectively, did not show prominent differences. However, for determination and breakpoint definition of CNVs, the data of all three arrays were considered, resulting in a total amount of 98 CNVs detected, with exactly half of them showing a copy number gain and loss in HAB mice, respectively. The list of CNVs including the genomic positions is shown in the supplementary table S1.

The CNVs ranged in size from 9,947 bp to 7,484,010 bp with a mean of 992,497 bp and a median of 544,637 bp, covering about 3.7% (97,264,699 bp) of the genome. The minimal distance between the different CNVs was shown to be 10,560 bp, the maximal 95,268,064 bp with a mean and a median of 20,361,585 bp and 7,251,905 bp, respectively. Out of 98 CNVs, 47 (= 48%) were smaller than 0.5 Mbp (megabasepairs), as shown in Figure IV-1. A more detailed frequency distribution of the CNVs with sizes less than 0.5 Mbp is depicted in Figure IV-2.

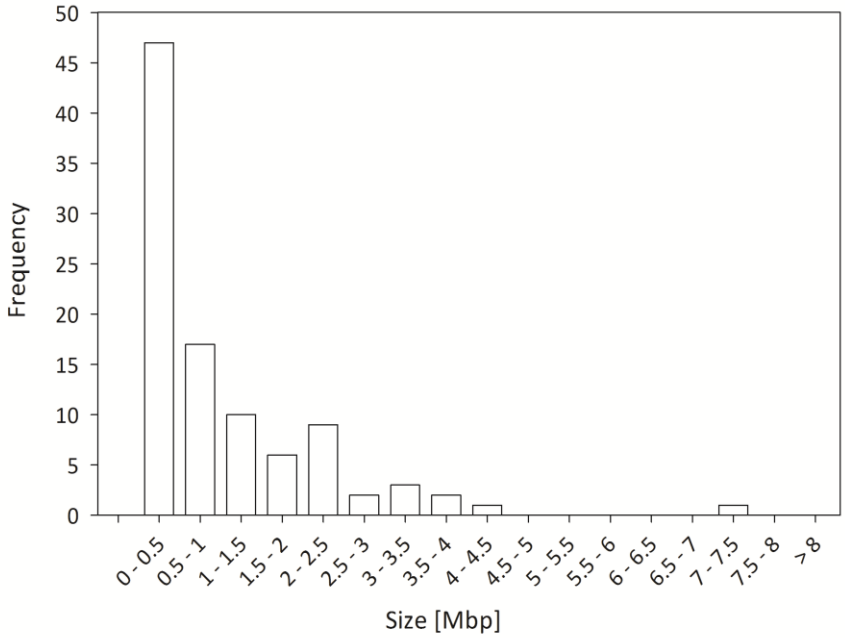

Figure IV-1: Frequency distribution of CNVs detected by aCGH in HAB *vs.* LAB mice, in size classes of 0.5 Mbp. The size classes shown are depicted in a range of 0.5 Mbp and are plotted against the number of CNVs in the respective size range.

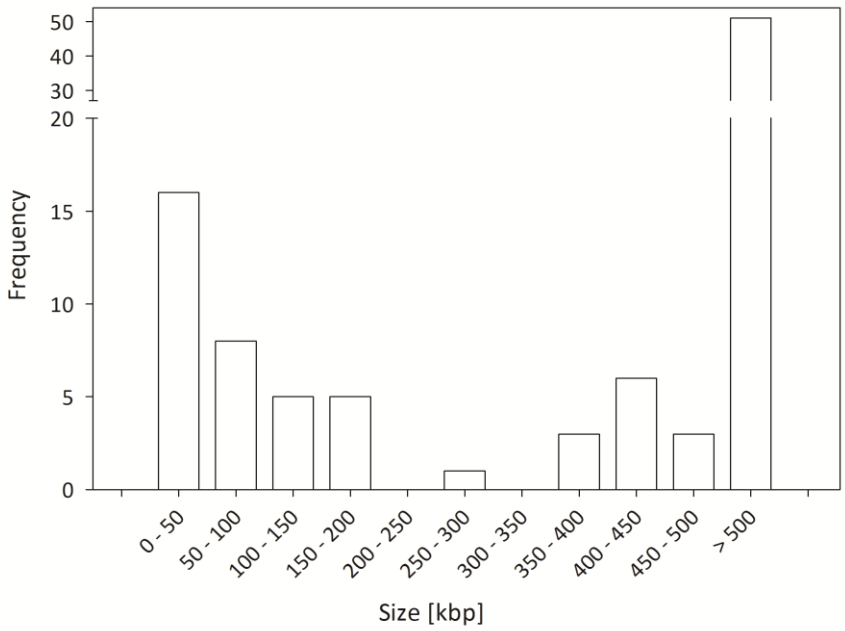

Figure IV-2: Detailed frequency distribution of CNVs < 0.5 Mbp, detected by aCGH in HAB *vs.* LAB mice. Size ranges of 50 kbp are plotted against the number of CNVs in the respective size range.

The detected CNVs are distributed all over the genome with most found on chromosome 7 (14 CNVs) and least on chromosome Y (1 CNV). Details on the chromosomal distribution of CNVs are shown in Figure IV-3 and Table IV-1.

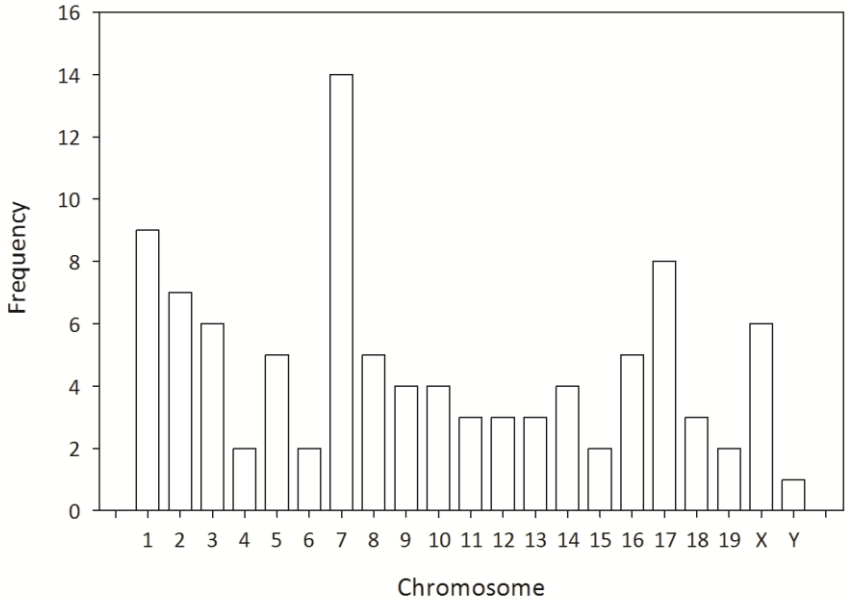

Figure IV-3: Chromosomal distribution of CNVs detected by aCGH in HAB *vs.* LAB mice. Chromosomes are shown on the x-axis. The y-axis depicts the frequency of CNVs in the respective chromosome. Most CNVs were detected on chromosome 7, least on chromosome Y.

Table IV-1: Detailed information on aCGH-detected CNVs per chromosome. The first column shows the chromosome. The other columns comprise the following information (from left to right): number of CNVs, size of the CNVs summed up, median and mean size of the CNVs, and the size of the largest and smallest CNV, respectively.

Chr	No. of CNVs	Sum [bp]	Median [bp]	Mean [bp]	Max [bp]	Min [bp]
1	9	8,649,071	882,700	961,008	2,436,586	15,782
2	7	6,172,682	541,689	881,812	2,686,812	21,312
3	6	3,608,317	665,839	601,386	915,869	35,096
4	2	3,277,833	1,638,917	1,638,917	2,263,464	1,014,369
5	5	5,558,001	945,478	1,111,600	2,158,583	26,505
6	2	1,226,380	613,190	613,190	1,090,185	136,195
7	14	22,516,119	802,785	1,608,294	7,484,010	46,243
8	5	4,927,061	1,142,405	985,412	2,344,409	17,576

Chr	No. of CNVs	Sum [bp]	Median [bp]	Mean [bp]	Max [bp]	Min [bp]
9	4	2,068,300	497,304	517,075	1,033,682	40,010
10	4	2,559,323	146,318	639,831	2,205,673	61,015
11	3	6,074,087	1,710,166	2,024,696	4,302,267	61,654
12	3	3,730,125	987,243	1,243,375	2,701,083	41,799
13	3	4,266,014	688,709	1,422,005	3,446,373	130,932
14	4	5,545,714	1,003,368	1,386,429	3,508,029	30,949
15	2	5,209,248	2,604,624	2,604,624	3,608,029	1,601,219
16	5	726,832	83,136	145,366	418,154	9,947
17	8	5,291,902	440,133	661,488	1,990,959	19,017
18	3	3,549,032	713,974	1,183,011	2,359,869	475,189
19	2	1,030,394	515,197	515,197	934,750	95,644
X	6	902,506	99,067	150,418	431,528	24,315
Y	1	375,758	375,758	375,758	375,758	375,758

1.2 Position of CNVs detected by JaxMDGA

The analysis of JaxMDGA data revealed a total of 180 CNVs, all listed in supplementary table S2, with a mean and median size of 81,842 bp and 8,906 bp, and a maximum and minimum size of 7,383,756 bp and 2 bp, respectively. The variants covered about 0.55% (14,731,644 bp) of the genome and arose at a minimal and maximal distance of 1,482 bp and 106,573,464 bp, with a mean and a median of 10,857,870 bp and 3,733,285 bp, respectively. Following the original definition of CNVs, thus considering only those regions larger than 1 kbp, the minimum size shifted to 1,195 bp, while the mean and median size of the total of 137 CNVs was 107,431 bp and 15,427 bp, respectively. Most of the CNVs, namely 82.2%, or 76.6% if only considering regions > 1 kbp, were smaller than 50 kbp. Just 2 CNVs had a size larger than 0.5 Mbp. The majority of CNVs (118 or 83) showed a gain in the HAB animal, however, only 62 (or 54) showed a loss. The frequency distribution is illustrated in Figure IV-4, and more detailed for smaller CNVs in Figure IV-5. Figure IV-6 depicts only CNVs of size < 1 kbp.

Different to the CNVs detected by aCGH, here no copy number change was found on chromosome Y, whereas chromosome 7 was again shown to harbor most of the CNVs. Details on the frequency distribution over all chromosomes are shown in Figure IV-7 and Table IV-2.

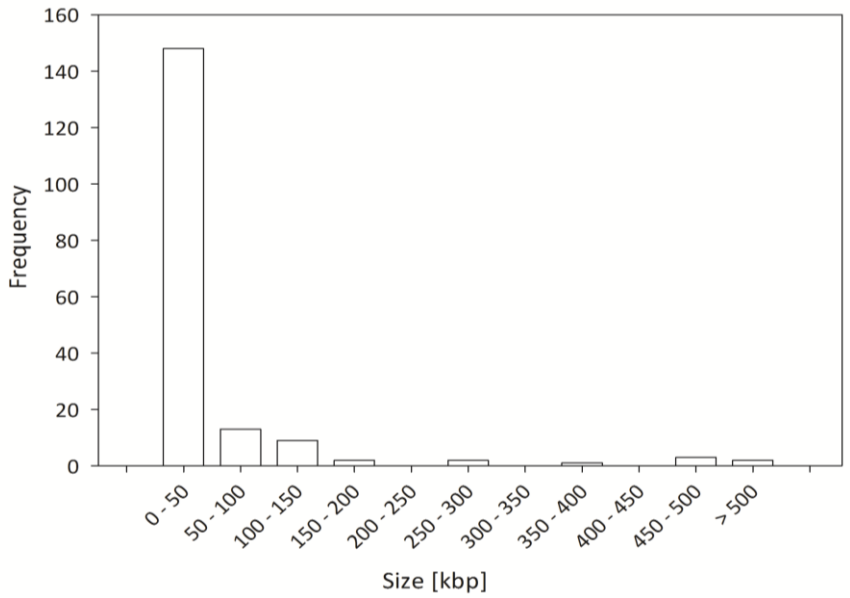

Figure IV-4: Frequency distribution of CNVs detected by JaxMDGA in HAB *vs.* LAB mice. The size classes are depicted on the xaxis and range from 50 kbp to 0.5 Mbp with a step size of 50 kbp. The y-axis shows the frequency of CNVs in the respective size class.

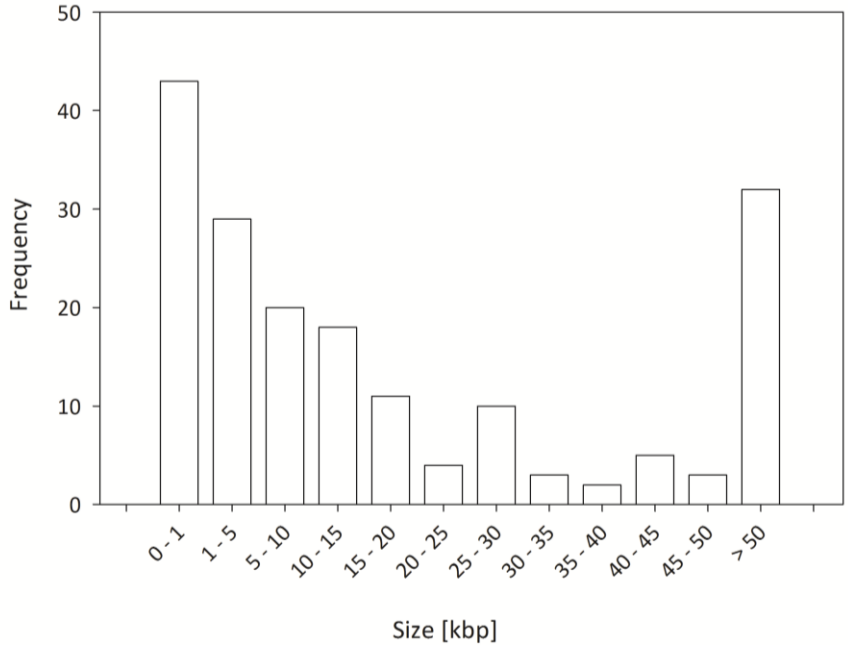

Figure IV-5: Detailed frequency distribution of CNVs < 50 kbp, detected by JaxMDGA in HAB *vs.* LAB mice. On the x-axis all bars but the first (1 kbp) depict a size range of 5 kbp. The last bar refers to all CNVs > 50 kbp. On the y-axis the frequency of CNVs in the respective size class is shown.

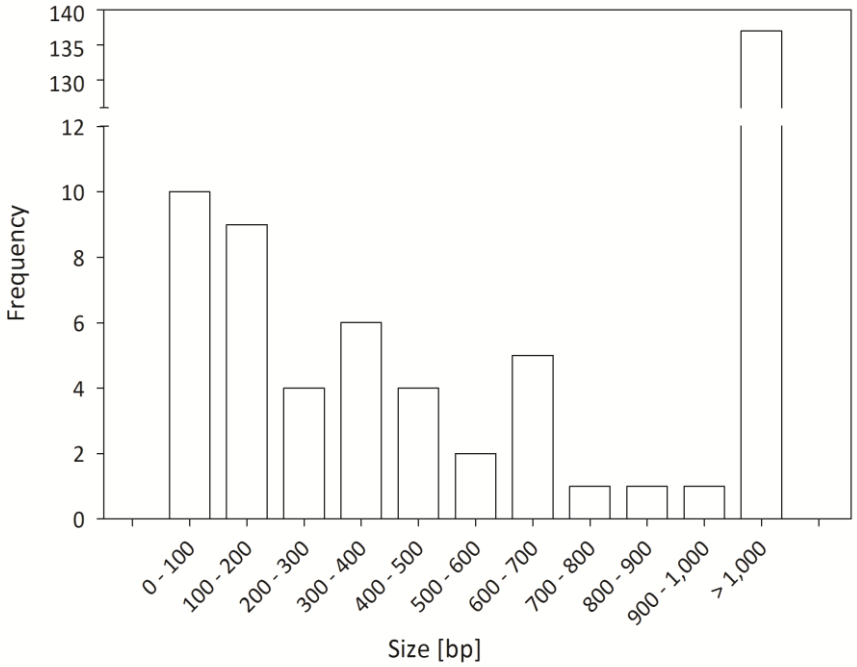

Figure IV-6: Frequency distribution of CNVs < 1 kbp, by JaxMDGA detected in HAB *vs.* LAB mice. On the x-axis the 43 CNVs of size < 1 kbp are depicted in size ranges of 100 bp. All other CNVs (137) are combined and referred to in the last bar. The y-axis shows the frequency of CNVs in the respective size class.

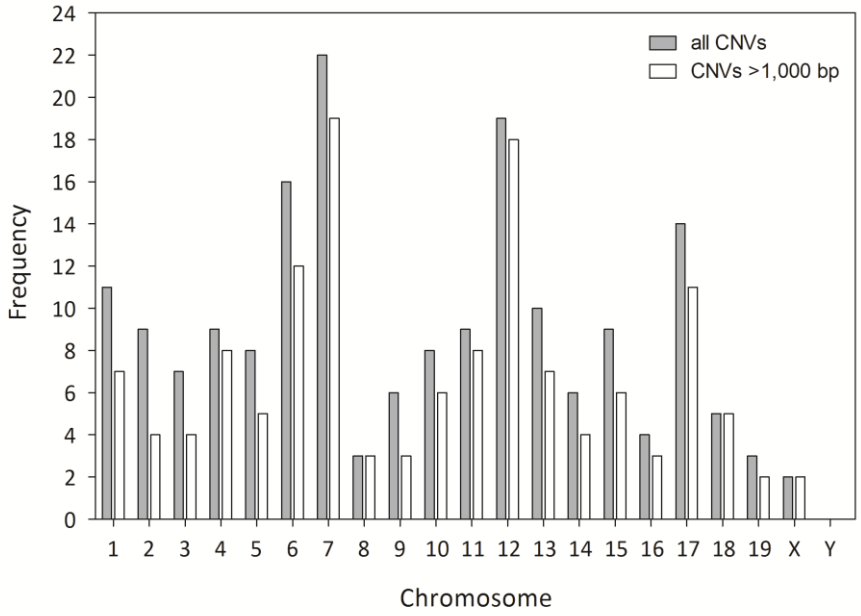

Figure IV-7: Chromosomal distribution of CNVs detected by JaxMDGA in HAB *vs.* LAB mice. For each chromosome the number of detected CNVs is depicted. Grey columns refer to the full set of revealed CNVs (180). White columns show only data of variants larger than 1 kbp.

Table IV-2: Detailed information on CNVs detected by JaxMDGA in HAB *vs.* LAB mice. For each chromosome (column 1) the median and mean size (columns 5 & 6) of all CNVs (number shown in column 3), as well as their maximum and minimum size (columns 7 & 8), are listed. The second column shows the set of CNVs to which the just mentioned parameters refer: "A" for the full set of 180 detected CNVs and "B" for only those 137 CNVs larger than 1 kbp.

Chr	Set	No. of CNVs	Sum [bp]	Median [bp]	Mean [bp]	Max [bp]	Min [bp]
1	A	11	294,171	7,448	26,743	106,453	2
1	B	7	293,389	27,743	41,913	106,453	3,594
2	A	9	59,780	706	6,642	33,825	97
2	B	4	58,062	9,831	14,516	33,825	4,575
3	A	7	522,573	7,641	74,653	479,350	483
3	B	4	520,866	16,938	130,217	479,350	7,641
4	A	9	1,791,498	15,427	199,055	1,116,514	333
4	B	8	1,791,165	21,034	223,896	1,116,514	2,288
5	A	8	501,157	11,531	62,645	286,323	52
5	B	5	500,334	49,355	100,067	286,323	11,398

Chr	Set	No. of CNVs	Sum [bp]	Median [bp]	Mean [bp]	Max [bp]	Min [bp]
6	A	16	574,612	12,960	35,913	156,205	19
6	B	12	573,518	28,712	47,793	156,205	3,224
7	A	22	8,003,249	12,242	363,784	7,383,756	145
7	B	19	8,002,512	13,324	421,185	7,383,756	1,381
8	A	3	63,564	16,848	21,188	45,480	1,236
8	B	3	63,564	16,848	21,188	45,480	1,236
9	A	6	9,849	1,241	1,642	3,412	651
9	B	3	7,546	2,617	2,515	3,412	1,517
10	A	8	169,490	4,612	21,186	91,918	111
10	B	6	169,191	7,262	28,199	91,918	1,391
11	A	9	118,816	16,140	13,202	29,831	203
11	B	8	118,613	16,566	14,827	29,831	1,195
12	A	19	1,130,124	36,415	59,480	358,926	80
12	B	18	1,130,044	38,416	62,780	358,926	2,097
13	A	10	173,208	12,190	17,321	67,023	164
13	B	7	171,628	25,909	24,518	67,023	1,696
14	A	6	58,511	10,528	9,752	22,496	243
14	B	4	57,823	14,063	14,456	22,496	7,201
15	A	9	130,189	7,311	14,465	76,451	29
15	B	6	129,652	12,543	21,609	76,451	1,700
16	A	4	116,773	4,936	29,193	106,576	325
16	B	3	116,448	5,919	38,816	106,576	3,953
17	A	14	826,845	8,820	59,060	465,283	74
17	B	11	826,538	9,628	75,140	465,283	1,661
18	A	5	108,886	13,141	21,777	59,779	6,131
18	B	5	108,886	13,141	21,777	59,779	6,131
19	A	3	27,592	1,308	9,197	26,255	29
19	B	2	27,563	13,782	13,782	26,255	1,308
X	A	2	50,757	25,379	25,379	33,025	17,732
X	B	2	50,757	25,379	25,379	33,025	17,732

1.3 Position of CNVs and structural variants (SVs) detected by NGS

A total of 5,851 CNVs was detected by NGS (supplementary table S3), with 3,489 regions showing a loss and 2,362 a gain of copy number in HAB mice. The CNVs ranged in size from 1,400 bp to 421,400 bp with a mean and a median of 4,654 bp and 2,400 bp, respectively, covering about 1.03% (27,233,200 bp) of the genome. The smallest distance between two variants was shown to be 200 bp, the largest 47,829,800 bp, with a mean and a median of 420,251 bp and 49,200 bp, respectively. Only 43 CNVs (0.73 %) were larger than 50 kbp, 4,862 (83.10 %) were even smaller than 5 kbp. The frequency distribution of all CNVs is shown in Figure IV-8 with size classes of 50 kbp. Figure IV-9 and Figure IV-10 illustrate the frequency distributions of CNVs smaller than 50 kbp and 5 kbp, with a step size of 5 kbp and 500 bp, respectively.

The CNVs are distributed all over the genome. In concordance to the other detection methods, most were found on chromosome 7 (646 CNVs) and least on chromosome Y (5 CNVs). Details on the chromosomal distribution of CNVs are shown in Figure IV-11 and Table IV-3.

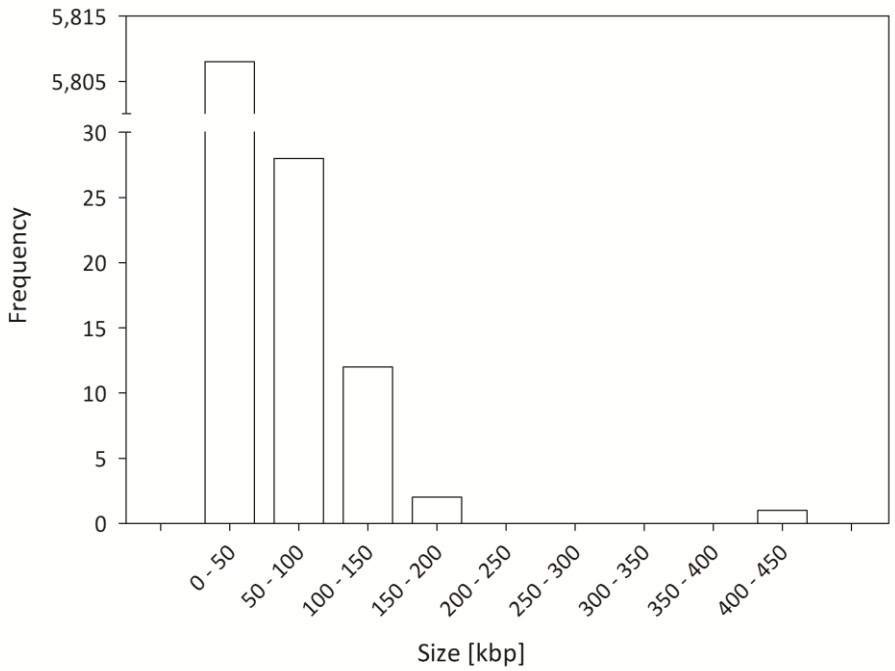

Figure IV-8: Frequency distribution of CNVs detected by NGS in HAB *vs.* LAB mice, in size classes of 50 kbp. The size classes are plotted against the number of CNVs in the respective range, with a step size of 50 kbp.

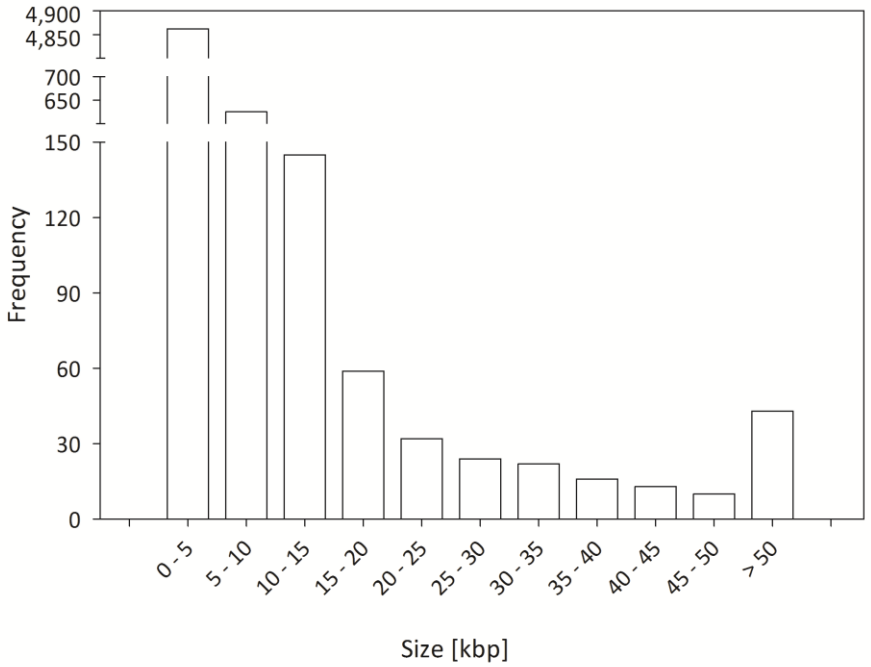

Figure IV-9: Detailed frequency distribution of CNVs < 50 kbp, detected by NGS in HAB *vs.* LAB mice. CNVs of size < 50 kbp are depicted in size ranges of 5 kbp. The last bar refers to all CNVs larger than 50 kbp.

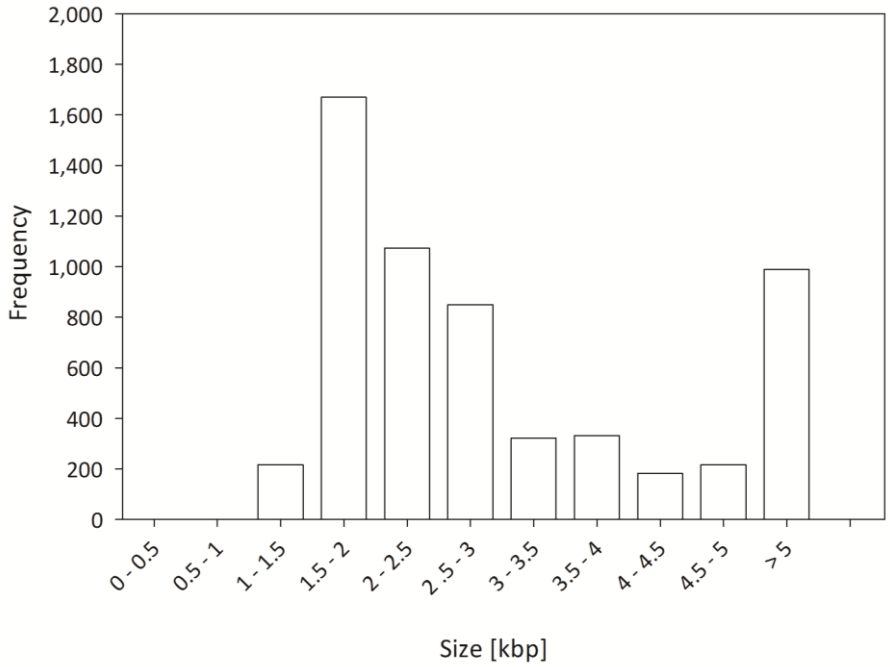

Figure IV-10: Detailed frequency distribution of CNVs < 5 kpb, detected by NGS in HAB *vs.* LAB mice. CNVs of size < 5 kbp are depicted in size ranges of 500 bp on the x-axis. The last bar represents all CNVs larger than 5 kbp. The frequency of CNVs in the respective size class is shown on the y-axis.

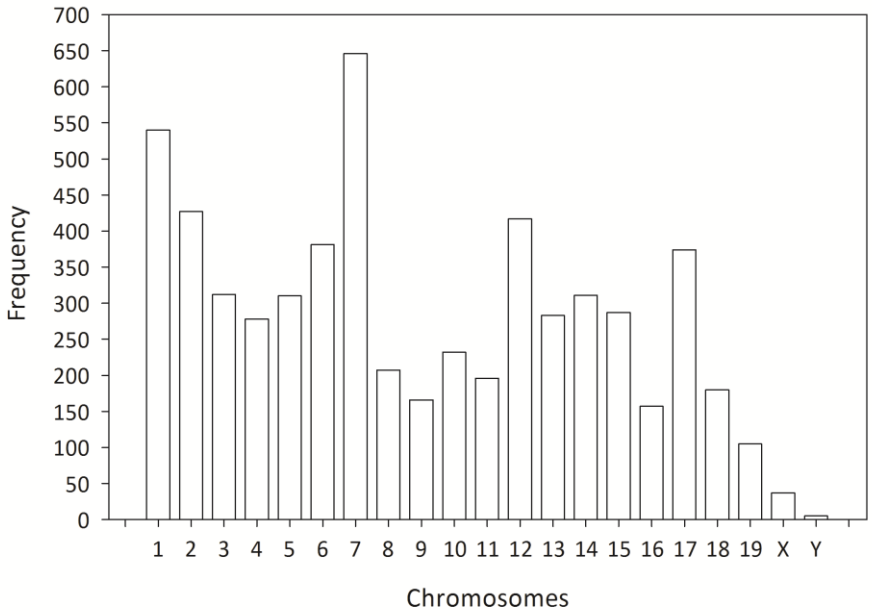

Figure IV-11: Frequency distribution of CNVs detected by NGS in HAB *vs.* LAB mice over all chromosomes. Most CNVs were found on chromosome 7, least on chromosome Y, with the frequencies plotted against the chromosomes.

Table IV-3: Detailed information on CNVs detected by NGS in HAB *vs.* LAB mice. The first column shows the chromosome. The other columns comprise the following information (from left to right): number of CNVs, size of the CNVs summed up, median and mean size of the CNVs, and the size of largest and smallest CNV, respectively.

Chr	No. of CNVs	Sum [bp]	Median [bp]	Mean [bp]	Max [bp]	Min [bp]
1	540	1,969,200	2,400	3,647	69,200	1,400
2	427	1,296,400	2,400	3,036	49,800	1,400
3	312	1,133,200	2,400	3,632	49,400	1,400
4	278	2,599,600	3,000	9,351	166,200	1,400
5	310	2,207,000	2,800	7,119	180,200	1,400
6	381	1,482,600	2,400	3,891	147,800	1,400
7	646	4,986,000	3,600	7,718	142,800	1,400
8	207	603,200	2,200	2,914	18,400	1,400
9	166	486,400	2,200	2,930	34,000	1,400
10	232	990,400	2,600	4,269	76,000	1,400
11	196	751,600	2,300	3,835	73,200	1,400
12	417	2,369,200	2,800	5,682	101,800	1,400

Chr	No. of CNVs	Sum [bp]	Median [bp]	Mean [bp]	Max [bp]	Min [bp]
13	283	843,000	2,400	2,979	14,400	1,400
14	311	939,800	2,400	3,022	22,200	1,400
15	287	824,000	2,400	2,871	9,600	1,400
16	157	493,000	2,400	3,140	28,400	1,400
17	374	2,267,200	2,600	6,062	421,400	1,400
18	180	469,600	2,200	2,609	8,200	1,400
19	105	284,600	2,200	2,710	8,800	1,400
X	37	225,400	3,800	6,092	55,200	1,600
Y	5	11,800	2,200	2,360	3,800	1,600

Besides the CNVs, 8,035 large deletions were detected in HAB and 7,208 in LAB mice, of which 6,451 were found to overlap at least partially between the two mouse lines. About 80% of the deletions did show an overlap of more than 80% (see figure 1.2 of supplementary report S1). There were 1,621 deletions only detected in HAB and 782 only in LAB. A list of all deletions can be found in the supplementary table S4. Furthermore, 236 SVs causing wrong mate order were found in HAB, 190 in LAB, with 140 overlapping between the lines. All SVs of that kind are listed in supplementary table S5. In addition to that, 81 insertions and 169 inversions were revealed in HAB, 36 and 118, respectively in LAB. Thereof, 82 inversions were unique in HAB and 31 in LAB. 33 of the insertions were overlapping between both lines. The respective data are outlined in supplementary table S6 (insertions) and table S7 (inversions).

1.4 Comparison of CNVs detected by aCGH, JaxMDGA and NGS

When comparing CNV data acquired by aCGH, JaxMDGA and NGS, a total of 4.8 Mbp were found to overlap between the data sets. This equates to approximately 5 %, 33 % and 18 % of the total length of CNVs detected by aCGH, JaxMDGA and NGS, respectively. Several CNVs did overlap, at least partially, in all three methods. Details on the comparison are shown in Table IV-4. Since it is possible that CNVs detected by one

method overlapped with just a single CNV defined by another method, the number of CNVs is shown with respect to just one detection method, e.g., 33 aCGH-CNVs overlapped with CNVs defined by both other methods. The positions of overlapping CNVs are outlined in detail in supplementary table S8 and depicted in Figure IV-12 to Figure IV-16. Some of the regions were shown to overlap in nearly full length between the different data sets, while others outlined the same region but were defined as one or multiple CNVs by different methods. Since, due to limitations in resolution, that effect might be difficult to recognize when displaying the full length of the chromosomes, parts of chromosome 17 were enlarged exemplarily (Figure IV-17). The first of the enlarged regions depicts the CNV with the *Glo1* locus, further investigated by the breeding approach described in chapter 1.6.

Table IV-4: Comparison of CNVs detected in HAB *vs.* LAB mice by three different methods. Upper part: numbers of respective CNVs. Part below: sizes of respective CNVs in basepairs. Line "overlap both others" shows the number of CNVs defined by the respective detection method that overlap with any other CNV detected by the other two methods.

	aCGH	**JaxMDGA**	**NGS**
No. of CNVs	98	180	5,851
Overlap aCGH (No. CNVs)	-	76	1,821
Overlap JaxMDGA (No. CNVs)	42	-	293
Overlap NGS (No. CNVs)	97	86	-
Overlap both others (No. CNVs)	33	53	220
Total size of CNVs [bp]	97,264,699	14,731,644	27,233,200
Mean size of CNVs [bp]	992,497	81,842	4,654
Median size of CNVs [bp]	544,637	8,906	2,400
Overlap in all methods [bp]		***4,824,789***	

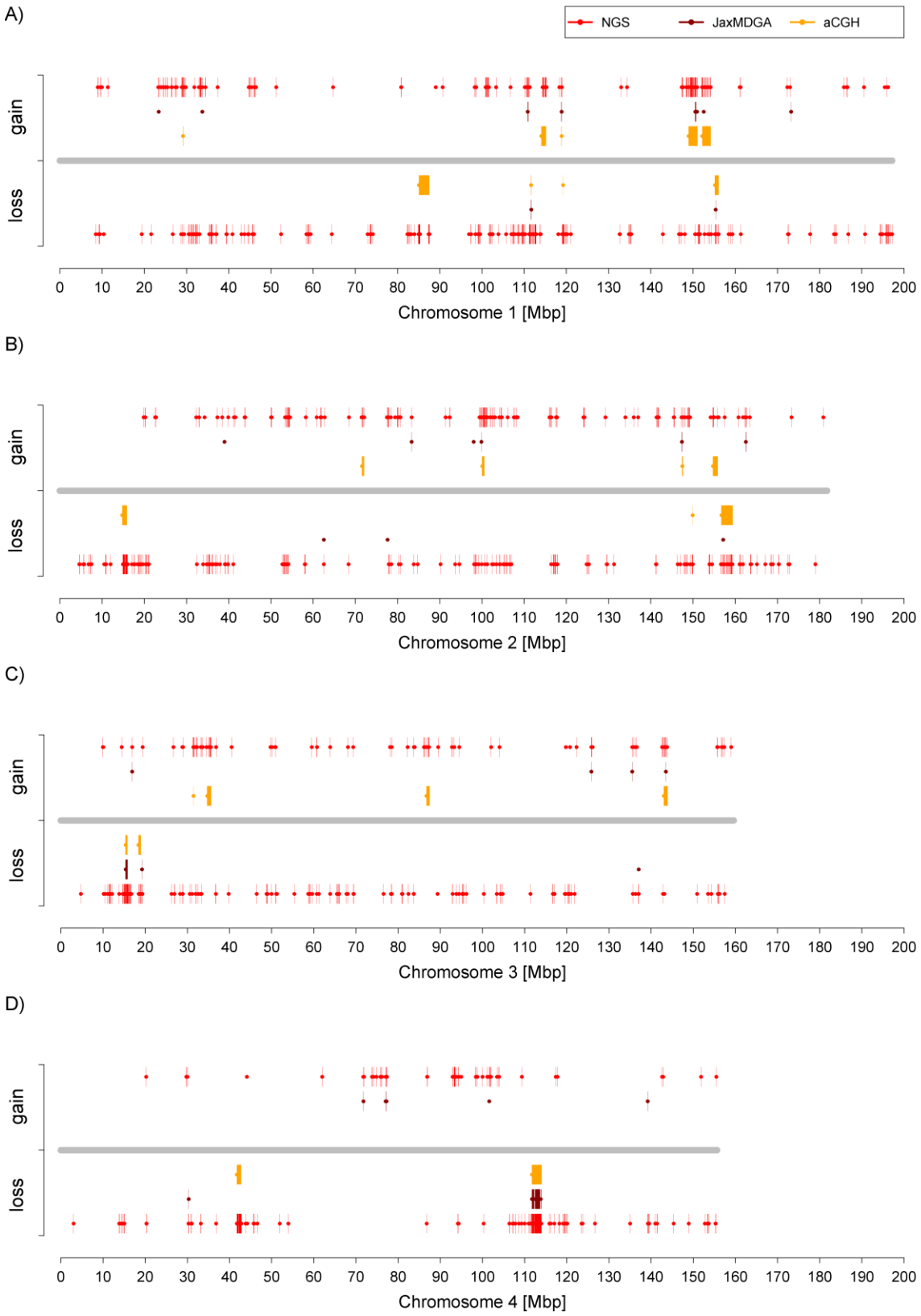

Figure IV-12 (A) – (D): Genomic positions of CNVs on chromosomes 1 – 4, detected in HAB *vs.* LAB mice. The chromosomes are indicated by grey lines. CNVs are shown in different colors depending on their detection method with red for NGS, dark red for JaxMDGA and orange for aCGH. Start points of CNVs are marked by dots and lines are drawn till the end points. Due to limitations in resolution small CNVs might appear as dots only. Lines and dots drawn above the chromosome indicate a copy number gain in HAB mice, while those below depict a copy number loss.

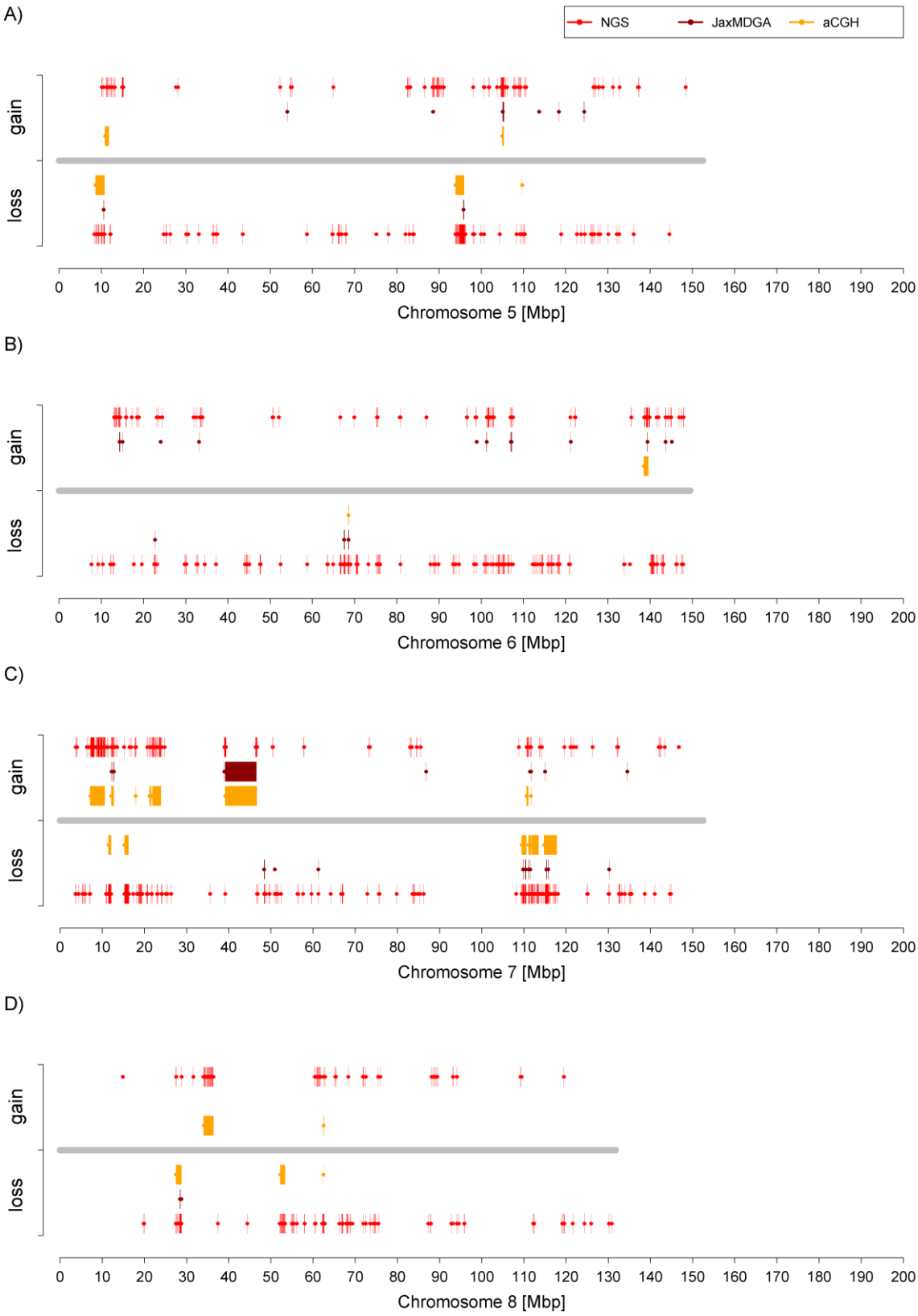

Figure IV-13 (A) – (D): Genomic positions of CNVs on chromosomes 5 – 8, detected in HAB *vs.* LAB mice. The chromosomes are indicated by grey lines. CNVs are shown in different colors depending on their detection method with red for NGS, dark red for JaxMDGA and orange for aCGH. Start points of CNVs are marked by dots and lines are drawn till the end points. Due to limitations in resolution small CNVs might appear as dots only. Lines and dots drawn above the chromosome indicate a copy number gain in HAB mice, while those below depict a copy number loss.

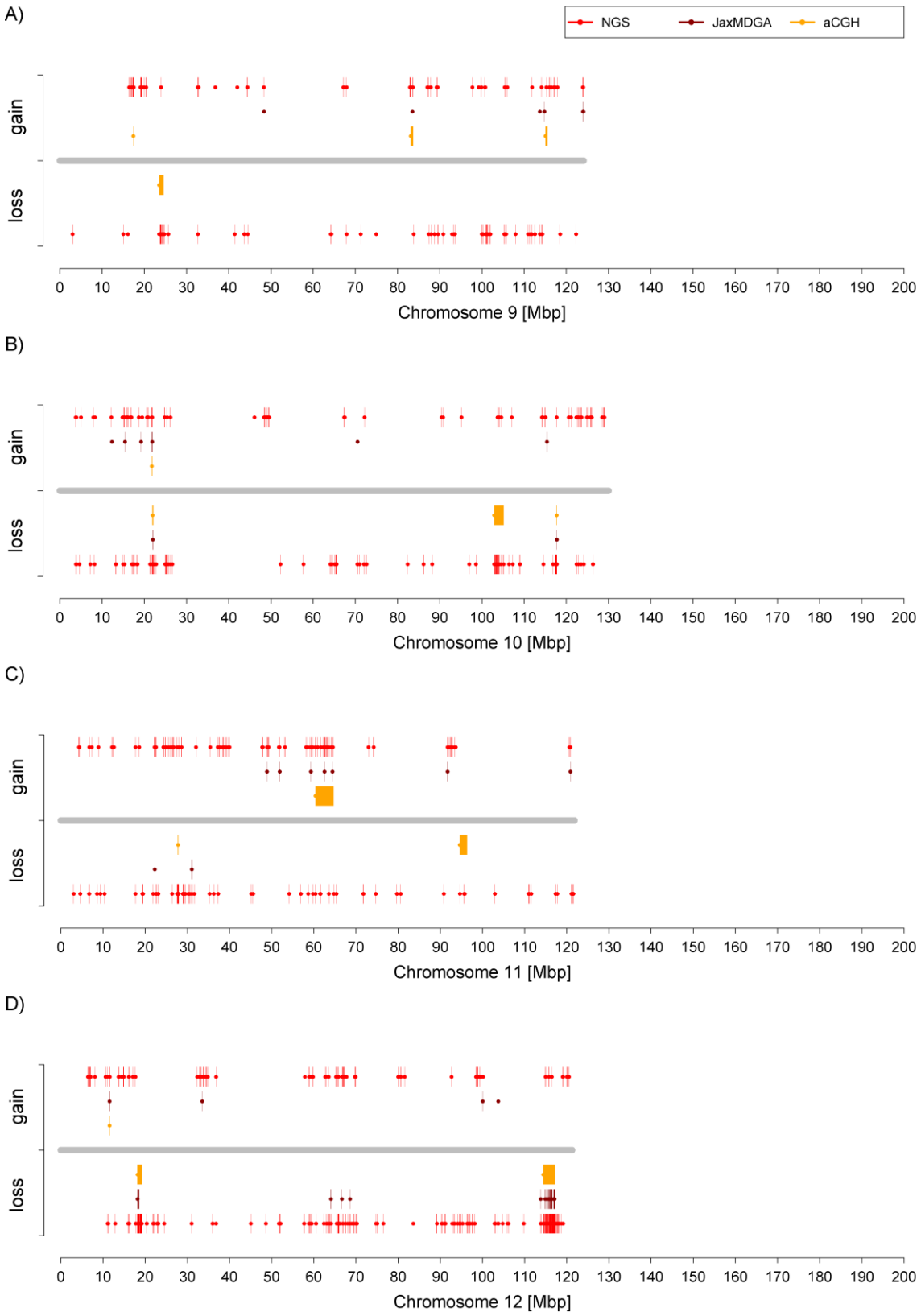

Figure IV-14 (A) – (D): Genomic positions of CNVs on chromosomes 9 – 12, detected in HAB *vs.* LAB mice. The chromosomes are indicated by grey lines. CNVs are shown in different colors depending on their detection method with red for NGS, dark red for JaxMDGA and orange for aCGH. Start points of CNVs are marked by dots and lines are drawn till the end points. Due to limitations in resolution small CNVs might appear as dots only. Lines and dots drawn above the chromosome indicate a copy number gain in HAB mice, while those below depict a copy number loss.

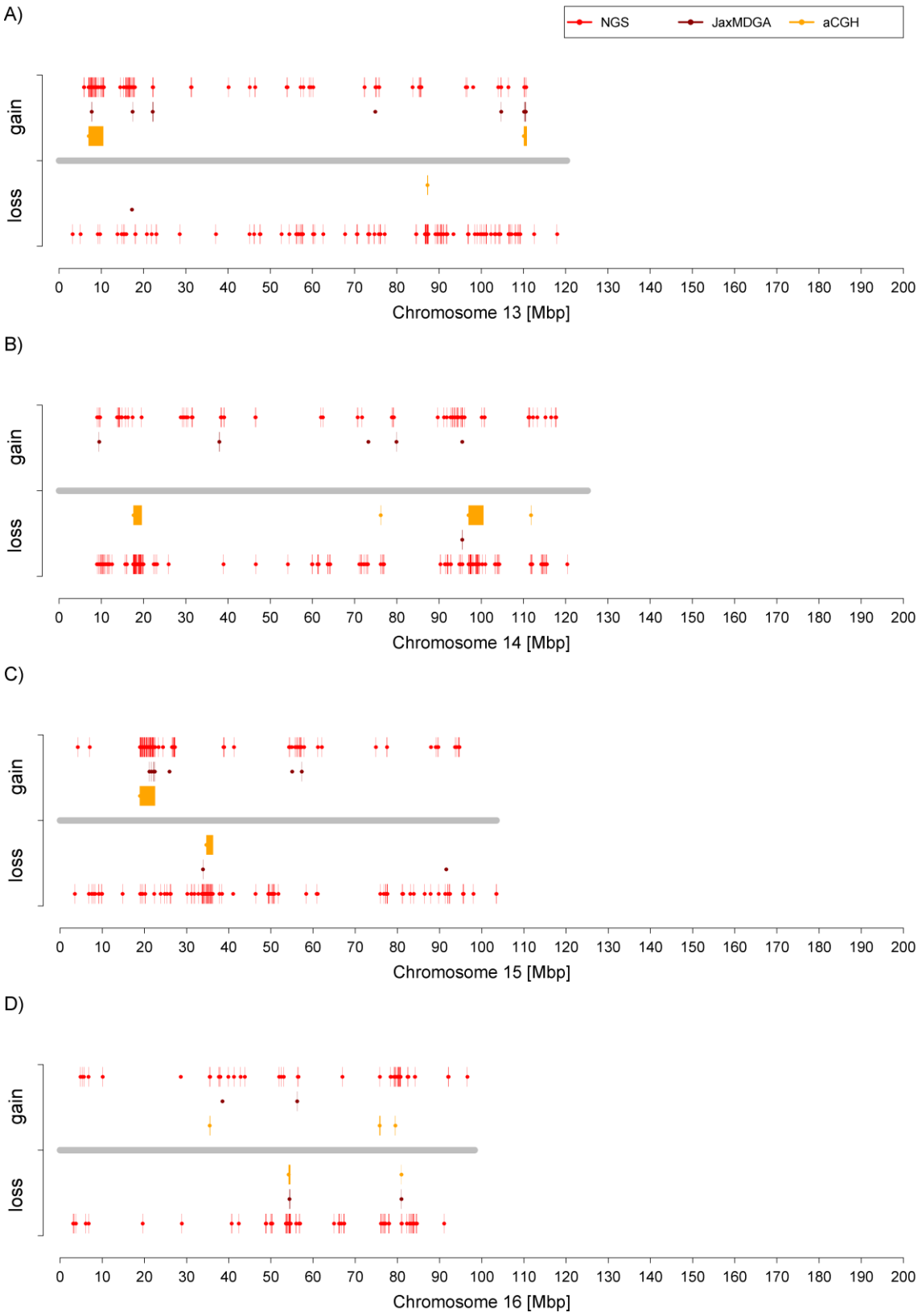

Figure IV-15 (A) – (D): Genomic positions of CNVs on chromosomes 13 – 16, detected in HAB *vs.* LAB mice. The chromosomes are indicated by grey lines. CNVs are shown in different colors depending on their detection method with red for NGS, dark red for JaxMDGA and orange for aCGH. Start points of CNVs are marked by dots and lines are drawn till the end points. Due to limitations in resolution small CNVs might appear as dots only. Lines and dots drawn above the chromosome indicate a copy number gain in HAB mice, while those below depict a copy number loss.

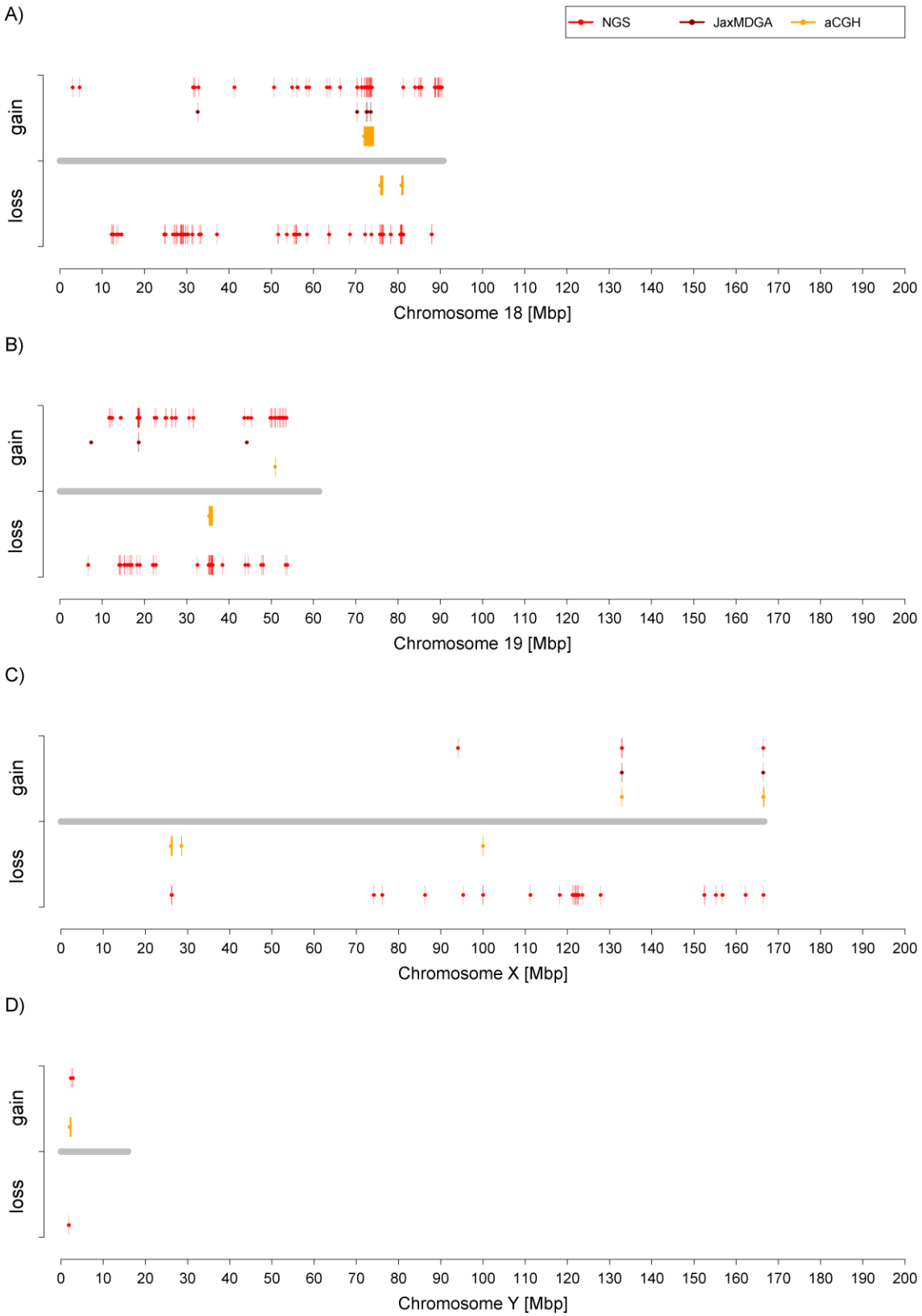

Figure IV-16 (A) – (D): Genomic positions of CNVs on chromosomes 18, 19, X & Y, detected in HAB *vs.* LAB mice. The chromosomes are indicated by grey lines. CNVs are shown in different colors depending on their detection method with red for NGS, dark red for JaxMDGA and orange for aCGH. Start points of CNVs are marked by dots and lines are drawn till the end points. Due to limitations in resolution small CNVs might appear as dots only. Lines and dots drawn above the chromosome indicate a copy number gain in HAB mice, while those below depict a copy number loss.

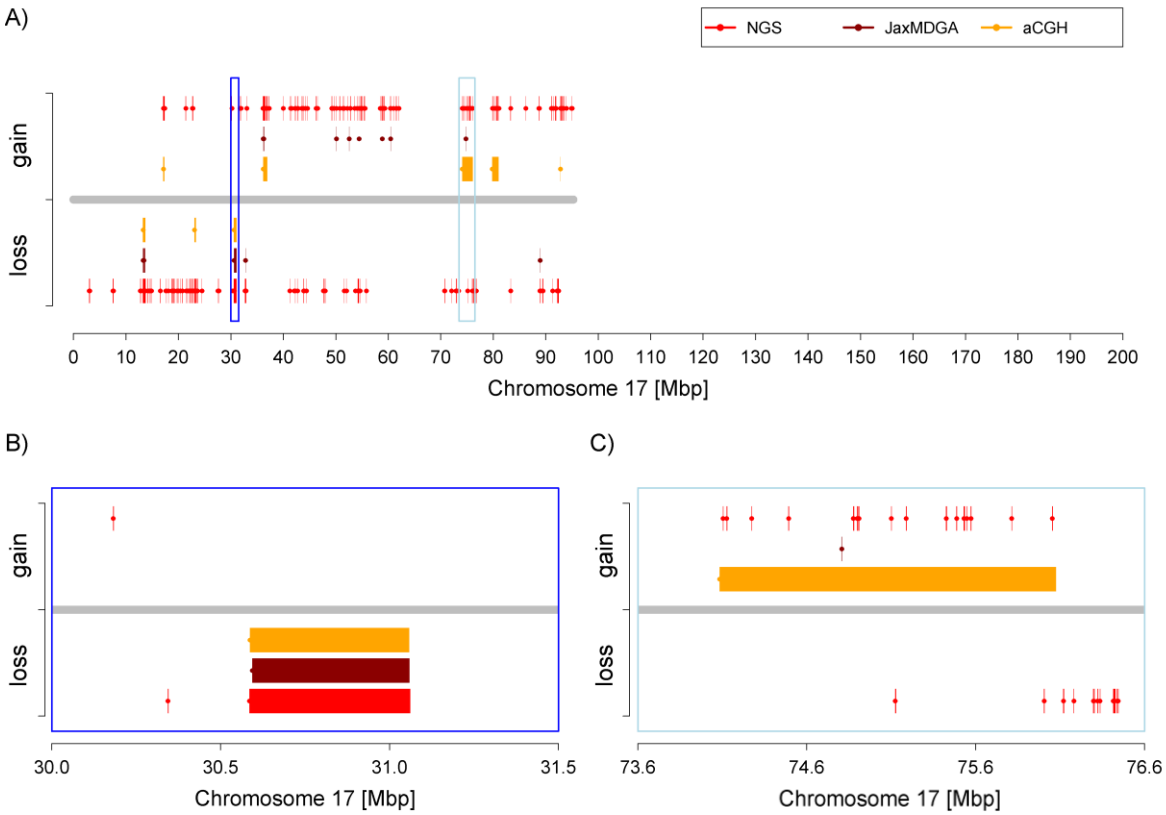

Figure IV-17: Genomic position of CNVs on chromosome 17, detected in HAB *vs.* LAB mice. The chromosome is indicated by a grey line. CNVs are shown in different colors depending on their detection method with red for NGS, dark red for JaxMDGA and orange for aCGH. Start points of CNVs are marked by dots and lines are drawn till the end points. Due to limitations in resolution small CNVs might appear as dots only. Lines and dots drawn above the chromosome indicate a copy number gain in HAB mice, while those below depict a copy number loss. Parts of the full-length chromosome (A) are enlarged below, highlighted by a dark and light blue square. Enlarged pictures show (B) a CNV including the *Glo1* locus with breakpoints defined similarly by all detection methods and (C) a region defined as single and multiple CNV(s), respectively with unequal breakpoints.

1.5 Expression versus CNV data

1.5.1 Differentially expressed genes detected by expression microarray

Out of 12,171 probes that passed the filtering process, the reappraisal of the expression microarray data showed 291 in the CeA, 117 in the BLA, 297 in the PVN and 254 in the Cg, respectively, to be differentially expressed between HAB and LAB mice. These probes represented a total of 8,981 protein coding genes, with 374 differentially expressed in at least one of the tested brain regions, of which 256 were detected in the CeA, 100 in the BLA, 260 in the PVN and 222 in the Cg. All 374 genes are listed in supplementary table S9, including information on how many probes representing them were found to feature

differences in expression levels, thus a p-value less than 0.05, and in which regions that difference occurred. Table IV-5 shows the p-values and the relative fold-change for all array probes of the 12 genes validated by qPCR.

Table IV-5: Expression differences of genes analyzed later on in qPCR, detected by expression microarray. A p-value less than 0.05 shows significant expression difference between HAB and LAB mice (written in bold letters), a p-value not indicating significance but being less than 0.1 indicates a trend (italic and bold letters). A negative fold-change indicates less, a positive more expression in HAB mice. If multiple probes target a gene, the result of each probe is shown separately in a distinct line. The genes that were targeted by more than one probe are marked by the probe number written in brackets behind the gene symbol.

Gene symbol	CeA pVal	CeA foldCh	BLA pVal	BLA foldCh	PVN pVal	PVN foldCh	Cg pVal	Cg foldCh
Alk	**0.0048**	-1.5218	0.9999	-1.2246	**0.0065**	-1.5067	1.0000	-1.1637
Epn2 (1)	1.0000	-1.0384	0.9999	1.0045	1.0000	1.0141	1.0000	1.0742
Epn2 (2)	0.6024	1.2571	0.9999	1.0822	1.0000	1.1726	1.0000	1.1457
Epn2 (3)	**0.0494**	1.7020	0.9999	1.2465	0.4560	1.4442	**0.0027**	2.0011
Epn2 (4)	1.0000	1.1404	0.9999	-1.0456	1.0000	1.0006	1.0000	1.0308
Epn2 (5)	1.0000	1.0689	0.9999	1.0131	1.0000	-1.0117	1.0000	1.0390
Fgfr2	1.0000	1.0801	0.9999	1.2349	1.0000	1.1086	1.0000	1.0080
Ghrh	**0.0000**	-2.3654	0.9999	-1.1687	**0.0000**	-9.8041	1.0000	1.0287
Glb1 (1)	0.9186	-1.1698	0.9999	-1.1616	**0.0012**	-1.5986	1.0000	-1.1765
Glb1 (2)	1.0000	1.0972	0.9999	-1.0182	1.0000	1.0730	1.0000	1.1376
Glo1	**0.0011**	-2.7181	***0.0792***	-2.2790	**0.0007**	-2.8015	**0.0017**	-2.6572
Gnaq (1)	**0.0000**	4.6921	**0.0006**	3.2946	**0.0000**	3.2645	**0.0000**	5.6461
Gnaq (2)	0.7608	1.2314	0.9999	1.0533	1.0000	-1.0328	1.0000	1.0024
Gnaq (3)	1.0000	-1.0290	0.9999	1.0662	1.0000	-1.0082	1.0000	-1.0542
Pdk2	**0.0001**	-1.9379	**0.0480**	-1.6630	**0.0000**	-2.1816	**0.0231**	-1.5839
Prkcdbp	0.1844	-1.4126	0.9999	-1.1686	**0.0157**	-1.6087	1.0000	-1.0019
Rgs16	1.0000	1.1484	0.9999	1.2215	1.0000	1.0822	**0.0267**	-1.6101
Rhoj	1.0000	-1.0978	0.9999	-1.0952	**0.0249**	-1.3600	1.0000	-1.0691
Slc30a6 (1)	1.0000	1.0182	0.9999	-1.2367	1.0000	-1.1919	1.0000	-1.1070
Slc30a6 (2)	**0.0000**	2.1748	**0.0019**	1.8883	**0.0000**	2.7410	**0.0000**	1.9532

1.5.2 Confirmation of differentially expressed genes by qPCR

For three of the twelve genes tested in qPCR, *Fgfr2, Ghrh* and *Prkcdbp*, no reliable results could be obtained by qPCR. Therefore, these genes were excluded from the statistical analysis. The qPCR results of the remaining nine genes are listed in Table IV-6. Comparing that list with Table IV-5 showed that most of the microarray results could be confirmed, as indicated by a (*) following the p-value in Table IV-6.

Table IV-6: Expression differences of genes successfully tested in qPCR. The first part shows relative expression rate with standard error (SEM) and p-value (calculated by Mann-Whitney-U test) for CeA and BLA, the second part for PVN and Cg. A p-value < 0.05 (bold letters) indicates a significant difference between HAB and LAB gene expression, a p-value < 0.1 (bold and italic letters) indicates a trend. * = confirmed microarray result; (*) = confirmed result of at least one microarray probe; *T = eventually confirmed microarray result, with one of the methods showing a trend.

Gene symbol	CeA relEx HAB +- SEM	CeA relEx LAB +- SEM	CeA pVal	BLA relEx HAB +- SEM	BLA relEx LAB +- SEM	BLA pVal
Alk	1.0 ± 0.09	0.91 ± 0.08	0.6203	1.0 ± 0.12	1.24 ± 0.17	0.3637*
Epn2	1.0 ± 0.07	0.72 ± 0.08	0.0132(*)	1.0 ± 0.06	0.88 ± 0.03	0.1864*
Glb1	1.0 ± 0.27	0.89 ± 0.12	0.4574*	1.0 ± 0.14	1.78 ± 0.22	0.0064
Glo1	1.0 ± 0.06	2.79 ± 0.20	0.0003*	1.0 ± 0.06	3.07 ± 0.22	0.0003*T
Gnaq	1.0 ± 0.08	0.95 ± 0.14	0.409(*)	1.0 ± 0.08	1.12 ± 0.22	0.8688(*)
Pdk2	1.0 ± 0.09	0.78 ± 0.03	0.0318	1.0 ± 0.12	0.93 ± 0.12	0.6203
Rgs16	1.0 ± 0.16	0.84 ± 0.15	0.409*	1.0 ± 0.09	0.85 ± 0.15	0.1864*
Rhoj	1.0 ± 0.10	1.18 ± 0.09	0.1604*	1.0 ± 0.09	1.31 ± 0.14	0.0475
Slc30a6	1.0 ± 0.18	0.64 ± 0.06	0.0575(*T)	1.0 ± 0.12	1.16 ± 0.10	0.2831(*)
	PVN relEx HAB +- SEM	**PVN relEx LAB +- SEM**	**PVN pVal**	**Cg relEx HAB +- SEM**	**Cg relEx LAB +- SEM**	**Cg pVal**
Alk	1.0 ± 0.14	1.66 ± 0.81	0.7412	1.0 ± 0.10	1.31 ± 0.09	0.0132
Epn2	1.0 ± 0.07	0.89 ± 0.20	0.1604*	1.0 ± 0.12	0.92 ± 0.07	0.8590(*)
Glb1	1.0 ± 0.07	1.94 ± 0.42	0.0166(*)	1.0 ± 0.11	1.23 ± 0.14	0.1167*
Glo1	1.0 ± 0.11	3.25 ± 0.36	0.0003*	1.0 ± 0.15	2.65 ± 0.25	0.0004*
Gnaq	1.0 ± 0.21	1.51 ± 0.27	0.1167(*)	1.0 ± 0.16	0.45 ± 0.04	0.0022(*)
Pdk2	1.0 ± 0.09	1.87 ± 1.04	0.5089	1.0 ± 0.17	0.91 ± 0.10	0.8688

Gene symbol	PVN relEx HAB +- SEM	PVN relEx LAB +- SEM	PVN pVal	Cg relEx HAB +- SEM	Cg relEx LAB +- SEM	Cg pVal
Rgs16	1.0 ± 0.12	1.08 ± 0.36	0.3218*	1.0 ± 0.09	0.85 ± 0.12	0.1604
Rhoj	1.0 ± 0.11	1.74 ± 0.61	0.039*	1.0 ± 0.13	0.90 ± 0.13	0.2477*
Slc30a6	1.0 ± 0.27	1.27 ± 0.54	0.6797(*)	1.0 ± 0.19	0.95 ± 0.10	0.8688(*)

1.5.3 Correlation of differentially expressed genes with CNV data

In total 8,981 protein coding genes were tested for expression differences in the microarray-based approach, of which 532 were located in regions where a CNV was shown either by aCGH, JaxMDGA or NGS (for details on CNV positions see chapter 1). The counting of genes that were differentially or equally expressed between HAB and LAB and those that fell into regions where a CNV was shown resulted in the numbers outlined in Table IV-7. Applying a two-proportion z-test on the data revealed a significant influence of CNVs on expression in all the tested brain regions, with a p-value of 1.17×10^{-69} for the comparison of the CNV data with the expression data including all genes that showed differences in at least one brain region. Considering each brain region separately resulted in p-values of 4.17×10^{-49} (CeA), 5.28×10^{-40} (BLA), 1.76×10^{-64} (PVN) and 3.33×10^{-54} (Cg).

Table IV-7: Count of genes that are differentially expressed between HAB and LAB mice and overlap with CNVs. The first column shows the respective brain region, the second column the count of genes that were found to be differentially expressed and the last column the number of genes that were not only found to show differences in expression level but also overlap with any CNV.

Brain region	Genes diff. expressed	Genes diff. ex. & in CNV
At least one	374	101
CeA	256	70
BLA	100	37
PVN	260	79
Cg	222	67

Applying the Pearson's product-moment correlation, a positive correlation between gene expression and the CNV status (i.e. a gain or loss in HAB *vs.* LAB animals) was shown for all four brain regions. Except for the BLA, the p-values did all reach significance, thus a value less than 0.05, before of the application of Bonferroni correction. Afterwards, however, the significance threshold was not reached any more. The respective values are listed in Table IV-8.

Table IV-8: Result of the correlation of CNVs with expression. Each brain is outlined in a separate line. Significant p-values (< 0.05) are marked with an asterisk (*). In the third column the nominal p-values are shown, in the last column the Bonferroni corrected p-values. A positive value in the second column indicates a positive correlation of copy number with expression.

Brain region	Correlation	p-value	p-value corrected
CeA	0.311	0.0021*	0.1998
BLA	0.208	0.1432	1.0000
PVN	0.309	0.0012*	0.1246
Cg	0.536	0.0057*	0.2831

1.6 Influence of the *Glo1* CNV on behavior

In order to test the influence of a single CNV that included the *Glo1* locus, the *Glo1* breeding approach was performed, mating animals showing the LAB-specific locus with HAB mice. Animals of the 6th generation were tested in different behavioral tests, the EPM, the OF, the TST and the FST. After determination of the phenotype, the animals were genotyped for the *Glo1* locus by PCR and subsequent gel electrophoresis. On the gel, a band indicating an existing product length of 150 bp (Figure IV-18) could only be seen, if more than one copy of the genomic region was present, thus exclusively in animals having the LAB-specific genotype. Out of 94 mice tested, 26 females and 30 males were identified to have that genotype, while 20 females and 18 males were shown to have the HAB-specific locus.

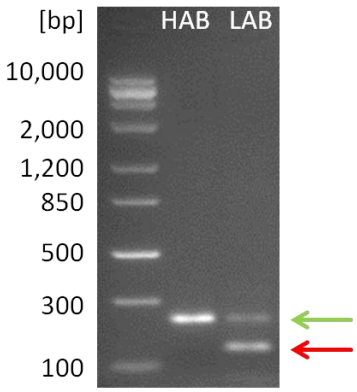

Figure IV-18: Picture of a gel showing both HAB- and LAB-specific genotypes. The control band of size 290 bp is indicated by a green arrow, the LAB-specific band (150 bp) by a red one. Marker: ZipRuler Express DNA Ladder 1.

The behavior of the animals was subsequently analyzed with respect to the genotype by discriminating a variety of behavioral parameters. The most important parameter was the time the mice spent on the open arm of the EPM (Figure IV-19A), since that is the key parameter of the original HAB/LAB breeding. For that behavioral aspect, no significant difference could be observed, neither when comparing all animals having the HAB-specific *Glo1* locus ("hab" animals) with all having the LAB-specific one ("lab" animals; $p = 0.939$), nor when these groups were analyzed for females ($p = 0.877$) and

males (p = 0.932) separately. The same was true for all parameters of other behavioral experiments analyzed, like the key parameters total distance animals travelled in the OF (p_{all} = 0.638, $p_{females}$ = 0.223, p_{males} = 0.717), the time animals spent floating in the FST (p_{all} = 0.338, $p_{females}$ = 0.278, p_{males} = 0.835) and the time animals were immobile during the TST (p_{all} = 0.864, $p_{females}$ = 0.629, p_{males} = 0.896), as depicted in Figure IV-19B-D.

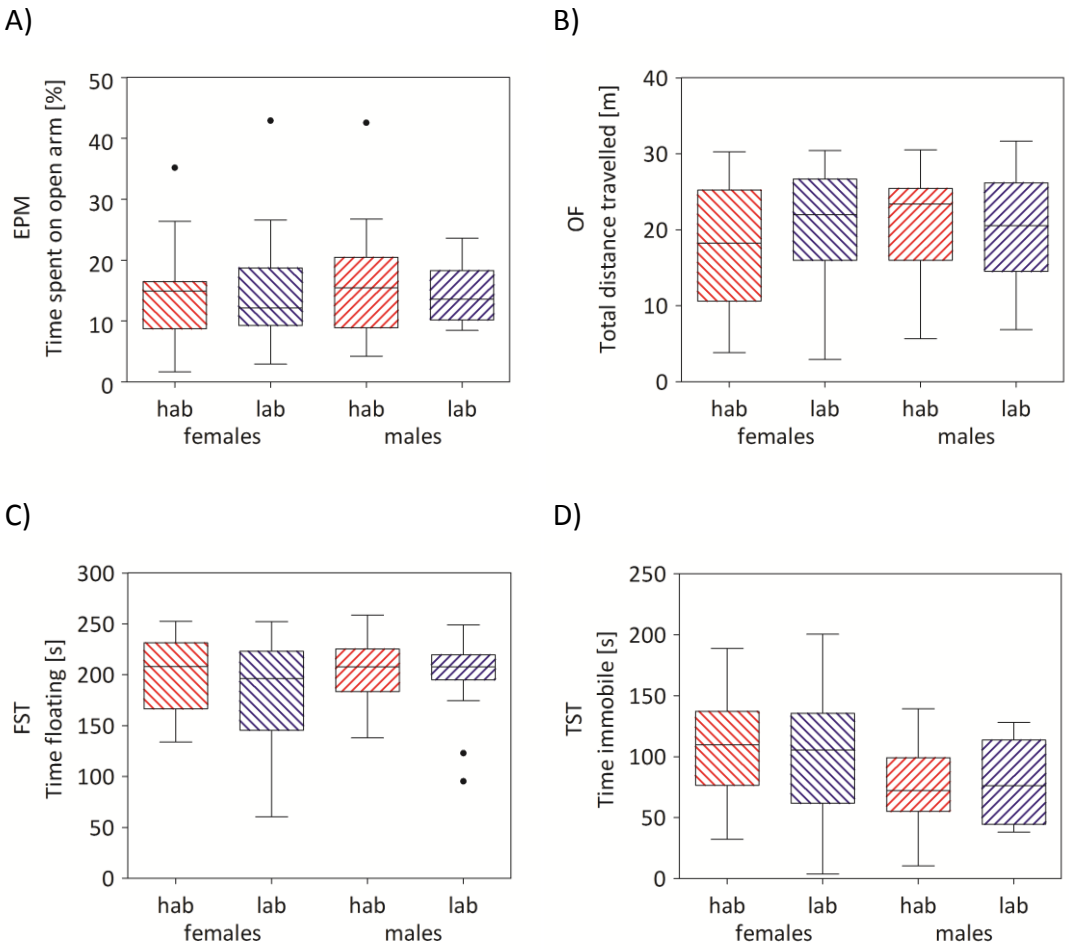

Figure IV-19: Results of behavioral testing in animals of the 6th *Glo1* breeding generation. Groups of "hab" and "lab" animals were compared to each other, gender-dependent and independent. No significant difference (p < 0.05) could be shown in any of the test parameters, neither in the time animals spent on the open arm of the EPM (A), in the total distance animals travelled in the OF (B), in the time animals floated during FST (C), nor in the time animals were immobile in the TST (D). Data of "hab" animals are shown in red (females N=26, males N = 30), data of "lab" in blue (females N = 20, males N = 18). Whiskers include values within a maximum of 1.5 interquartile range (IQR). Outliers (values not within 1.5 IQR) are indicated by dots.

2 Position and impact of CNVs in CD-1 mice

2.1 CNVs examined by MLPA

The measurement values of 33 probes targeting 11 CNVs previously detected in CD-1 mice by JaxMDGA were analyzed using the Coffalyser software. For all samples a warning was shown that DNA concentration in the MLPA run was too low. However, the results for all probes (1:100-dilution of MLPA probes) are outlined in supplementary table S10. The results of the 1:50-dilution were not shown, as they seemed less reliable than the measurement values of the 1:100-dilution. The data were further analyzed probe-wise for differences between the two groups of each 27 animals showing least and most anxiety-related behavior on the EPM (percent time on the open arm), respectively, by applying the Mann-Whitney-U test. Only one probe (no. 15) showed a significant difference between the two groups ($p = 0.039$) and two other (no. 21 and 14) a trend ($p = 0.077$ and 0.086). In Table IV-9, the result for the HAB sample is compared to the CNV status detected by aCGH, JaxMDGA and NGS.

Table IV-9: Comparison of the CNV status of 33 different loci determined by MLPA with the CNV status detected by aCGH, JaxMDGA and NGS in a HAB sample. The first column shows the MLPA probe number, the second the MLPA result. In the other columns the status of each CNV, detected by the respective method, in which the MLPA probe can be mapped, is listed. A minus (-) indicates that no CNV was detected. Probes mapping into different CNVs are separated by black lines.

MLPA Probe no.	MLPA (CNV status in HAB)	aCGH (CNV status in HAB)	JaxMDGA (CNV status in HAB)	NGS (CNV status in HAB)
1	-	loss	loss	-
2	-	loss	-	-
3	-	loss	-	-
4	loss	-	loss	loss
5	loss	-	loss	loss
6	-	-	gain	-

MLPA Probe no.	MLPA (CNV status in HAB)	aCGH (CNV status in HAB)	JaxMDGA (CNV status in HAB)	NGS (CNV status in HAB)
7	-	-	gain	-
8	-	-	gain	-
9	-	-	gain	-
10	-	-	gain	-
11	-	-	gain	-
12	gain	-	gain	-
13	loss	loss	loss	loss
14	loss	loss	loss	loss
15	loss	loss	loss	loss
16	-	-	-	-
17	loss	-	gain	gain
18	gain	-	gain	-
19	-	loss	-	loss
20	loss	loss	loss	loss
21	loss	loss	loss	loss
22	gain	loss	loss	loss
23	loss	loss	loss	-
24	-	loss	loss	loss
25	-	-	gain	-
26	-	-	gain	-
27	-	-	-	-
28	loss	loss	loss	-
29	-	loss	-	-
30	-	loss	-	-
31	-	-	gain	-
32	-	-	gain	-
33	-	-	gain	-

2.2 Behavior of CD-1 mice screened for CNVs by JaxMDGA

In the "CD-1 panel" 384 mice were phenotyped. Here, only selective behavioral data of the 64 mice tested for CNVs by JaxMDGA are shown. For the full data set and a more detailed analysis see Widner-Andrä (2011). The animals to be examined on the JaxMDGA were chosen based on their phenotypes averaged over all behavioral tests, with 16 animals showing intermediate and each 24 high and low anxiety-related behavior, respectively. However, that clear distinction could not be drawn with regard to single behavioral parameters (Figure IV-20). As mentioned above, with the percent time the animals spent on the open arm of the EPM, the distance they travelled in the OF, the time they floated in the FST, the Cort increase after the SRT and the time they spent immobile in the TST, only the key parameters of behavioral data are shown, since they were used for further analysis.

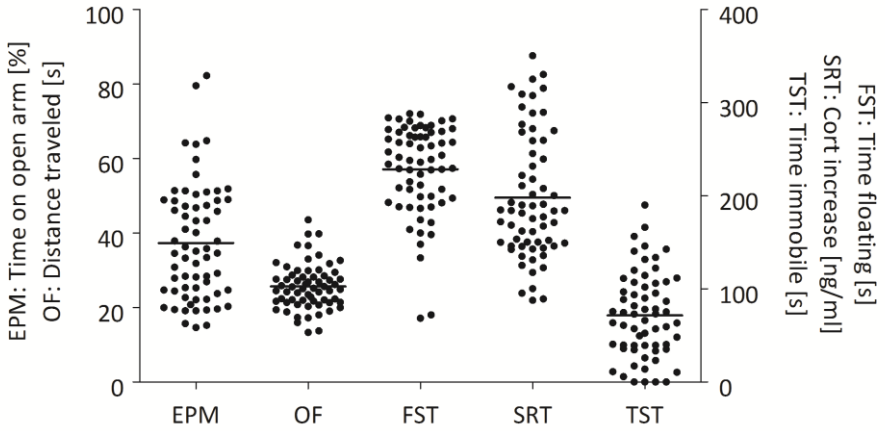

Figure IV-20: Behavior of 64 CD-1 mice screened for CNVs by JaxMDGA. On the x-axis the key parameters of five different behavioral tests are shown. Each dot represents the respective measurement value of a single animal, with the height drawn on the y-axes. EPM and OF data refer to the left y-axis, FST, SRT and TST to the right. Black lines indicate the respective mean values of all animals.

2.3 Position of CNVs detected in CD-1 mice by JaxMDGA

For CNV detection in the raw data of the JaxMDGA, the probe measurement values of a sample animal has to be compared to those of a reference animal. Since it was not useful to determine one animal to serve as reference sample (as described in chapter III-4.2.1.2),

all 32 animals per array were compared to each other by running multiple sample-reference combinations. Afterwards the data of the two arrays were merged. Thus, 764 regions were shown to differ in copy number in at least one pair of the 64 animals screened for CNVs. All the detected CNVs are listed in supplementary table S11 and their genomic position is depicted in Figure IV-21.

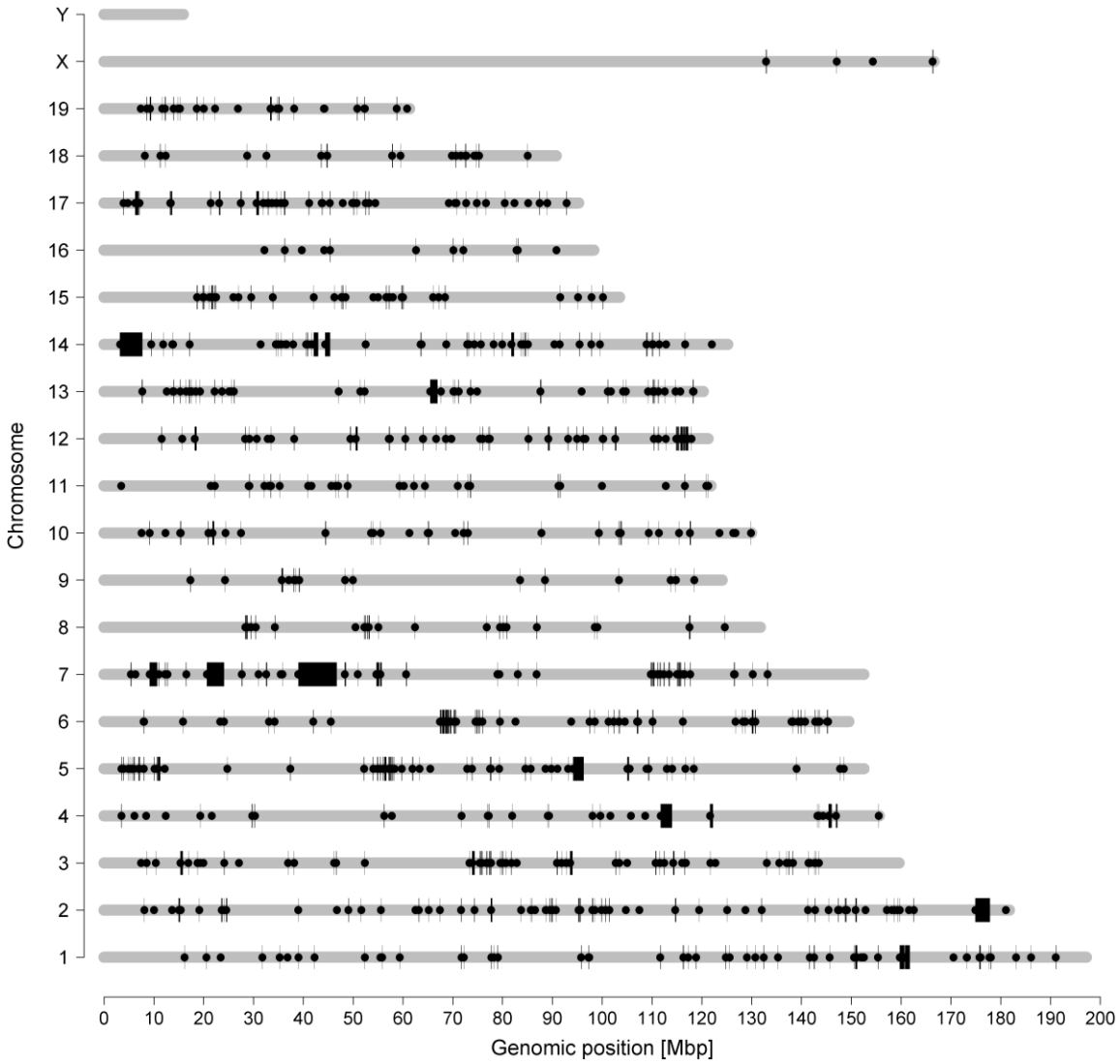

Figure IV-21: Genomic positions of CNVs detected in 64 CD-1 mice by JaxMDGA. Chromosomes are indicated by grey lines with the basepairs shown on the x-axis and the chromosomes on the yaxis. Start points of CNVs are marked by dots and lines are drawn till the end points. Due to limitations in resolution small CNVs might appear as dots only.

Most of the CNVs (61%) ranged in size between 1 kbp and 50 kbp, 18.6% (622) were smaller than 1 kbp, with a mean size of 77,249 bp and a median of 14,377 bp. The

frequency distribution of all CNVs is shown in Figure IV-22. The distribution of regions smaller than 50 kbp is outlined in more detail in Figure IV-23, the distribution of those smaller than 1 kbp in Figure IV-24. Considering only the 622 CNVs whose size exceeded 1 kbp, the smallest region was shown to have 1,003 bp and the largest 7,568,999 bp, with a mean of 94,796 bp and a median of 20,322 bp. Most of the CNVs were found on chromosome 2 and chromosome 6, considering all regions or only those larger than 1 kbp, respectively, whereas none could be revealed on chromosome Y. Details on the frequency distribution of CNVs over all chromosomes are shown in Figure IV-25 and Table IV-10.

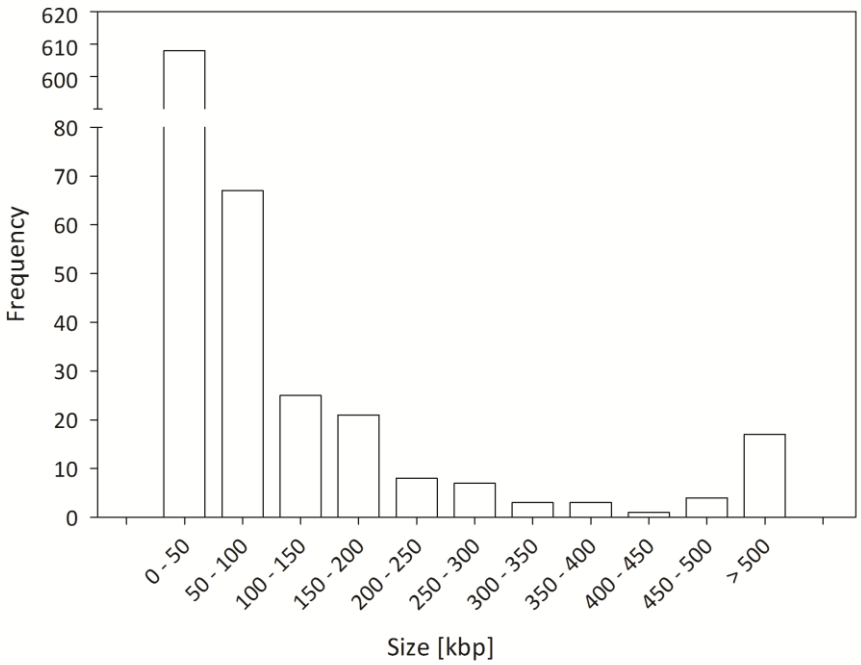

Figure IV-22: Frequency distribution of CNVs detected in 64 CD-1 mice by JaxMDGA. The size classes are plotted against the number of CNVs in the respective range, with a step size of 50 kbp.

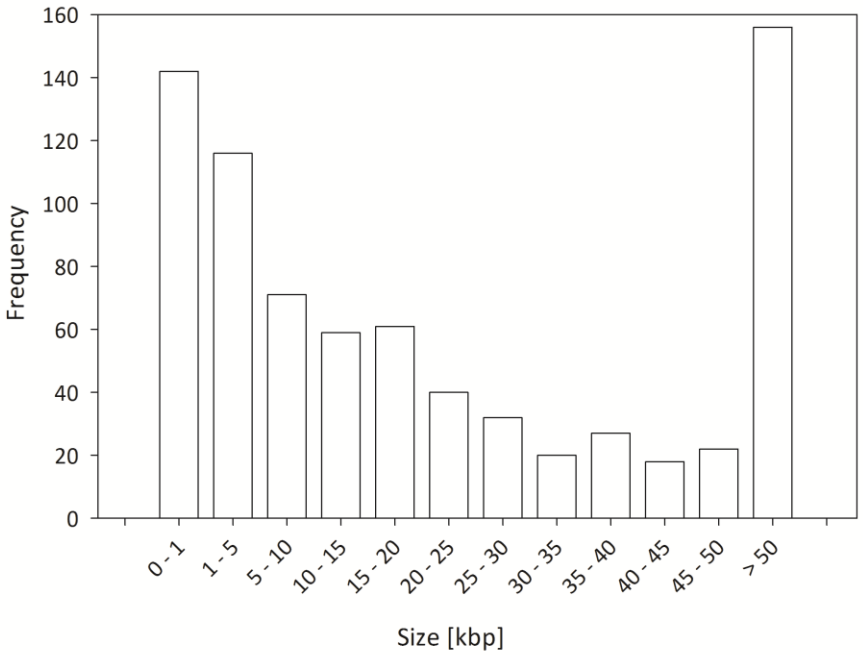

Figure IV-23: Detailed frequency distribution of CNVs < 50 kbp detected in CD-1 mice. All bars but the first (1 kbp) comprise a size range of 5 kbp, indicated on the x-axis. The last bar includes all CNVs > 50 kbp. The number of CNVs in the respective size range is shown on the y-axis.

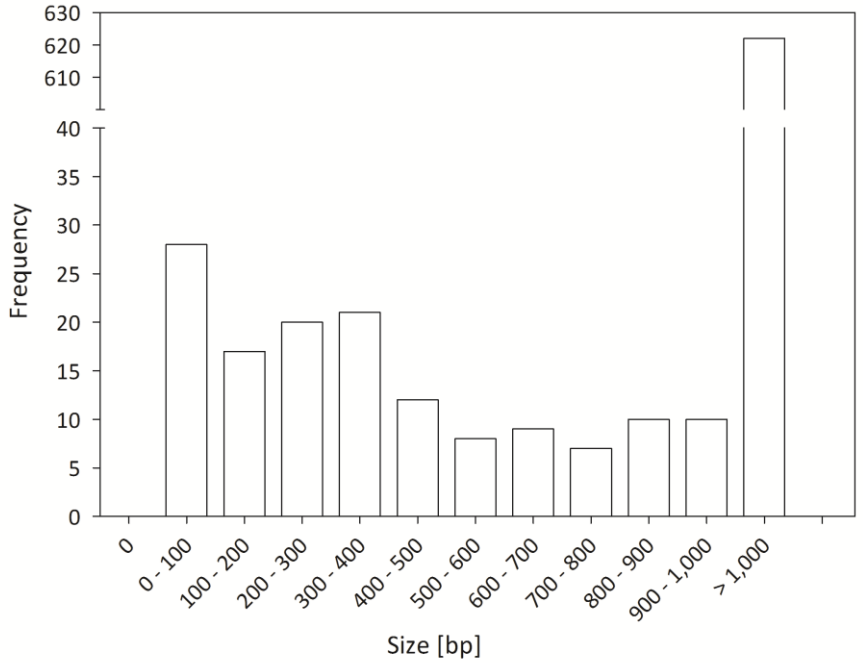

Figure IV-24: Frequency distribution of CNVs < 1 kbp detected in 64 CD-1 mice by JaxMDGA. 142 CNVs of size < 1 kbp are depicted in size ranges of 100 bp, indicated on the x-axis. All other CNVs (622) are referred to in the last column. The number of CNVs in the respective size range is shown on the y-axis.

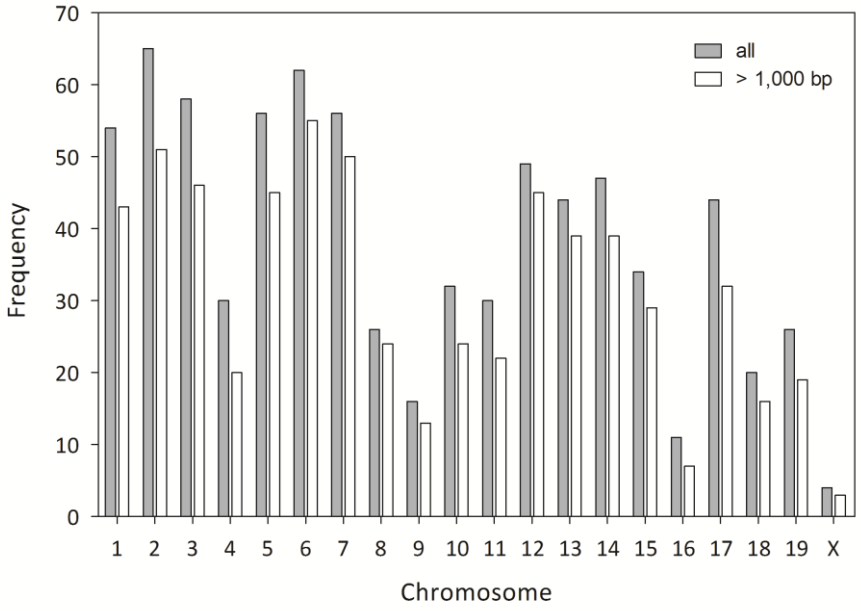

Figure IV-25: Chromosomal distribution of CNVs detected in 64 CD-1 mice. For each chromosome (x-axis) the number of detected CNVs (y-axis) is depicted. Grey columns refer to the full set of revealed CNVs (764). White columns show only data of variants larger than 1 kbp (622).

Table IV-10: Detailed information on CNVs detected in 64 CD-1 mice by JaxMDGA. For each chromosome (column 1) the median and mean size (columns 5 & 6) of all CNVs (quantity shown in column 3), as well as their maximum and minimum size (columns 7 & 8), are listed. The second column shows the set of CNVs to which the just mentioned parameters refer: "A" for the full set of 764 detected CNVs and "B" for only those 622 CNVs larger than 1 kbp.

Chr	Set	No. of CNVs	Sum [bp]	Median [bp]	Mean [bp]	Max [bp]	Min [bp]
1	A	54	3,320,987	15,407	61,500	946,665	13
1	B	43	3,317,251	24,755	77,145	946,665	2,653
2	A	65	5,036,683	13,644	77,487	2,916,531	10
2	B	51	5,030,235	18,828	98,632	2,916,531	1,030
3	A	58	2,627,902	8,842	45,309	479,349	12
3	B	46	2,623,363	17,297	57,030	479,349	1,225
4	A	30	3,853,017	4,337	128,434	1,619,401	137
4	B	20	3,848,386	17,915	192,419	1,619,401	1,280
5	A	56	4,946,924	17,423	88,338	2,100,801	8
5	B	45	4,942,885	25,792	109,842	2,100,801	1,025

Chr	Set	No. of CNVs	Sum [bp]	Median [bp]	Mean [bp]	Max [bp]	Min [bp]
6	A	62	2,791,483	17,489	45,024	376,782	1
6	B	55	2,788,702	21,494	50,704	376,782	1,033
7	A	56	14,929,513	22,530	266,598	7,568,999	182
7	B	50	14,927,087	30,770	298,542	7,568,999	1,112
8	A	26	946,602	16,442	36,408	158,972	4
8	B	24	946,215	18,676	39,426	158,972	1,132
9	A	16	519,584	9,233	32,474	276,328	391
9	B	13	518,145	15,310	39,857	276,328	1,516
10	A	32	1,160,360	11,465	36,261	324,148	76
10	B	24	1,158,370	24,581	48,265	324,148	1,003
11	A	30	544,634	10,062	18,154	95,599	7
11	B	22	541,694	16,444	24,622	95,599	1,599
12	A	49	3,188,781	20,603	65,077	386,118	25
12	B	45	3,187,642	26,299	70,836	386,118	1,174
13	A	44	2,371,123	12,736	53,889	1,434,939	82
13	B	39	2,367,750	17,070	60,712	1,434,939	1,071
14	A	47	7,879,612	13,385	167,651	4,470,511	60
14	B	39	7,876,836	15,576	201,970	4,470,511	1,599
15	A	34	1,101,538	20,574	32,398	271,706	28
15	B	29	1,099,434	29,596	37,912	271,706	1,948
16	A	11	201,041	6,651	18,276	76,999	1
16	B	7	200,766	16,163	28,681	76,999	4,084
17	A	44	2,375,102	9,181	53,980	653,222	34
17	B	32	2,370,932	19,713	74,092	653,222	1,119
18	A	20	298,590	7,860	14,930	82,183	63
18	B	16	296,422	13,752	18,526	82,183	1,171
19	A	26	759,115	14,055	29,197	187,300	77
19	B	19	755,669	18,287	39,772	187,300	1,307
X	A	4	165,544	38,952	41,386	87,242	398
X	B	3	165,146	70,035	55,049	87,242	7,869

2.4 CNVs associated with behavior

Each of the 764 CNVs detected by JaxMDGA was associated with different behavioral parameters using a generalized linear model. The resulting p-values are outlined in supplementary tables S12 to S16. Below, the result for the key parameter of each behavioral test is summarized.

For the most important key parameter, the time animals spent on the open arm of the EPM, 55 CNVs were shown to be significantly associated ($p < 0.05$) with nominal p-values between 0.0005 and 0.0489. Another 35 CNVs showed a trend with values between 0.0503 and 0.0993. However, significances got lost after applying Bonferroni correction. Therefore, only nominal significances were considered for further analysis. In Table IV-11 all CNVs having a significant nominal p-value are outlined. Exemplarily for a significant p-value, a trend and none significance, three different CNVs associated with the time animals spent on the open arm of the EPM are depicted in Figure IV-26. There, the behavioral data of all 64 animals are plotted against the mean normalized intensities of all JaxMDGA probes in the respective CNV as an indicator for different copy numbers.

Table IV-11: List of CNVs detected in 64 CD-1 mice that were significantly associated with the time animals spent on the open arm of the EPM. Significance ($p < 0.05$) was only reached with regard to nominal p-value.

RegNo	Chr	Start	End	Size	No. of probes in CNV	nominal p-value	p-value corrected
9	1	52,348,330	52,357,761	9,431	3	0.0173	13.2277
14	1	72,198,891	72,210,724	11,833	11	0.0048	3.6427
52	1	183,060,113	183,072,518	12,405	24	0.0351	26.8540
65	2	46,736,100	46,736,483	383	3	0.0457	34.9074
106	2	147,572,128	147,572,911	783	5	0.0265	20.2351
110	2	152,838,130	152,838,140	10	3	0.0447	34.1463
113	2	158,641,258	158,643,203	1,945	5	0.0139	10.6138
115	2	159,730,233	159,749,061	18,828	7	0.0297	22.7034
123	3	15,340,258	15,819,607	479,349	47	0.0109	8.3530

RegNo	Chr	Start	End	Size	No. of probes in CNV	nominal p-value	p-value corrected
134	3	52,324,222	52,329,358	5,136	6	0.0131	9.9925
137	3	75,378,209	75,414,973	36,764	15	0.0481	36.7160
163	3	116,043,731	116,044,155	424	2	0.0457	34.9099
165	3	116,765,381	116,765,393	12	3	0.0049	3.7257
177	3	143,512,892	143,520,968	8,076	4	0.0144	10.9773
179	4	6,095,005	6,095,833	828	2	0.0364	27.7772
199	4	111,745,396	112,286,229	540,833	112	0.0338	25.8107
200	4	112,348,832	113,968,233	1,619,401	160	0.0203	15.5093
237	5	59,737,484	59,830,718	93,234	21	0.0331	25.3247
245	5	79,306,300	79,320,284	13,984	5	0.0151	11.5637
253	5	105,034,912	105,359,320	324,408	61	0.0236	18.0645
254	5	105,511,760	105,511,768	8	3	0.0281	21.4800
263	5	148,493,203	148,495,504	2,301	12	0.0368	28.1326
265	6	8,041,854	8,049,012	7,158	12	0.0359	27.4472
267	6	23,304,438	23,304,526	88	2	0.0351	26.8289
302	6	107,165,345	107,215,744	50,399	14	0.0401	30.6273
311	6	137,984,036	137,984,036	1	1	0.0024	1.8134
312	6	138,231,553	138,237,439	5,886	11	0.0221	16.9189
313	6	138,345,026	138,345,242	216	3	0.0056	4.2408
317	6	139,939,182	139,950,758	11,576	9	0.0303	23.1715
397	8	79,438,032	79,453,386	15,354	17	0.0088	6.6910
401	8	80,806,213	80,834,051	27,838	15	0.0414	31.6039
408	9	17,303,897	17,358,876	54,979	15	0.0374	28.5508
442	10	72,164,250	72,231,618	67,368	11	0.0093	7.1281
443	10	73,046,724	73,090,099	43,375	9	0.0286	21.8730
463	11	33,359,372	33,375,067	15,695	4	0.0261	19.9380
483	11	116,602,360	116,630,414	28,054	33	0.0089	6.8325
486	12	11,545,659	11,610,212	64,553	13	0.0391	29.8354
497	12	57,140,406	57,157,776	17,370	11	0.0039	3.0090
498	12	57,366,059	57,387,621	21,562	8	0.0005	0.3516

RegNo	Chr	Start	End	Size	No. of probes in CNV	nominal p-value	p-value corrected
508	12	77,397,670	77,403,999	6,329	18	0.0489	37.3403
516	12	100,077,912	100,080,008	2,096	7	0.0013	1.0050
517	12	100,214,407	100,217,042	2,635	6	0.0034	2.5618
518	12	102,582,454	102,828,642	246,188	96	0.0440	33.5929
519	12	110,355,216	110,412,139	56,923	8	0.0456	34.8510
561	13	73,546,675	73,574,553	27,878	11	0.0461	35.2310
600	14	63,715,836	63,731,412	15,576	7	0.0317	24.1878
607	14	78,207,014	78,207,745	731	3	0.0208	15.9274
625	14	122,006,064	122,006,350	286	6	0.0375	28.6396
636	15	29,502,036	29,545,764	43,728	18	0.0072	5.5338
673	17	6,255,961	6,909,183	653,222	59	0.0012	0.8981
695	17	45,324,397	45,349,233	24,836	10	0.0026	2.0003
701	17	53,145,663	53,184,954	39,291	15	0.0464	35.4150
731	18	74,681,698	74,696,062	14,364	9	0.0098	7.5124
732	18	74,750,012	74,750,945	933	4	0.0114	8.6793
733	18	75,257,879	75,261,870	3,991	6	0.0426	32.5340

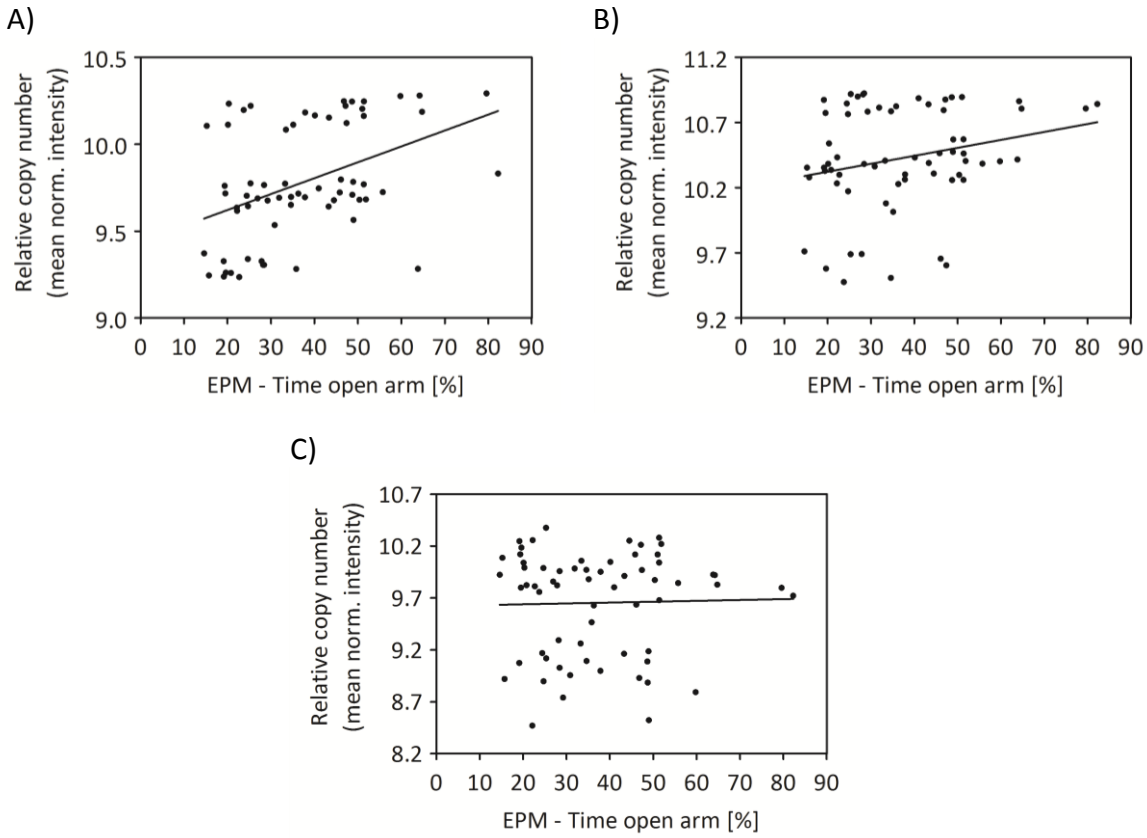

Figure IV-26: Association of copy number with behavior. Each dot represents a single animal (N=64). On the x-axes the time animals spent on the open arm is plotted. The y-axes show the mean normalized intensities of JaxMDGA probes in the respective CNV, representing the relative copy number. (A) CNV no. 498 is shown as example for a region significantly associated with behavior (nominal p = 0.0005). Regression line: y = 0.0091x + 9.4389. (B) CNV no. 164 is depicted as example for a region with a trend in the association p-value (nominal p = 0.0509). Regression line: y = 0.0061x + 10.201. (C) CNV no. 453 is shown as example for a region not significantly associated with behavior (nominal p = 0.9969). Regression line: y = 0.0008x + 9.6225.

For the key parameter of the TST, the time animals were immobile, the nominal p-values reflected a trend in 43 CNVs and were significant in 25 CNVs, ranging from 0.0019 to 0.0496. The time animals floated in the FST was significantly associated with 69 CNVs (nominal p-values between 0.0003 and 0.0492) and associated with a trend in 40 CNVs. 35 CNVs could be significantly associated with the total distance the animals travelled in the OF with a minimum nominal p-value of 0.0023 and a maximum of 0.0494. Here, 28 CNVs had nominal p-values less than 0.1 but greater than 0.05. Finally, for the SRT parameter “Cort increase” the association was shown to be nominal significant for 44 CNVs with nominal p-values ranging from 0.0014 to 0.0497. For 51 CNVs a trend could be shown.

3 Candidate genes of anxiety-related behavior

3.1 Candidate genes in HAB/LAB mice

In total, 73 of 98 CNVs detected by aCGH, 101 of 180 CNVs detected by JaxMDGA and 1,581 of 5,851 CNVs detected by NGS were shown to include at least one protein coding gene. Since CNVs were revealed as a potential influencing factor on gene expression in HAB/LAB mice (see chapter 1.5.3), all protein coding genes in regions where CNVs were detected are candidate genes that potentially influence the phenotypic expression of anxiety-related behavior. Out of 998 (aCGH), 145 (JaxMDGA) and 1,085 (NGS) genes, respectively, 68 were found to be the most promising candidates as they appeared in all three data sets obtained by the different CNV detection methods. The latter ones are listed in Table IV-12, the others in supplementary tables S17 to S19.

Of special interest are those genes that did not only lay in regions where CNVs were detected in HAB/LAB, but also overlapped CNVs of CD-1 mice found to be associated with the time animals spent on the open arm of the EPM (see chapter 2.4). These 15 genes are outlined in Table IV-13.

Table IV-12: Protein coding genes in CNVs detected by aCGH, JaxMDGA and NGS in HAB/LAB mice. The table shows all 68 genes found in data of all three detection methods, sorted by genomic position.

Gene symbol	MGI ID	Chr	Gene start	Gene end
Hmcn1	2685047	1	152,409,654	152,840,565
Npl	1921341	1	155,350,146	155,396,844
Sirpb1a	2444824	3	15,371,819	15,426,504
Sirpb1b	3779828	3	15,495,754	15,575,065
Sirpb1c	3807521	3	15,695,145	15,748,528
Skint4	2444425	4	111,744,621	111,840,681
Skint3	3045331	4	111,904,850	111,973,073
Skint9	3045341	4	112,058,574	112,106,590
Skint6	3649262	4	112,908,844	112,959,568
Skint5	3650151	4	113,613,249	113,672,102

Gene symbol	MGI ID	Chr	Gene start	Gene end
Skint11	2685415	4	113,835,989	113,917,633
Gm3259	3781437	5	95,734,834	95,772,936
Gm7963	3643214	5	95,854,546	95,857,656
D5Ertd577e	1261918	5	95,885,825	95,914,583
C230055K05Rik	2441896	5	105,242,188	105,288,830
Vmn1r72	2182256	7	12,254,948	12,255,868
Gm6605	3644762	7	39,231,315	39,237,190
Gm5591	3648692	7	39,303,158	39,313,212
Gm4454	3782638	7	39,351,859	39,356,187
Gm1988	3780157	7	46,425,170	46,435,208
Gm6124	3779556	7	46,474,468	46,480,344
Gm16387	3646111	7	46,511,994	46,519,904
Gm6833	3645553	7	46,562,898	46,572,951
Dub2a	3051372	7	110,398,251	110,401,019
Olfr597	3030431	7	110,468,915	110,469,874
Olfr635	3030469	7	111,127,690	111,128,655
Trim34-1	2137359	7	111,392,973	111,489,207
Trim5	3646853	7	111,411,900	111,436,608
Gm15134	3805550	7	111,448,744	111,485,423
Trim12-2	4821183	7	111,487,268	111,501,876
Olfr486	3030320	7	115,315,312	115,316,256
Olfr487	3030321	7	115,355,097	115,356,041
Tex24	1921539	8	28,454,866	28,459,659
E030030I06Rik	2442914	10	21,833,938	21,868,629
Raet1c	109431	10	21,878,375	22,093,945
H60b	3649078	10	21,993,152	22,008,655
Raet1d	1861032	10	22,082,077	22,093,945
Iltifb	2151139	10	117,726,685	117,732,094
Gm10480	3642435	12	18,223,364	18,224,634
Gm9590	3779999	12	18,394,561	18,413,667
5730507C01Rik	1917882	12	18,521,544	18,540,997
Gm16948	4439872	12	114,849,395	114,849,828

Gene symbol	MGI ID	Chr	Gene start	Gene end
Gm16999	4439923	12	115,272,016	115,272,487
Gm16697	4439621	12	115,821,780	115,822,213
Gm16746	4439670	12	117,100,078	117,100,511
Gm16812	4439736	12	117,106,984	117,107,413
Pde4d	99555	13	109,240,142	110,743,669
Cdh12	109503	15	21,041,207	21,519,288
Tcp10b	98542	17	13,253,977	13,275,092
Gm10512	3642173	17	13,397,908	13,399,077
Smok2a	1351487	17	13,414,054	13,420,524
Smok2b	3037705	17	13,421,718	13,430,055
Gm9880	3711246	17	13,547,438	13,569,717
Btbd9	1916625	17	30,357,046	30,667,310
Gm9874	3642006	17	30,622,456	30,622,908
Glo1	95742	17	30,729,806	30,749,539
1700097N02Rik	1914772	17	30,759,394	30,762,030
Dnahc8	107714	17	30,763,936	31,012,209
Gm9937	3642051	17	31,011,751	31,012,146
Glp1r	99571	17	31,038,812	31,073,455
H2-T22	95956	17	36,175,352	36,179,692
H2-T9	95965	17	36,175,352	36,179,692
Gm6034	3646212	17	36,179,906	36,195,590
H2-T10	95942	17	36,254,035	36,258,389
Gm10500	3641966	17	36,258,775	36,265,692
Gm10499	3702919	17	36,278,703	36,282,868
Mid1	1100537	X	166,123,131	166,428,730
G530011O06Rik	3603513	X	166,412,975	166,416,849

Table IV-13: Protein coding genes in CNVs detected in HAB/LAB and CD-1 mice. Genes listed here were shown to be in regions where CNVs were detected in HAB/LAB mice by three different methods. The genes were also found to lay in CNVs detected in CD-1 mice that were associated with the time animals spent on the open arm of the EPM, with nominal p-values indicating significance or a trend. (* = nominal p-value < 0.05; ° = nominal p-value < 0.1.

Gene symbol	MGI ID	Chr	Gene start	Gene end
*Sirpb1a**	2444824	3	15,371,819	15,426,504
*Sirpb1b**	3779828	3	15,495,754	15,575,065
*Sirpb1c**	3807521	3	15,695,145	15,748,528
*Skint4**	2444425	4	111,744,621	111,840,681
*Skint3**	3045331	4	111,904,850	111,973,073
*Skint9**	3045341	4	112,058,574	112,106,590
*Skint6**	3649262	4	112,908,844	112,959,568
*Skint5**	3650151	4	113,613,249	113,672,102
*Skint11**	2685415	4	113,835,989	113,917,633
*C230055K05Rik**	2441896	5	105,242,188	105,288,830
Tcp10b°	98542	17	13,253,977	13,275,092
Gm10512°	3642173	17	13,397,908	13,399,077
Smok2a°	1351487	17	13,414,054	13,420,524
Smok2b°	3037705	17	13,421,718	13,430,055
Gm9880°	3711246	17	13,547,438	13,569,717

3.2 Candidate genes in CD-1 mice

All protein coding genes that were in regions where CNVs associated with any behavioral parameter were revealed are of potential interest for further studies. Here, the focus is on the genes in CNVs that were significantly associated (i.e. nominal $p < 0.05$) with the time animals spent on the open arm of the EPM (see chapter 2.4). These 53 genes are listed in Table IV-14. Considering all 764 CNVs, a total amount of 911 protein coding genes was found to overlap with 389 of them. These genes are outlined in supplementary table S20.

Table IV-14: Protein coding genes in CNVs detected by JaxMDGA in 64 CD-1 mice. The symbols, MGI IDs and genomic positions of the genes are shown. The last column includes the number of the respective CNV the gene is overlapping with (for details on CNV position see chapter 2.3). If more than one CNV number is mentioned, the gene comprises a genomic region with multiple CNVs.

Gene symbol	MGI ID	Chr	Gene start	Gene end	CNV no.
Mreg	2151839	1	72,206,020	72,258,881	14
Nvl	1914709	1	183,023,554	183,074,288	52
Pdrg1	1915809	2	152,834,626	152,841,163	110
Dhx35	1918965	2	158,620,543	158,683,950	113
Sirpb1a	2444824	3	15,371,819	15,426,504	123
Sirpb1b	3779828	3	15,495,754	15,575,065	123
Sirpb1c	3807521	3	15,695,145	15,748,528	123
Serpini1	1194506	3	75,361,469	75,447,417	137
Cdc14a	2442676	3	115,975,471	116,126,950	163
Skint4	2444425	4	111,744,621	111,840,681	199
Skint3	3045331	4	111,904,850	111,973,073	199
Skint9	3045341	4	112,058,574	112,106,590	199
Skint2	3649629	4	112,286,202	112,320,445	199
Skint10	2685416	4	112,383,752	112,447,471	200
Skint6	3649262	4	112,908,844	112,959,568	200
Skint5	3650151	4	113,613,249	113,672,102	200
Skint11	2685415	4	113,835,989	113,917,633	200
C230055K05Rik	2441896	5	105,242,188	105,288,830	253
Gbp9	3605620	5	105,508,317	105,539,297	254
Flt1	95558	5	148,373,180	148,537,587	263
Col28a1	2685312	6	7,947,808	8,142,617	265
Cadps2	2443963	6	23,212,773	23,789,420	267
Slc15a5	3607714	6	137,932,107	138,021,717	311
Lmo3	102810	6	138,311,439	138,530,489	313
Plcz1	2150308	6	139,938,197	139,989,978	316
Nr3c2	99459	8	79,423,341	79,768,911	397
Ttc29	1920551	8	80,737,196	80,918,225	401
Pcdh15	1891428	10	72,562,090	74,112,482	443

Gene symbol	MGI ID	Chr	Gene start	Gene end	CNV no.
Ranbp17	1929706	11	33,111,795	33,413,746	463
Gm11735	3713525	11	116,599,248	116,611,242	483
BC018473	3039625	11	116,613,481	116,620,687	483
St6galnac1	1341826	11	116,626,339	116,636,821	483
Mthfd1	1342005	12	77,356,395	77,420,789	508
Eml5	2442513	12	100,024,814	100,139,694	516
Ttc8	1923510	12	100,158,784	100,221,448	517
Catsperb	2443988	12	102,642,892	102,864,160	518
Dock9	106321	14	121,941,261	122,097,639	625
Tmem181a	1924356	17	6,270,475	6,305,783	673
Gm2792	3780960	17	6,407,216	6,410,679	673
Dynlt1b	98643	17	6,429,260	6,435,444	673
Gm2802	3780971	17	6,491,318	6,493,657	673
Gm2808	3780977	17	6,506,579	6,582,263	673
Dynlt1f	3780996	17	6,600,835	6,609,656	673
Dynlt1c	3807476	17	6,646,959	6,655,064	673
Gm2833	3781005	17	6,658,217	6,679,001	673
Gm2827	3780999	17	6,661,442	6,672,834	673
Gm2839	3781011	17	6,753,404	6,803,325	673
Dynlt1d	3781039	17	6,797,760	6,818,159	673
Dynlt1e	3781053	17	6,851,795	6,860,617	673
Gm2867	3781044	17	6,874,264	6,878,036	673
Sytl3	1933367	17	6,877,808	6,942,391	673
Myo5b	106598	18	74,600,590	74,931,147	731, 732
Dym	1918480	18	75,178,426	75,446,620	733

3.3 Identified clusters in the candidate genes

The list of 15 candidate genes, defined in HAB/LAB mice by comparing the CNV data of three different detection methods with the above mentioned candidate genes of the

CD-1 mice, was submitted to the DAVID Bioinformatics Database (http://david.abcc.ncifcrf.gov) for functional annotation clustering. Thus, two significant cluster could be identified that could be linked to the immune system and to signal transduction, respectively. The genes involved in these clusters are outlined in Table IV-15.

Table IV-15: Identified clusters in 15 candidate genes of anxiety-related behavior in HAB/LAB mice. Genes were determined comparing their genomic position with the position of CNVs detected in HAB/LAB mice by three different methods and afterwards with the position of CNVs in CD-1 mice associated with anxiety-related behavior. Clusters were detected using the functional annotation tool of the DAVID Bioinformatics Database. Genes are sorted in alphabetical order.

Cluster	Enrichment score	No. of genes in cluster	Symbols of genes in cluster
„Immune system“	2.21	10	*Sirpb1a, Sirpb1c, Skint3, Skint4, Skint5, Skint6, Skint9, Skint11, Tcp10b, Tcp10c (= Gm9880)*
„Signal transduction“	1.56	4	*Smok2a, Smok2b, Tcp10b, Tcp10c (= Gm9880)*

The same analysis was performed for the 53 candidate genes that were shown to overlap with CNVs associated with anxiety-related behavior in CD-1 mice (see chapter 3.2). Here again, two significantly enriched gene clusters with enrichment scores > 1.3 were identified (Table IV-16). The first cluster was named “Signal transduction“, since the 3 genes embraced in that cluster are all involved in processes mediating signaling. The second significantly enriched cluster comprises 26 genes, which can be connected to the immune system.

Table IV-16: Identified clusters in 53 candidate genes of anxiety-related behavior in CD-1 mice. Genes were determined comparing their genomic position with the position of CNVs in CD-1 mice associated with anxiety-related behavior. Clusters were detected using the functional annotation tool of the DAVID Bioinformatics Database. Genes are sorted in alphabetical order.

Cluster	Enrichment score	No. of genes in cluster	Symbols of genes in cluster
„Signal transduction“	2.09	3	*Dock9, Ranbp17, Sytl3*
„Immune system“	1.95	26	*Cadps2, Catsperb, Col28a1, Dock9, Flt1, Gm11735, Mreg, Nr3c2, Pcdh15, Ranbp17, Serpini1, Sirpb1a, Sirpb1c, Skint10, Skint11, Skint2, Skint3, Skint4, Skint5, Skint6, Skint9, Slc15a5, St6galnac1, Sytl3, Tmem181a, Ttc8*

V Discussion

1 The detection of CNVs

Multiple methods have been developed recently to detect CNVs either in a genome-wide approach or at specific loci. But still, the CNV detection remains a challenging task since the results depend not only on the method used but also on the algorithm applied on the data and no "gold standard" algorithm has been established so far (Alkan *et al.*, 2011; Pinto *et al.*, 2011; van de Wiel *et al.*, 2010; Warden *et al.*, 2011; Zhang *et al.*, 2011). That is why in the current study three different methods, two probe-based high-density genotyping arrays (aCGH, JaxMDGA) and one whole-genome sequencing approach (NGS), were used to detect CNVs in the HAB/LAB mouse model to overcome the limitations of a single approach and thus increase the reliability of the data. In total, four different methods to reveal CNVs in the genome of HAB/LAB or CD-1 mice have been employed. As previously mentioned, aCGH, JaxMDGA and NGS were used for screening the genome of HAB/LAB mice, while JaxMDGA was used to examine the genome of CD-1 mice as well. Additionally, another test screening specific loci (MLPA) was performed. One LAB and one HAB animal were also tested in that approach to serve as reference and positive control samples, respectively.

1.1 CNVs in the HAB/LAB mouse model

For CNV detection by means of aCGH, the DNA was purified from two HAB and LAB mice each. It was extracted from tail tips of all four individuals, as well as from brain tissue of just one HAB and LAB mouse, respectively. Thus, three independent data sets were generated. The results of the different data sets, i.e. the probes' $\log_2$ ratios of measurement values from the HAB against the LAB samples, were compared to each other. A very high level of concordance was observed in the data gathered from tail tissue of both HAB/LAB mouse pairs, indicating that the animals of one mouse line did not show differences in their genome with respect to CNVs. Since the mouse lines were bred for over 40 generations and thus, feature a high inbreeding factor, this outcome was not

surprising and confirms a high degree of genetic homogeneity among the individuals of one line. The DNA extracted from different tissues did also not display diverging variation, which was again expected since copy number dependency on tissue was described before only for mitochondrial (Fuke *et al.*, 2011), but not genomic DNA. However, since DNA of only two animal pairs and two tissues were compared, further studies should be conducted to clearly show the extent of variation between the individuals of a single breeding line.

As mentioned above, the definition of CNV breakpoints strongly depends on the algorithm applied. There are multiple algorithms available for analyzing aCGH data (Benelli *et al.*, 2010; Cahan *et al.*, 2008; Picard *et al.*, 2005; van de Wiel *et al.*, 2010), each having advantages and disadvantages. In order to simplify the processing, Peter Park and others (Lai *et al.*, 2008) developed the CGHweb online tool to analyze aCGH raw data with multiple algorithms simultaneously. However, in this study it was refrained from applying several algorithms since CNVs were confirmed by two additional techniques. Thus, just the segMNT algorithm was applied on the aCGH data, recommended and performed by NimbleGen, the array-providing company.

In total, 98 CNVs have been detected in HAB/LAB mice by means of aCGH, covering about 3.7% of the genome, with 52% of the CNVs being larger than 500 kbp, about 32% differing in size between 50 and 500 kbp and none being smaller than 5 kbp. It is difficult to estimate the reliability of these findings, since only a few studies were published comparing CNVs in different mouse lines based on aCGH data. These studies mainly detected CNVs between different mouse strains and there the occurrence of variants is supposed to be enlarged compared to those found in two individuals of distinct lines that were bred based on the same mouse strain like HAB/LAB (Cutler *et al.*, 2007; Graubert *et al.*, 2007). Besides, in the published literature there is not always a clear distinction drawn between segmental duplications and CNVs, which makes comparisons more difficult. However, the size range of detected CNVs from some thousands to several mega basepairs was similar in all studies using aCGH (Cutler *et al.*, 2007; Graubert *et al.*, 2007; Henrichsen *et al.*, 2009a; Henrichsen *et al.*, 2009b), and is thus in concordance with the size range of CNVs found in HAB/LAB mice.

The second detection method, JaxMDGA, resembled the first in also being a high-density genotyping array. However, it provided a greater coverage than the first array. Analogues to the aCGH analysis, multiple algorithms were available for the processing of raw data (Banerjee *et al.*, 2011; Teo *et al.*, 2011; van de Wiel *et al.*, 2010; Wang *et al.*, 2007; Winchester and Ragoussis, 2012; Zhang *et al.*, 2011). However, the application of those was desisted, since here again, the validation should occur by the implementation of the other detection methods, aCGH and NGS. Thus, the CNV data were generated just by means of one algorithm, which was included in the R package provided by the Jackson Laboratory (Yang, 2010).

In contrast to the results of the aCGH assay, the analysis of JaxMDGA data revealed in total more variable regions (180) but none of them was shown to be on the Y chromosome. Missing CNVs on the Y chromosome might be explained by the fact that only the first 16.6% of that chromosome were covered by probes of the JaxMDGA chip, since for the rest of the chromosome no “chippable” probes according to stringent selection criteria could be defined and thus added to the chip by the supplier (Yang *et al.*, 2009). Consequently, if CNVs would occur in the remaining 83.4% of the chromosome they would not be detected. The size of all CNVs summed up comprised just about 0.6% of the genome as the single regions were smaller than those found by aCGH, with only about 1% exceeding 500 kbp, nearly 17% ranging in size between 50 and 500 kbp and 40% being smaller than 5 kbp. It remains to be clarified, which of the data reflect the real situation in the animals more effective. According to a current study of Pinto and colleagues (Pinto *et al.*, 2011) the results of the JaxMDGA analysis should yield more accurate data, not only because of the increased resolution due to the higher number of probes on the chip, but also since a better individual probe performance was shown. However, these are not the only aspects to be considered, since the choice of the computational method could also strongly influence the outcome (Pinto *et al.*, 2011). Since all array probes were designed based on the C57BL/6J mouse strain it could happen that for a specific region of the CD-1 mouse strain-based HAB/LAB genome the binding probability of distinct probes is reduced due to changes in the nucleotide sequence. Depending of the algorithm applied on the raw data, this reduced probe binding might be wrongly detected as a change in copy number or not. Additionally, since different probes

were used in aCGH and JaxMDGA, the binding probability might be reduced in probes of just one of the two arrays. In consequence, an evaluation of the two created data sets remains difficult. Hence a third experiment, the NGS, was performed to encircle the position of CNVs in the HAB/LAB mouse genome.

Even though NGS has a great potential and is clearly the most sophisticated whole-genome sequencing approach existing to date, the thereby generated data do not answer the claim for completeness. Besides the unavoidable technical-induced variance that exists for each experimental setup, problems might have been occurred during the alignment of the single reads derived from HAB/LAB mice (having the genomic background of the CD-1 mouse strain) against the reference genome, which is the commonly used genome of the C57BL/6J mouse strain. Thus, DNA segments not existing in the reference genome could get lost since the respective reads could not be aligned. For sequences with slight differences to the reference genome, specific reads might not have passed the alignment threshold and therefore the coverage at some positions could be reduced and with it the reliability of the data at the respective position. Furthermore, the performance of a complex computational analysis was necessary to gather the final CNV positions, in which several variables, such as the bin size or the number of allowed mismatches had to be estimated. Although the calculation was done to best knowledge, false positive or negative findings are likely to occur also in the CNV data generated by NGS.

A total of 5,851 CNVs were detected by means of NGS. Their mean size was much smaller than the one of variants detected by other methods. About 83% of the variable regions were shown to be smaller than 5 kbp, while 0.7% ranged in size between 50 and 500 kbp and none exceeded 500 kbp. Nonetheless, in considering the total size, they covered about 1.03% of the whole genome, which is in between the coverage of JaxMDGA- (0.6%) and aCGH-defined CNVs (3.7%).

Considering the pitfalls of all performed assays, it seems likely that the real number and size of the CNVs indeed existing in HAB and LAB mice is somewhere in between the ones discovered. One way to define the real breakpoints of the variants could be to estimate their exact position by means of qPCR and follow-up PCR as it was done for the *Glo1* locus

by Williams and colleagues (Williams *et al.*, 2009), and finally use Sanger sequencing to uncover the sequence of the region. However, if data gathered from three different detection methods point to the same region to vary, this region might already be accepted as a present CNV.

The sum of all basepairs overlapping between CNVs defined by the three applied detection methods exceeded 4.8 Mbp. This reflects about 0.18% of the whole genome and thus a relatively large part of it. For comparison only, HAB/LAB mice were estimated to harbor about 800,000 SNPs, which reflects nearly 0.03% of the genome (Ludwig Czibere, personal communication). Considering their relatively large extent alone, it seems likely that they affect the expression of several genes, directly by gene dosage alteration or indirectly (see chapter II-6) and thus contribute to the distinct phenotypes of HAB and LAB mice.

1.2 CNVs in CD-1 mice

The detection of CNVs in 64 CD-1 mice by JaxMDGA appeared to be more challenging than the examination in HAB/LAB mice using the same method, since a completely new determination approach had to be proposed to discover the variants as discussed below.

In contrast to the detection of SNP data, where each probe provides an absolute signal indicating the presence or absence of a specific basepair, the probe signals for CNV detection need to be set in relation to a reference. To date, there is no commonly available algorithm that would be easily applicable on the raw data of a flexible number of array runs, to analyze the data with respect to each other. Therefore, the same function (“simpleCNV”) used before to detect HAB/LAB mouse CNVs had to be implemented (Yang, 2010). This led to the first problem, the definition of the reference sample, since in contrast to the study of HAB/LAB mouse DNA, multiple animals of variable phenotypes had to be examined. If just one of the animals had been chosen to serve as reference sample, information on variable positions between two separate animals might have gotten lost, as explained in more detail in chapter III-4.2.1.2 (Figure III-10). For this reason, the raw data of each animal were separately compared to the raw data of each other animal tested in the approach. The outcome of that multiple

application of the "simple CNV" function was a list of all regions assumed to harbor CNVs in any of the tested pairs. In the next step, overlapping regions were defined as one, with breakpoints as far away from each other as possible (Figure III-11). That "strategy of the largest size" was chosen since, in case of one pair of animals showing a CNV of remarkable smaller size at the specific locus than all the other pair-wise comparisons, the opposite strategy would not only clearly underestimate the extent of the CNV but also could in worst case (e.g., if the region would comprise just a few probes) lead to an increased false-positive rate, obtained by the subsequent calculation. Of course, similar problems might occur by using the strategy applied, but in this particular case the subsequent calculation would lead to an increased false-negative rate, which can be accepted since it increases the reliability of the detected CNVs. After the new breakpoint definition the recalculation of the respective intensity values as a measurement parameter of the copy number had to be performed. These intensities were required for the successive association of CNVs with the animals' behavior.

Using the procedure described above, 764 CNVs were detected, of which about 2% exceeded 500 kbp, around 18% were in a size range of 50 to 500 kbp and 34% appeared to be smaller than 5 kbp. The size distribution is in concordance to the one found in HAB/LAB mice by the same method (i.e. JaxMDGA), and hence does not differ from the expectation. Here, the CNVs covered about 5.2% of the genome. Even though that is 10-fold more than the same method revealed in HAB/LAB mice, it seems to be acceptable since not only two inbred but 64 outbred animals were compared and not all of the CNVs were found in each sample-reference pair. Here again, no variants were detected on the Y chromosome, which could be explained as above, that is, by the fact that the JaxMDGA probes cover only the first part of the chromosome.

MLPA, a second method to screen for CNVs associated with behavior in CD-1 mice was used with the intention of confirming the findings of the JaxMDGA approach for specific regions. Unfortunately, the probes for the MLPA experiment had to be designed based on preliminary data, but the final CNV data determined by JaxMDGA differed remarkably. As a consequence, none of the regions reviewed by MLPA showed significant association with anxiety-related behavior according to the final JaxMDGA data. However, 12 of 33 MLPA probes mapped to CNVs showing a trend in the nominal association p-values. In

order to test the reliability of measurement values of MLPA probes with respect to the JaxMDGA data (HAB/LAB) a HAB sample was included in the analysis. Thereby, the results of 18 probes could be confirmed. However, as long as the definite breakpoints of the HAB/LAB mouse CNVs are not verified, this cannot be an absolute confirmation of reliability, since data gathered from aCGH and NGS might show different results. Besides, it needs to be considered that there is a bias in the detection of CNVs between different CD-1 mice since measurements of a single individual had to be analyzed with respect to the reference sample, which was a LAB animal. Therefore, existing differences between CD-1 animals might be masked. However, the findings of the JaxMDGA (CD-1) approach could be confirmed by 19 of 33 probes targeting 11 CNVs. The analysis of three probes (no. 14, 15 and 21) did hint towards an association of the respective CNV with the animals' behavior on the EPM. Since animals used for MLPA were different from those tested in JaxMDGA and they showed more extreme behavior, these regions indeed could be of interest for further analyses, even though they were not occurring in the JaxMDGA-detected CNVs associated with behavior. But, as long as the functionality of the probes is not approved and since the analysis software warned of less DNA amount used in the MLPA PCR run, the MLPA results need to be considered with caution.

2 The contribution of CNVs to phenotypic variation

A study published some years ago concluded that common CNVs that can be typed on existing platforms are unlikely to have a major role in the genetic basis of human diseases (Wellcome Trust Case Control Consortium, 2010). However, this conclusion has to be considered with caution since it was drawn from a genome-wide association study. Although GWAS are commonly used as a powerful tool to examine the genetics of complex traits, Lee and colleagues claimed that their design is better poised to detect common variants of common diseases (Lee *et al.*, 2012). Hence, the genetic architecture of complex diseases and phenotypes that consists of multiple rare variants of modest effect, seems to be difficult to detect by GWAS (Lee *et al.*, 2012; McClellan *et al.*, 2007). Therefore, CNVs should not be neglected as contributors to complex phenotypes like anxiety (Gamazon *et al.*, 2011).

There are multiple molecular mechanisms conceivable to explain how CNVs and other SVs might contribute to respective phenotypes and diseases. Some of the effects are more evident than others, like the change of gene dosage by the alteration of copy number (Lupski *et al.*, 1992), while others are of more complex nature. CNVs might also convey the effects even when they do not harbor any coding region, for instance by the duplication or deletion of regulatory elements. Enhancers and repressors were shown to act as *cis*-regulatory domains even though they extend long distances outside the gene (Kleinjan and van Heyningen, 2005) and hence CNVs could regulate the transcription of genes far outside their breakpoints. Furthermore, CNVs could also mediate their effects by physically impairing the access of the genes to the transcription machinery or by influencing transvection (Henrichsen *et al.*, 2009a; Stankiewicz and Lupski, 2010; Wu and Morris, 1999). In this study, the focus is on the influence of CNVs on the expression of genes located inside the variant regions. Further studies are supposed to examine the remaining effects mentioned above.

2.1 The influence of CNVs on expression levels in HAB/LAB mice

About 25% of all autosomal genes were shown to overlap with CNVs in mice (Henrichsen *et al.*, 2009b) and therefore a considerable contribution of this type of variation on the gene products, i.e. the expression level, is to be expected. Effects of CNVs on expression levels have previously been reported with a positive correlation in some cases and a negative in others (Cahan *et al.*, 2009; Stranger *et al.*, 2007; Williams *et al.*, 2009). As a first step to test for the impact of CNVs on gene expression in HAB/LAB mice, a reappraisal of a previously conducted expression microarray experiment (Czibere, 2008; Czibere *et al.*, 2011) was performed. The expression differences between HAB and LAB animals were newly generated as better computational approaches have been developed since the array was analyzed for the first time. Moreover, for better comparability, the data sets of expression values and CNVs should refer to the most actual version of the mouse genome, which was UCSC mm9 in terms of CNVs, while formerly generated expression data referred to version mm8 and thus had to be updated. In total, 374 protein coding genes were found to be significantly differentially expressed in at least one of the four tested brain regions. Of those genes, 12 were carefully chosen to be validated

by qPCR based on two selection criteria. First, they had to feature different expression levels as indicated by microarray results and second, they should be part of a functional protein association network, shown by means of the STRING online software. Most of the qPCR results confirmed the expression differences detected by means of the expression microarray (see Table IV-6), thereby emphasizing the reliability of the data created by the array. The reasons for potential problems occurring during the microarray performance, the qPCR process and the probe design leading to contradictory outcome of the two methods are manifold and are not discussed here, since the focus of this study is on the functional impact of CNVs and a discourse about limitations of methods to detect expression differences would go beyond the scope of this dissertation. These topics have already been in the focus of multiple studies and were discussed in recent publications (Roberts, 2008; Vagner *et al.*, 2013).

The correlation procedure of CNVs with the expression profile in HAB/LAB mice was done in two separate steps. First, the influence of CNVs on expression was tested by the application of a two-proportion z-test. For all analyzed brain regions (CeA, BLA, Cg, PVN) it was demonstrated that a significant amount of genes showing expression differences were hosted in regions of genomic variance, with p-values up to 1.76×10^{-64}. Thus, the evidence of CNVs being able to regulate the expression of genes located inside their breakpoints was adduced. The second step focused on the question if an elevated copy number entailed an increase, and a decreased copy number a decline in expression levels, respectively. For CeA, Cg and PVN, but not for BLA a significant positive correlation was shown with respect to the nominal p-values (Pearson's product-moment correlation). In contrast, the Bonferroni corrected p-values were not significant for any of the four brain regions. This outcome is in concordance with a previous study showing only a weakly significant correlation between gene dosage and relative expression levels (Henrichsen *et al.*, 2009b) and could be explained by the CNVs' modes of action themselves (as outlined in chapter II-6), for instance by a negative feedback loop. Besides, several environmental and genetic factors like SNPs, epigenetic factors and others are known to influence gene expression and thereby might interfere with the effects mediated by CNVs.

Even though the most simple conclusion, namely that more gene copies reflect more gene expression and *vice versa*, cannot be drawn, the convincing p-values of the applied

z-test make clear that CNVs have to be considered as a serious influencing factor of gene expression and hence phenotypic variations. Therefore, it seems worthwhile to focus on CNVs and their influence on phenotypic behavior in further studies of the HAB/LAB mouse model.

2.2 A breeding approach to examine the impact of a single CNV comprising the *Glo1* locus on anxiety-related behavior

The idea of the breeding approach was to breed animals having the full genetic background of HAB mice, with exception of one locus harboring a LAB-specific CNV. The advantage of a breeding approach over biotechnological methods is the conservation of potential genetic linkage of the CNV. To go for the CNV comprising the *Glo1* and the *Dnahc8* locus was decided since *Glo1* was shown before to be linked to anxiety and depression (Hovatta *et al.*, 2005; Tanna *et al.*, 1989; Williams *et al.*, 2009) and was previously detected as a candidate gene in the HAB/LAB mouse model (Hambsch *et al.*, 2010; Kromer *et al.*, 2005; Landgraf *et al.*, 2007). The HAB- and LAB-specific loci could be easily discriminated by means of a PCR analysis using primers targeting the breakpoints of the CNV that induced a specific product only if more than one copy of the locus was present, which was just the case for the LAB-specific locus.

Animals of the sixth breeding generation, having a calculated similarity coefficient to HAB mice of about 0.98, were considered to have sufficient HAB-specific genetic background to develop the high-anxious phenotype, and therefore to examine potential effects of the CNV. If the CNV would contribute remarkably to the phenotype, the animals harboring the LAB-specific locus were expected to behave less anxious than those with the HAB-specific locus. In none of the analyzed parameters of the four performed behavioral tests (EPM, OF, FST, TST) a significant difference could be found, indicating that the examined CNV alone is not causing a modification of anxiety-related behavior. But that does not mean that the duplicated region has no effect at all. Following the idea of the formerly described Two Hit Model (Canales and Walz, 2011; Girirajan and Eichler, 2010; Girirajan *et al.*, 2010) it might be that the CNV (“first hit”) does only influence the phenotype if a “second hit” occurs, for instance another CNV or any other genetic factor.

Furthermore, it has to be considered that the phenotyped animals, that were shown to harbor the LAB-specific CNV, were heterozygous for that locus, since half of their genome was inherited from a HAB individual. Therefore, it remains possible that the effect mediated by the LAB-specific allele is masked by the HAB-specific one. Hence, in a follow up study, the animals of the next generation should be bred by intercrossing mice from the 6^{th} generation to generate individuals homozygous for the CNV locus (HAB- or LAB-specific), which should subsequently be analyzed for differences in their behavioral patterns.

The breeding approach could not only reveal a potential influence of the examined CNV, it could additionally be used to test if the phenotype of the HAB mouse line can be sustained, even though intercrossed with the LAB line in one generation. For that purpose, the time animals spent on the open arms of the EPM (the behavioral parameter the HAB/LAB mouse lines are bred based on) was compared between animals harboring the HAB-specific and LAB-specific locus, respectively. The behavior of animals with the most HAB-like genetic background (i.e. mice with the HAB-specific locus) was observed to be in the same range as the behavior of HAB mice, which usually spend about 10 to 20 seconds on the open arms (Figure IV-19A and Figure II-1). Thus, the approach emphasizes the stability of the lineage specific phenotype of HAB animals.

2.3 The effect of CNVs on anxiety-related behavior in CD-1 mice

By means of the JaxMDGA 764 CNVs could be detected in 64 CD-1 mice, as discussed in chapter 1.2. The aim of the subsequent analysis was to find those CNVs that affect the anxiety-related behavior in these mice. For this purpose, the variable regions were correlated with different behavioral parameters. But before that correlation could be performed, the problem of defining the respective copy number status, i.e. the amount of copies of a single CNV in all animals with respect to each other, had to be solved. Thus, as first step, the “simple CNV” function of the “MouseDivGeno” R package was modified to achieve the normalized intensities of each JaxMDGA probe (outlined in supplementary report S5). Then all intensity values of the probes falling into a newly defined CNV were averaged. The resulting mean intensity values were used as measurement variable of the

relative copy number of the respective region. Finally, these values were associated with multiple behavioral parameters in a statistical approach using a generalized linear model. The most important parameter was the time the animals spent on the open arm of the EPM, for which a total of 55 CNVs were shown to be significantly associated with when only considering the nominal p-value (nominal p-value < 0.05). Other 35 CNVs still exhibited a trend in the nominal p-value ($p < 0.1$). The key parameter of the EPM was termed as more important than that of other behavioral tests since the HAB/LAB breeding is based on, enabling a comparison of CNVs found to be associated in CD-1 mice to those detected in HAB/LAB mice. Nevertheless, other tests were performed and statistically analyzed, in order to complete the picture and provide the basis for further studies. Here only the results for the respective key parameters showing nominal p-values that reach significance ($p < 0.05$) are mentioned, with 25 CNVs for the time immobile in the TST, 69 CNVs for the floating time in the FST, 35 CNVs for the distance travelled in the OF, and 44 CNVs for the Cort increase in the SRT. However, all significances got lost after the application of Bonferroni correction. But does that mean that the effect of CNVs on anxiety-related behavior is negligible?

The answer is a clear “no”. Considering the full spectrum of literature available, there should be no doubt that anxiety is induced by complex molecular mechanisms that in turn are influenced by multiple genetic and environmental factors. It was postulated before that the heritability of complex traits is not likely due to some single genes but to multiple genes of small effect size (Plomin *et al.*, 2009). Thus, specific genes and genetic factors in and of itself might be less strongly associated with complex traits and diseases than particular patterns of genetic variation and environmental interaction (Rucker *et al.*, 2011). Not including at least parts of the patterns when analyzing the influence of a single factor to complex behavior might easily lead to p-values not reaching significance, which does not necessarily signify that the single factor *per se* is irrelevant. Therefore, future studies should focus on the revelation of the variation patterns. In concordance to this claim, the authors of a current review postulated that “the implication of CNV on [human] health will have to wait several large-scale correlation studies not only with one CNV but also with permutations and combinations of various likely [genetic and environmental] variations” (Almal and Padh, 2012). To date, more complex calculations of associations

including all or at least most of the possible influencing factors are not feasible since the number of these factors pushes the physical limits of calculable capacity. Recently, computational methods were developed mapping phenotypes not just to single loci in the genome but to pairs of genetic loci, thereby systematically searching for epistatic interactions (Kam-Thong *et al.*, 2012; 2011). Even though these methods are heading in the right direction, it will take several years until more complex patterns than just two interacting factors can be calculated in a cost- and time-effective manner. By then, research is constrained to the available methods and should aim for collecting as many information on influencing factors of complex traits as possible. In consequence, the goal of the here performed association study was to detect the CNVs that are most likely to contribute to the phenotype and not to prove their contribution *per se*. Therefore, even though significances of p-values got lost after correction, the CNVs with a significant nominal p-value could be accepted as interesting candidates affecting anxiety-related behavior. The fact that CNVs can indeed contribute to phenotypic variation was previously shown by others (e.g., Malhotra and Sebat, 2012) as well as in the current study of CNVs in the HAB/LAB mouse model as described above (chapter 2.1).

3 Candidate genes of anxiety-related behavior

3.1 Candidate genes in the HAB/LAB mouse model

CNVs were shown to alter gene expression in HAB/LAB mice, as discussed above (chapter 2.1). At least to some extent this effect is mediated by changes in copy number of the protein coding genes themselves. Therefore, the CNVs defined by three different detection methods were checked for harboring genes, using a list of all known mouse genes referring to NCBI build m37. For the 98 CNVs revealed by aCGH 998 genes were found to map into. Even though more CNVs were detected by JaxMDGA, fewer genes (145) overlapped these regions, which is in concordance with the fact that the total size of the CNVs was smaller as well (~ 27 Mbp against ~ 97 Mbp). The total size of the variants detected by NGS (~ 15 Mbp) was more similar to the one revealed by JaxMDGA, however the number of genes found in these CNVs resembled with 1,085 the amount of

genes overlapping the aCGH-detected CNVs. Even though it is likely that the size of the variants is overestimated by aCGH and maybe underestimated by NGS, the true situation of genomic variation is difficult to assess and therewith the definition of candidate genes. However, if genes mapped into regions where CNVs were revealed by all three methods, they were considered to be reliable candidates of influencing the behavior in HAB/LAB mice. That condition was fulfilled by 68 genes.

Besides the limitations in detection methods, another factor has to be considered when defining candidate genes. Due to a population bottleneck effect that potentially occurred by choosing a restricted proportion of the CD-1 mouse population as basis for the HAB/LAB mouse model and by the subsequent inbreeding protocol, some of the genetic variants emerging between the HAB and LAB mouse lines might have been established by chance and do not contribute to the phenotypical expression. Therefore, the 68 candidate genes were compared to the genes that mapped into CNVs of 64 CD-1 mice showing a significant association with the time the animals spent on the open arms of the EPM. In doing so, the list of candidate genes was further narrowed down to 15 genes. These 15 candidate genes were analyzed for functional annotation my means of the software tool provided at the DAVID Bioinformatics Database (da Huang *et al.*, 2009a; 2009b).

The resulting two clusters were related to the immune system and to signal transduction, respectively. It is not a new finding that the immune system can be linked to anxiety and depression (Leonard and Song, 1996; Stein, 1989). For example, it is known that there exists a relationship between anxiety/depression and organ or stem cell transplantation, which could, at least partially, be explained by the immunosuppressive medication the patients get (Cukor *et al.*, 2008; Tecchio *et al.*, 2012). Furthermore, it was observed that patients with anxiety disorder had a reduced count in $CD8^+$ T cells (T suppressor cytotoxic lymphocytes), cells that contribute to the immune response by recognizing antigen presenting cells via binding of the Class I MHC (major histocompatibility complex) molecule by the cells' T cell receptor (Atanackovic *et al.*, 2004; Van Laethem *et al.*, 2012). In Wistar rats a genetically determined relationship was suggested to exist between low activity of the immune response and a high level of active anxiety (Loskutova *et al.*, 2007). Nautiyal and colleagues found an elevated anxiety-like behavior in mast cell deficient

mice, thereby providing evidence for the behavioral importance of neuroimmune links (Nautiyal *et al.*, 2008). Although there are plenty more findings on the correlation of anxiety and/or depression and the immune system than the just mentioned, the underlying molecular processes are of a complex nature and not understood yet. Thus the question of the involvement of the candidate genes forming the here defined immune system-related cluster is not easy to answer. The function of the respective genes is described below. However, further studies need to be performed to enlighten the genes' contribution to the anxiety phenotype and the respective role the CNVs might play.

Concerning the second cluster, a connection to an anxiety mouse model was easily conceivable, at least at first glance, as signal transduction is clearly needed for the expression of any phenotypical behavior. However, checking the genes forming the cluster, one could notice that they are all linked to male fertility. Considering that there was no difference in the fertility of HAB and LAB males observed so far, this fact was a bit puzzling. Hence, a closer look at the characteristic of the respective genes was taken and is outlined below. Briefly summarized, the genes belong to two families, both detected by examinations on fertility (Hammer *et al.*, 1989; Herrmann *et al.*, 1999). Their exact function still remains ambiguous. However, their proximity to the major histocompatibility complex (MHC), a cell surface molecule involved in the immune response that is encoded by a large gene family, as well as their structure of protein kinase domains (Ma *et al.*, 2002; Redkar *et al.*, 2000) led to the assumption that the genes might be involved in important, not fertility-related molecular processes that could contribute to the immune response or manipulate behavioral expression. That hypothesis lacks proof, though. Subsequent studies should be performed to enlighten the contribution of these genes to the HAB/LAB phenotyes.

Sirpb1a and *Sirpb1c*

The signal-regulatory protein beta 1A (*Sirpb1a*), also called *Sirpb1* or *Sirpβ1*, and the signal-regulatory protein beta 1C (*Sirpb1c*) are genes whose human equivalents belong to the so called "*SIRP* family", a gene family, which was detected in 1997 with at least 15 members all encoding proteins involved in the regulation of signals defining different physiological and pathological processes (Kharitonenkov *et al.*, 1997). The *SIRP* family

belongs to the immunoglobulin (Ig) superfamily since its members possess an Ig-like domain. This domain was shown to be involved in cell-surface recognition and binding functions and appears not only in immunoglobulins but also in non-immunoglobulins. Even though most proteins with Ig-folds were found to play a role in the adaptive immune system, the abundance of the domain type reflect their usefulness in protein interactions in many cell types (Barclay, 2003; Bork *et al.*, 1994; Williams and Barclay, 1988). Although most of the studies hint to an involvement of the *SIRP* family members in the immune system it might be that it executes other functions as well, possibly through the activation of the MAPK pathway (Barclay and Brown, 2006; Hayashi *et al.*, 2004). This pathway is known to play a role in important cellular processes like cell differentiation and survival, growth control and cellular adaptation to chemical and physical stress (Chang and Karin, 2001; Cobb, 1999; Orton *et al.*, 2005; Widmann *et al.*, 1999). Additionally, the MAPK pathway was linked to anxiety and depression before (Di Benedetto *et al.*, 2009; Jurek *et al.*, 2012; Wefers *et al.*, 2012), thus, further studies on the connection of *Sirp* genes and the MAPK pathway seem promising to take the research on uncovering the molecular processes of anxiety a step forward.

Skint

Not much is known about the members of the *Skint* (selection and upkeep of intraepithelial T cells) gene family yet, whereat *Skint1* is still the best examined. Ten intact paralogous genes to *Skint1* (*Skint2 – Skint11)* were found in mice that all show expression in the thymus and/or the skin (Boyden *et al.*, 2008). To date, no detailed information on the *Skint* family members defined as candidate genes in HAB/LAB mice (*Skint3*, *Skint4*, *Skint5*, *Skint6*, *Skint9* and *Skint11*) is available. However, *Skint1* was shown to be involved in the determination of functional programs of the murine T cell development (Barbee *et al.*, 2011; Havran and Jameson, 2010; Turchinovich and Hayday, 2011). Hence, it was considered to define a critical axis of communication between thymic stroma and γδ T cell progenitors (Hayday, 2009). The γδ T cells are the prototype of "unconventional" T lymphocytes and were shown to recognize "stress antigens" like the *Rae-1* (RAE1 RNA export 1 homolog) gene product that was linked to skin carcinoma, and thus initiate a lymphoid stress-surveillance response (Girardi *et al.*, 2001; Hayday, 2000, 2009).

Considering that a response to stress must be rapid, the precursor frequency of T cells to be suitable for stress surveillance should be high, which could be indirectly achieved by the mediation of *Skint1*. Unfortunately there are no data available for the *Skint* family members detected in HAB/LAB mice. Besides, there is no indication that the lymphoid stress surveillance affects anxiety or depression, although it was supposed to be involved in tumor immunology, inflammation, allergy, autoimmunity and infectious disease (Hayday, 2009). However, the involvement of "stress antigens" and a surveillance mechanism including *Skint* genes in the phenotypical expression of anxiety is conceivable and might open a field for further investigations.

Furthermore, the *Skint* genes are, like the *Sirpb* genes, members of the immunoglobulin superfamily (Boyden *et al.*, 2008). Therefore, these genes possess an Ig-like domain through which they could act on protein interactions in many cell types (Barclay, 2003; Bork *et al.*, 1994; Williams and Barclay, 1988) and, thus, might have yet undetected functions, related or non-related to the immune system, that could somehow influence the behavioral phenotype.

Tcp10 and *Smok2*

The *Tcp10* (t-complex protein 10) genes map to the so called t-complex, a proximal part of the mouse chromosome 17 that spans about 15-20 cM and reflects a contiguous region of the MHC, an important contributor of the immune system. The t-complex is known to appear naturally in closely related variants, also referred to as t-haplotypes. These variants are held together as unified entities by four neighbouring inversions (Hammer *et al.*, 1989; Redkar *et al.*, 2000). The functions of the genes located in the t-complex are largely unknown, although some of the genes were linked to male fertility (Harrison *et al.*, 1998; Huw *et al.*, 1995; Ma *et al.*, 2002; Mazarakis *et al.*, 1991; Olds-Clarke and Johnson, 1993). It has been suggested that multiple genes within the t-complex contribute to male sterility by altering the sperm function, including motility, capacitation (a process of interaction with the oocyte) and binding to the *zona pellucida* (a glycoprotein membrane surrounding the oocyte) (Redkar *et al.*, 2000). Although not shown before, it might be that genes involved in the just mentioned mechanisms could also contribute to other molecular processes that require cell-cell interactions and thus alter e.g., the immune

response or any other phenotypical expression like anxiety-related behavior. Of course, that hypothesis is highly speculative and needs to be pursued by further experiments.

One element of the t-complex is the so called t-complex responder (Tcr) locus that influences the transmission ratio distortion (TRD). The TRD is the phenomenon that male mice heterozygous for a complete t-haplotype (t/+) preferentially transmit the t-haplotype chromosome to their progeny. However, the precise location of the responder region is not evident. A member of the *Tcp10* gene family, *Tcp10b*, was initially proposed as a candidate for Tcr (Bullard and Schimenti, 1991; Davies and Willison, 1991; Pilder *et al.*, 1992; Schimenti *et al.*, 1988) but was later on rejected due to the results of further examination (Ewulonu *et al.*, 1996). In another study large parts of the region supposed to contain the Tcr were isolated and appeared in *in vitro* experiments to be an odd serine/threonine protein kinase, created from a rearrangement of a gene designated as *Smok* (sperm motility kinase) with its neighboring gene *Rsk3* (ribosome S6 kinase). In the same study *Smok* was shown to represent a new gene family. Thus, Tcr was supposed to represent a mutant form of the *Smok* gene family and was therefore abbreviated by the symbol $Smok^{Tcr}$. Another two family members were suggested to form a subfamily, termed *Smok2*, since different to the other genes, their open reading frame was extended (Herrmann *et al.*, 1999; Redkar *et al.*, 2000). Even though no further data are available, the Smok2 proteins contain conserved residues of the serine/threonine protein kinase consensus sequence (Herrmann *et al.*, 1999) and thus a highly speculative hypothesis could be that Smok2 participates in signal transduction or other important cellular processes through the phosphorylation of distinct proteins.

3.2 Candidate genes in CD-1 mice

As described in chapter II-6, CNVs might influence behavioral phenotypes by the alteration of gene expression in multiple ways. One possibility is by directly changing the amount of specific genes *per se*. Thus, genes mapping into regions of detected CNVs were considered to be of primary interest and these candidate genes were not only revealed in HAB/LAB mice, as discussed in chapter 3.1, but also in CD-1 mice. Of the 764 CNVs detected by JaxMDGA in 64 CD-1 mice only a part could be associated to anxiety-related

behavior, measured by different parameters in several behavioral tests. Thus, 55 CNVs were associated with the percentage of time the animals spent on the open arm of the EPM. These CNVs were chosen to provide the basis for further analysis as HAB/LAB mice were bred for that behavioral parameter and therefore, the resulting genes were comparable to the HAB/LAB derived candidates. In total, 53 genes were found to map into the genomic regions of these CNVs and they were subsequently studied to identify functional clusters, using the functional annotation tool provided at the DAVID Bioinformatics database (da Huang *et al.*, 2009a; 2009b).

The same two types of clusters that were found in HAB/LAB animals were also defined here, i.e. one cluster that could be linked to the immune system and another one comprising genes that were shown to be involved in signal transduction. Many of the genes found to form the immune system cluster were also defined as candidate genes in HAB/LAB mice. A detailed description of these genes and the discussion about the relationship between the immune system and anxiety/depression could be found above (chapter 3.1).

Although designated equally, the signal transduction cluster was formed by three genes different to the ones building the equivalent cluster in HAB/LAB animals. Below, the characteristics of these genes are just mentioned briefly to give an idea on how they might influence signal transduction and thereby the behavior. A deeper insight into their molecular mechanisms requires extensive investigations and is thus subjected to advanced studies.

One of the genes, *Ranbp17* (RAN binding protein 17), was described to belong to the importin β superfamily, having a high sequence similarity to the importin β-related transport receptors, which are important mediators of the transport between nucleus and cytoplasm (Kutay *et al.*, 2000; Lee *et al.*, 2010). The role of *Ranbp17* in nuclear transport though remains to be reported. Interestingly, the transcript(s) expression of *Ranbp17* has been described to be testis-enriched in mouse tissue samples (Koch *et al.*, 2000; Lee *et al.*, 2010). That is striking, since the candidate genes of the analogues cluster in HAB/LAB mice could be linked to male fertility as well. Although it might be a coincidence, genes being involved in the alteration of fertility that also play a role in other

molecular processes altering the behavior, could indeed be a driving factor of anxiety evolution and further studies should focus on that topic.

The second gene of the signal transduction cluster, *Dock9* (dedicator of cytokinesis 9), which is also part of the immune system cluster, was shown to activate Cdc42, a Rho family small GTPase. That family is known to regulate multiple molecular processes like cell migration, cell cycle progression, gene expression, innate immunity, and bacterial and viral infections (Kwofie and Skowronski, 2008). Therefore, multiple ways on how the gene might be involved in signal transduction and thereby manipulate anxiety-related behavior are conceivable and might be the topic of future examinations.

Another gene also appearing in both clusters is *Sytl3* (synaptotagmin-like 3). Its product, a peripheral membrane protein, might be involved in vesicular trafficking (Fukuda and Mikoshiba, 2001; Kuroda *et al.*, 2002). Although it was not shown to be associated with any lentivirus infection in mammals so far, it could interact with *Rab27a,* a member of the Ras oncogene family, reported to be required for enveloped virus assembly of human cytomegalovirus (Fraile-Ramos *et al.*, 2010; White *et al.*, 2012). Furthermore, *Rab27a* was reported to regulate phagocytosis and to be involved in exosome synthesis (Ostrowski *et al.*, 2010; White *et al.*, 2012; Yokoyama *et al.*, 2011). Thus *Sytl3* could indirectly exert influence on those processes and thereby on behavioral patterns.

4 Conclusion

Despite newly developed detection methods, the generation of reliable genetic data, as basis for subsequent investigations on genetic variation influencing behavioral phenotypes, remains a huge challenge in current research. This is not only because each of the available techniques is biased in some way, but also due to the dependency of results from the applied computational method for processing the raw data. In this study, employing three of the most advanced techniques, a comprehensive catalogue of CNVs in the HAB/LAB mouse model was successfully generated.

Based on this catalogue, it could be demonstrated that CNVs contribute to the expression pattern of protein coding genes mapping into the variant regions. By the correlation of

CNVs with anxiety-related behavior in CD-1 mice, further genes of interest were discovered, thereby verifying some of the candidate genes found in HAB/LAB mice. These confirmed candidate genes could be linked to the immune system and to signal transduction by use of a functional annotation clustering method. Hence, this thesis provides not only a comprehensive catalogue of CNVs in HAB/LAB mice but also a first insight into the functionality of CNVs with respect to anxiety-related behavior, and therefore forms an important basis for the conductance of advanced studies.

VI Perspectives

This study focused on the detection of CNVs, but also examined their influence on gene expression and behavior, thereby providing a profound basis for advanced examinations. Multiple follow-up studies are conceivable and the prioritized are described below.

First, CNV data could be checked against information on SNPs in HAB/LAB mice. Although associated with behavior before, the role of SNPs, especially for the ones provoking a silent mutation, is not yet clear. One hypothesis might be that SNPs pave the way for the formation of CNVs and other structural variants when occurring at so called “hotspots”, genomic regions of high sequence similarity leading to misalignments and thus rearrangements of the genome (Gu *et al.*, 2008). A potential finding of an accumulation of SNPs near CNV breakpoints could support that hypothesis. In this context the examination of mitochondrial DNA (mtDNA) would be of interest, too, since this type of DNA could vary in copy number depending of the tissue examined (Fuke *et al.*, 2011). Thus, the detection of an enlarged occurrence of both SNPs and CNVs in the mtDNA of a specific tissue but no SNPs and no CNVs in another tissue of the same animals, would provide further evidence for SNPs being a precursor of CNV formation.

Furthermore, knowing the exact CNV breakpoints of HAB/LAB mice could ease the detection of these variants in animals of future generations. Besides, it could help to indicate, which of the applied techniques and algorithms for CNV detection performed best and thus give a hint on the reliability of variances detected in the 64 CD-1 mice by JaxMDGA. The most important, but not easiest thing to do could be to encircle the exact breakpoints of the CNVs using a PCR-based strategy, that is by designing qPCR primers mapping in the proximity of the assumed breakpoint thereby narrowing down the region to a size that could be targeted by PCR primers. A similar strategy was applied by Williams and colleagues to verify the CNV harboring the *Glo1* locus (Williams *et al.*, 2009). Subsequently performed Sanger sequencing could finally reveal the exact breakpoints. This strategy sounds easy to follow, however, successful primer design could be a time consuming task and, though a single run might not be expensive, the sum of all costs could quickly reach financial limits. Thus, it is recommended to focus on those CNVs that

seem to be of major interest, either because they are likely to be involved in the expression alteration of genes found to be of interest by other studies, or because they are likely to interact with other genetic factors like SNPs or epigenetic factors, thereby influencing the mice's anxiety-related behavior.

Further work could also focus on the detection of new candidate genes. In the here performed study only the protein coding genes overlapping regions of suggested CNVs were listed as candidate genes. Other types of genetic features like (non-) coding RNA genes or known transcription factor binding sites could be added. Additionally, since cis-acting regulators were found to act over a distance of several megabases (Kalari *et al.*, 2010; Lettice *et al.*, 2002; Nobrega *et al.*, 2003), all protein coding genes and other genetic feature types occurring at a defined distance from the CNV breakpoints could be of interest and included in the list.

Finally, a continuation of the so called "*Glo1* breeding approach" could be performed. Although no difference in the anxiety-related behavior of animals with a HAB genetic background harboring the LAB-specific *Glo1* CNV compared to those animals harboring the HAB-specific locus was shown in this study, it should be considered that the tested animals were heterozygous for that locus. It might be that behavioral changes would occur when the mice were homozygous for the respective CNV. Thus, a cross-breeding approach of the next generation of animals could result in the required homozygous animals and the performance of behavioral test might reveal a less anxious phenotype.

VII References

Alkan C., Coe B.P. and Eichler E.E. (2011): "Genome structural variation discovery and genotyping". *Nat Rev Genet* 12 (5): p. 363-376

Almal S.H. and Padh H. (2012): "Implications of gene copy-number variation in health and diseases". *J Hum Genet* 57 (1): p. 6-13

American Psychiatric Association (1994): "Diagnostic and statistical manual of mental disorders" (4th ed.).

Andrews P.W., Gangestad S.W. and Matthews D. (2002): "Adaptationism--how to carry out an exaptationist program". *Behav Brain Sci* 25 (4): p. 489-504; discussion 504-453

Andrews P.W. and Thomson J.A., Jr. (2009): "The bright side of being blue: depression as an adaptation for analyzing complex problems". *Psychol Rev* 116 (3): p. 620-654

Arlt M.F., Wilson T.E. and Glover T.W. (2012): "Replication stress and mechanisms of CNV formation". *Curr Opin Genet Dev* 22 (3): p. 204-210

Atanackovic D., Kroger H., Serke S. and Deter H.C. (2004): "Immune parameters in patients with anxiety or depression during psychotherapy". *J Affect Disord* 81 (3): p. 201-209

Bailey J.A., Church D.M., Ventura M., Rocchi M. and Eichler E.E. (2004): "Analysis of segmental duplications and genome assembly in the mouse". *Genome Res* 14 (5): p. 789-801

Banerjee S., Oldridge D., Poptsova M., Hussain W.M., Chakravarty D. and Demichelis F. (2011): "A computational framework discovers new copy number variants with functional importance". *PLoS One* 6 (3): p. e17539

Barbee S.D., Woodward M.J., Turchinovich G., Mention J.J., Lewis J.M., Boyden L.M., Lifton R.P., Tigelaar R. and Hayday A.C. (2011): "Skint-1 is a highly specific, unique selecting component for epidermal T cells". *Proc Natl Acad Sci U S A* 108 (8): p. 3330-3335

Barclay A.N. (2003): "Membrane proteins with immunoglobulin-like domains--a master superfamily of interaction molecules". *Semin Immunol* 15 (4): p. 215-223

Barclay A.N. and Brown M.H. (2006): "The SIRP family of receptors and immune regulation". *Nat Rev Immunol* 6 (6): p. 457-464

Barlow D.H. (2000): "Unraveling the mysteries of anxiety and its disorders from the perspective of emotion theory". *Am Psychol* 55 (11): p. 1247-1263

Beck A.T. and Emery G. (2005): "Anxiety Disorders and Phobias: A Cognitive Perspective" ed.). *Basic Books*, New York

Belzung C. and Griebel G. (2001): "Measuring normal and pathological anxiety-like behaviour in mice: a review". *Behav Brain Res* 125 (1-2): p. 141-149

Belzung C. and Philippot P. (2007): "Anxiety from a phylogenetic perspective: is there a qualitative difference between human and animal anxiety?" *Neural Plast* 2007 p. 59676

Benelli M., Marseglia G., Nannetti G., Paravidino R., Zara F., Bricarelli F.D., Torricelli F. and Magi A. (2010): "A very fast and accurate method for calling aberrations in array-CGH data". *Biostatistics* 11 (3): p. 515-518

Blanchard D.C., Griebel G. and Blanchard R.J. (2003): "The Mouse Defense Test Battery: pharmacological and behavioral assays for anxiety and panic". *Eur J Pharmacol* 463 (1-3): p. 97-116

Bochukova E.G., Huang N., Keogh J., Henning E., Purmann C., Blaszczyk K., Saeed S., Hamilton-Shield J., Clayton-Smith J., O'Rahilly S., Hurles M.E. and Farooqi I.S. (2010): "Large, rare chromosomal deletions associated with severe early-onset obesity". *Nature* 463 (7281): p. 666-670

Bork P., Holm L. and Sander C. (1994): "The immunoglobulin fold. Structural classification, sequence patterns and common core". *J Mol Biol* 242 (4): p. 309-320

Boyden L.M., Lewis J.M., Barbee S.D., Bas A., Girardi M., Hayday A.C., Tigelaar R.E. and Lifton R.P. (2008): "Skint1, the prototype of a newly identified immunoglobulin superfamily gene cluster, positively selects epidermal gammadelta T cells". *Nat Genet* 40 (5): p. 656-662

Brasch-Andersen C., Christiansen L., Tan Q., Haagerup A., Vestbo J. and Kruse T.A. (2004): "Possible gene dosage effect of glutathione-S-transferases on atopic asthma: using real-time PCR for quantification of GSTM1 and GSTT1 gene copy numbers". *Hum Mutat* 24 (3): p. 208-214

Bullard D.C. and Schimenti J.C. (1991): "Molecular structure of Tcp-10 genes from the t complex responder locus". *Mamm Genome* 1 (4): p. 228-234

Cahan P., Godfrey L.E., Eis P.S., Richmond T.A., Selzer R.R., Brent M., McLeod H.L., Ley T.J. and Graubert T.A. (2008): "wuHMM: a robust algorithm to detect DNA copy number variation using long oligonucleotide microarray data". *Nucleic Acids Res* 36 (7): p. e41

Cahan P., Li Y., Izumi M. and Graubert T.A. (2009): "The impact of copy number variation on local gene expression in mouse hematopoietic stem and progenitor cells". *Nat Genet* 41 (4): p. 430-437

Canales C.P. and Walz K. (2011): "Copy number variation and susceptibility to complex traits". *EMBO Mol Med* 3 (1): p. 1-4

Chaignat E., Yahya-Graison E.A., Henrichsen C.N., Chrast J., Schutz F., Pradervand S. and Reymond A. (2011): "Copy number variation modifies expression time courses". *Genome Res* 21 (1): p. 106-113

Chang L. and Karin M. (2001): "Mammalian MAP kinase signalling cascades". *Nature* 410 (6824): p. 37-40

Chen D., Katdare A. and Lucas N. (2006): "Chemosignals of fear enhance cognitive performance in humans". *Chem Senses* 31 (5): p. 415-423

Chiong W. (2011): "The self: from philosophy to cognitive neuroscience". *Neurocase* 17 (3): p. 190-200

Clement Y., Joubert C., Kopp C., Lepicard E.M., Venault P., Misslin R., Cadot M. and Chapouthier G. (2007): "Anxiety in mice: a principal component analysis study". *Neural Plast* 2007 p. 35457

Cobb M.H. (1999): "MAP kinase pathways". *Prog Biophys Mol Biol* 71 (3-4): p. 479-500

Coffa J. and van den Berg J. (2011): "Analysis of MLPA Data Using Novel Software Coffalyser.NET by MRC-Holland". In *Modern Approaches To Quality Control* (edited by Badr Eldin A.), p. *InTech*,

Cowen P.J. (2009): "Not fade away: the HPA axis and depression". *Psychol Med* p. 1-4

Cryan J.F. and Holmes A. (2005): "The ascent of mouse: advances in modelling human depression and anxiety". *Nat Rev Drug Discov* 4 (9): p. 775-790

Cukor D., Newville H. and Jindal R. (2008): "Depression and immunosuppressive medication adherence in kidney transplant patients". *Gen Hosp Psychiatry* 30 (4): p. 386-387

Cutler G., Marshall L.A., Chin N., Baribault H. and Kassner P.D. (2007): "Significant gene content variation characterizes the genomes of inbred mouse strains". *Genome Res* 17 (12): p. 1743-1754

Czibere L. (2008): "Assessing the complex nature of behavior: Sequence-based and transcriptomic analyses in a mouse model of extremes in trait anxiety". *Doctoral thesis.* Ludwig-Maximilians-Universität, Munich, Germany

Czibere L., Baur L.A., Wittmann A., Gemmeke K., Steiner A., Weber P., Putz B., Ahmad N., Bunck M., Graf C., Widner R., Kuhne C., Panhuysen M., Hambsch B., Rieder G., Reinheckel T., Peters C., Holsboer F., Landgraf R. and Deussing J.M. (2011): "Profiling trait anxiety: transcriptome analysis reveals cathepsin B (Ctsb) as a novel candidate gene for emotionality in mice". *PLoS One* 6 (8): p. e23604

da Huang W., Sherman B.T. and Lempicki R.A. (2009a): "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources". *Nat Protoc* 4 (1): p. 44-57

da Huang W., Sherman B.T., Zheng X., Yang J., Imamichi T., Stephens R. and Lempicki R.A. (2009b): "Extracting biological meaning from large gene lists with DAVID". *Curr Protoc Bioinformatics* Chapter 13 p. Unit 13 11

Davies P.O. and Willison K.R. (1991): "Sequence of the t complex Tcp-10at gene and examination of the Tcp-10t gene family". *Mamm Genome* 1 (4): p. 235-241

Deussing J.M. and Wurst W. (2005): "Dissecting the genetic effect of the CRH system on anxiety and stress-related behaviour". *C R Biol* 328 (2): p. 199-212

Di Benedetto B., Kallnik M., Weisenhorn D.M., Falls W.A., Wurst W. and Holter S.M. (2009): "Activation of ERK/MAPK in the lateral amygdala of the mouse is required for acquisition of a fear-potentiated startle response". *Neuropsychopharmacology* 34 (2): p. 356-366

Ditzen C., Jastorff A.M., Kessler M.S., Bunck M., Teplytska L., Erhardt A., Kromer S.A., Varadarajulu J., Targosz B.S., Sayan-Ayata E.F., Holsboer F., Landgraf R. and Turck C.W. (2006): "Protein biomarkers in a mouse model of extremes in trait anxiety". *Mol Cell Proteomics* 5 (10): p. 1914-1920

Ditzen C., Varadarajulu J., Czibere L., Gonik M., Targosz B.S., Hambsch B., Bettecken T., Kessler M.S., Frank E., Bunck M., Teplytska L., Erhardt A., Holsboer F., Muller-Myhsok B., Landgraf R. and Turck C.W. (2009): "Proteomic-based genotyping in a mouse model of trait anxiety exposes disease-relevant pathways". *Mol Psychiatry* 15 (7): p. 702 - 711

Duncan I.W. (2002): "Transvection effects in Drosophila". *Annu Rev Genet* 36 p. 521-556

El Yacoubi M. and Vaugeois J.M. (2007): "Genetic rodent models of depression". *Curr Opin Pharmacol* 7 (1): p. 3-7

Ewulonu U.K., Schimenti K., Kuemerle B., Magnuson T. and Schimenti J. (1996): "Targeted mutagenesis of a candidate t complex responder gene in mouse t haplotypes does not eliminate transmission ratio distortion". *Genetics* 144 (2): p. 785-792

Faravelli C., Lo Sauro C., Lelli L., Pietrini F., Lazzeretti L., Godini L., Benni L., Fioravanti G., Talamba G.A., Castellini G. and Ricca V. (2012): "The Role of Life Events and HPA Axis in Anxiety Disorders: A Review". *Curr Pharm Des* 18 (35): p. 5663-5674

Feuk L., Carson A.R. and Scherer S.W. (2006): "Structural variation in the human genome". *Nat Rev Genet* 7 (2): p. 85-97

Finn D.A., Rutledge-Gorman M.T. and Crabbe J.C. (2003): "Genetic animal models of anxiety". *Neurogenetics* 4 (3): p. 109-135

Flicek P., Amode M.R., Barrell D., Beal K., Brent S., Chen Y., Clapham P., Coates G., Fairley S., Fitzgerald S., Gordon L., Hendrix M., Hourlier T., Johnson N., Kahari A., Keefe D., Keenan S., Kinsella R., Kokocinski F., Kulesha E., Larsson P., Longden I., McLaren W., Overduin B., Pritchard B., Riat H.S., Rios D., Ritchie G.R., Ruffier M., Schuster M., Sobral D., Spudich G., Tang Y.A., Trevanion S., Vandrovcova J., Vilella A.J., White S., Wilder S.P., Zadissa A., Zamora J., Aken B.L., Birney E., Cunningham F., Dunham I., Durbin R., Fernandez-Suarez X.M., Herrero J., Hubbard T.J., Parker A., Proctor G., Vogel J. and Searle S.M. (2011): "Ensembl 2011". *Nucleic Acids Res* 39 (Database issue): p. D800-806

Fraile-Ramos A., Cepeda V., Elstak E. and van der Sluijs P. (2010): "Rab27a is required for human cytomegalovirus assembly". *PLoS One* 5 (12): p. e15318

Fuke S., Kubota-Sakashita M., Kasahara T., Shigeyoshi Y. and Kato T. (2011): "Regional variation in mitochondrial DNA copy number in mouse brain". *Biochim Biophys Acta* 1807 (3): p. 270-274

Fukuda M. and Mikoshiba K. (2001): "Synaptotagmin-like protein 1-3: a novel family of C-terminal-type tandem C2 proteins". *Biochem Biophys Res Commun* 281 (5): p. 1226-1233

Gamazon E.R., Nicolae D.L. and Cox N.J. (2011): "A study of CNVs as trait-associated polymorphisms and as expression quantitative trait loci". *PLoS Genet* 7 (2): p. e1001292

Girardi M., Oppenheim D.E., Steele C.R., Lewis J.M., Glusac E., Filler R., Hobby P., Sutton B., Tigelaar R.E. and Hayday A.C. (2001): "Regulation of cutaneous malignancy by gammadelta T cells". *Science* 294 (5542): p. 605-609

Girirajan S., Campbell C.D. and Eichler E.E. (2011): "Human Copy Number Variation and Complex Genetic Disease". *Annu Rev Genet* p.

Girirajan S. and Eichler E.E. (2010): "Phenotypic variability and genetic susceptibility to genomic disorders". *Hum Mol Genet* 19 (R2): p. R176-187

Girirajan S., Rosenfeld J.A., Cooper G.M., Antonacci F., Siswara P., Itsara A., Vives L., Walsh T., McCarthy S.E., Baker C., Mefford H.C., Kidd J.M., Browning S.R., Browning B.L., Dickel D.E., Levy D.L., Ballif B.C., Platky K., Farber D.M., Gowans G.C., Wetherbee J.J., Asamoah A., Weaver D.D., Mark P.R., Dickerson J., Garg B.P., Ellingwood S.A., Smith R., Banks V.C., Smith W., McDonald M.T., Hoo J.J., French B.N., Hudson C., Johnson J.P., Ozmore J.R., Moeschler J.B., Surti U., Escobar L.F., El-Khechen D., Gorski J.L., Kussmann J., Salbert B., Lacassie Y., Biser A., McDonald-McGinn D.M., Zackai E.H., Deardorff M.A., Shaikh T.H., Haan E., Friend K.L., Fichera M., Romano C., Gecz J., DeLisi L.E., Sebat J., King M.C., Shaffer L.G. and Eichler E.E. (2010): "A recurrent 16p12.1 microdeletion supports a two-hit model for severe developmental delay". *Nat Genet* 42 (3): p. 203-209

Graubert T.A., Cahan P., Edwin D., Selzer R.R., Richmond T.A., Eis P.S., Shannon W.D., Li X., McLeod H.L., Cheverud J.M. and Ley T.J. (2007): "A high-resolution map of segmental DNA copy number variation in the mouse genome". *PLoS Genet* 3 (1): p. e3

Gross C. and Hen R. (2004): "The developmental origins of anxiety". *Nat Rev Neurosci* 5 (7): p. 545-552

Gu W., Zhang F. and Lupski J.R. (2008): "Mechanisms for human genomic rearrangements". *Pathogenetics* 1 (1): p. 4

Haegler K., Zernecke R., Kleemann A.M., Albrecht J., Pollatos O., Bruckmann H. and Wiesmann M. (2010): "No fear no risk! Human risk behavior is affected by chemosensory anxiety signals". *Neuropsychologia* 48 (13): p. 3901-3908

Hambsch B., Chen B.G., Brenndorfer J., Meyer M., Avrabos C., Maccarrone G., Liu R.H., Eder M., Turck C.W. and Landgraf R. (2010): "Methylglyoxal-mediated anxiolysis

involves increased protein modification and elevated expression of glyoxalase 1 in the brain". *J Neurochem* 113 (5): p. 1240-1251

Hammer M.F., Schimenti J. and Silver L.M. (1989): "Evolution of mouse chromosome 17 and the origin of inversions associated with t haplotypes". *Proc Natl Acad Sci U S A* 86 (9): p. 3261-3265

Harrison A., Olds-Clarke P. and King S.M. (1998): "Identification of the t complex-encoded cytoplasmic dynein light chain tctex1 in inner arm I1 supports the involvement of flagellar dyneins in meiotic drive". *J Cell Biol* 140 (5): p. 1137-1147

Hastings P.J., Ira G. and Lupski J.R. (2009): "A microhomology-mediated break-induced replication model for the origin of human copy number variation". *PLoS Genet* 5 (1): p. e1000327

Havran W.L. and Jameson J.M. (2010): "Epidermal T cells and wound healing". *J Immunol* 184 (10): p. 5423-5428

Hayashi A., Ohnishi H., Okazawa H., Nakazawa S., Ikeda H., Motegi S., Aoki N., Kimura S., Mikuni M. and Matozaki T. (2004): "Positive regulation of phagocytosis by SIRPbeta and its signaling mechanism in macrophages". *J Biol Chem* 279 (28): p. 29450-29460

Hayday A.C. (2000): "[gamma][delta] cells: a right time and a right place for a conserved third way of protection". *Annu Rev Immunol* 18 p. 975-1026

Hayday A.C. (2009): "Gammadelta T cells and the lymphoid stress-surveillance response". *Immunity* 31 (2): p. 184-196

Henrichsen C.N., Chaignat E. and Reymond A. (2009a): "Copy number variants, diseases and gene expression". *Hum Mol Genet* 18 (R1): p. R1-8

Henrichsen C.N., Vinckenbosch N., Zollner S., Chaignat E., Pradervand S., Schutz F., Ruedi M., Kaessmann H. and Reymond A. (2009b): "Segmental copy number variation shapes tissue transcriptomes". *Nat Genet* 41 (4): p. 424-429

Herrmann B.G., Koschorz B., Wertz K., McLaughlin K.J. and Kispert A. (1999): "A protein kinase encoded by the t complex responder gene causes non-mendelian inheritance". *Nature* 402 (6758): p. 141-146

Hettema J.M., Kuhn J.W., Prescott C.A. and Kendler K.S. (2006): "The impact of generalized anxiety disorder and stressful life events on risk for major depressive episodes". *Psychol Med* 36 (6): p. 789-795

Hettema J.M., Prescott C.A. and Kendler K.S. (2003): "The effects of anxiety, substance use and conduct disorders on risk of major depressive disorder". *Psychol Med* 33 (8): p. 1423-1432

Hogg S. (1996): "A review of the validity and variability of the elevated plus-maze as an animal model of anxiety". *Pharmacol Biochem Behav* 54 (1): p. 21-30

Holmes P.V. (2003): "Rodent models of depression: reexamining validity without anthropomorphic inference". *Crit Rev Neurobiol* 15 (2): p. 143-174

Hovatta I. and Barlow C. (2008): "Molecular genetics of anxiety in mice and men". *Ann Med* 40 (2): p. 92-109

Hovatta I., Tennant R.S., Helton R., Marr R.A., Singer O., Redwine J.M., Ellison J.A., Schadt E.E., Verma I.M., Lockhart D.J. and Barlow C. (2005): "Glyoxalase 1 and glutathione reductase 1 regulate anxiety in mice". *Nature* 438 (7068): p. 662-666

Hranov L.G. (2007): "Comorbid anxiety and depression: illumination of a controversy". *International Journal of Psychiatry in Clinical Practice* 11 (3): p. 171-189

Huber W., von Heydebreck A., Sultmann H., Poustka A. and Vingron M. (2002): "Variance stabilization applied to microarray data calibration and to the quantification of differential expression". *Bioinformatics* 18 Suppl 1 p. S96-104

Huw L.Y., Goldsborough A.S., Willison K. and Artzt K. (1995): "Tctex2: a sperm tail surface protein mapping to the t-complex". *Dev Biol* 170 (1): p. 183-194

Ionescu I.A., Dine J., Yen Y.C., Buell D.R., Herrmann L., Holsboer F., Eder M., Landgraf R. and Schmidt U. (2012): "Intranasally administered neuropeptide S (NPS) exerts anxiolytic effects following internalization into NPS receptor-expressing neurons". *Neuropsychopharmacology* 37 (6): p. 1323-1337

Jeon J.P., Shim S.M., Nam H.Y., Ryu G.M., Hong E.J., Kim H.L. and Han B.G. (2010): "Copy number variation at leptin receptor gene locus associated with metabolic traits and the risk of type 2 diabetes mellitus". *BMC Genomics* 11 p. 426

Jurek B., Slattery D.A., Maloumby R., Hillerer K., Koszinowski S., Neumann I.D. and van den Burg E.H. (2012): "Differential contribution of hypothalamic MAPK activity to anxiety-like behaviour in virgin and lactating rats". *PLoS One* 7 (5): p. e37060

Kalari K.R., Hebbring S.J., Chai H.S., Li L., Kocher J.P., Wang L. and Weinshilboum R.M. (2010): "Copy number variation and cytidine analogue cytotoxicity: a genome-wide association approach". *BMC Genomics* 11 p. 357

Kallen V.L., Tulen J.H., Utens E.M., Treffers P.D., De Jong F.H. and Ferdinand R.F. (2008): "Associations between HPA axis functioning and level of anxiety in children and adolescents with an anxiety disorder". *Depress Anxiety* 25 (2): p. 131-141

Kam-Thong T., Azencott C.A., Cayton L., Putz B., Altmann A., Karbalai N., Samann P.G., Scholkopf B., Muller-Myhsok B. and Borgwardt K.M. (2012): "GLIDE: GPU-based linear regression for detection of epistasis". *Hum Hered* 73 (4): p. 220-236

Kam-Thong T., Czamara D., Tsuda K., Borgwardt K., Lewis C.M., Erhardt-Lehmann A., Hemmer B., Rieckmann P., Daake M., Weber F., Wolf C., Ziegler A., Putz B., Holsboer F., Scholkopf B. and Muller-Myhsok B. (2011): "EPIBLASTER-fast exhaustive two-locus epistasis detection strategy using graphical processing units". *Eur J Hum Genet* p.

Kawamura Y., Otowa T., Koike A., Sugaya N., Yoshida E., Yasuda S., Inoue K., Takei K., Konishi Y., Tanii H., Shimada T., Tochigi M., Kakiuchi C., Umekage T., Liu X., Nishida N., Tokunaga K., Kuwano R., Okazaki Y., Kaiya H. and Sasaki T. (2011): "A genome-wide CNV association study on panic disorder in a Japanese population". *J Hum Genet* 56 (12): p. 852-856

Kendler K.S. and Prescott C.A. (2006): "Genes, environment, and psychopathology: Understanding the causes of psychiatric and substance use disorders" ed.). *The Guilford Press*, New York

Kennison J.A. and Southworth J.W. (2002): "Transvection in Drosophila". *Adv Genet* 46 p. 399-420

Kent W.J., Sugnet C.W., Furey T.S., Roskin K.M., Pringle T.H., Zahler A.M. and Haussler D. (2002): "The human genome browser at UCSC". *Genome Res* 12 (6): p. 996-1006

Kessler M.S., Murgatroyd C., Bunck M., Czibere L., Frank E., Jacob W., Horvath C., Muigg P., Holsboer F., Singewald N., Spengler D. and Landgraf R. (2007): "Diabetes insipidus and, partially, low anxiety-related behaviour are linked to a SNP-associated vasopressin deficit in LAB mice". *Eur J Neurosci* 26 (10): p. 2857-2864

Kessler R.C., Chiu W.T., Demler O., Merikangas K.R. and Walters E.E. (2005): "Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication". *Arch Gen Psychiatry* 62 (6): p. 617-627

Kharitonenkov A., Chen Z., Sures I., Wang H., Schilling J. and Ullrich A. (1997): "A family of proteins that inhibit signalling through tyrosine kinase receptors". *Nature* 386 (6621): p. 181-186

Kim P.M., Lam H.Y., Urban A.E., Korbel J.O., Affourtit J., Grubert F., Chen X., Weissman S., Snyder M. and Gerstein M.B. (2008): "Analysis of copy number variants and segmental duplications in the human genome: Evidence for a change in the process of formation in recent evolutionary history". *Genome Res* 18 (12): p. 1865-1874

Kiyokawa Y., Shimozuru M., Kikusui T., Takeuchi Y. and Mori Y. (2006): "Alarm pheromone increases defensive and risk assessment behaviors in male rats". *Physiol Behav* 87 (2): p. 383-387

Kleinjan D.A. and van Heyningen V. (2005): "Long-range control of gene expression: emerging mechanisms and disruption in disease". *Am J Hum Genet* 76 (1): p. 8-32

Klopocki E. and Mundlos S. (2011): "Copy-number variations, noncoding sequences, and human phenotypes". *Annu Rev Genomics Hum Genet* 12 p. 53-72

Koch P., Bohlmann I., Schafer M., Hansen-Hagge T.E., Kiyoi H., Wilda M., Hameister H., Bartram C.R. and Janssen J.W. (2000): "Identification of a novel putative Ran-binding protein and its close homologue". *Biochem Biophys Res Commun* 278 (1): p. 241-249

Krishnan V. and Nestler E.J. (2008): "The molecular neurobiology of depression". *Nature* 455 (7215): p. 894-902

Kromer S.A., Kessler M.S., Milfay D., Birg I.N., Bunck M., Czibere L., Panhuysen M., Putz B., Deussing J.M., Holsboer F., Landgraf R. and Turck C.W. (2005): "Identification of glyoxalase-I as a protein marker in a mouse model of extremes in trait anxiety". *J Neurosci* 25 (17): p. 4375-4384

Krömer S.A., Kessler M.S., Milfay D., Birg I.N., Bunck M., Czibere L., Panhuysen M., Putz B., Deussing J.M., Holsboer F., Landgraf R. and Turck C.W. (2005): "Identification of glyoxalase-I as a protein marker in a mouse model of extremes in trait anxiety". *J Neurosci* 25 (17): p. 4375-4384

Kuroda T.S., Fukuda M., Ariga H. and Mikoshiba K. (2002): "The Slp homology domain of synaptotagmin-like proteins 1-4 and Slac2 functions as a novel Rab27A binding domain". *J Biol Chem* 277 (11): p. 9212-9218

Kutay U., Hartmann E., Treichel N., Calado A., Carmo-Fonseca M., Prehn S., Kraft R., Gorlich D. and Bischoff F.R. (2000): "Identification of two novel RanGTP-binding proteins belonging to the importin beta superfamily". *J Biol Chem* 275 (51): p. 40163-40168

Kwofie M.A. and Skowronski J. (2008): "Specific recognition of Rac2 and Cdc42 by DOCK2 and DOCK9 guanine nucleotide exchange factors". *J Biol Chem* 283 (6): p. 3088-3096

Lachman H.M., Pedrosa E., Petruolo O.A., Cockerham M., Papolos A., Novak T., Papolos D.F. and Stopkova P. (2007): "Increase in GSK3beta gene copy number variation in bipolar disorder". *Am J Med Genet B Neuropsychiatr Genet* 144B (3): p. 259-265

Lacoste A., Malham S.K., Cueff A. and Poulet S.A. (2001): "Stress-induced catecholamine changes in the hemolymph of the oyster Crassostrea gigas". *Gen Comp Endocrinol* 122 (2): p. 181-188

Lai W., Choudhary V. and Park P.J. (2008): "CGHweb: a tool for comparing DNA copy number segmentations from multiple algorithms". *Bioinformatics* 24 (7): p. 1014-1015

Lam K.W. and Jeffreys A.J. (2006): "Processes of copy-number change in human DNA: the dynamics of {alpha}-globin gene deletion". *Proc Natl Acad Sci U S A* 103 (24): p. 8921-8927

Landgraf R., Kessler M.S., Bunck M., Murgatroyd C., Spengler D., Zimbelmann M., Nussbaumer M., Czibere L., Turck C.W., Singewald N., Rujescu D. and Frank E. (2007): "Candidate genes of anxiety-related behavior in HAB/LAB rats and mice: focus on vasopressin and glyoxalase-I". *Neurosci Biobehav Rev* 31 (1): p. 89-102

Landgraf R. and Wigger A. (2002): "High vs low anxiety-related behavior rats: an animal model of extremes in trait anxiety". *Behav Genet* 32 (5): p. 301-314

Lane R.D., Reiman E.M., Axelrod B., Yun L.S., Holmes A. and Schwartz G.E. (1998): "Neural correlates of levels of emotional awareness. Evidence of an interaction between emotion and attention in the anterior cingulate cortex". *J Cogn Neurosci* 10 (4): p. 525-535

Lawford B.R., Young R., Noble E.P., Kann B. and Ritchie T. (2006): "The D2 dopamine receptor (DRD2) gene is associated with co-morbid depression, anxiety and social dysfunction in untreated veterans with post-traumatic stress disorder". *Eur Psychiatry* 21 (3): p. 180-185

Lee J.A., Carvalho C.M. and Lupski J.R. (2007): "A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders". *Cell* 131 (7): p. 1235-1247

Lee J.H., Zhou S. and Smas C.M. (2010): "Identification of RANBP16 and RANBP17 as novel interaction partners for the bHLH transcription factor E12". *J Cell Biochem* 111 (1): p. 195-206

Lee K.W., Woon P.S., Teo Y.Y. and Sim K. (2012): "Genome wide association studies (GWAS) and copy number variation (CNV) studies of the major psychoses: what have we learnt?" *Neurosci Biobehav Rev* 36 (1): p. 556-571

Leonard B.E. and Song C. (1996): "Stress and the immune system in the etiology of anxiety and depression". *Pharmacol Biochem Behav* 54 (1): p. 299-303

Lettice L.A., Horikoshi T., Heaney S.J., van Baren M.J., van der Linde H.C., Breedveld G.J., Joosse M., Akarsu N., Oostra B.A., Endo N., Shibata M., Suzuki M., Takahashi E., Shinka T., Nakahori Y., Ayusawa D., Nakabayashi K., Scherer S.W., Heutink P., Hill R.E. and Noji S. (2002): "Disruption of a long-range cis-acting regulator for Shh causes preaxial polydactyly". *Proc Natl Acad Sci U S A* 99 (11): p. 7548-7553

Levy D., Ronemus M., Yamrom B., Lee Y.H., Leotta A., Kendall J., Marks S., Lakshmi B., Pai D., Ye K., Buja A., Krieger A., Yoon S., Troge J., Rodgers L., Iossifov I. and Wigler M. (2011): "Rare de novo and transmitted copy-number variation in autistic spectrum disorders". *Neuron* 70 (5): p. 886-897

Lewis E.B. (1954): "The theory and application of a new method of detecting chromosomal rearrangements in Drosophila melanogaster". *The American Naturalist* 88 (841): p. 225-239

Lieber M.R., Ma Y., Pannicke U. and Schwarz K. (2003): "Mechanism and regulation of human non-homologous DNA end-joining". *Nat Rev Mol Cell Biol* 4 (9): p. 712-720

Liebsch G., Montkowski A., Holsboer F. and Landgraf R. (1998): "Behavioural profiles of two Wistar rat lines selectively bred for high or low anxiety-related behaviour". *Behav Brain Res* 94 (2): p. 301-310

Lister R.G. (1987): "The use of a plus-maze to measure anxiety in the mouse". *Psychopharmacology (Berl)* 92 (2): p. 180-185

Liu H., Huang J., Wang J., Jiang S., Bailey A.S., Goldman D.C., Welcker M., Bedell V., Slovak M.L., Clurman B., Thayer M., Fleming W.H. and Epner E. (2008): "Transvection mediated by the translocated cyclin D1 locus in mantle cell lymphoma". *J Exp Med* 205 (8): p. 1843-1858

Livak K.J. and Schmittgen T.D. (2001): "Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method". *Methods* 25 (4): p. 402-408

Loskutova L.V., Idova G.V. and Gevorgyan M.M. (2007): "Immune response in Wistar rats with high and low level of situational anxiety". *Bull Exp Biol Med* 144 (5): p. 706-708

Lupski J.R., de Oca-Luna R.M., Slaugenhaupt S., Pentao L., Guzzetta V., Trask B.J., Saucedo-Cardenas O., Barker D.F., Killian J.M., Garcia C.A., Chakravarti A. and Patel P.I. (1991): "DNA duplication associated with Charcot-Marie-Tooth disease type 1A". *Cell* 66 (2): p. 219-232

Lupski J.R., Wise C.A., Kuwano A., Pentao L., Parke J.T., Glaze D.G., Ledbetter D.H., Greenberg F. and Patel P.I. (1992): "Gene dosage is a mechanism for Charcot-Marie-Tooth disease type 1A". *Nat Genet* 1 (1): p. 29-33

Ma Y., Zhang S., Xia Q., Zhang G., Huang X., Huang M., Xiao C., Pan A., Sun Y., Lebo R. and Milunsky A. (2002): "Molecular characterization of the TCP11 gene which is the human homologue of the mouse gene encoding the receptor of fertilization promoting peptide". *Mol Hum Reprod* 8 (1): p. 24-31

Malhotra D., McCarthy S., Michaelson J.J., Vacic V., Burdick K.E., Yoon S., Cichon S., Corvin A., Gary S., Gershon E.S., Gill M., Karayiorgou M., Kelsoe J.R., Krastoshevsky O., Krause V., Leibenluft E., Levy D.L., Makarov V., Bhandari A., Malhotra A.K., McMahon F.J., Nothen M.M., Potash J.B., Rietschel M., Schulze T.G. and Sebat J. (2011): "High frequencies of de novo CNVs in bipolar disorder and schizophrenia". *Neuron* 72 (6): p. 951-963

Malhotra D. and Sebat J. (2012): "CNVs: harbingers of a rare variant revolution in psychiatric genetics". *Cell* 148 (6): p. 1223-1241

Maschwitz U.W. (1966): "Alarm substances and alarm behavior in social insects". *Vitam Horm* 24 p. 267-290

Mazarakis N.D., Nelki D., Lyon M.F., Ruddy S., Evans E.P., Freemont P. and Dudley K. (1991): "Isolation and characterisation of a testis-expressed developmentally regulated gene from the distal inversion of the mouse t-complex". *Development* 111 (2): p. 561-571

McClellan J.M., Susser E. and King M.C. (2007): "Schizophrenia: a common disease caused by multiple rare alleles". *Br J Psychiatry* 190 p. 194-199

Merali Z., Khan S., Michaud D.S., Shippy S.A. and Anisman H. (2004): "Does amygdaloid corticotropin-releasing hormone (CRH) mediate anxiety-like behaviors?

Dissociation of anxiogenic effects and CRH release". *Eur J Neurosci* 20 (1): p. 229-239

Mervis C.B., Dida J., Lam E., Crawford-Zelli N.A., Young E.J., Henderson D.R., Onay T., Morris C.A., Woodruff-Borden J., Yeomans J. and Osborne L.R. (2012): "Duplication of GTF2I results in separation anxiety in mice and humans". *Am J Hum Genet* 90 (6): p. 1064-1070

Mills R.E., Walter K., Stewart C., Handsaker R.E., Chen K., Alkan C., Abyzov A., Yoon S.C., Ye K., Cheetham R.K., Chinwalla A., Conrad D.F., Fu Y., Grubert F., Hajirasouliha I., Hormozdiari F., Iakoucheva L.M., Iqbal Z., Kang S., Kidd J.M., Konkel M.K., Korn J., Khurana E., Kural D., Lam H.Y., Leng J., Li R., Li Y., Lin C.Y., Luo R., Mu X.J., Nemesh J., Peckham H.E., Rausch T., Scally A., Shi X., Stromberg M.P., Stutz A.M., Urban A.E., Walker J.A., Wu J., Zhang Y., Zhang Z.D., Batzer M.A., Ding L., Marth G.T., McVean G., Sebat J., Snyder M., Wang J., Eichler E.E., Gerstein M.B., Hurles M.E., Lee C., McCarroll S.A. and Korbel J.O. (2011): "Mapping copy number variation by population-scale genome sequencing". *Nature* 470 (7332): p. 59-65

Mineka S., Watson D. and Clark L.A. (1998): "Comorbidity of anxiety and unipolar mood disorders". *Annu Rev Psychol* 49 p. 377-412

Molina E., Cervilla J., Rivera M., Torres F., Bellon J.A., Moreno B., King M., Nazareth I. and Gutierrez B. (2011): "Polymorphic variation at the serotonin 1-A receptor gene is associated with comorbid depression and generalized anxiety". *Psychiatr Genet* 21 (4): p. 195-201

Nautiyal K.M., Ribeiro A.C., Pfaff D.W. and Silver R. (2008): "Brain mast cells link the immune system to anxiety-like behavior". *Proc Natl Acad Sci U S A* 105 (46): p. 18053-18057

Nobrega M.A., Ovcharenko I., Afzal V. and Rubin E.M. (2003): "Scanning human gene deserts for long-range enhancers". *Science* 302 (5644): p. 413

Olds-Clarke P. and Johnson L.R. (1993): "t haplotypes in the mouse compromise sperm flagellar function". *Dev Biol* 155 (1): p. 14-25

Orozco L.D., Cokus S.J., Ghazalpour A., Ingram-Drake L., Wang S., van Nas A., Che N., Araujo J.A., Pellegrini M. and Lusis A.J. (2009): "Copy number variation influences gene expression and metabolic traits in mice". *Hum Mol Genet* 18 (21): p. 4118-4129

Orton R.J., Sturm O.E., Vyshemirsky V., Calder M., Gilbert D.R. and Kolch W. (2005): "Computational modelling of the receptor-tyrosine-kinase-activated MAPK pathway". *Biochem J* 392 (Pt 2): p. 249-261

Ostrowski M., Carmo N.B., Krumeich S., Fanget I., Raposo G., Savina A., Moita C.F., Schauer K., Hume A.N., Freitas R.P., Goud B., Benaroch P., Hacohen N., Fukuda M., Desnos C., Seabra M.C., Darchen F., Amigorena S., Moita L.F. and Thery C. (2010): "Rab27a and Rab27b control different steps of the exosome secretion pathway". *Nat Cell Biol* 12 (1): p. 19-30; sup pp 11-13

Ottaviani E. and Franceschi C. (1996): "The neuroimmunology of stress from invertebrates to man". *Prog Neurobiol* 48 (4-5): p. 421-440

Palkovits M. (1973): "Isolated removal of hypothalamic or other brain nuclei of the rat". *Brain Res* 59 p. 449-450

Paxinos G. and Franklin K.B.J. (2001): "The Mouse Brain in Stereotaxic Coordinates" (2nd ed.). *Academic Press*, San Diego (USA), London (UK)

Pellow S., Chopin P., File S.E. and Briley M. (1985): "Validation of open:closed arm entries in an elevated plus-maze as a measure of anxiety in the rat". *J Neurosci Methods* 14 (3): p. 149-167

Picard F., Robin S., Lavielle M., Vaisse C. and Daudin J.J. (2005): "A statistical approach for array CGH data analysis". *BMC Bioinformatics* 6 p. 27

Pilder S.H., Decker C.L., Islam S., Buck C., Cebra-Thomas J.A. and Silver L.M. (1992): "Concerted evolution of the mouse Tcp-10 gene family: implications for the functional basis of t haplotype transmission ratio distortion". *Genomics* 12 (1): p. 35-41

Pinto D., Darvishi K., Shi X., Rajan D., Rigler D., Fitzgerald T., Lionel A.C., Thiruvahindrapuram B., Macdonald J.R., Mills R., Prasad A., Noonan K., Gribble S., Prigmore E., Donahoe P.K., Smith R.S., Park J.H., Hurles M.E., Carter N.P., Lee C., Scherer S.W. and Feuk L. (2011): "Comprehensive assessment of array-based platforms and calling algorithms for detection of copy number variants". *Nat Biotechnol* 29 (6): p. 512-520

Plomin R., Haworth C.M. and Davis O.S. (2009): "Common disorders are quantitative traits". *Nat Rev Genet* 10 (12): p. 872-878

Porsolt R.D., Bertin A. and Jalfre M. (1977): "Behavioral despair in mice: a primary screening test for antidepressants". *Arch Int Pharmacodyn Ther* 229 (2): p. 327-336

Porsolt R.D., Bertin A. and Jalfre M. (1978): ""Behavioural despair" in rats and mice: strain differences and the effects of imipramine". *Eur J Pharmacol* 51 (3): p. 291-294

Prut L. and Belzung C. (2003): "The open field as a paradigm to measure the effects of drugs on anxiety-like behaviors: a review". *Eur J Pharmacol* 463 (1-3): p. 3-33

R Development Core Team (2010): "R: A language and environment for statistical computing". Published by R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Rassoulzadegan M., Magliano M. and Cuzin F. (2002): "Transvection effects involving DNA methylation during meiosis in the mouse". *Embo J* 21 (3): p. 440-450

Redkar A.A., Si Y., Twine S.N., Pilder S.H. and Olds-Clarke P. (2000): "Genes in the first and fourth inversions of the mouse t complex synergistically mediate sperm capacitation and interactions with the oocyte". *Dev Biol* 226 (2): p. 267-280

Reymond A., Henrichsen C.N., Harewood L. and Merla G. (2007): "Side effects of genome structural changes". *Curr Opin Genet Dev* 17 (5): p. 381-386

Ricard G., Molina J., Chrast J., Gu W., Gheldof N., Pradervand S., Schutz F., Young J.I., Lupski J.R., Reymond A. and Walz K. (2010): "Phenotypic consequences of copy number variation: insights from Smith-Magenis and Potocki-Lupski syndrome mouse models". *PLoS Biol* 8 (11): p. e1000543

Roberts P.C. (2008): "Gene expression microarray data analysis demystified". *Biotechnol Annu Rev* 14 p. 29-61

Roberts W.A. and Feeney M.C. (2009): "The comparative study of mental time travel". *Trends Cogn Sci* 13 (6): p. 271-277

Rovelet-Lecrux A., Hannequin D., Raux G., Le Meur N., Laquerriere A., Vital A., Dumanchin C., Feuillette S., Brice A., Vercelletto M., Dubas F., Frebourg T. and Campion D. (2006): "APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy". *Nat Genet* 38 (1): p. 24-26

Rubnitz J. and Subramani S. (1984): "The minimum amount of homology required for homologous recombination in mammalian cells". *Mol Cell Biol* 4 (11): p. 2253-2258

Rucker J.J., Breen G., Pinto D., Pedroso I., Lewis C.M., Cohen-Woods S., Uher R., Schosser A., Rivera M., Aitchison K.J., Craddock N., Owen M.J., Jones L., Jones I., Korszun A., Muglia P., Barnes M.R., Preisig M., Mors O., Gill M., Maier W., Rice J., Rietschel M., Holsboer F., Farmer A.E., Craig I.W., Scherer S.W. and McGuffin P. (2011): "Genome-wide association analysis of copy number variation in recurrent depressive disorder". *Mol Psychiatry* p.

Sah A., Schmuckermair C., Sartori S.B., Gaburro S., Kandasamy M., Irschick R., Klimaschewski L., Landgraf R., Aigner L. and Singewald N. (2012): "Anxiety- rather than depression-like behavior is associated with adult neurogenesis in a female mouse model of higher trait anxiety- and comorbid depression-like behavior". *Transl Psychiatry* 2 p. e171

Sajithlal G., Huttunen H., Rauvala H. and Munch G. (2002): "Receptor for advanced glycation end products plays a more important role in cellular survival than in neurite outgrowth during retinoic acid-induced differentiation of neuroblastoma cells". *J Biol Chem* 277 (9): p. 6888-6897

Sandhu K.S., Shi C., Sjolinder M., Zhao Z., Gondor A., Liu L., Tiwari V.K., Guibert S., Emilsson L., Imreh M.P. and Ohlsson R. (2009): "Nonallelic transvection of multiple imprinted loci is organized by the H19 imprinting control region during germline development". *Genes Dev* 23 (22): p. 2598-2603

Schimenti J., Cebra-Thomas J.A., Decker C.L., Islam S.D., Pilder S.H. and Silver L.M. (1988): "A candidate gene family for the mouse t complex responder (Tcr) locus responsible for haploid effects on sperm function". *Cell* 55 (1): p. 71-78

Sexton T., Umlauf D., Kurukuti S. and Fraser P. (2007): "The role of transcription factories in large-scale structure and dynamics of interphase chromatin". *Semin Cell Dev Biol* 18 (5): p. 691-697

Seyfarth R.M., Cheney D.L. and Marler P. (1980): "Monkey responses to three different alarm calls: evidence of predator classification and semantic communication". *Science* 210 (4471): p. 801-803

Sharp A.J., Locke D.P., McGrath S.D., Cheng Z., Bailey J.A., Vallente R.U., Pertz L.M., Clark R.A., Schwartz S., Segraves R., Oseroff V.V., Albertson D.G., Pinkel D. and Eichler E.E. (2005): "Segmental duplications and copy-number variation in the human genome". *Am J Hum Genet* 77 (1): p. 78-88

Shlien A. and Malkin D. (2009): "Copy number variations and cancer". *Genome Med* 1 (6): p. 62

Singleton A.B., Farrer M., Johnson J., Singleton A., Hague S., Kachergus J., Hulihan M., Peuralinna T., Dutra A., Nussbaum R., Lincoln S., Crawley A., Hanson M., Maraganore D., Adler C., Cookson M.R., Muenter M., Baptista M., Miller D., Blancato J., Hardy J. and Gwinn-Hardy K. (2003): "alpha-Synuclein locus triplication causes Parkinson's disease". *Science* 302 (5646): p. 841

Smyth G.K. (2004): "Linear models and empirical bayes methods for assessing differential expression in microarray experiments". *Stat Appl Genet Mol Biol* 3 p. Article3

Smyth G.K. (2005): "Limma: linear models for microarray data". In *Bioinformatics and Computational Biology Solutions using R and Bioconductor - Statistics for Biology and Health* (edited by Gentleman R., Carey V.J., Dudoit S., Irizarry R.A. and Huber W.), p. p. 397-420, *Springer*, New York

Smyth G.K., Michaud J. and Scott H.S. (2005): "Use of within-array replicate spots for assessing differential expression in microarray experiments". *Bioinformatics* 21 (9): p. 2067-2075

Sommers J.M. (2006): "Prevalence and incidence studies of anxiety disorders: a systematic review of the literature". *Can J Psychiatry* 51 (2): p. 100 - 113

Sotnikov S.V., Markt P.O., Umriukhin A.E. and Landgraf R. (2011): "Genetic predisposition to anxiety-related behavior predicts predator odor response". *Behav Brain Res* 225 (1): p. 230-234

Stankiewicz P. and Lupski J.R. (2002): "Genome architecture, rearrangements and genomic disorders". *Trends Genet* 18 (2): p. 74-82

Stankiewicz P. and Lupski J.R. (2010): "Structural variation in the human genome and its role in disease". *Annu Rev Med* 61 p. 437-455

Stein M. (1989): "Stress, depression, and the immune system". *J Clin Psychiatry* 50 Suppl p. 35-40; discussion 41-32

Stein M.B. and Heimberg R.G. (2004): "Well-being and life satisfaction in generalized anxiety disorder: comparison to major depressive disorder in a community sample". *J Affect Disord* 79 (1-3): p. 161-166

Steru L., Chermat R., Thierry B., Mico J.A., Lenegre A., Steru M., Simon P. and Porsolt R.D. (1987): "The automated Tail Suspension Test: a computerized device which differentiates psychotropic drugs". *Prog Neuropsychopharmacol Biol Psychiatry* 11 (6): p. 659-671

Steru L., Chermat R., Thierry B. and Simon P. (1985): "The tail suspension test: a new method for screening antidepressants in mice". *Psychopharmacology (Berl)* 85 (3): p. 367-370

Stranger B.E., Forrest M.S., Dunning M., Ingle C.E., Beazley C., Thorne N., Redon R., Bird C.P., de Grassi A., Lee C., Tyler-Smith C., Carter N., Scherer S.W., Tavare S., Deloukas P., Hurles M.E. and Dermitzakis E.T. (2007): "Relative impact of nucleotide and copy number variation on gene expression phenotypes". *Science* 315 (5813): p. 848-853

Suddendorf T. and Corballis M.C. (2007): "The evolution of foresight: What is mental time travel, and is it unique to humans?" *Behav Brain Sci* 30 (3): p. 299-313; discussion 313-251

Sylvers P., Lilienfeld S.O. and LaPrairie J.L. (2011): "Differences between trait fear and trait anxiety: implications for psychopathology". *Clin Psychol Rev* 31 (1): p. 122-137

Tanna V.L., Wilson A.F., Winokur G. and Elston R.C. (1989): "Linkage analysis of pure depressive disease". *J Psychiatr Res* 23 (2): p. 99-107

Tecchio C., Bonetto C., Bertani M., Cristofalo D., Lasalvia A., Nichele I., Bonani A., Andreini A., Benedetti F., Ruggeri M. and Pizzolo G. (2012): "Predictors of anxiety and depression in hematopoietic stem cell transplant patients during protective isolation". *Psychooncology* p.

Tecott L.H. (2003): "The genes and brains of mice and men". *Am J Psychiatry* 160 (4): p. 646-656

Teo S.M., Pawitan Y., Kumar V., Thalamuthu A., Seielstad M., Chia K.S. and Salim A. (2011): "Multi-platform segmentation for joint detection of copy number variants". *Bioinformatics* 27 (11): p. 1555-1561

Thornalley P.J. (2006): "Unease on the role of glyoxalase 1 in high-anxiety-related behaviour". *Trends Mol Med* 12 (5): p. 195-199

Tooby J. and Cosmides L. (1990): "The past explains the present: Emotional adaptations and the structure of ancestral environments". *Ethology and Sociobiology* 11 (4-5): p. 375 - 424

Touma C., Bunck M., Glasl L., Nussbaumer M., Palme R., Stein H., Wolferstatter M., Zeh R., Zimbelmann M., Holsboer F. and Landgraf R. (2008): "Mice selected for high

versus low stress reactivity: a new animal model for affective disorders". *Psychoneuroendocrinology* 33 (6): p. 839-862

Tulving E. (1985): "Memory and consciousness". *Can. Psychol.* 26 p. 1 - 12

Turchinovich G. and Hayday A.C. (2011): "Skint-1 identifies a common molecular mechanism for the development of interferon-gamma-secreting versus interleukin-17-secreting gammadelta T cells". *Immunity* 35 (1): p. 59-68

Turner D.J., Miretti M., Rajan D., Fiegler H., Carter N.P., Blayney M.L., Beck S. and Hurles M.E. (2008): "Germline rates of de novo meiotic deletions and duplications causing several genomic disorders". *Nat Genet* 40 (1): p. 90-95

Vagner B., Pinto L.F.R. and Albano R.M. (2013): "Gene expression analysis by real-time PCR: Experimental demonstration of PCR detection limits". *Analytical Biochemistry* 432 (2): p. 131-133

van de Wiel M.A., Picard F., van Wieringen W.N. and Ylstra B. (2010): "Preprocessing and downstream analysis of microarray DNA copy number profiles". *Brief Bioinform* 12 (1): p. 10-21

Van Laethem F., Tikhonova A.N. and Singer A. (2012): "MHC restriction is imposed on a diverse T cell receptor repertoire by CD4 and CD8 co-receptors during thymic selection". *Trends Immunol* 33 (9): p. 437-441

van Veen T., Goeman J.J., Monajemi R., Wardenaar K.J., Hartman C.A., Snieder H., Nolte I.M., Penninx B.W. and Zitman F.G. (2012): "Different gene sets contribute to different symptom dimensions of depression and anxiety". *Am J Med Genet B Neuropsychiatr Genet* 159B (5): p. 519-528

Vrijenhoek T., Buizer-Voskamp J.E., van der Stelt I., Strengman E., Sabatti C., Geurts van Kessel A., Brunner H.G., Ophoff R.A. and Veltman J.A. (2008): "Recurrent CNVs disrupt three candidate genes in schizophrenia patients". *Am J Hum Genet* 83 (4): p. 504-510

Waldman A.S. and Liskay R.M. (1988): "Dependence of intrachromosomal recombination in mammalian cells on uninterrupted homology". *Mol Cell Biol* 8 (12): p. 5350-5357

Walsh R.N. and Cummins R.A. (1976): "The Open-Field Test: a critical review". *Psychol Bull* 83 (3): p. 482-504

Wang K., Li M., Hadley D., Liu R., Glessner J., Grant S.F., Hakonarson H. and Bucan M. (2007): "PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data". *Genome Res* 17 (11): p. 1665-1674

Warden M., Pique-Regi R., Ortega A. and Asgharzadeh S. (2011): "Bioinformatics for copy number variation data". *Methods Mol Biol* 719 p. 235-249

Wefers B., Hitz C., Holter S.M., Trumbach D., Hansen J., Weber P., Putz B., Deussing J.M., de Angelis M.H., Roenneberg T., Zheng F., Alzheimer C., Silva A., Wurst W. and Kuhn R. (2012): "MAPK signaling determines anxiety in the juvenile mouse brain but depression-like behavior in adults". *PLoS One* 7 (4): p. e35035

Weiger W.A. (1997): "Serotonergic modulation of behaviour: a phylogenetic overview". *Biol Rev Camb Philos Soc* 72 (1): p. 61-95

Wellcome Trust Case Control Consortium W. (2010): "Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls". *Nature* 464 (7289): p. 713-720

Weterings E. and van Gent D.C. (2004): "The mechanism of non-homologous end-joining: a synopsis of synapsis". *DNA Repair (Amst)* 3 (11): p. 1425-1435

Wheeler M.A., Stuss D.T. and Tulving E. (1997): "Toward a theory of episodic memory: the frontal lobes and autonoetic consciousness". *Psychol Bull* 121 (3): p. 331-354

White S.N., Mousel M.R., Herrmann-Hoesing L.M., Reynolds J.O., Leymaster K.A., Neibergs H.L., Lewis G.S. and Knowles D.P. (2012): "Genome-wide association identifies multiple genomic regions associated with susceptibility to and control of ovine lentivirus". *PLoS One* 7 (10): p. e47829

Widmann C., Gibson S., Jarpe M.B. and Johnson G.L. (1999): "Mitogen-activated protein kinase: conservation of a three-kinase module from yeast to human". *Physiol Rev* 79 (1): p. 143-180

Widner-Andrä R. (2011): "Assignment of functional impact on genetic data in two mouse models of affective disorders". *Doctoral thesis.* Ludwig-Maximilians-Universität, Munich, Germany

Williams A.F. and Barclay A.N. (1988): "The immunoglobulin superfamily--domains for cell surface recognition". *Annu Rev Immunol* 6 p. 381-405

Williams R., Lim J.E., Harr B., Wing C., Walters R., Distler M.G., Teschke M., Wu C., Wiltshire T., Su A.I., Sokoloff G., Tarantino L.M., Borevitz J.O. and Palmer A.A. (2009): "A common and unstable copy number variant is associated with differences in Glo1 expression and anxiety-like behavior". *PLoS One* 4 (3): p. e4649

Winchester L. and Ragoussis J. (2012): "Algorithm implementation for CNV discovery using Affymetrix and Illumina SNP array data". *Methods Mol Biol* 838 p. 291-310

Woodward K.J., Cundall M., Sperle K., Sistermans E.A., Ross M., Howell G., Gribble S.M., Burford D.C., Carter N.P., Hobson D.L., Garbern J.Y., Kamholz J., Heng H., Hodes M.E., Malcolm S. and Hobson G.M. (2005): "Heterogeneous duplications in patients with Pelizaeus-Merzbacher disease suggest a mechanism of coupled homologous and nonhomologous recombination". *Am J Hum Genet* 77 (6): p. 966-987

Wu C.T. and Morris J.R. (1999): "Transvection and other homology effects". *Curr Opin Genet Dev* 9 (2): p. 237-246

Wyttenbach R.A., May M.L. and Hoy R.R. (1996): "Categorical perception of sound frequency by crickets". *Science* 273 (5281): p. 1542-1544

Yang H. (2010): "MouseDivGeno: Tools for analyzing Mouse Diversity Array". R package version 1.0.0. http://genomedynamics.org/tools/MouseDivGeno/CDFfiles

Yang H., Ding Y., Hutchins L.N., Szatkiewicz J., Bell T.A., Paigen B.J., Graber J.H., de Villena F.P. and Churchill G.A. (2009): "A customized and versatile high-density genotyping array for the mouse". *Nat Methods* 6 (9): p. 663-666

Yen Y.C., Mauch C.P., Dahlhoff M., Micale V., Bunck M., Sartori S.B., Singewald N., Landgraf R. and Wotjak C.T. (2012): "Increased levels of conditioned fear and avoidance behavior coincide with changes in phosphorylation of the protein kinase B (AKT) within the amygdala in a mouse model of extremes in trait anxiety". *Neurobiol Learn Mem* 98 (1): p. 56-65

Yerevanian B.I., Koek R.J. and Ramdev S. (2001): "Anxiety disorders comorbidity in mood disorder subgroups: data from a mood disorders clinic". *J Affect Disord* 67 (1-3): p. 167-173

Yokoyama K., Kaji H., He J., Tanaka C., Hazama R., Kamigaki T., Ku Y., Tohyama K. and Tohyama Y. (2011): "Rab27a negatively regulates phagocytosis by prolongation of the actin-coating stage around phagosomes". *J Biol Chem* 286 (7): p. 5375-5382

Zhang D., Qian Y., Akula N., Alliey-Rodriguez N., Tang J., Gershon E.S. and Liu C. (2011): "Accuracy of CNV Detection from GWAS Data". *PLoS One* 6 (1): p. e14511

Zhang F., Gu W., Hurles M.E. and Lupski J.R. (2009): "Copy number variation in human health, disease, and evolution". *Annu Rev Genomics Hum Genet* 10 p. 451-481

Zhi J. (2010): "MAPD: a probe design suite for multiplex ligation-dependent probe amplification assays". *BMC Res Notes* 3 p. 137

VIII Supplementary data

All supplementary data can be found on the enclosed DVD. Here, the whole content of the DVD, including short descriptions of the data, is listed.

Report S1: **Analysis of structural variants in HAB/LAB mice based on NGS data.** This preliminary report was written by André Altmann (former member of RG Binder, MPI of Psychiatry) who was in charge of data processing, as the NGS was done in collaboration with RG Binder.

Report S2: **Analysis of expression microarray data.** A detailed description of expression microarray data analysis to detect differentially expressed genes between HAB and LAB mice.

Report S3: **STRING analysis of candidate genes.** All genes found to be differentially expressed in at least one of the brain regions BLA, CeA, Cg and PVN via expression microarray were checked for building of networks. Here the results of the analysis using STRING online software v9.0 are shown.

Report S4: **Breeding scheme of *Glo1*-selective breeding approach.** A family tree of all animals that were mated is shown. For each animal the number and generation is shown.

Report S5: **Rewritten version of the Simple CNV function.** The modified "simpleCNV" function of the R package "MouseDivGeno", used to analyze data of the JaxMDGA approach screening for CNVs in CD-1 mice. All parts that were rewritten are highlighted in this report. The notes not belonging to the R code itself were marked by symbol (#).

Table S1: **CNVs detected by aCGH.** All 98 CNVs detected by aCGH are listed including information on the genomic position.

Table S2: **CNVs detected by JaxMDGA.** All 180 CNVs detected by JaxMDGA are listed including information on the genomic position.

Table S3: **CNVs detected by NGS.** All 5,851 CNVs detected by NGS are listed including information on the genomic position.

Table S4: **Deletions detected by NGS.** All deletions detected by NGS either in HAB or LAB mice are listed including information on the genomic position.

Table S5: **Sequences of wrong orientation detected by NGS.** All regions detected by NGS are listed, where the R3 read was aligned downstream to the F3 read during analysis, indicating a complex structural variant.

Table S6: **Insertions detected by NGS.** All insertions detected by NGS either in HAB or LAB mice are listed including information on their genomic position.

Table S7: **Inversions detected by NGS.** All inversions detected by NGS either in HAB or LAB mice are listed including information on the genomic position.

Table S8: **Comparison of CNVs detected by three different methods.** List of all CNVs detected by aCGH, JaxMDGA and NGS in HAB/LAB with overlapping CNVs written in one line.

Table S9: **Genes differentially expressed between HAB and LAB mice.** Genomic positions of genes found to be differentially expressed in HAB and LAB mice by expression microarray.

Table S10: **Result of MLPA analysis.** For all 33 probes targeting 11 potential CNVs in 54 CD-1 mice and one HAB mouse the copy number status in respect to a LAB sample is shown.

Table S11: **CNVs detected in CD-1 mice by JaxMDGA.** CNVs listed were revealed in at least one pair of the 64 animals screened.

Table S12 Association of CNVs, detected in CD-1 mice, with behavior on the EPM**.** For five different parameters of the EPM test the nominal p-value and the p-value after Bonferroni correction are shown.

Table S13 Association of CNVs, detected in CD-1 mice, with behavior in the TST**.** For three different parameters of the TST the nominal p-value and the p-value after Bonferroni correction are shown.

Table S14 Association of CNVs, detected in CD-1 mice, with behavior in the FST**.** For seven different parameters of the FST the nominal p-value and the p-value after Bonferroni correction are shown.

Table S15 Association of CNVs, detected in CD-1 mice, with behavior in the OF**.** For seven different parameters of the OF the nominal p-value and the p-value after Bonferroni correction are shown.

Table S16 Association of CNVs, detected in CD-1 mice, with Cort concentrations in SRT**.** For three different Cort concentrations the nominal p-value and the p-value after Bonferroni correction are shown.

Table S17 Protein coding genes in CNVs detected by aCGH in HAB/LAB mice. List of all protein coding genes of which the genomic position overlapped with any CNV detected in HAB/LAB by aCGH.

Table S18 Protein coding genes in CNVs detected by JaxMDGA in HAB/LAB mice. List of all protein coding genes of which the genomic position overlapped with any CNV detected in HAB/LAB by JaxMDGA.

Table S19 Protein coding genes in CNVs detected by NGS in HAB/LAB mice. List of all protein coding genes of which the genomic position overlapped with any CNV detected in HAB/LAB by NGS.

Table S20 Protein coding genes in CNVs detected by JaxMDGA in 64 CD-1 mice. Genes overlapping with any of the detected CNVs are shown, not considering if an association with behavioral data was found.

Acknowledgements/ Danksagung

Als erstes geht mein aufrichtiger Dank an meinen Doktorvater, Herrn Prof. Rainer Landgraf, der mir nicht nur die Möglichkeit gegeben hat, meine Doktorarbeit in seiner Arbeitsgruppe anzufertigen, sondern mir auch stets großes Vertrauen in meine Fähigkeiten und meine Arbeit entgegenbrachte. Er ließ mir die Freiheit, meine eigenen Ideen zu entwickeln und seine Tür stand stets offen, wann immer ich seinen Rat und seine Hilfe benötigte. Vielen Dank dafür!

Ich bedanke mich sehr bei Prof. John Parsch, der sich bereit erklärt hat die arbeitsreiche Aufgabe des Zweitgutachters zu übernehmen. Danke auch an alle anderen Mitglieder der Prüfungskommission.

Vielen Dank an meinen Betreuer, Dr. Ludwig Czibere. Er nahm sich immer die Zeit, wenn ich seine Unterstützung benötigte und war mir vor allem in der stressigen Endphase meiner Arbeit eine große Hilfe. Mit kaum einem anderen konnte ich so gut über wissenschaftliche Themen diskutieren wie mit ihm.

Weiterhin bedanke ich mich bei Dr. Elisabeth Binder und ihrer Arbeitsgruppe, insbesondere bei Dr. Peter Weber und Monika Rex Haffner, für die praktische Durchführung des next-generation sequencing Projekts. Vielen Dank auch an Dr. André Altmann der eine tolle Arbeit bei der Analyse der NGS-Daten geleistet hat und mit dem ich sehr gerne zusammengearbeitet habe.

Ein großer Dank auch an alle, die mich mit Rat und Tat unterstützen, allen voran Prof. Bertram Müller-Myhsok und seine Gruppenmitglieder Dr. Benno Pütz (der den Expressions-microarray neu ausgewertet hat), Tony Kam Thong, Darina Czamara, Nazanin Karbalai und Christiane Wolf. Außerdem danke an Regina Widner-Andrä für die Zusammenarbeit beim CD-1 panel Projekt und an alle anderen helfenden Hände und Köpfe, Dr. Thomas Bettecken, Rebekka Diepold, Lisa Tietze, Silja McIlwrick, Markus Nußbaumer, Marina Zimbelmann und Dr. Chadi Touma.

Für die praktische Durchführung der MLPA und für ihre Hilfe bedanke ich mich bei Dr. Gertrud Eckstein, Jelena Golic und Marion Schieweg vom Helmholtz Zentrum München.

Desweiteren bedanke ich mich bei den noch nicht genannten aktuellen und ehemaligen Mitgliedern der AG Landgraf und anderer Arbeitsgruppen, Nadine Brehm, Anke Wittmann, Yi-Chun Yen, Sebastian Scharf, Katja Stangl, Alana Knapman, Jan-Michael Heinzmann, Gabriele Mattos, Sergey Sotnikov, Roshan Naik, Patrick Markt, Natalia Checkmareva und Victoria Malik. Sie alle haben dazu beigetragen eine Arbeitsatmosphäre zu schaffen, in der ich mich sehr wohlgefühlt habe.

Mein größter Dank gilt jedoch den wichtigsten Menschen in meinem Leben, ohne deren Unterstützung ich niemals so weit gekommen wäre. Vor allem meinen Eltern, die tollsten der Welt, die immer für mich da sind wenn ich sie brauche. Meinem Bruder, der mich

immer zum Lachen bringen kann, auch wenn mir eigentlich nicht danach ist. Meiner Schwester, auf deren Unterstützung ich immer bauen kann. Valentin und Paulina, die mich daran erinnern was im Leben wichtig ist. Und natürlich meinem Freund Oskar, der mir so gut es ging den Rücken frei hielt, mir zugehört und mich unterstützt hat und während der stressigsten Zeit an meiner Seite war. Danke euch allen!

Curriculum Vitae

Name	Julia Brenndörfer
Geburtsdatum	02.11.1981
Geburtsort	Dillingen (Donau)

Abschlüsse	Diplom Biologie (September 2009)
	Abitur (Juni 2001)

Ausbildung	Seit Dezember 2009 Doktorarbeit (Ludwig-Maximilians-Universität München) angefertigt am Max-Planck-Institut für Psychiatrie
	Dezember 2008 – September 2009 Diplomarbeit am Max-Planck-Institut für Psychiatrie
	April 2005 – September 2009 Studium an der Julius-Maximilians-Universität Würzburg Studiengang Diplom Biologie
	Oktober 2002 – April 2005 Studium an der Julius-Maximilians-Universität Würzburg Studiengang Lehramt an Gymnasien Biologie und Chemie
	Oktober 2001 – Februar 2002 Studium an der Universität Ulm Studiengang Zahnmedizin
	Juni 2001 Schulabschluss (Abitur) am Albertus-Gymnasium Lauingen

Declaration/ Erklärung

Ich versichere hiermit an Eides statt, dass die vorgelegte Dissertation von mir selbständig und ohne unerlaubte Hilfe angefertigt ist.

München, den ....

(Unterschrift)

Hiermit erkläre ich,

dass die Dissertation nicht ganz oder in wesentlichen Teilen einer anderen Prüfungskommission vorgelegt worden ist und

dass ich mich anderweitig einer Doktorprüfung ohne Erfolg **nicht** unterzogen habe.

München, den ....

(Unterschrift)