# Connecting Anxiety and Genomic Copy Number Variation: A Genome-Wide Analysis in CD-1 Mice

Julia Brenndörfer[1]*, André Altmann[2¤], Regina Widner-Andrä[1], Benno Pütz[2], Darina Czamara[2], Erik Tilch[4,5], Tony Kam-Thong[2], Peter Weber[3], Monika Rex-Haffner[3], Thomas Bettecken[1], Andrea Bultmann[1], Bertram Müller-Myhsok[2], Elisabeth E. Binder[3], Rainer Landgraf[1], Ludwig Czibere[1]

1 Department of Behavioral Neuroendocrinology, Max Planck Institute of Psychiatry, Munich, Germany, 2 Department of Statistical Genetics, Max Planck Institute of Psychiatry, Munich, Germany, 3 Department of Molecular Genetics of Affective Disorders, Max Planck Institute of Psychiatry, Munich, Germany, 4 Institute of Human Genetics, Helmholtz Zentrum München, Munich, Germany, 5 Institute of Human Genetics, Technische Universität München, Munich, Germany

¤ Current address: Department of Neurology and Neurological Sciences, Stanford University, Stanford, California, United States of America

* jbrenndoerfer@mpipsykl.mpg.de

## Abstract

Genomic copy number variants (CNVs) have been implicated in multiple psychiatric disorders, but not much is known about their influence on anxiety disorders specifically. Using next-generation sequencing (NGS) and two additional array-based genotyping approaches, we detected CNVs in a mouse model consisting of two inbred mouse lines showing high (HAB) and low (LAB) anxiety-related behavior, respectively. An influence of CNVs on gene expression in the central (CeA) and basolateral (BLA) amygdala, paraventricular nucleus (PVN), and cingulate cortex (Cg) was shown by a two-proportion $Z$-test ($p = 1.6 \times 10^{-31}$), with a positive correlation in the CeA ($p = 0.0062$), PVN ($p = 0.0046$) and Cg ($p = 0.0114$), indicating a contribution of CNVs to the genetic predisposition to trait anxiety in the specific context of HAB/LAB mice. In order to confirm anxiety-relevant CNVs and corresponding genes in a second mouse model, we further examined CD-1 outbred mice. We revealed the distribution of CNVs by genotyping 64 CD 1 individuals using a high-density genotyping array (Jackson Laboratory). 78 genes within those CNVs were identified to show nominally significant association (48 genes), or a statistical trend in their association (30 genes) with the time animals spent on the open arms of the elevated plus-maze (EPM). Fifteen of them were considered promising candidate genes of anxiety-related behavior as we could show a significant overlap (permutation test, $p = 0.0051$) with genes within HAB/LAB CNVs. Thus, here we provide what is to our knowledge the first extensive catalogue of CNVs in CD-1 mice and potential corresponding candidate genes linked to anxiety-related behavior in mice.

## Introduction

With the advances in genome-wide screening arrays and sequencing technologies, scientists were enabled to examine genetic variations and their effect on behavioral phenotypes. In recent years a new type of variation became increasingly important: the copy number variants (CNVs). Not only have CNVs already been associated with common disorders and metabolic diseases like asthma, type 2 diabetes, obesity and cancer [1–4], they also have been reported to affect disease susceptibility of neurological disorders including Parkinson's disease, Alzheimer's disease, autism, schizophrenia, bipolar disorders and anxiety disorders [5–17]. Further, a large and common CNV in mice including the Glyoxalase 1 (*Glo1*) locus has been associated with anxiety-like behavior [18]. Although the general impact of CNVs on (disease) phenotypes is not clear yet, there is evidence from many distinct studies pointing to their involvement in phenotypic expression. Thus, considering the mechanisms by which CNVs might act on gene expression and their high abundance across the genome, their contribution is likely to be of importance [19].

There are multiple potential mechanisms that explain how CNVs might contribute to distinct diseases and behavioral phenotypes such as anxiety-related behavior. For instance, a direct change of gene dosage following copy number alterations [20] is an obvious mode of action. More complex, however, are effects mediated by changes in copy numbers of enhancers and repressors, which were shown to act as *cis*-regulatory domains even though they extend long distances outside the coding region itself [21]. Furthermore, it is conceivable that CNVs mediate their effects by physically impairing the access of genes to the transcription machinery or by influencing transvection [22–24]. Hence, theoretically, CNVs could regulate the transcription of genes beyond their breakpoints. Although different detection methods exist and many studies have been performed so far, the detection of CNVs and the analysis of their effects still remain challenging.

Animal models such as the inbred HAB/LAB (high/low anxiety-related behavior) mouse model [25–27] represent an ideal tool for revealing the complex impact of CNVs on behavioral phenotypes. In order to provide a large-scale analysis of CNVs influencing anxiety-related behavior, we decided to not only apply three different detection methods to screen for CNVs in the HAB/LAB mouse model, but also to analyze CNVs in the genomic context of a second mouse model (CD-1 outbred mice). Thus, here we provide, first, an extensive study of anxiety-relevant CNVs and corresponding genes and, second, a catalogue of CNVs in CD-1 mice that might serve as basis for subsequent studies on the effects of CNVs.

## Material and Methods

### Ethics statement and general remarks

All animal experiments were conducted in accordance with the current regulations for animal experimentation in Germany and the European Union (Council of the European Communities Directive 86/609/EEC) and were approved by the Government of Upper Bavaria.

Information on genomic positions all refer to the UCSC genome browser assembly mm9 (http://genome.ucsc.edu/) [28].

### Animals and behavioral experiments

All animals were housed in the animal facility of the Max Planck Institute of Psychiatry under standard conditions, i.e., a temperature of 23 ± 2°C, a relative air humidity of 60 ± 5% and a 12/12-hour light-dark cycle with beginning of light phase at 8 a.m. Animals 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, Germany) and tap water *ad libitum*.

Male HAB and LAB mice used in this study were selected from generations 35 to 41, bred in the animal facility of the Max Planck Institute of Psychiatry (Munich, Germany). The HAB/ LAB mouse model was described before by Krömer *et al.* [25]. Briefly, the HAB and LAB mouse lines were bred following bidirectional breeding protocols successfully applied in rats before [26, 29, 30]. Depending on their performance on the elevated plus-maze (EPM) [31], individuals of a population of more than 250 animals from over 25 litter of outbred Swiss CD-1 mice were chosen to either found the HAB or LAB mouse line. After nine generations of outbreeding across families but within behavioral restrictions, a strict inbreeding protocol was followed, resulting in two mouse lines showing a stable anxiety-related phenotype, with HAB mice spending less than 20% and LAB mice more than 50% of the test time on the open arm of the EPM.

Male outbred Swiss CD-1 mice used in this study were purchased from Charles River (Sulzfeld, Germany) and delivered at the age of eight weeks in eight different batches of 48 animals each (in total 384 animals). Deliveries of discrete batches were made at an interval of at least one week. The CD-1 mice of each batch were phenotypically characterized in a series of five tests covering different facets of anxiety (Fig 1): after arrival and a 4-day-habituation, half of the animals were tested on the EPM [31], the other half in the open field (OF) [32], and two days later vice versa. On day eight after arrival, the forced-swim test (FST) [33] was performed, followed by stress-reactivity test (SRT) [34] and tail suspension test (TST) [35] on day 12 and 14, respectively. Behavioral tests were conducted between 9 a.m. and 1 p.m. under standard housing conditions. EPM and OF were analyzed by means of the tracking software Any-maze v4.72 (Stoelting, Wood Dale, IL, USA). Other tests were recorded and analyzed by experienced researchers using Eventlog v1.0.

EPM: The apparatus we used was made of grey polyvinylchloride (PVC). It consisted of two open arms (30 × 5 cm, 300 lx) and two closed arms (30 × 5 × 15 cm, 10 lx) extending from a central platform (5 × 5 cm) and elevated by four legs (40 cm). Each mouse was placed in the central zone facing a closed arm. Its behavior was recorded for 5 min.

OF: The test apparatus consisted of a round PVC wall of 40 cm height, framing a field of 60 cm in diameter. Intensity of light differed between brighter inner zone and darker outer zone of about 15 lx. Each mouse was placed in the inner zone.
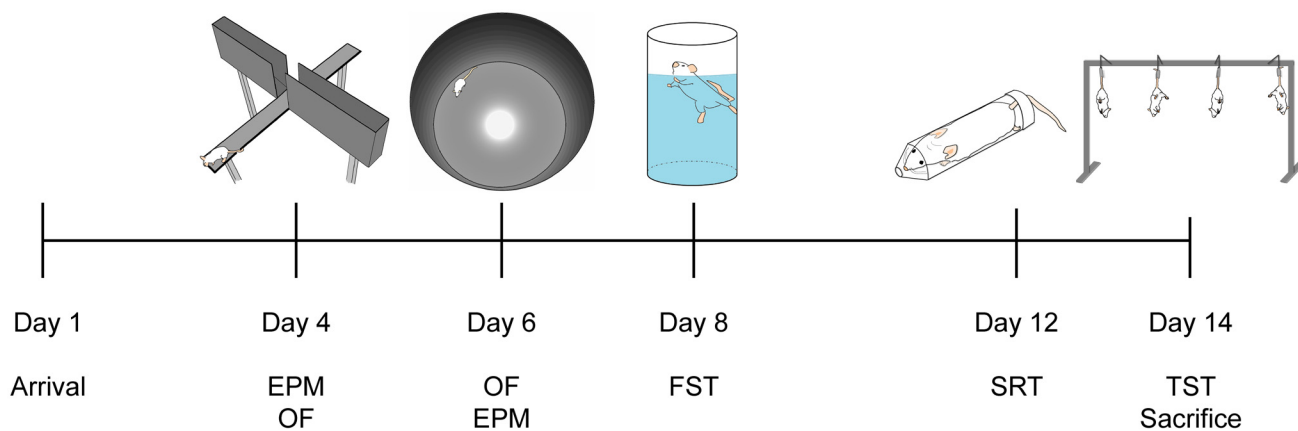


**Fig 1. Test series to phenotypically characterize CD-1 mice.** Each animal had to perform five tests (EPM, OF, FST, SRT, TST) in a row as shown. Animals tested on the EPM on day 4 were tested in the open field on day 6 and vice versa.

doi:10.1371/journal.pone.0128465.g001

FST: Mice were placed in a 2 l glass beaker filled with 1.75 l tap water (23°C). Behavior was recorded for 6 min and analyzed with a customized Eventlog program, differentiating between freezing, floating, struggling and swimming. Floating was defined as not showing any movements except very slight balancing movements.

SRT: The increase of corticosterone (Cort) levels in the blood was detected after a 15-minute period of restrained stress, where animals were placed into a 50 ml plastic tube (11.4 x 2.8 cm²) with holes for ventilation. Cort levels were measured by radioimmunoassay.

TST: Four mice were suspended in parallel from a metal frame (height: 37 cm) by fixing their tail tips with an adhesive tape. Behavior was videotaped for 6 min and analyzed for immobility (no movement at all) and struggling, using a customized Eventlog program.

## DNA extraction

For DNA extraction from tail tips of both HAB/LAB and CD-1 mice, a NucleoSpin Tissue kit (Macherey-Nagel, Düren, Germany) was used, following the manufacturer's instructions provided in the "standard protocol for human and animal tissue and cultured cells". In case of NGS only, the DNA was extracted using a DNeasy blood & tissue kit (Qiagen, Hilden, Germany), following the guideline provided with the kit. DNA concentration was measured on a NanoPhotometer (Implen, München, Germany) and DNA quality was checked by gel electrophoresis using 1.0% agarose gels (or 0.7% for NGS samples).

## CNV detection. . .

We screened the genome of HAB/LAB mice for CNVs using three different detection methods, (1) array comparative genomic hybridization (aCGH), (2) the Jax Mouse Diversity Genotyping Array (JaxMDGA), and (3) next-generation sequencing (NGS), as described below.

. . . **by aCGH.** DNA extracted from tail tips of two 16-week-old male HAB and LAB mice each, and from brain tissue of only one pair of these mice, respectively, was sent to Roche NimbleGen (Madison, WI, USA) to access NimbleGen's full CGH microarray service (Mouse CGH 3x720 K Whole-Genome Tiling Array; probes: 50- to 75-mers; median probe spacing: 3.5 kb). Briefly, according to NimbleGen's supplied information sheet (NimbleGen CGH Services: Guide to Your CGH Data v3.0; 2009), the signal intensity was spatially corrected based on the X and Y coordinate position on the array using locally weighted polynomial regression. The Cy3 (HAB sample) and Cy5 (LAB sample) signal intensities were normalized to one another using qspline normalization. Roche NimbleGen applied the segMNT algorithm that identifies copy number changes using dynamic programming to globally minimize the sum of squares error relative to the segment means. Assessing the provided data, we defined the final set of high confidence CNV calls if segMNT-defined segments had a mean $\log_2$ signal ratio (Cy3 signal/Cy5 signal) greater/less than +/-0.09 in all three assays.

. . . **by JaxMDGA.** The high-density Jax Mouse Diversity Genotyping Array (The Jackson Laboratory, Bar Harbor, ME, USA), containing 623,124 SNP and 916,269 invariant genomic probes [36], was applied to screen the genome of HAB/LAB and CD-1 mice for CNVs. We accessed the basic service offered by The Jackson Laboratory, including DNA extraction, sample preparation, array hybridization and provision of raw data. We provided tail tips of one male HAB and LAB mouse each, and of 64 male CD-1 mice, respectively, to analyze their genome on a high-density Jax Mouse Diversity Genotyping Array (The Jackson Laboratory) [36]. The CD-1 mice, a subgroup of the 384 animals described in Materials and Methods, were chosen based on their behavior averaged over all tests, with 24 mice showing highest, 24 lowest, and 16 intermediate anxiety-related behavior (for details on selection criteria see reference [37]). The array was performed in two batches of 32 samples each.

With LAB defined as reference sample, we performed CNV calling in the raw data of HAB/ LAB mice by applying the Hidden Markov Model-based function "simple CNV" implemented in the "MouseDivGeno" R package [38]. This function integrates normalized intensities from SNPs and exons. To infer the most likely state from three possible states (loss, normal, or gain compared to the reference sample) the function uses "HiddenMarkov", an existing HMM R package. [39]

CNV calling in the raw data of CD-1 mice was performed by applying the "simple CNV" function for each pair of animals, i.e. each animal was once declared as reference sample. Thus, we increased the sensitivity for CNV calls. By contrast, a selection of just one single sample as reference would lead to a loss of information about potential CNVs (for a detailed explanation see S1 Text). However, discrepancies in breakpoint definition between different pairs of animals could occur. We solved this problem by unifying the breakpoints of respective CNVs, i.e. we assessed the starting and end points in a way the CNV was defined as large as possible (see S1 Fig). By means of the "normalizeForSimpleCNV" function, a subfunction of the "simple CNV" function, we were able to calculate the mean normalized intensities of all probes within those unified CNVs for each sample separately. These intensity values were required for the association study described below.

. . . by NGS. The DNA of six male HAB and LAB mice each was pooled to form a single HAB and LAB sample, respectively, to be sequenced on a SOLiD 4 System (Applied Biosystems, Foster City, CA, USA). After shearing the genomic DNA to an average size of 2,000 bp using a Covaris S2 system (Covaris, Woburn, MA, USA), 2x60 bp mate-paired libraries were prepared following the Applied Biosystems' Mate-Paired Library Preparation guide (part # 4460958 Rev. A, revision date: March 2011). A quality control of libraries was performed using an Agilent 2100 Bioanalyzer (Agilent Technologies, Böblingen, Germany) and quantification was done by qPCR using a SOLiD Library TaqMan Quantification kit (Applied Biosystems, Cat. # A12127). For both libraries a 2 x E80 bead preparation scale was chosen, using an input of 1 pM each. Each E80 preparation was then loaded onto one full slide and sequencing was performed in two runs with one HAB and one LAB library at a time. Subsequent analysis of sequencing data was done as follows:

In a first step all reads with an average quality score (Phred-like score) below 10 were removed. For aligning the paired sequences to the mouse reference genome (UCSC genome browser version mm9) the two burrows-wheeler aligners bowtie (v0.12.7) [40] and BWA (v0.5.7) [41] were used in a step-wise procedure. First, bowtie was used to align read pairs in correct orientation (i.e., expected orientation and distance) to the reference genome. Second, the slower but more flexible BWA software was used for aligning previously unmapped read pairs in alternative orientations (i.e., unexpected orientations and distances). Third, BWA was used to align single reads (i.e., reads where the mate did not pass the quality control). The rationale behind this stepwise procedure was to leverage the computational performance of bowtie as well as the flexibility of BWA. Allowing reads to be aligned in alternative orientations and distances is crucial for the detection of large structural variants, since large deletions and insertions cause an increase and a reduction, respectively, in the distance between the mates, while inversions are detectable with unexpected orientations of the two mates (see Xi et al. [42] for a review). The alignments were sorted according to genomic locations and successfully mapped reads from the three steps above were merged into a single alignment file. For post-processing we used the tools Picard (http://picard.sourceforge.net/) and SAMtools [43].

CNV calling was based on a depth-of-coverage (DOC) approach [42]. Briefly, this approach works by comparing the effective coverage in one region in the LAB sample to the effective coverage in the same region in the HAB sample: if, for instance, in HAB this region shows four-times the coverage compared to the LAB sample, then this particular region is likely to be a

CNV with four-times the copies. The CNV calling was performed using our newly developed software CNVfinder (http://cnvfinder.sourceforge.net/). Briefly, CNVfinder divides the genome into equally sized bins. Next, the coverage for each bin is computed as the sum of (i) the number of reads that are aligned to that bin and (ii) the number of read pairs whose insert size are covering that bin. This was done separately for the HAB and LAB samples. Finally, CNVfinder works at one chromosome at a time and compares the coverage for each bin on that chromosome in the two samples using Fisher's exact test. The derived p-values are corrected for multiple testing using the method by Benjamini and Hochberg [44] resulting in a q-value for a false discovery rate. Candidate CNVs were identified by consecutive stretches of bins all exceeding a q-value threshold; candidate CNVs were extended to the left and right by including bins exceeding a lower (extension) q-value threshold.

For the analysis, we restricted the alignments to uniquely mapped mate pairs as indicated by a mapping score of 20 or better. Next, from the alignment, we extracted the information of chromosome, alignment position, insert size, and alignment flag (which contains information about the orientations of the two mates). Then, we removed PCR artifacts by removing all but one aligned read with the same start position and insert length. For the analysis with CNVfinder, the bin size was set to 200 bp, the initial CNV finding threshold was set to—$\log_{10}(q) > 12$ with a window size of 8 and the CNV extension threshold was set to—$\log_{10}(q) > 10$. Hence, the lower limit on the length for detectable CNV was 1,600 bp (8 x 200 bp). There were no limits regarding the maximal length of a CNV.

## Association analysis of CNV and behavior data

Prior to performing the association analysis of CNVs detected in 64 CD-1 mice (as described above) with all behavioral parameters tested, we aimed to take the relationship between those mice into account. Thus, behavioral traits were transformed with GenABEL's function "polygenic" and its GRAMMAR+ transformation output [45] in R. The association analysis was then performed based on a generalized linear model (i.e., the R function "glm"), using the transformed behavioral data. For measuring values of CNVs, the mean normalized intensities of JaxMDGA probes within the respective CNV were used. A likelihood-ratio test as implemented in the "anova" function of the "stats" R package was applied to calculate $p$-values. Distribution of $p$-values was checked by Q-Q plots (shown in S2 Fig). The $p$-values were adjusted for multiple testing using a correction method developed by Holm [46].

## Gene expression analysis

RNA extracted from basolateral (BLA) and central amygdala (CeA), hypothalamic paraventricular nucleus (PVN) and anterior part of cingulate cortex (Cg) of eight HAB and eight LAB animals was tested on a MouseWG-6 v1.1 Expression Bead Chip (Illumina, San Diego, CA, USA), containing 48,318 probes. The samples were not pooled (for details see reference [47]). A reanalysis of raw data from that earlier expression microarray experiment [47] was conducted (for raw date see GEO database GSE29015, http://www.ncbi.nlm.nih.gov/gds/). Data were normalized using the R function "vsn" [48]. Probes which were not sampled on at least one array were dropped and the remaining probes (N = 46,657) were further filtered by three criteria: 1. Probes had to map into genes with an EntrezGene-ID; 2. A unique alignment of probes to the genome (mm9) with a maximum of two mismatches was required; 3. Genes having a detection p-value greater than $1x10^{-4}$ were excluded. Probes passing the filtering conditions (N = 12,171) were used for subsequent analysis, using the R package "Limma" [49]. Significantly regulated genes were ranked using an empirical Bayes method [50]. Multiple testing was corrected for using the false discovery rate (FDR) approach [44]. Finally, significant

expression differences were indicated by adjusted p-values less than 0.05, obtained from the performance of a contrast analysis.

We confirmed the results of the gene expression microarray by quantitative real-time PCR (qPCR). Brains of 11 HAB and 8 LAB males were cut in slices of 200 μm using a Microm HM560 cryostat (Microm, Walldorf, Germany). The brain areas of interest, CeA (bregma [51] 1.46 to 1.82), BLA (bregma 1.22 to 1.58), Cg (bregma 1.34 to 0.22), and PVN (bregma 0.58 to 0.94), were obtained by micropuncture, using sample corers of 0.5 mm (BLA, CeA) and 1.0 mm (Cg, PVN) in diameter (Fine Science Tools, Heidelberg, Germany). RNA was isolated by means of an RNeasy Plus Micro Kit (Qiagen) following the kit's protocol. cDNA was transcribed from 0.5 μg RNA each, using a High Capacity cDNA Reverse Transcriptase Kit (Applied Biosystems), and further mixed with SYBR Green MasterMix (Qiagen) to be tested in duplicates on a LightCycler 480 (Roche Diagnostics, Mannheim; Germany). Primers (S1 Table) targeting two housekeeper and nine candidate genes shown to be both differentially expressed between HAB and LAB mice and part of functional protein association networks (STRING online software v9.0; http://string-db.org) were designed using the online tool Primer-Blast (www.ncbi.nlm.nih.gov/tools/primer-blast), and were purchased from Sigma-Aldrich (Taufkirchen, Germany). Data analysis was performed using the absolute quantification fit points method, provided with the LightCycler software. Sample data were analyzed relative to the housekeeper using the ΔΔCT-method [52], and normalized to the mean value of all HAB samples. Expression differences between HAB and LAB mice were determined applying the Mann-Whitney U test (SPSS software v16.0.1), assuming a significance threshold of 0.05.

## Correlation analysis of CNV and expression data

We tested the hypothesis that CNVs have an influence on gene expression using a two proportion Z-test, as described in S1 Text. Next, we examined the question of an underlying positive or negative correlation, starting with the creation of a list including all genes in copy number variable regions that were found to be differentially expressed between HAB and LAB mice in at least one of the tested brain regions (CeA, BLA, Cg, PVN). Information on copy number status was added by assigning a value of +1 (or -1) to all genes in regions of copy number gain (or loss, respectively) in HAB compared to LAB mice. The same was done for information on expression status, with a value of +1 (or -1) assigned to genes showing increased (or reduced, respectively) expression in HAB mice. Genes were assigned to both +1 and -1 when showing contradictory CNV status based on the three applied CNV detection methods, or contradictory expression status based on different microarray probes targeting the gene, respectively. We tested for a correlation between expression and CNV status by applying the Cohen's weighted Kappa [53], using the "cohen.kappa" function of the "psych" library in R. The p-values were corrected for multiple testing using Holm's correction method [46].

## Results

### CNVs related to anxiety

In total, we found 98 CNVs in HAB vs. LAB mice using aCGH, 180 and 5,851 CNVs by means of JaxMDGA and NGS, respectively. Their median size was calculated to be 2.4 kb (NGS), 8.9 kb (JaxMDGA) and 544.6 kb (aCGH). The size range of CNVs detected by the two arrays was in a range to be expected. [23, 54–58] All CNVs taken together, their total size of about 97.3 Mb (aCGH), 14.7 Mb (JaxMDGA) and 27.2 Mb (NGS) represent about 3.7% (aCGH), 0.6% (JaxMDGA) and 1.0% (NGS) of the whole genome, respectively. Sequences revealed by all three detection methods to differ in copy number covered about 4.8 Mb, which reflects 0.18% of the genome. In most cases, these findings were not contradictory, however, for about 168 kb

altogether a copy number gain in HAB vs. LAB mice was shown by one method while a loss was discovered by the other methods, or vice versa. S2 Table details all CNVs detected by aCGH, JaxMDGA and NGS, including their genomic position and copy number status with respect to HAB mice. Further details on total and mean CNV size and overlap between results of distinct detection methods can be found in S3 Table. Examples of CNVs and their log2 signal intensity ratios (aCGH, JaxMDGA), or fold changes (NGS), respectively, are plotted in S3–S5 Figs.

With the intention to confirm the relevance of specific CNVs for anxiety-related behavior in a second mouse model, we genotyped 64 CD-1 mice using JaxMDGA. Comparing the raw data of all 64 animals to each other we revealed a total number of 764 CNVs with a median size of 14 kb. In order to define CNVs linked to anxiety-related behavior, we ran association analyses of CNVs with multiple behavioral parameters. Before correcting for multiple testing, we found 47 CNVs to be significantly associated (nominal $p$-value < 0.05) with the time the animals spent on the open arms of the EPM. For another 40 CNVs and the same behavioral parameter a trend (nominal $p$-value < 0.1) was shown. However, none of these effects survived the correction for multiple testing. The mean normalized intensities of all 764 CNVs and the time the 64 animals spent on the open arm of the EPM are shown in S4a Table. Data of three distinct CNVs are depicted exemplarily in Fig 2. In S4b and S4c Tables behavioral data of all tests are shown before and after GRAMMAR+ transformation, respectively. All CNVs including nominal and corrected $p$-values of the association analyses with all behavioral tests performed are outlined in S5 Table. Fig 3 illustrates the chromosomal distribution of CNVs found in CD-1 mice, with regions having nominal $p$-values less than 0.1 highlighted in color.

For the comparison of CNVs in CD-1 with those found in HAB/LAB mice, the distribution of CNVs in both mouse models is depicted in S6 Fig, and exemplarily for chromosome 3 in Fig 4.

## CNVs and gene expression

Out of 12,171 expression microarray probes that passed the filtering process, we found 291 probes in CeA, 117 in BLA, 297 in PVN, and 254 in Cg, respectively, to show expression differences between HAB and LAB mice. S6a Table details the probes' genomic positions and respective results of microarray analysis. The 12,171 probes represented 8,981 protein coding genes of which 374 appeared to be differentially expressed in at least one of the tested brain regions. These 374 genes, as well as information on the microarray probes representing them, are listed
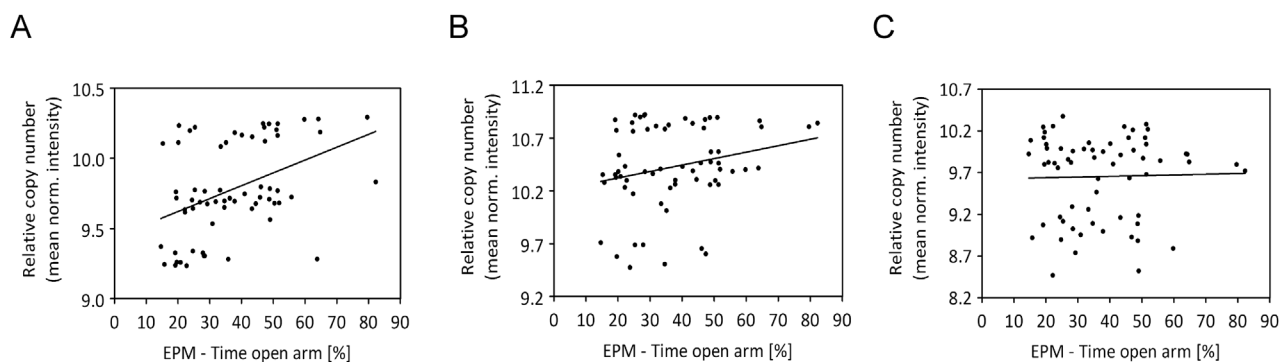


**Fig 2. Association of copy number with anxiety-related behavior in CD-1 mice.** Exemplarily, data of three associations resulting in nominal $p$-values reaching significance ($p < 0.05$), a trend ($p < 0.1$), and not reaching significance ($p > 0.05$), respectively, are shown. Each dot represents data of a single animal (N = 64). The relative copy number is represented by the mean normalized intensities of JaxMDGA probes within the respective CNV. **(A)** CNV no. 498; $P_{nom} = 0.0009$; regression line: y = 0.0091x + 9.4389. **(B)** CNV no. 164; $P_{nom} = 0.0554$; regression line: y = 0.0061x + 10.201. **(C)** CNV no. 453; $P_{nom} = 0.9791$; regression line: y = 0.0008x + 9.6225.
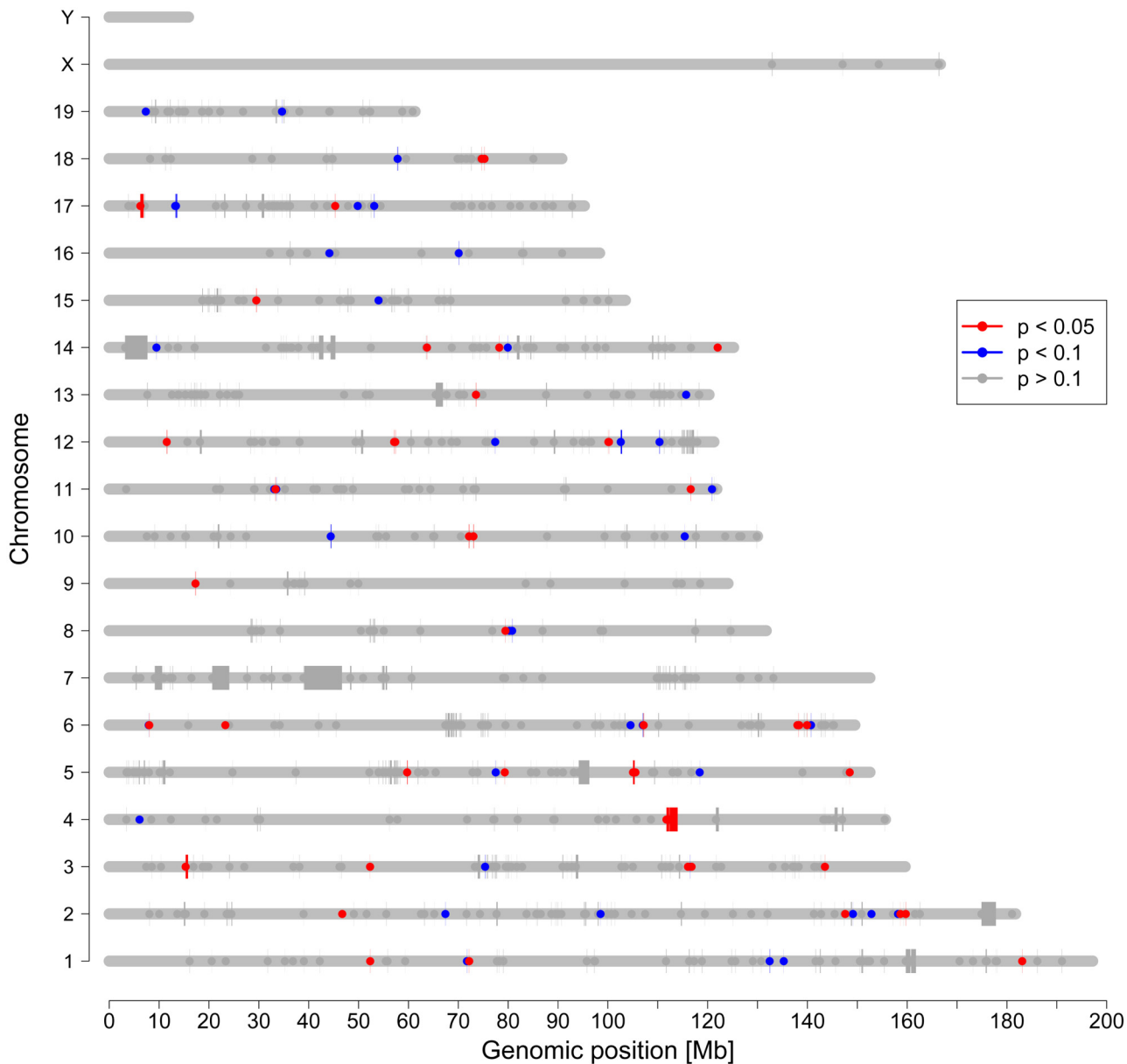
doi:10.1371/journal.pone.0128465.g002

**Fig 3. Distribution of CNVs in CD-1 mice.** Chromosomes are indicated by grey horizontal lines. Start points of CNVs are marked by dots and lines are drawn to the end points. Due to limitations in resolution, a small CNV might appear as dot only. CNVs highlighted in blue or red were associated with anxiety-related behavior (time on the open arm of the EPM) with a nominal *p*-value less than 0.1 or 0.05, respectively.

doi:10.1371/journal.pone.0128465.g003

in S6b Table. We confirmed the reliability of gene expression changes revealed by microarray analysis in a small subset of nine genes using qPCR. The data are shown in S7 Table.

We then statistically analyzed the overlap of genes that were tested in the expression micro-array with CNVs which were confirmed by three detection methods, i.e. aCGH, JaxMDGA and NGS, by applying a two-proportion $Z$-test. All four brain regions individually (CeA, $p = 5.9 \times 10^{-16}$; BLA, $p = 1.2 \times 10^{-21}$; PVN, $p = 3.0 \times 10^{-21}$; Cg, $p = 3.1 \times 10^{-33}$) and all four brain regions taken together ($p = 1.6 \times 10^{-31}$) show highly significant $p$-values. Thus, we demonstrated that loci with variable copy number between HAB and LAB mice include significantly more differentially expressed genes than it is to be expected if CNVs would not influence expression levels.
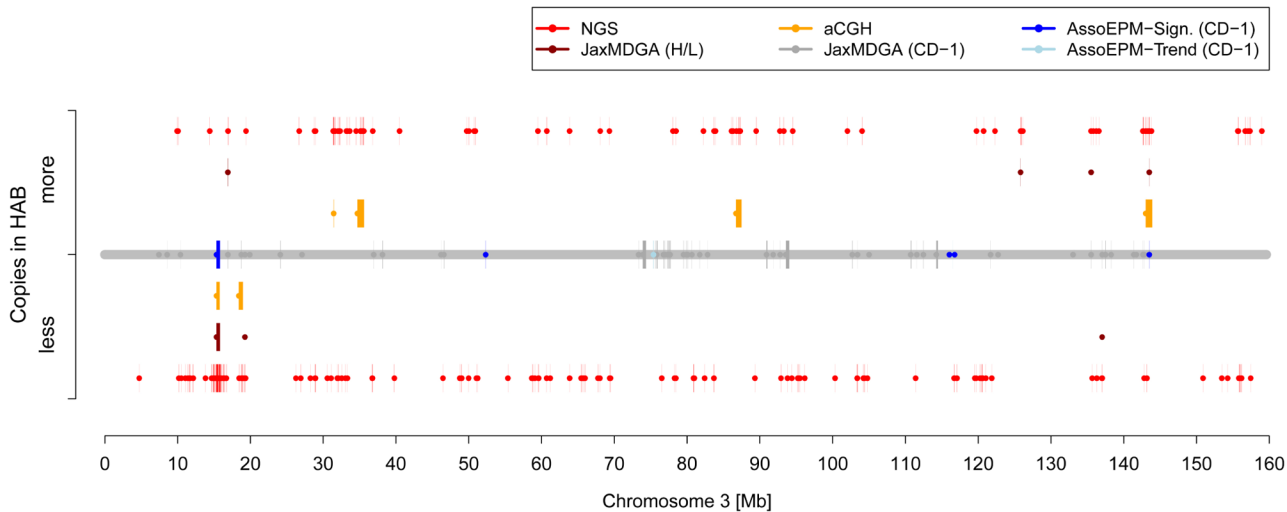
**Fig 4. Genomic positions of CNVs on chromosome 3.** The chromosome is indicated by a thick horizontal line (grey). Depending on the detection method, CNVs in HAB/LAB mice are depicted in orange (aCGH), dark red (JaxMDGA) and red (NGS), respectively. Data displayed above the grey line represent a copy number gain in HAB vs. LAB animals, data below a copy number loss. Data printed on the grey line show CNVs in 64 CD-1 mice, with those highlighted in color that could be associated with anxiety-related behavior (time on the open arm of EPM) with a nominal p-value less than 0.1 (light blue) or less than 0.05 (blue). Start points of CNVs are marked by dots and lines are drawn to the end points.

doi:10.1371/journal.pone.0128465.g004

In other words, CNVs affect the expression level of genes located in the CNVs. Further, the subsequently performed correlation analysis (Cohen's weighted Kappa) revealed a significantly ($p < 0.05$) positive correlation between copy number and gene expression in three of the four brain regions examined (CeA: $K = 0.260$, $p = 0.0028$; BLA: $K = 0.124$, $p = 0.0815$; PVN: $K = 0.270$, $p = 0.0028$; Cg: $K = 0.230$, $p = 0.0056$). That is, more copies of a gene result in higher expression levels.

## Candidate genes of anxiety

After having shown that CNVs are likely to influence gene expression, we identified all protein coding genes in genomic regions revealed to differ in copy number between HAB and LAB mice. The resulting 998 (aCGH), 145 (JaxMDGA), and 1,085 (NGS) genes, respectively, are listed in S8a–S8c Table. In order to improve the reliability of candidate genes, we kept only those genes found by all three methods. The remaining 68 genes (S8d Table) were further compared with protein coding genes within CNVs best associated with anxiety-related behavior in CD-1 mice. Therefore, in a first step, all genes (N = 911) found to overlap with CNVs in 64 CD-1 mice were identified (S8e Table). Next, considering only those CNVs that were best associated with the time animals spent on the open arm of the EPM (nominal p-values < 0.1, or < 0.05), the latter list was reduced to 78 genes (S8f Table). Finally, we compared those 78 genes with the 68 genes found inside regions of HAB/LAB CNVs. Among the two sets 15 genes overlapped. Based on a permutation test with 10,000 permutations, the overlap was shown to be significant ($p = 0.0051$). These 15 genes were considered promising candidate genes of anxiety-related behavior and are shown in Table 1.

Detailed data not shown in this manuscript are available upon request.

## Discussion

In recent years, multiple methods have been developed to screen for CNVs in genome-wide approaches. Nevertheless, CNV detection remains a challenging task since results do not only

**Table 1. Protein coding genes in genomic regions of CNVs detected in HAB/LAB and CD-1 mice.**

| Gene | | | | | CNV (in CD-1 mice) | | | Association |
|---|---|---|---|---|---|---|---|---|
| Symbol | MGI ID | Chr | Start | End | Chr | Start | End | nominal *p*-value |
| Sirpb1a | 2444824 | 3 | 15371819 | 15426504 | 3 | 15340258 | 15819607 | 0.01213 |
| Sirpb1b | 3779828 | 3 | 15495754 | 15575065 | 3 | 15340258 | 15819607 | 0.01213 |
| Sirpb1c | 3807521 | 3 | 15695145 | 15748528 | 3 | 15340258 | 15819607 | 0.01213 |
| Skint4 | 2444425 | 4 | 111744621 | 111840681 | 4 | 111745396 | 112286229 | 0.04514 |
| Skint3 | 3045331 | 4 | 111904850 | 111973073 | 4 | 111745396 | 112286229 | 0.04514 |
| Skint9 | 3045341 | 4 | 112058574 | 112106590 | 4 | 111745396 | 112286229 | 0.04514 |
| Skint6 | 3649262 | 4 | 112908844 | 112959568 | 4 | 112348832 | 113968233 | 0.02412 |
| Skint5 | 3650151 | 4 | 113613249 | 113672102 | 4 | 112348832 | 113968233 | 0.02412 |
| Skint11 | 2685415 | 4 | 113835989 | 113917633 | 4 | 112348832 | 113968233 | 0.02412 |
| C230055K05Rik | 2441896 | 5 | 105242188 | 105288830 | 5 | 105034912 | 105359320 | 0.03593 |
| Tcp10b | 98542 | 17 | 13253977 | 13275092 | 17 | 13259210 | 13301725 | 0.07376 |
| Gm10512 | 3642173 | 17 | 13397908 | 13399077 | 17 | 13398037 | 13649185 | 0.07250 |
| Smok2a | 1351487 | 17 | 13414054 | 13420524 | 17 | 13398037 | 13649185 | 0.07250 |
| Smok2b | 3037705 | 17 | 13421718 | 13430055 | 17 | 13398037 | 13649185 | 0.07250 |
| Gm9880 | 3711246 | 17 | 13547438 | 13569717 | 17 | 13398037 | 13649185 | 0.07250 |

All genes listed overlap both CNVs in HAB/LAB mice detected with aCGH, JaxMDGA and NGS, and CNVs in CD-1 mice which were best associated with the time the animals spent on the open arm of the EPM (nominal *p*-value < 0.1).

doi:10.1371/journal.pone.0128465.t001

depend on the method employed but also on the applied algorithm, and no "gold standard" algorithm has been established so far (see, for example, refs. [57, 59–62]). We consider it very important that the reader is aware of the fact that there is neither a generally accepted "gold standard" for a CNV detection method nor for an algorithm, which means that all data generated by just a single method should be regarded and handled with care. This applies not only to our study but also to others. We would like to point out that our association study in CD-1 mice is based on CNV data generated by JaxMDGA solely. Although results were generated thoroughly, they depend on the bias of the applied method and algorithm. Using two probe-based high-density genotyping arrays (aCGH, JaxMDGA) and one whole-genome next-generation sequencing approach (NGS) to reveal CNVs in the HAB/LAB mouse model, we could overcome the limitations of a single approach and thus increase the reliability of the CNV data obtained (S2 Table). The use of multiple processing pipelines and/or data sources is also frequently employed in SNP calling from NGS data in order to improve the reliability of called SNPs and to reduce the chance of false-positive calls [63]. However, the precise breakpoints of CNVs in HAB/LAB mice are likely to be somewhere in between the results of the three applied detection methods.

After having detected the CNVs in HAB and LAB mice, we were further interested in their functional impact. Since effects of CNVs on gene expression levels have been reported before, showing a positive correlation in some cases and a negative one in others [18, 64, 65], we wanted to know, first, if CNVs mediate an effect on gene expression in HAB/LAB mice and, second, if there is a positive or negative correlation between them. Therefore, we compared CNV data with expression data resulting from a genome-wide gene expression analysis. The reliability of expression data was demonstrated by qPCR, with most qPCR outcomes confirming the results of the microarray analysis. Reasons for contradictory findings in array and qPCR data due to a potential method-specific bias are manifold and have already been discussed before by others [66, 67]. We, first, were able to demonstrate that a significant number of genes

showing expression differences between HAB and LAB mice were located in CNV regions and, second, found a positive correlation in CeA, Cg and PVN. This outcome is in accordance with a study of Henrichsen and colleagues [56], reporting that the expression of genes within CNVs tends to correlate with changes in copy number. The authors suggested CNVs to "play an even more important role with respect to normal phenotypic variation and risk of complex disease than previously anticipated". Our study clearly supports this suggestion. Consequently, CNVs should be considered as an influencing factor of gene expression and hence phenotypic variations with respect to anxiety phenotypes. The fact that we could not show a significant correlation between gene dosage and relative expression levels in the BLA might be explained by the CNVs' modes of action themselves, as reviewed in a recent publication [23]. For example, CNVs might, on one hand, increase gene expression simply by altering gene dosage, and, on the other hand, decrease expression by a negative feedback loop following an increase of the gene products due to the duplication of coding regions [23]. Besides, several environmental and genetic factors such as SNPs, epigenetic factors and others are known to influence gene expression. Thus, especially in case of complex phenotypes, the effects mediated by these factors are likely to interfere with those mediated by CNVs. However, as we could confirm in three of four brain regions tested, the impact of CNVs on gene expression and hence behavioral phenotypes is substantial.

In order to confirm anxiety-relevant CNVs and corresponding genes, we examined CNVs in a second mouse model, the CD-1 outbred mouse model. CNV detection in these mice was performed using an array-based genotyping approach (JaxMDGA). Although multiple computational methods are available to detect CNVs in raw data of genotyping arrays, all commonly applied algorithms compare raw data of each sample against a single reference sample. One consequence of this procedure is an increase in false-negative calls and thus a potential loss of information on CNVs between non-reference samples. We improved CNV calling in the tested 64 CD-1 mice by analyzing the data using each sample once as reference, and thereby conducting all pair-wise comparisons. The mean signal intensities of all JaxMDGA probes within a defined CNV were calculated. In order to keep the number of analysis steps and the related bias as low as possible, we did not define clusters of relative copy numbers, but directly used the mean signal intensities to perform an association analysis of CNVs with anxiety-related behavior. Considering the comparability with HAB/LAB CNV data, the most interesting behavioral parameter is the selection parameter of the HAB/LAB mouse model, that is, the percent time animals spent on the open arm of the EPM. Indeed, we found significant associations of CNVs with this parameter at the nominal $p$-values but not at the threshold corrected for multiple comparisons. This is a well-known phenomenon in genome-wide association studies of complex phenotypes, since common variants with small effect sizes and rare variants are difficult to catch, and multiple-testing correction would often require larger sample sizes than feasible in order to be able to detect these effects [68–70]. 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 [71]. Thus, specific genes and genetic factors might be less strongly associated with complex traits and diseases than particular patterns of genetic variation and environmental interaction [72]. As a consequence, an association study of a single factor of small effect size with a complex trait leads to significant $p$-values only if other factors are considered in the analysis as well. If such a factor is analyzed solely, significance might get lost, which does not necessarily mean that the single factor *per se* is irrelevant. Almal and Padh [73] have recently 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". To date, however, more complex calculations of associations including all or at least most of the suggested influencing factors are not feasible since the number of these factors pushes the limits of computing capacity. Even though computational methods like those recently developed by Kam-Thong and colleagues [74, 75] facilitate at least the calculation of pair-wise interactions, it will take time until factors of more complex patterns can be offset against each other in a cost- and time-effective manner. Therefore, we decided to use a permutation test to successfully confirm the relevance of genes within CNVs showing nominally significant association, or a statistical trend in their association with anxiety-related behavior.

To our knowledge, the 15 candidate genes revealed in our study were not shown to be linked to anxiety before. However, two of the genes, *Sirpb1a* and *Sirpb1c*, have human equivalents that belong to the so called *SIRP* family. Genes of that large family encode proteins involved in the regulation of signals defining different physiological and pathological processes [76]. *SIRP* family members were suggested to be involved in the activation of the MAPK pathway [77, 78], which was not only shown to play a role in cell differentiation and survival, growth control and cellular adaptation to chemical and physical stress [79–82], but also to be linked to anxiety and depression [83–85]. Thus, further studies on the connection of *Sirpb* genes, the MAPK pathway, and anxiety might be promising.

Interestingly, the *Glo1* gene, a gene within a large and common CNV [18, 64], was not amongst our 15 candidate genes. The *Glo1* gene has been described to influence anxiety-related behavior before [18, 25, 86–89], as discussed in a review by Distler and Palmer [90]. Although we found the CNV including, amongst others, the *Glo1* locus in CD-1 mice and HAB/LAB mice, with more copies and higher expression of *Glo1* in LAB mice, our association study in CD-1 mice did not reveal any influence on anxiety-related behavior. This might be explained by an insufficient statistical power using 64 animals in our study. Using 64 animals only, the power for any QTL analysis is limited, however, the power to show an increase in the respective QTLs based on 64 CD-1 mice and 764 CNVs from an expected number (as done here by a permutation test) is much higher, but maybe not high enough. Since anxiety is a complex trait, the effect of the CNV in this particular experimental setup could be offset by other factors and, thus, many more animals would be required to reveal the effect in an association study. However, similarly, a recent study also could not show an effect of the respective CNV and anxiety-related behavior [91, 92]. Aim of this study was to analyze the influence of *Fkbp5* (FK506 binding protein 5) deficiency on the physiological stress response in *Fkbp5*$^{-/-}$ mice [91]. In a follow-up study analyzing the expression profile of *Fkbp5*$^{-/-}$ and *Fkbp5*$^{+/+}$ mice, an increased expression of *Glo1* mRNA in *Fkbp5*$^{-/-}$ mice was observed, which was shown to result from a co-selection of the *Glo1* duplication with the *Fkbp5*$^{-/-}$allele [92]. Thus, no influence of *Glo1* expression on anxiety-related behavior could be observed in *Fkbp5*$^{-/-}$ and *Fkbp5*$^{+/+}$ mice [91, 92]. Another study using BAC transgenic mice to overexpress *Glo1* demonstrated increased anxiety-related behavior, however, this effect was only observed in mouse lines with the highest copy numbers [88]. In conclusion, Glo1 seems to play a role in the regulation of anxiety-related phenotypes, however its precise effect and the influence of the respective CNV remains to be discovered. As discussed above, complex phenotypes are influenced by a variety of distinct factors; thus, a protein such as Glo1 might be of importance only in the context of a specific genomic background or a certain metabolic state. For example, it was hypothesized that Glo1 affects anxiety-related behavior by controlling levels of methylglyoxal [88, 89]. However, further investigations are required to shed light on the link between Glo1 and anxiety. Although we do not want to extend the discussion at this point, it might also be of interest to examine the relationship between *Glo1* and depression-like behavior in future, since our association study of CD-1 mice showed promising results in the TST (see S5 Table, CNV No. 680).

Taken together, our study provides an extensive catalogue of CNVs and corresponding genes potentially linked to anxiety-related behavior in CD-1 mice. Even though their precise role remains to be investigated, we suggest that these loci might be of interest for future studies focusing on biomarkers of anxiety. Furthermore, with the revelation of CNVs in CD-1 mice, we provide the basis for further investigations of the effects of CNVs in general.

## Supporting Information

**S1 Fig. Breakpoint definition of CNVs in CD-1 mice.** If regions defined as CNVs (orange lines) by applying the „simple CNV"function showed a huge overlap between several sample comparisons, their breakpoints (black dashed lines) were unitized to consider the region as one CNV only. New breakpoints are indicated by green lines. S1 = sample 1, S2 = sample 2, S3 = sample 3.
(TIF)

**S2 Fig. Q-Q plots of the nominal *p*-values resulting from the associations of CNVs with distinct behavioral parameters.** Each behavioral parameter is shown in a distinct plot. Expected *p*-values (x-axis) are plotted against observed *p*-values (y-axis) in logarithmic scale.
(PDF)

**S3 Fig. Log2 signal intensity ratios of aCGH probes in HAB/LAB mice.** The signal ratio of each probe (black dots) refers to the signal intensity of HAB versus LAB sample. Segments defined by "segMNT" are indicated in red. Genomic position on chromosome 17 is shown on the x-axis.
(TIF)

**S4 Fig. Log2 signal intensity ratios of JaxMDGA probes in HAB/LAB mice.** The signal ratio of each probe (black dots) refers to the signal intensity of HAB versus LAB sample. CNVs defined by "simpleCNV" are indicated in red. Genomic position on chromosome 17 is shown on the x-axis.
(TIF)

**S5 Fig. Fold change of CNVs (NGS) plotted against genomic positions on chromosome 17.** CNVs in HAB/LAB mice discovered by "CNVfinder" on Chromosome 17. The x-axis marks the genomic location of the CNV on Chromosome 17. The y-axis corresponds to the log2(fold change). Positive values indicate more copies in HAB than in LAB. Likewise, negative values indicate more copies in LAB compared to HAB.
(TIF)

**S6 Fig. Genomic positions of CNVs on chromosomes 1–19, X and Y.** Chromosomes are indicated by thick grey lines with basepair information shown on the x-axis. Depending on the detection method, CNVs in HAB/LAB mice are depicted in orange (aCGH), dark red (JaxMDGA) and red (NGS), respectively. Data displayed above the grey line represent a copy number gain in HAB vs. LAB animals, data below a copy number loss. Data printed on the grey line show CNVs in 64 CD-1 mice, with those highlighted in color that could be associated with anxiety-related behavior (time on open arm of the EPM) with a nominal *p*-value less than 0.1 (light blue) or less than 0.05 (blue). Start points of CNVs are marked by dots and lines are drawn to the end points.
(PDF)

**S1 Table. Information on primers used for qPCR.** The table is sorted by chromosome. Columns show (left to right): chromosome, gene represented by primer, primer orientation,

primer sequence (5' to 3'), melting temperature and size of the resulting PCR product.
(DOC)

**S2 Table. CNVs in HAB/LAB mice.** Table of all CNVs detected by aCGH, JaxMDGA and NGS in HAB/LAB mice. Data are sorted by genomic position. A CNV detected by one method is shown repeatedly if overlapping with more than one CNV detected by another method. The copy number status (gain/ loss) is shown with respect to HAB animals.
(XLS)

**S3 Table. Comparison of CNV detection methods.** *Upper part*: counts of respective CNVs found in HAB/LAB mice. Numbers in parentheses indicate contradictory findings of distinct methods, that is, a copy number loss found by one method and a gain found by at least one other method. 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. *Part below*: size of respective CNVs in basepairs (bp).
(DOC)

**S4 Table. Mean normalized intensities of all CNVs in 64 CD-1 mice and results of behavioral tests. (a)** In the first line the percentage of time the animals spent on the open arm of the EPM is shown. Below the mean normalized intensities of all CNVs (position information in columns A-E) are shown for the respective animals. **(b)** Original data of performed behavioral tests of all 64 animals are shown, Those data after GRAMMAR transformation are provided in **(c)**.
(XLS)

**S5 Table. Results of the association analysis of CNVs with behavior in CD-1 mice.** For each test parameter of all performed behavioral tests (EPM, FST, OF, TST, SRT), the nominal $p$-values and $p$-values corrected for multiple testing using Holm's correction method are shown. Further columns show information on CNVs' chromosomal positions and the number of probes on the array (JaxMDGA) targeting the respective CNV.
(XLS)

**S6 Table. Results of expression microarray. (a)** All probes of the expression microarray (12,171) that passed the filtering process are listed, including information about their genomic position. Up to three positions are shown if probes could be mapped to the reference genome multiple times. The $p$-values shown indicate differences in expression between HAB and LAB mice (significant if $p < 0.05$) in the respective brain area (CeA, BLA, Cg, PVN). The relative difference in expression is given as "fold change" (foldCh), with positive values indicating more expression in HAB mice. **(b)** All protein coding genes revealed to be differentially expressed between HAB and LAB mice by expression microarray are listed. The columns contain information about (from left to right): gene symbol, gene number (MGI), chromosome, gene start and end position, the number of microarray probes targeting the gene and showing significant expression differences ($p < 0.05$), and, finally, the number of probes showing more and less expression in HAB mice in the respective brain region (CeA, BLA, Cg and PVN).
(XLS)

**S7 Table. Expression differences of genes tested in qPCR.** The first part shows the 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 in gene expression between HAB and LAB mice; a $p$-value $< 0.1$ (bold and italic letters) indicates a trend.
(DOC)

**S8 Table. Protein coding genes in genomic regions of CNVs detected in HAB/LAB (a - d) and CD-1 mice (e,f).** Genes in regions of CNVs detected in HAB/LAB mice by **(a)** aCGH, **(b)** JaxMDGA and **(c)** NGS, respectively, are listed and the position information of genes and correspondent CNVs are shown. Genes are listed multiple times if overlapping with more than one CNV. In **(d)** those genes overlapping CNVs detected in HAB/LAB mice by all of the three detection methods are displayed. **(e)** Genes in genomic regions of CNVs detected in 64 CD-1 mice by JaxMDGA are listed and the position information of genes and correspondent CNVs are shown. **(f)** Reduction of table (e) to those genes overlapping CNVs that could be best associated with the time the animals spent on the open arm of the EPM (nominal $p$-values $< 0.1$, or $< 0.05$).
(XLS)

**S1 Text. Supporting Material and Methods.**
(DOC)

## Acknowledgments

## Author Contributions

## References

1. Brasch-Andersen C, Christiansen L, Tan Q, Haagerup A, Vestbo J, Kruse TA. 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 2004; 24: 208–14. PMID: 15300848

2. Jeon JP, Shim SM, Nam HY, Ryu GM, Hong EJ, Kim HL, et al. Copy number variation at leptin receptor gene locus associated with metabolic traits and the risk of type 2 diabetes mellitus. BMC Genomics 2010; 11: 426. doi: 10.1186/1471-2164-11-426 PMID: 20624279

3. Bochukova EG, Huang N, Keogh J, Henning E, Purmann C, Blaszczyk K, et al. Large, rare chromosomal deletions associated with severe early-onset obesity. Nature 2010; 463: 666–70. doi: 10.1038/nature08689 PMID: 19966786

4. Shlien A, Malkin D. Copy number variations and cancer. Genome Med 2009; 1: 62. doi: 10.1186/gm62 PMID: 19566914

5. Singleton AB, Farrer M, Johnson J, Singleton A, Hague S, Kachergus J, et al. alpha-Synuclein locus triplication causes Parkinson's disease. Science 2003; 302: 841. PMID: 14593171

6. Rovelet-Lecrux A, Hannequin D, Raux G, Le Meur N, Laquerriere A, Vital A, et al. APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy. Nat Genet 2006; 38: 24–6. PMID: 16369530

7. Levy D, Ronemus M, Yamrom B, Lee YH, Leotta A, Kendall J, et al. Rare de novo and transmitted copy-number variation in autistic spectrum disorders. Neuron 2011; 70: 886–97. doi: 10.1016/j.neuron.2011.05.015 PMID: 21658582

8. Vrijenhoek T, Buizer-Voskamp JE, van der Stelt I, Strengman E, Sabatti C, Geurts van Kessel A, et al. Recurrent CNVs disrupt three candidate genes in schizophrenia patients. Am J Hum Genet 2008; 83: 504–10. doi: 10.1016/j.ajhg.2008.09.011 PMID: 18940311

9. Lachman HM, Pedrosa E, Petruolo OA, Cockerham M, Papolos A, Novak T, et al. Increase in GSK3beta gene copy number variation in bipolar disorder. Am J Med Genet B Neuropsychiatr Genet 2007; 144B: 259–65. PMID: 17357145

10. Malhotra D, McCarthy S, Michaelson JJ, Vacic V, Burdick KE, Yoon S, et al. High frequencies of de novo CNVs in bipolar disorder and schizophrenia. Neuron 2011; 72: 951–63. doi: 10.1016/j.neuron.2011.11.007 PMID: 22196331

11. Kawamura Y, Otowa T, Koike A, Sugaya N, Yoshida E, Yasuda S, et al. A genome-wide CNV association study on panic disorder in a Japanese population. J Hum Genet 2011; 56: 852–6. doi: 10.1038/jhg.2011.117 PMID: 22011818

12. Levy RJ, Xu B, Gogos JA, Karayiorgou M. Copy number variation and psychiatric disease risk. Methods Mol Biol 2012; 838: 97–113. doi: 10.1007/978-1-61779-507-7_4 PMID: 22228008

13. Perlis RH, Ruderfer D, Hamilton SP, Ernst C. Copy number variation in subjects with major depressive disorder who attempted suicide. PLoS One 2012; 7: e46315. doi: 10.1371/journal.pone.0046315 PMID: 23029476

14. St Clair D. Copy number variation and schizophrenia. Schizophr Bull 2009; 35: 9–12. doi: 10.1093/schbul/sbn147 PMID: 18990708

15. Malhotra D, Sebat J. CNVs: harbingers of a rare variant revolution in psychiatric genetics. Cell 2012; 148: 1223–41. doi: 10.1016/j.cell.2012.02.039 PMID: 22424231

16. Grayton HM, Fernandes C, Rujescu D, Collier DA. Copy number variations in neurodevelopmental disorders. Prog Neurobiol 2012; 99: 81–91. doi: 10.1016/j.pneurobio.2012.07.005 PMID: 22813947

17. Fujimoto M, Uchida S, Watanuki T, Wakabayashi Y, Otsuki K, Matsubara T, et al. Reduced expression of glyoxalase-1 mRNA in mood disorder patients. Neurosci Lett 2008; 438: 196–9. doi: 10.1016/j.neulet.2008.04.024 PMID: 18455873

18. Williams R, Lim JE, Harr B, Wing C, Walters R, Distler MG, et al. A common and unstable copy number variant is associated with differences in Glo1 expression and anxiety-like behavior. PLoS One 2009; 4: e4649. doi: 10.1371/journal.pone.0004649 PMID: 19266052

19. Feuk L, Carson AR, Scherer SW. Structural variation in the human genome. Nat Rev Genet 2006; 7: 85–97. PMID: 16418744

20. Lupski JR, Wise CA, Kuwano A, Pentao L, Parke JT, Glaze DG, et al. Gene dosage is a mechanism for Charcot-Marie-Tooth disease type 1A. Nat Genet 1992; 1: 29–33. PMID: 1301995

21. Kleinjan DA, van Heyningen V. Long-range control of gene expression: emerging mechanisms and disruption in disease. Am J Hum Genet 2005; 76: 8–32. PMID: 15549674

22. Wu CT, Morris JR. Transvection and other homology effects. Curr Opin Genet Dev 1999; 9: 237–46. PMID: 10322135

23. Henrichsen CN, Chaignat E, Reymond A. Copy number variants, diseases and gene expression. Hum Mol Genet 2009; 18: R1–8. doi: 10.1093/hmg/ddp011 PMID: 19297395

24. Stankiewicz P, Lupski JR. Structural variation in the human genome and its role in disease. Annu Rev Med 2010; 61: 437–55. doi: 10.1146/annurev-med-100708-204735 PMID: 20059347

25. Krömer SA, Kessler MS, Milfay D, Birg IN, Bunck M, Czibere L, et al. Identification of glyoxalase-I as a protein marker in a mouse model of extremes in trait anxiety. J Neurosci 2005; 25: 4375–84. PMID: 15858064

26. Landgraf R, Kessler MS, Bunck M, Murgatroyd C, Spengler D, Zimbelmann M, et al. Candidate genes of anxiety-related behavior in HAB/LAB rats and mice: focus on vasopressin and glyoxalase-I. Neurosci Biobehav Rev 2007; 31: 89–102. PMID: 16934871

27. Sartori SB, Landgraf R, Singewald N. The clinical implications of mouse models of enhanced anxiety. Future Neurol 2011; 6: 531–571. PMID: 21901080

28. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. Genome Res 2002; 12: 996–1006. PMID: 12045153

29. Landgraf R, Wigger A. High vs low anxiety-related behavior rats: an animal model of extremes in trait anxiety. Behav Genet 2002; 32: 301–14. PMID: 12405513

30. Liebsch G, Montkowski A, Holsboer F, Landgraf R. Behavioural profiles of two Wistar rat lines selectively bred for high or low anxiety-related behaviour. Behav Brain Res 1998; 94: 301–10. PMID: 9722280

31. Pellow S, Chopin P, File SE, Briley M. Validation of open:closed arm entries in an elevated plus-maze as a measure of anxiety in the rat. J Neurosci Methods 1985; 14: 149–67. PMID: 2864480

32. Prut L, Belzung C. The open field as a paradigm to measure the effects of drugs on anxiety-like behaviors: a review. Eur J Pharmacol 2003; 463: 3–33. PMID: 12600700

33. Porsolt RD, Bertin A, Jalfre M. Behavioral despair in mice: a primary screening test for antidepressants. Arch Int Pharmacodyn Ther 1977; 229: 327–36. PMID: 596982

34. Touma C, Bunck M, Glasl L, Nussbaumer M, Palme R, Stein H, et al. Mice selected for high versus low stress reactivity: a new animal model for affective disorders. Psychoneuroendocrinology 2008; 33: 839–62. doi: 10.1016/j.psyneuen.2008.03.013 PMID: 18502051

35. Steru L, Chermat R, Thierry B, Simon P. The tail suspension test: a new method for screening antidepressants in mice. Psychopharmacology (Berl) 1985; 85: 367–70. PMID: 3923523

36. Yang H, Ding Y, Hutchins LN, Szatkiewicz J, Bell TA, Paigen BJ, et al. A customized and versatile high-density genotyping array for the mouse. Nat Methods 2009; 6: 663–6. doi: 10.1038/nmeth.1359 PMID: 19668205

37. Widner-Andrä R. Assignment of functional impact on genetic data in two mouse models of affective disorders. Doctoral thesis, Ludwig-Maximilians-Universität München. 2011. Available: http://edoc.ub.uni-muenchen.de/14380/

38. Yang H. MouseDivGeno: Tools for analyzing Mouse Diversity Array; 2010. Accessed: http://cgd.jax.org/tools/mousedivgeno/

39. Yang H, Sheppard K. MouseDivGeno Vignette; 2011. Accessed: http://cgd.jax.org/tools/mousedivgeno/MouseDivGeno.pdf

40. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 2009; 10: R25. doi: 10.1186/gb-2009-10-3-r25 PMID: 19261174

41. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009; 25: 1754–60. doi: 10.1093/bioinformatics/btp324 PMID: 19451168

42. Xi R, Kim TM, Park PJ. Detecting structural variations in the human genome using next generation sequencing. Brief Funct Genomics 2010; 9: 405–15. doi: 10.1093/bfgp/elq025 PMID: 21216738

43. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009; 25: 2078–9. doi: 10.1093/bioinformatics/btp352 PMID: 19505943

44. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society. Series B (Methodological). 1995; 57: 289–300.

45. Belonogova NM, Svishcheva GR, van Duijn CM, Aulchenko YS, Axenovich TI. Region-based association analysis of human quantitative traits in related individuals. PLoS One 2013; 8: e65395. doi: 10.1371/journal.pone.0065395 PMID: 23799013

46. Holm S. A Simple Sequentially Rejective Multiple Test Procedure. Scand J Statist 1979; 6: 65–70.

47. Czibere L, Baur LA, Wittmann A, Gemmeke K, Steiner A, Weber P, et al. Profiling trait anxiety: transcriptome analysis reveals cathepsin B (Ctsb) as a novel candidate gene for emotionality in mice. PLoS One 2011; 6: e23604. doi: 10.1371/journal.pone.0023604 PMID: 21897848

48. Huber W, von Heydebreck A, Sultmann H, Poustka A, Vingron M. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. Bioinformatics 2002; 18 Suppl 1: S96–104. PMID: 12169536

49. Smyth GK. Limma: linear models for microarray data. In: Gentleman R, Carey VJ, editors. Bioinformatics and Computational Biology Solutions using R and Bioconductor—Statistics for Biology and Health. New York: Springer; 2005. pp. 397–420. doi: 10.1093/bioinformatics/btp578 PMID: 22847932

50. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol 2004; 3: Article3. PMID: 16646809

51. Paxinos G, Franklin KBJ. The Mouse Brain in Stereotaxic Coordinates. 2nd ed. San Diego (USA), London (UK): Academic Press; 2001.

52. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 2001; 25: 402–8. PMID: 11846609

53. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. Psychol Bull 1968; 70: 213–20. PMID: 19673146

54. Cutler G, Marshall LA, Chin N, Baribault H, Kassner PD. Significant gene content variation characterizes the genomes of inbred mouse strains. Genome Res 2007; 17: 1743–54. PMID: 17989247

55. Graubert TA, Cahan P, Edwin D, Selzer RR, Richmond TA, Eis PS, et al. A high-resolution map of segmental DNA copy number variation in the mouse genome. PLoS Genet 2007; 3: e3. PMID: 17206864

56. Henrichsen CN, Vinckenbosch N, Zollner S, Chaignat E, Pradervand S, Schutz F, et al. Segmental copy number variation shapes tissue transcriptomes. Nat Genet 2009; 41: 424–9. doi: 10.1038/ng.345 PMID: 19270705

57. Pinto D, Darvishi K, Shi X, Rajan D, Rigler D, Fitzgerald T, et al. Comprehensive assessment of array-based platforms and calling algorithms for detection of copy number variants. Nat Biotechnol 2011; 29: 512–20. doi: 10.1038/nbt.1852 PMID: 21552272

58. Haraksingh RR, Abyzov A, Gerstein M, Urban AE, Snyder M. Genome-wide mapping of copy number variation in humans: comparative analysis of high resolution array platforms. PLoS One 2011; 6: e27859. doi: 10.1371/journal.pone.0027859 PMID: 22140474

59. Alkan C, Coe BP, Eichler EE. Genome structural variation discovery and genotyping. Nat Rev Genet 2011; 12: 363–76. doi: 10.1038/nrg2958 PMID: 21358748

60. van de Wiel MA, Picard F, van Wieringen WN, Ylstra B. Preprocessing and downstream analysis of microarray DNA copy number profiles. Brief Bioinform 2010; 12: 10–21. doi: 10.1093/bib/bbq004 PMID: 20172948

61. Warden M, Pique-Regi R, Ortega A, Asgharzadeh S. Bioinformatics for copy number variation data. Methods Mol Biol 2011; 719: 235–49. doi: 10.1007/978-1-61779-027-0_11 PMID: 21370087

62. Zhang D, Qian Y, Akula N, Alliey-Rodriguez N, Tang J, Gershon ES, et al. Accuracy of CNV Detection from GWAS Data. PLoS One 2011; 6: e14511. doi: 10.1371/journal.pone.0014511 PMID: 21249187

63. Altmann A, Weber P, Bader D, Preuss M, Binder EB, Muller-Myhsok B. A beginners guide to SNP calling from high-throughput DNA-sequencing data. Hum Genet 2012; 131: 1541–54. doi: 10.1007/s00439-012-1213-z PMID: 22886560

64. Cahan P, Li Y, Izumi M, Graubert TA. The impact of copy number variation on local gene expression in mouse hematopoietic stem and progenitor cells. Nat Genet 2009; 41: 430–7. doi: 10.1038/ng.350 PMID: 19270704

65. Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, Thorne N, et al. Relative impact of nucleotide and copy number variation on gene expression phenotypes. Science 2007; 315: 848–53. PMID: 17289997

66. Bernardo V, Ribeiro Pinto LF, Albano RM. Gene expression analysis by real-time PCR: experimental demonstration of PCR detection limits. Anal Biochem 2013; 432: 131–3. doi: 10.1016/j.ab.2012.09.029 PMID: 23026780

67. Roberts PC. Gene expression microarray data analysis demystified. Biotechnol Annu Rev 2008; 14: 29–61. doi: 10.1016/S1387-2656(08)00002-1 PMID: 18606359

68. Diekstra FP, Saris CG, van Rheenen W, Franke L, Jansen RC, van Es MA, et al. Mapping of gene expression reveals CYP27A1 as a susceptibility gene for sporadic ALS. PLoS One 2012; 7: e35333. doi: 10.1371/journal.pone.0035333 PMID: 22509407

69. Farber CR. Systems-level analysis of genome-wide association data. G3 (Bethesda) 2013; 3: 119–29. doi: 10.1534/g3.112.004788 PMID: 23316444

70. Altshuler D, Daly MJ, Lander ES. Genetic mapping in human disease. Science 2008; 322: 881–8. doi: 10.1126/science.1156409 PMID: 18988837

71. Plomin R, Haworth CM, Davis OS. Common disorders are quantitative traits. Nat Rev Genet 2009; 10: 872–8. doi: 10.1038/nrg2670 PMID: 19859063

72. Rucker JJ, Breen G, Pinto D, Pedroso I, Lewis CM, Cohen-Woods S, et al. Genome-wide association analysis of copy number variation in recurrent depressive disorder. Mol Psychiatry 2013; 18: 183–9. doi: 10.1038/mp.2011.144 PMID: 22042228

73. Almal SH, Padh H. Implications of gene copy-number variation in health and diseases. J Hum Genet 2012; 57: 6–13. doi: 10.1038/jhg.2011.108 PMID: 21956041

74. Kam-Thong T, Czamara D, Tsuda K, Borgwardt K, Lewis CM, Erhardt-Lehmann A, et al. EPIBLASTER-fast exhaustive two-locus epistasis detection strategy using graphical processing units. Eur J Hum Genet 2011; 19: 465–71. doi: 10.1038/ejhg.2010.196 PMID: 21150885

75. Kam-Thong T, Azencott CA, Cayton L, Putz B, Altmann A, Karbalai N, et al. GLIDE: GPU-based linear regression for detection of epistasis. Hum Hered 2012; 73: 220–36. PMID: 22965145

76. Kharitonenkov A, Chen Z, Sures I, Wang H, Schilling J, Ullrich A. A family of proteins that inhibit signalling through tyrosine kinase receptors. Nature 1997; 386: 181–6. PMID: 9062191

77. Barclay AN, Brown MH. The SIRP family of receptors and immune regulation. Nat Rev Immunol 2006; 6: 457–64. PMID: 16691243

78. Hayashi A, Ohnishi H, Okazawa H, Nakazawa S, Ikeda H, Motegi S, et al. Positive regulation of phagocytosis by SIRPbeta and its signaling mechanism in macrophages. J Biol Chem 2004; 279: 29450–60. PMID: 15123631

79. Chang L, Karin M. Mammalian MAP kinase signalling cascades. Nature 2001; 410: 37–40. PMID: 11242034

80. Cobb MH. MAP kinase pathways. Prog Biophys Mol Biol 1999; 71: 479–500. PMID: 10354710

81. Orton RJ, Sturm OE, Vyshemirsky V, Calder M, Gilbert DR, Kolch W. Computational modelling of the receptor-tyrosine-kinase-activated MAPK pathway. Biochem J 2005; 392: 249–61. PMID: 16293107

82. Widmann C, Gibson S, Jarpe MB, Johnson GL. Mitogen-activated protein kinase: conservation of a three-kinase module from yeast to human. Physiol Rev 1999; 79: 143–80. PMID: 9922370

83. Di Benedetto B, Kallnik M, Weisenhorn DM, Falls WA, Wurst W, Holter SM. Activation of ERK/MAPK in the lateral amygdala of the mouse is required for acquisition of a fear-potentiated startle response. Neuropsychopharmacology 2009; 34: 356–66. doi: 10.1038/npp.2008.57 PMID: 18432190

84. Jurek B, Slattery DA, Maloumby R, Hillerer K, Koszinowski S, Neumann ID, et al. Differential contribution of hypothalamic MAPK activity to anxiety-like behaviour in virgin and lactating rats. PLoS One 2012; 7: e37060. doi: 10.1371/journal.pone.0037060 PMID: 22615888

85. Wefers B, Hitz C, Holter SM, Trumbach D, Hansen J, Weber P, et al. MAPK signaling determines anxiety in the juvenile mouse brain but depression-like behavior in adults. PLoS One 2012; 7: e35035. doi: 10.1371/journal.pone.0035035 PMID: 22529971

86. Thornalley PJ. Unease on the role of glyoxalase 1 in high-anxiety-related behaviour. Trends Mol Med 2006; 12: 195–9. PMID: 16616641

87. Hovatta I, Tennant RS, Helton R, Marr RA, Singer O, Redwine JM, et al. Glyoxalase 1 and glutathione reductase 1 regulate anxiety in mice. Nature 2005; 438: 662–6. PMID: 16244648

88. Distler MG, Plant LD, Sokoloff G, Hawk AJ, Aneas I, Wuenschell GE, et al. Glyoxalase 1 increases anxiety by reducing GABAA receptor agonist methylglyoxal. J Clin Invest 2012; 122: 2306–15. doi: 10.1172/JCI61319 PMID: 22585572

89. Hambsch B, Chen BG, Brenndorfer J, Meyer M, Avrabos C, Maccarrone G, et al. Methylglyoxal-mediated anxiolysis involves increased protein modification and elevated expression of glyoxalase 1 in the brain. J Neurochem 2010; 113: 1240–51. doi: 10.1111/j.1471-4159.2010.06693.x PMID: 20345757

90. Distler MG, Palmer AA. Role of Glyoxalase 1 (Glo1) and methylglyoxal (MG) in behavior: recent advances and mechanistic insights. Front Genet 2012; 3: 250. doi: 10.3389/fgene.2012.00250 PMID: 23181072

91. Touma C, Gassen NC, Herrmann L, Cheung-Flynn J, Bull DR, Ionescu IA, et al. FK506 binding protein 5 shapes stress responsiveness: modulation of neuroendocrine reactivity and coping behavior. Biol Psychiatry 2011; 70: 928–36. doi: 10.1016/j.biopsych.2011.07.023 PMID: 21907973

92. Kollmannsberger LK, Gassen NC, Bultmann A, Hartmann J, Weber P, Schmidt MV, et al. Increased glyoxalase-1 levels in Fkbp5 knockout mice caused by glyoxalase-1 gene duplication. G3 (Bethesda) 2013; 3: 1311–3. doi: 10.1534/g3.113.006445 PMID: 23733890