DOI: 10.1111/mec.16076

ORIGINAL ARTICLE

Check for updates

Natural copy number variation of tandemly repeated regulatory SNORD RNAs leads to individual phenotypic differences in mice

Maryam Keshavarz 💿 📔 Yoland Savriama 💿 📔 Peter Refki 📋 R. Guy Reeves 💿 📔

Max-Planck Institute for Evolutionary Biology, Plön, Germany

Diethard Tautz 💿

Correspondence

Diethard Tautz, Max-Planck Institute for Evolutionary Biology, Plön, Germany. Email: tautz@evolbio.mpg.de

Funding information Open access funding enabled and organized by projektDEAL

Abstract

Genic copy number differences can have phenotypic consequences, but so far this has not been studied in detail in natural populations. Here, we analysed the natural variation of two families of tandemly repeated regulatory small nucleolar RNAs (SNORD115 and SNORD116) in the house mouse (Mus musculus). They are encoded within the Prader-Willi Syndrome gene region, known to be involved in behavioural, metabolic, and osteogenic functions in mammals. We determined that the copy numbers of these SNORD RNAs show substantial natural variation, both in wild-derived mice as well as in an inbred mouse strain (C57BL/6J). We show that copy number differences are subject to change across generations, making them highly variable and resulting in individual differences. In transcriptome data from brain samples, we found SNORD copy-number correlated regulation of possible target genes, including *Htr2c*, a predicted target gene of SNORD115, as well as Ankrd11, a predicted target gene of SNORD116. Ankrd11 is a chromatin regulator, which has previously been implicated in regulating the development of the skull. Based on morphometric shape analysis of the skulls of individual mice of the inbred strain, we show that shape measures correlate with SNORD116 copy numbers in the respective individuals. Our results suggest that the variable dosage of regulatory RNAs can lead to phenotypic variation between individuals that would typically have been ascribed to environmentally induced variation, while it is actually encoded in individual differences of copy numbers of regulatory molecules.

KEYWORDS

copy number variation, digital PCR, geometric morphometrics, small nucleolar RNAs, tandem repeats

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited. © 2021 The Authors. *Molecular Ecology* published by John Wiley & Sons Ltd.

1 | INTRODUCTION

Phenotypic variation between individuals of a population is usually ascribed to a combination of genetic and environmental influences (Walsh & Lynch, 2018). The genetic component is identified through its heritability across generations, while the nonheritable component of the phenotype is ascribed to environmentally induced variation. The molecular assessment of genetic variation is usually based on SNPs that are not expected to mutate so fast that they would change between generations, that is, they conform with this heritability expectation. However, if a genetic locus influencing a phenotype mutates very fast, with changes in every generation, the distinction between genetically and environmentally determined variation blurs. Such fast mutations can be expected in tandemly repeated regions, where unequal crossover during meiosis or mitosis (Schlotterer & Tautz, 1994) can lead to new copy numbers in every generation.

Tandem repeat variation and its influence on phenotypes have mainly been studied in the context of copy number variation of ribosomal DNA genes (Bughio & Maggert, 2019), or copy number variation of short tandem repeats within coding regions of genes. The latter was initially discovered to cause human genetic diseases (Ashley & Warren, 1995; Sutherland & Richards, 1995) but has also been implicated in morphological evolution (Fondon & Garner, 2004) and genetic plasticity (Hannan, 2018). In addition to these short repeats, there are also several tandem repeat arrays that code for genes and that are copy number polymorphic in populations (Pezer et al., 2015). However, so far, little is known about the phenotypic consequences of copy number variation of such large arrays of gene coding repeats, although some have been implicated in the reinforcement of sexual isolation or transmission distortion (Didion et al., 2015; Moretti et al., 2020; North et al., 2020; Pezer et al., 2017)

We have previously identified the region between *Ube3a* and *Snrpn* in the Prader–Willi syndrome (PWS) gene region as one of two parentally imprinted regions that evolve particularly fast between

- MOLECULAR ECOLOGY - WILFY

natural populations of mice (Lorenc et al., 2014). In humans, deletions in the PWS region cause abnormalities in cognitive behaviours, metabolic syndromes, and craniofacial shape changes (Cassidy et al., 2012). The PWS region between *Ube3a* and *Snrpn* includes two small nucleolar RNA (SNORD) gene families, which are organized in large, mostly tandemly repeated clusters known as SNORD115 and SNORD116 (Cavaille, 2017) (Figure 1). The functional SNORD RNAs are processed from introns of a large transcript (*Snhg14*) that spans the whole region between *Snrpn* and *Ube3a* (Runte et al., 2001).

SNORDs are part of a large group of small, metabolically stable snoRNAs that regulate post-transcriptional modification of their target genes (Falaleeva et al., 2017; Kiss, 2001). One mechanism for snoRNA interaction is to bind via a complementary antisense region to a target RNA (Kishore et al., 2010). The SNORD115 antisense element exhibits complementarity to the alternatively spliced exon V of the serotonin receptor *Htr2c*, which is part of the serotonin regulatory pathway (Kishore & Stamm, 2006). Human SNORD116 shows a complementary sequence to an *Ankrd11* exon, which suggests a direct regulation (Bazeley et al., 2008). ANKRD11 is an ankyrin repeat domain-containing protein involved in craniofacial development (Barbaric et al., 2008) and acts as a transcriptional cofactor with multiple possible target genes (Gallagher et al., 2015).

Here, the copy number variation of the tandemly repeated SNORD115 and SNORD116 clusters in the two natural populations included in our original study (Lorenc et al., 2014) were studied. These populations were kept under outbreeding conditions in the laboratory (Harr et al., 2016). In addition, we also studied mice from the standard laboratory inbred strain C57BL/6J, expected to be isogenic. We sought to determine whether the copy number variation between individuals also leads to expression changes of their predicted target genes, implying that this should also lead to individual phenotypic differences. We found major differences in copy number in the two wild-type strains but surprisingly also in the inbred strain. Using transcriptome analyses and PCR assays, we show that the predicted regulation of *Htr2c* and *Ankrd11* can indeed be observed. Our data suggest further that new alleles with different copy numbers are



FIGURE 1 Structure of the PWS region around the SNORD clusters and copy numbers. The scheme depicts the genes as arrows pointing in the direction of transcription. Note that the transcript structures are more complex and not fully resolved (Cavaille, 2017; Lorenc et al., 2014). *Ipw* is, for example, a shorter version of the *Snhg14* transcript. The tandem clusters coding for the SNORD115 and SNORD116 snoRNAs are depicted by smaller block arrows (not drawn to size). The table below the graph provides the measures and ranges of copy numbers of the clusters in the different mouse stocks in the present study (data are provided in Table S1). The coefficients of determination (R^2) between the numbers in the two clusters are listed to the right

WILEY-MOLECULAR ECOLOGY

generated at every generation. Given the known effect of *Ankrd11* on skull development, we applied morphometric skull shape analysis to individuals from the inbred strain. We show that skull shape measures indeed correlate with SNORD116 copy number, implying that the corresponding change of *Ankrd11* leads to differences in phenotype among the otherwise genetically isogenic individuals. Therefore, our results suggest that the SNORD115/SNORD116 variation can cause individual phenotypic variation that would usually be ascribed to environmental variation, albeit derived from DNA-based variation. We discuss the implications of our findings for the evolution of a locus that maintains a phenotypic polymorphism in the population, possibly as part of a bet-hedging strategy effect.

2 | MATERIALS AND METHODS

2.1 | Mouse strains

Wild-type mice (*Mus musculus domesticus*) used in this study were offspring of mice that originated from wild populations sampled in 2005 and 2006 in the Massif Central region of France (MC) and the Cologne/ Bonn region of Germany (CB) and then held under outbreeding conditions that were started with 10 unrelated pairs, as described in detail in (Harr et al., 2016). Individuals of the C57BL/6J inbred strain were purchased at the age of three weeks from Charles River Laboratories, Research Models and Services, Germany GmbH (Sandhofer Weg 7, 97633 Sulzfeld) and further bred in our in-house facility.

All mice were bred in type III cages (Bioscape) and weaned at the age of 3–4 weeks. At weaning, the sexes were separated. Enrichment, including wood wool, toilet paper, egg cartons, and a spinning wheel (Plexx), was provided in each cage. Mice were fed a standard diet 1324 (Altromin) and provided water ad libitum. Housing was at 20–24°C, 50%–65% humidity, and on a 12:12 light-dark schedule with lights on at 7:00 AM.

Animals were kept according to FELASA (Federation of European Laboratory Animal Science Association) guidelines, with a permit from the Veterinäramt Kreis Plön: 1401-144/PLÖ-004697. Following the legal requirements for the use of animals in the context of the data presented in this paper, the animal welfare officer (Professor Schultheiß, University of Kiel) was informed of the animal sacrifice for molecular analyses.

2.2 | DNA and RNA extraction and cDNA synthesis

All dissections were done following standardized protocols. Prepared tissues were immediately frozen and kept at -70°C until DNA/RNA preparation. DNA extraction was performed according to a standard salt extraction protocol. Briefly, samples were lysed by using HOM buffer (80 mM EDTA, 100 mM Tris and 0.5% SDS) with Proteinase K (0.2 mg/ml) for 16 h in Thermomixer (Eppendorf) at 55°C. Then, 500 μ l sodium chloride (4.5 M) was added to each sample and incubated on ice for 10 min. Chloroform was then added, mixed, and

spun for 10 min at 10,000 rpm. The upper aqueous phase was separated, mixed well with isopropanol (0.7 volume), and spun for 10 min at 13,000 rpm. The pellet was washed with ethanol (70%), air-dried, and dissolved in TE buffer (10 mM Tris, 0.1 mM EDTA). DNA concentration was measured on a NanoDrop 3300 fluorospectrometer using Quant-iT dsDNA BR Assay kit (Invitrogen) reagent.

RNA extraction from whole brains was done by using Trizol reagent. Samples were lysed by tissue lyser II (Qiagen) at 30 Hertz for 5 min. Homogenized samples were incubated at room temperature for 5 min and 200 µl chloroform (per 1 ml TRIzol) was added to each 1 ml sample. This was then shaken vigorously by hand for 15 s, followed by 3 min incubation at room temperature, and spun at 12,000 g for 15 min at 4°C. The aqueous phase was transferred to a new tube, and 0.5 volumes of isopropanol added, incubated at room temperature for 10 min, and spun at 12,000 g at 4°C. The supernatant was removed, and the pellet was washed with 75% EtOH (made with RNAse-free water). Samples were mixed by hand several times and then spun at 7,500 g for 5 min at 4°C. The supernatant was removed and the pellet dried shortly at room temperature, dissolved in 200 µl RNAse free water, and stored at -20°C overnight. An equal volume of LiCL (5 M) was added to the crude RNA extract, mixed by hand, and incubated for 1 h at -20°C. Samples were spun at 16,000 g for 30 min. The supernatant was removed; samples were washed twice with EtOH 70% and spun at 10,000 at 4°C. The pellet was dried at room temperature, dissolved in RNAse free water, and kept for long-term storage at -70°C. The sample quality of the RNA was measured with Bio-Analyzer chips, and samples with RIN values below 7.5 were discarded. cDNA was synthesized using the MMLV High Performance Reverse Transcriptase kit according to the manufacturer's instructions (Epicenter).

2.3 Small RNA extraction and cDNA synthesis

Total RNA was extracted from the whole brain using the mirVana miRNA isolation kit, which enriches small RNAs. The RNA quality was measured with BioAnalyzer chips, and samples with RIN values below eight were discarded. The Illumina TruSeq Small RNA Library Prep kit was used for small RNA cDNA synthesis, which was then used for the ddPCR experiments (see below). The protocol in the Illumina kit takes advantage of the standard natural structure in most known small RNA molecules. Most mature small RNAs have a 5'-phosphate and a 3'-hydroxyl group. The Illumina adapters in this kit are directly ligated to the small RNAs. Reverse transcriptase with a primer for this adaptor was then used to synthesize cDNA from small RNA, and 1 μ g of total RNA, enriched in small RNA, was used as input.

2.4 | RNAseq analysis

Poly-A⁺ RNA was used for cDNA synthesis and Illumina library preparation using the Truseq stranded RNA HT kit. The libraries passing the standard quality control were subjected to sequencing on

-MOLECULAR ECOLOGY -WILEY

an Illumina NextSeq 500 sequencing system, and paired-end reads were 150 bases long. Raw sequence reads were quality trimmed using Trimmomatic (Bolger et al., 2014). The quality trimming was performed basewise, removing bases below a quality score of 20 (Q20) and keeping reads whose average quality was at least Q60. Reads were mapped to the mouse mm10 reference genome by using Hisat2 (Kim et al., 2015). Between 20.1 to 28.7 million mapped reads were obtained per sample (average 23.8 million). Differential expression analysis was performed with the DESeq2 package (Love et al., 2014) in the R environment. To perform the alternatively spliced isoforms analysis, BAM files from Hisat2 were used as input for SAMtools (Li et al., 2009) by using option –c for total read from each exon and option –q 60 for total read of each sample.

2.5 | Droplet digital PCR for copy number quantification

Digital PCR is a method which enables absolute quantification of DNA targets (either genomic DNA or cDNA synthesized from RNA preparations - see above) without the need to construct a calibration curve as used in qPCR (Zhao et al., 2016). It requires a reference to calculate gene copy numbers and to normalize gene expression levels. For the latter, we used β -actin for mRNAs and SNORD66 for standardizing SNORD RNAs. SNORD 66 is a single copy gene located in an intron of the eukaryotic translation initiation factor 4 and was used as the reference gene for copy number calculations. All reactions were done as technical triplicates, which yielded highly congruent results (Table S1). Averages of the triplicates were used to calculate copy numbers. Values for copy numbers represent the total number of copies per diploid genome. The values for expression represent the number of RNA copies per μ of input material.

Primers (Table S2) were designed aiming for a balanced GC content <50% and with a low potential for primer-dimer structure. For all primers, we verified via gel electrophoresis that they amplified only the expected length fragment.

The SNORD115/116 primers were placed within the coding sequences of the respective snoRNAs. Two pairs were tested for an initial set of 19 animals and yielded highly congruent results ($R^2 = .99$, Table S3). Hence, we used only the first primer set (labelled with -a in Table S2) for the remainder of the analyses.

The QX100 Droplet Digital PCR System (Bio-Rad) was used according to the manufacturer's instructions. Briefly, fluorescent PCR reactions for each sample were prepared in a 23 μ l volume containing 12 μ l 2 X EvaGreen supermixes, 200 nM of each forward and reverse primers, 1 ng DNA, and water. For tandem copy separation, sample viscosity reduction and improved template accessibility, DNA digestion was done by different restriction enzymes (Table S2, Thermo Fisher Scientific). Five units of restriction enzymes were added to each sample. The samples were kept for 20 min at room temperature for complete digestion.

Droplets were generated using a droplet generator (DG) with an 8-channel DG8 cartridge and cartridge holder with 70 μl of DG oil/

well, 20 µl of fluorescent PCR reaction mixture and a DG8 gasket. The prepared droplets were transferred to corresponding wells of a 96-well PCR plate (Eppendorf). The PCR plate was subsequently heat-sealed with pierceable foil using a PX1 PCR plate sealer (Bio-Rad) and then amplified in a LifeEco thermal cycler (Bioer). The thermocycling protocol was: initial denaturation at 95°C for 5 min, then 40 cycles of denaturation at 95°C for 30 s, annealing at 60°C for 45 s and, finally, incubation first at 4°C for 5 min and then at 90°C for 5 min. After cycling, the 96-well plate was fixed into a plate holder and placed into the droplet reader. Droplets of each sample were analysed sequentially, and fluorescent signals of each droplet were measured individually by a detector. Copy numbers were calculated according to the procedures suggested by BioRad using QUANTASOFT 1.7.

2.6 | Droplet digital PCR for expression comparison

We used the same general setup as for DNA quantification, with the following specific adjustments. As input, we used first strand cDNA from the small RNA extraction procedure described above. The cDNA was diluted to about 0.2 ng/µl RNA equivalents. For each fluorescent PCR reaction, 1 ng (5 µl of 0.2 ng/µl) of RNA-equivalent cDNA was used. The PCR reaction for each sample was prepared for the target transcript and the ß-actin reference in two separate tubes. After PCR amplification, each droplet provided a fluorescent positive or negative signal indicating the target RNA was present or not after partitioning. Each droplet provided an independent digital measurement. Positive and negative droplets were counted by the QX100 Droplet Digital PCR System and the concentration of target RNA as copies/µl was calculated using QuantaSoft. The ß-actin concentration was used to normalize between the samples. To control for DNA contamination, we included a tube with all ingredients apart from the reverse transcriptase in the initial cDNA synthesis step. The results obtained from this tube were used as gDNA contamination to correct the gene expression data for each sample.

2.7 | Craniofacial shape analysis

Landmarks for the morphometric analysis were obtained as previously described (Pallares et al., 2015). We used two batches of C57BL/6J mice from our breeding colony, for which we determined the SNORD116 copy numbers. All were male mice, 85% from different litters, and represented the spread of copy number variation (details in Table S9 and Table S10). Heads were scanned using a computer tomograph (micro-CT-vivaCT 40; Scanco) at a resolution of 21 cross-sections per millimetre. A total of 44 three-dimensional landmarks were positioned in the skull (the landmarks are described in Pallares et al., 2015). The raw 3D landmark coordinates were obtained with 3DSlicer (Fedorov et al., 2012) (http://www.slicer. org). We analysed the component of symmetric shape variation WILEY-MOLECULAR ECOLOGY

while discarding the asymmetry via the method of object symmetry (Mardia et al., 2000; Savriama & Klingenberg, 2011). A generalized Procrustes analysis (GPA) was used to extract shape from the landmark coordinates by removing extraneous effects of location, orientation, and position using functions gpagen and bilat.symmetry from the geomorph R package. A principal component analysis (PCA) was used on the Procrustes coordinates' covariance matrix to extract the PC scores later used as shape data in subsequent analyses. The multivariate regression of the shape data (Procrustes coordinates) on SNORD116 copy number was carried out via function procD. Im from the geomorph R package v3.0.7 (Adams & Otarola-Castillo, 2013).

To visualize the patterns of predicted shape differences for minimum and maximum values of SNORD116 copy numbers, the closest specimen to the overall mean shape in multidimensional space was first identified via the function FINDMEANSPEC in the geomorph package v3.0.7 (Adams & Otarola-Castillo, 2013) and a surface mesh extracted from a CT-scan of the specimen closest to the mean shape was warped to this overall mean shape using the Thin Plate Spline (TPS) method via the function tps3d in R-package MORPHO v2.6. Heatmap visualizations were made with the meshDist function from MORPHO (Schlager, 2017) and the colour palette from R-package VIRIDIS v0.5.1 (https://CRAN.R-project.org/package=viridis).

2.8 | Assessment of linkage disequilibrium

We used PLINK v1.90 (Chang et al., 2015) to calculate r2 for autosomes using the command./plink --file [filename] --r2 yes-really --ld-window-r2 0 --ld-window-kb 500 --ld-window 300000 --maf 0.13 --not-chr 23 24 0 (or --chr 7). This was done for 24 individuals each of the current generation of MC and CB mice randomly drawn from the breeding colony and commercially genotyped with the 143,000 SNP Giga Mouse Universal Genotyping Array (GigaMUGA, Neogen Genomics). The total genotyping rate was 0.96, and only SNPs with a >0 bp distance between them were analysed. For plotting, all pairs of SNPs with a highly similar physical distance (within a 1 kb range) were binned into groups, followed by a calculation of mean r^2 for each bin.

3 | RESULTS

3.1 | Digital PCR to measure repeat numbers

To assess repeat number variation for the SNORD115 and SNORD116 genes, we were interested in developing a reliable measurement for copy number variation within the respective clusters. In the mouse reference genome, the repeat unit length for SNORD115 is 1.97 kb, including the region encoding the mature RNA with a length of 79 nt. For SNORD116, the repeat unit length is 2.54 kb, and the RNA has a length of 94 nt. Not all copies are annotated in the reference genome, but in the genome sequence itself, SNORD115 is represented with 143 copies and SNORD116 with 70 copies, with a 50 kb sequence gap (Figure S1).

Typing copy numbers in this range has long been a technical challenge, explaining why natural copy number variation in this region has received scant attention. The classic quantitative PCR approaches do not have enough power to distinguish, for example, 150 from 160 copies. Also, long-read sequencing procedures (e.g., PacBio or Oxford Nanopore) are limited in this region since they stretch across several hundred kb, beyond the standard limits of these technologies. However, droplet digital PCR (ddPCR) can potentially achieve the necessary resolution (Pinheiro et al., 2012; Quan et al., 2018). We have previously validated this technology by comparing predicted copy numbers based on sequencing read depth from whole mouse genome sequences with measured copy numbers based on ddPCR assays (Pezer et al., 2015). Figure S2 shows a very good correspondence between ddPCR results and read depth calculation for different CNV loci with high copy numbers. A possible confounding factor for ddPCR would be that primers may not bind equally in all repeat units due to polymorphisms in the primer binding sequence. To reduce the chance of polymorphism in the primer binding sites, we placed them inside the regions encoding the mature SNORD RNAs. However, a survey of all described SNORD variants suggested that some polymorphism also exists in these regions. Hence, we tested two different sets of primers and found that both yielded highly congruent results (see Methods). Therefore we conducted the bulk of the experiments with only one set of primers.

To control the technical replicability, we ran the ddPCR reactions for the individuals in triplicates. These showed a very high correspondence, with average standard deviations at least 10-fold lower than the natural variation (Table S1). Hence, ddPCR is a method that has shown sufficient power and reproducibility to quantify the copy numbers in tandem repeat regions, which is the key to the results described below (see also discussion).

3.2 | SNORD copy number variation

Using the ddPCR assay, we tested different mouse strains for copy number variation. First, we focused on the variability in two M. m. domesticus outbred stocks, where the mice were initially caught in the wild and then kept under outbreeding conditions (stocks MC and CB). The two populations from which these are derived are separated for about 3,000 years. We have previously shown that these populations are genetically highly differentiated from each other (Harr et al., 2016; Linnenbrink et al., 2018, 2020), especially for the imprinted region studied here (Ube3a-Snrpn) (Lorenc et al., 2014). We used genome-wide SNP analysis to confirm that no inbreeding effects had occurred since these populations were introduced into the breeding facility, and found that they still mainly represent the wild-type LD patterns described in Staubach et al. (2012) (Figure S3). We found an average of 202 SNORD115 copies in CB animals and 168 in MC. A similar difference exists for SNORD116 copies, with 229 in CB and 188 in MC (Figure 1). Although there is a considerable

Intriguingly, we also observed a strong covariation of copy number between the two SNORD clusters, with highly significant correlation coefficients (Figure 1). Given that the level of linkage disequilibrium expected at a distance of 100 kb or larger is extremely low (Figure S3), and the clusters are about 300 kb apart from each other, this effect is very unlikely to be the result of simple genetic linkage, especially given the known enhanced meiotic recombination rate in imprinted regions (Paigen & Petkov, 2010). Furthermore, if unequal crossover causes the copy number variation at each locus, their copy numbers should evolve independently since their repeat units share no sequence similarities. The mechanism that maintains this covariation is presently unclear.

To test whether a possibly unknown technical factor led to this correlation, we also typed the two other CNV loci in the mouse genome that showed exceptionally high copy-number ranges (Sfi1 and Cwc22 - see Figure S2) for the same set of individuals. Sfi1 had between 18-98 copies in the CB population and 18-121 copies in the MC population; Cwc22 showed high copy number variation in the CB population only (18-99 copies) (Pezer et al., 2015). We found that neither of these two loci correlated with SNORD115 or SNORD116 copy numbers (Table S4). As a further control for the validity of the correlation between SNORD115 and SNORD116 copy numbers, we compared read coverage from the whole genome sequence data set of Mus musculus populations included in Harr et al. (Harr et al., 2016). We also found a high correlation (r = 0.64) between read coverage for SNORD115 and SNORD116 across all sequenced samples (Figure S4) as well as RNA expression (Figure S5), supporting the observations from the ddPCR data.

3.3 | Copy numbers and RNA expression

Given the proven efficiency of ddPCR for copy-number determination, we also used it to quantify RNA expression (Quan et al., 2018) in individuals from the wild-type stocks. The measured SNORD gene copy numbers showed a highly significant regression with the respective SNORD RNA expression in the whole brain (Figure 2a,c), suggesting a direct relationship between copy number and expression level.

Further, we asked whether the predicted target RNA expressions correlate with their respective SNORD expression. The target of SNORD115 is the alternatively spliced exon V of *Htr2cr* (Kishore & Stamm, 2006), which is split into two halves through the alternative splice – exonVa and exonVb (Figure 3a). We used primers that specifically span the exonVb to exonVI junction, and found a significant -MOLECULAR ECOLOGY -WILEY

positive correlation for this variant with SNORD115 copy numbers in both wild-type populations (Figure 2c). The predicted target of SNORD116 is exon IX of *Ankrd11* (Bazeley et al., 2008) (Figure 3b). Since no alternative splice has so far been described for this, we placed the primers close to the predicted SNORD116 binding site, and found a significant correlation with SNORD116 copy numbers (Figure 2d). Note that these patterns are confirmed and extended by the RNASeg analysis described below.

3.4 | Inheritance pattern

Given the high variability of SNORD gene copy numbers in the mouse populations, one can ask whether they are subject to frequent changes, such as unequal crossover events during meiosis. Given the total length of the repeat regions, it is difficult to test this directly since the haplotypes cannot be distinguished. However, we approached this question by comparing the inheritance of copy numbers in nine families of the MC population. We found that the offspring showed a considerable variation, exceeding the spread of numbers measured for the parents in seven out of the nine families (Figure 4). Given that the individual haplotypes cannot be resolved, an exact prediction of the expected copy numbers of the offspring is not possible. However, if the alleles of the offspring were only combinations of the parental alleles, one should expect at most four allele classes in the offspring. However, two families with more than four offspring (families 1 and 7) showed additional allele classes (even when one takes the standard deviation of measurement errors into account: Table S6), suggesting that new length alleles were generated in a single generation.

3.5 | Target gene analysis from transcriptome data

To investigate whole transcriptome changes in response to copy number variation of the SNORD genes, we chose 10 C57BL/6J individuals, representing a spread of copy numbers from low to high and sequenced RNA from their brains. First, we sought to confirm that the target genes that we had tested in the wild-type background by ddPCR (see above) would show the same patterns in the RNASeq data from the inbred strain.

The first analysis focused on *Htr2c* as the target of SNORD115. The expected alternative splice was in exon V, which results in a 95nt long exon part (exon Vb) specific for one splice variant, while the rest of the mRNA is shared between both (Figure 3). Therefore, we determined RNASeq read coverage for the exon Vb part and exon VI and compared each to SNORD115 copy numbers from the respective animals. We found a significant positive correlation for exon Vb, but none for exon VI ($R^2 = .43$ vs. -.12; p = .04 vs. .3) (Table S7). This confirms our results from the ddPCR experiments in the wild-type populations reported above. We also further analysed the edited sites in exon Vb in our RNASeq data and found the same previously described sites (Burns et al., 1997; Kishore & Stamm, 2006)



FIGURE 2 Regression between SNORD115 and SNORD116 copy numbers and expression for the outbred populations (N = 8 for each). (a) and (c) Regression with the respective SNORD RNA expression values. (b) and (d) Regression between the respective SNORD expression and the predicted target genes. All values were determined by ddPCR in whole-brain RNA preparations from eight outbred animals each, from the CB (dots) and MC (triangles) populations. The y-axis shows the RNA molecule copy number per μ l input (see Methods) after normalization. The animals constituted a subset of the animals used for the overall copy number variation reported in Figure 1 and were chosen to reflect the spread of the variation. The data for the graphs are provided in Table S5

(Figure 3), but we did not find the additional edited sites reported in the ectopic SNORD115 expression study (Raabe et al., 2019).

Computational analysis has predicted TAF1, RALGPS1, PBRM1, CRHR1, and DPM2 as five additional possible targets for SNORD115 (Kishore et al., 2010). However, our RNAseq data analysis showed no significant correlation between SNORD115 copy number with the expression of these five genes, only a tendency for a negative correlation for DPM2 ($R^2 = .32$, p = .09) (Table S7).

For Ankrd11, we correlated read coverage for the exon IX 5'part, including the PCR fragment used for the ddPCR study above, as well as the SNORD116 binding site (Figure 3). This showed a significant positive correlation to SNORD116 copy number ($R^2 = .5$; p = .02) (Table S7). We also analysed read coverage of the exons preceding exon IX to explore possible alternative splicing. Interestingly, exon VIII showed no significant correlation with SNORD116 copy number ($R^2 = .009$; p = .79), while exon VII showed a strongly negative one ($R^2 = -.53$; p = .01) (Table S7). This suggests that alternative splicing happens in this region, but the current data did not allow us to resolve this unequivocally. Also, in contrast to *Htr2c*, we found no evidence for RNA editing around the predicted SNORD116 binding site in *Ankrd11*. Hence, we conclude that the mechanism for regulation is different, but this is not unexpected given the diversity of regulatory interactions known for C/D-box snoRNAs (Falaleeva et al., 2017).

3.6 | Whole transcriptome analysis

Given that we tested a range of copy number variants rather than two states, it is impossible to perform a conventional analysis of significantly differentially expressed genes across the whole genome, where only two states are compared. Hence, we used correlation coefficients (*R*) as a basis for identifying possible target genes. Since SNORD116 appears to regulate *Ankrd11* and since ANKRD11 is a chromatin regulator that influences the expression of downstream genes by binding chromatin-modifying enzymes like histone deacetylases, we started with a list of previously identified target genes of ANKRD11 (Gallagher et al., 2015). We were able to match the data for 635 predicted target genes from the (Gallagher et al., 2015) study with our data from the whole transcriptome analysis. We used these to correlate the expression data of these genes from our whole transcriptome analysis to both the *Ankrd11*-exon IX 5'-part



FIGURE 3 Transcript structures of *Htr2c* and *Ankrd11* genes with annotated sites. The structure details with exon arrangements are taken from the UCSC browser. Exon numbers are indicated in roman numerals. For *Htr2c*, exon V and VI are enlarged; for *Ankrd11*, exon IX is enlarged. SNORD binding sites and primer sites for the ddPCR experiment for determining copy numbers are marked with boxes. The read alignment of RNASeq reads to Exon V of *Htr2c* show the edited sites

ddPCR and the SNORD116 copy numbers. When applying a cutoff of R > 0.4 for both correlations, we found 72 positively correlated and 32 negatively correlated genes (Table S8).

We applied the same correlation analysis to all genes in the transcriptome in a second step, using a correlation with either SNORD115 or SNORD116 copy numbers. We found an additional 67 candidate genes for SNORD115 and 89 for SNORD116 (Table S8). This suggests that the transcriptome effects of the two SNORD genes go beyond the known target pathways, although this could, of course, be a secondary effect. Note that the *Ankrd11* level correlates as a whole with the SNORD116 copy number, as predicted from the ddPCR on the *Ankrd11*-exon IX 5`-part results above, that is, the exon which represents most of the full-length RNA (Figure 3). In contrast, the level of the full *Htr2c* transcript does not correlate with

the SNORD115 copy number since the alternative splice changes are only a minor part of it.

3.7 | SNORD116 copy number and craniofacial features

A previous study (Barbaric et al., 2008) found that a dominant ENUinduced mutation had craniofacial effects, including a wider skull and parietal bone versus skull length effect. This phenotype was traced to a missense mutation in *Ankrd*11-exon IX, the exon affected by the SNORD116 levels. Hence, we tested whether SNORD116 copy number variation would also correlate with craniofacial features in the C57BL/6J mouse strain. We used a 3D landmarking approach



FIGURE 4 Inheritance patterns of SNORD gene copy numbers in families. Nine families of the MC outbred stock were tested, and the copy numbers were determined for each individual (left for SNORD115 and right for SNORD116). The parental copy numbers are marked with a star. The complete data are provided in Table S6

to quantify differences in skull shapes based on CT scans from 33 animals, following the procedures described in Pallares et al. (2015) and the landmarks depicted in Figure 5a. Multivariate regression of the shape data (Procrustes coordinates) on SNORD116 copy number indicated a strong relationship (percentage predicted = 7.97%; p = .0002; 10,000 rounds of permutation - Table S10; Figure 5b). We also tested age and centroid size as alternative variables and found no dependence on age but a weak dependence on centroid size (Table S10). When corrected for size, the dependence on SNORD116 copy numbers remained highly significant (percentage predicted = 6.52%; p = .0013; Table S10).

The variation in SNORD116 copy number is associated with skull shape changes predominantly localized at the back of the skull (i.e., braincase, squamosal, and occipital), with a low copy number being linked to more slender and straight skulls while a high copy number is related to more rounded and bent skulls with an enlarged braincase (Figure 5c,d; video S1). This phenotype is directly comparable to the phenotypic effects described by Barbaric et al. (2008), especially for the parietal bone length and skull width effects.

4 | DISCUSSION

Here, we showed a substantial variation in copy number for the SNORD115 and SNORD116 tandemly repeated genes, both in wildtype mouse populations as well as in an inbred mouse strain. This variation correlates with the expression of predicted target genes, of which we validate one that was so far only computationally predicted. The gene, *Ankrd11*, was previously found to be involved in craniofacial shape differences, and a similar phenotypic effect correlated to the copy number of SNORD116, which itself correlates with *Ankrd11* expression is reported here. Our findings thus suggest that copy number variation of a regulatory RNA can lead to individual phenotypic differences.

4.1 | Origin of variation

The high level of copy number variation in the SNORD115 and SNORD116 clusters can be ascribed to their tandem repeat organization, facilitating unequal crossover and thus length variation of the cluster. Both the family study in the wild-type mice and the observation of many alleles in the otherwise isogenic strain C57BL/6J suggest an exceptionally high rate of generation of new repeat number variants. C57BL/6J mice were initially generated from a single pair of mice in 1921 and kept as a well-identifiable strain with currently minimal SNP variation (Zurita et al., 2011). Since 2003 they have been managed as part of the genetic stability programme (GSP) developed by the Jackson Laboratory, whereby such strains are maintained as a cryopreserved colony nucleus derived from a single pair (Taft et al., 2006). New mice are periodically generated from the colony nucleus, and new families are started with these. While this ensures a much better strain homogeneity, animals supplied to endusers can still have a distance of about 10 generations from each other (Chebib et al., 2021).

We found at least 25 copy number variants for each SNORD cluster (when binned in distances of five) among the 57 animals that were initially provided by a single supplier operating under the GSP protocol (Charles River Laboratories) and then bred for another few generations in our facility. While the individuals in our sample are expected to be no more than maximally 15 generations away from each other, there could still have been up to four alleles in their founder pair, each of which could have given rise to new alleles. Given that we found at least 25 alleles among them (and



WILFY-MOLECULAR ECOLOGY

FIGURE 5 Morphometric analysis of mouse (C57BL/6J) crania in dependence to SNORD116 copy number variation. (a) Depiction of the landmarks used in the study. (b) Shape scores obtained from the multivariate regression plotted against SNORD116 copy number (data in Table S10). (c) Shape change associated with a low SNORD116 copy number (top) and high copy number (bottom) visualized by warped 3D surfaces of mouse skulls representing the deviation of each predicted shape and the overall mean shape (warm colours indicate high deviation while cold colours stand for low deviation measured in the Procrustes distance). (d) Same as (b), but differences between predicted shape change for a low and high SNORD116 copy number. See S1 video for an animation of shape changes across the range of copy numbers

this can be considered a conservative estimate), this suggests at minimum the generation of one new variant every few generations. This is in line with our direct test in families and suggests that the copy number variation of the SNORD115 and SNORD116 loci contributes substantially to differences between individuals of a given population or breeding stock.

4.2 | Craniofacial trait mapping

The high mutation rate in the tandem repeat regions presents a problem for approaches that try to identify phenotypic trait loci via association mapping. These rely on the linkage of usually neutral markers to variants of loci that influence the trait. However, when the trait locus is subject to changes almost every generation, it would inherently become unmappable. We have previously used genome-wide association mapping to identify loci that affect skull shape, both in wild-type mice as well as laboratory strains, using different SNP sets, including some flanking the repeat region, but we found no association with the Ube3a/Snrpn region (Pallares et al., 2014, 2015). Interestingly, however, mapping in the Diversity Outbred panel that combines genome regions from different subspecies shows a major QTL including the Ube3a/Snrpn region (Katz et al., 2020). This QTL corresponds to an allele from a mouse strain of a different subspecies (PWD derived from *M. m.* musculus) that might have a general overall difference in SNORD copy numbers that override the individual effects. To confirm this, we compared overall genomic read coverage data from the M. m. musculus population from the Czech Republic (CZE in Harr et al., 2016) that was the source of PWD and found that SNORD115 and SNORD116 had similar numbers in contrast to the M. m. domesticus populations, where SNORD116 has higher average numbers than SNOD115. However, this still requires direct experimental confirmation.

4.3 | Target genes

Our transcript analysis data in response to SNORD115 copy number variation confirm the previously described regulation of the alternatively spliced exon Vb of the serotonin receptor *Htr2c* (Kishore & Stamm, 2006). However, it has recently been shown that alternative splicing of this RNA also occurs in the absence of SNORD115 (Hebras et al., 2020), implying that SNORD115 would only act as a modifier of splicing efficiency. The same may apply to the predicted interaction of SNORD116 with *Ankrd*11. Note that the exact bp

match between SNORD116 and Ankrd11 is only found in humans, while the corresponding region in mice showed three mismatches in the same region. On the other hand, it has been suggested that up to three mismatches between the antisense box and the target region could be tolerated (Kishore et al., 2010). C/D-box snoRNAS can function in many different ways, including regulating pre-mRNA alternative splicing and mRNA abundance, and they may be processed into shorter ncRNAs resembling miRNAs and piRNAs (Falaleeva et al., 2017). The exact mechanistic interaction between SNORD115 and SNORD116 with their target genes is at present unsolved and needs to be studied further.

4.4 | Evolutionary implications

The copy number of tandem repeats is generally known to mutate much faster than single nucleotide substitutions. The shortest tandem repeats in eukaryotic genomes are simple sequences or microsatellites (Ellegren, 2004; Tautz & Renz, 1984), and their copy number variation can be used to distinguish individuals and determine parentage (Ellegren, 2004; Tautz, 1989). However, most of the microsatellite variation is in intergenic regions and thought to be usually neutral, although it may affect the regulation of genes (Gemayel et al., 2010). On the other hand, short tandem repeats that occur within coding regions of proteins can influence the protein function and thus the phenotype, often identified as a disease effect in humans (Hannan, 2018), but it has also been implicated in skull shape differences between dog breeds (Fondon & Garner, 2004, 2007) and to some extent in other mammals, including primates (Pointer et al., 2012; Ritzman et al., 2017). However, these studies have not yet looked systematically into the phenotypic effects within populations, or even inbred strains.

Copy number variation of tandem repeats of whole genes is also expected to lead to phenotypic differences. For example, a tandemly repeated gene set with relevance for environmental adaptations are the amylase genes. Copy number variation of amylase genes has been suggested to correlate with diet preferences in mammals (Pajic et al., 2019; Perry et al., 2007). For the mouse populations studied here (Fra and Ger), we have recently shown that their amylase variants are subject to adaptive introgression, suggesting indeed a functional role (Linnenbrink et al., 2020).

The SNORD115 and SNORD116 copy number variation found in the current study goes beyond such single-gene effects since it involves regulatory RNAs that influence a whole network of target genes, especially via the chromatin regulator *Ankrd*11. In our transcriptome analysis we found that more than 70 previously

- MOLECULAR ECOLOGY – WI

identified target genes of Ankrd11 showed a correlated response to SNORD116 copy number variation. In humans, loss of function mutations in Ankrd11 have a very pleiotropic phenotype (the KBG syndrome, characterized by intellectual disability, autism spectrum disorder, and craniofacial abnormalities (Sirmaci et al., 2011)). This suggests that additional phenotypic traits may be influenced by the copy-number variation of SNORD115 and SNORD116. Ankrd11 is also known to be involved in dendrite differentiation in the cerebral cortex via the BDNF/TrkB signalling pathway (Ka & Kim, 2018), and a subunit of the GABA A receptor (Gabrg1) shows the strongest positive correlation with SNORD 116 copy number. Gabrg1 plays a crucial role in modulating brain circuits that regulate emotional responses to potentially threatening stimuli (Nuss, 2015). The SNORD115 target gene *Htr2c* is part of the serotonin regulatory pathway. Activation of the receptor by serotonin inhibits dopamine and norepinephrine release in some brain regions, which regulates mood, feeding, and motoneuron functions (Stamm et al., 2017). Therefore, it will be interesting also to assess behavioral traits for correlation with SNORD115 and SNORD116 copy numbers.

The arrangement of the two tandem repeat clusters of SNORD115 and SNORD116 between the Ube3a and Snrpn gene (see Figure 1) is conserved in mammals, although overall copy numbers vary to a large extent between them (Zhang et al., 2014). Hence, this locus is apparently evolutionarily maintained not only in its overall structure, but at the same time also in its propensity to vary between individuals. This could suggest that the generation of individual variation is actually a generic function of the locus. Individual variation is relevant for adaptation in unpredictably changing environments, a strategy called bet-hedging (Simons, 2011). The above discussed pleiotropic effects of the copy number variation would further support such a conclusion, especially when behavioral phenotypes are involved. In this context, it is of particular note that the clusters are part of an imprinted region thought to earmark genetic functions with a potential to create conflicts between sexes or mother-infant relationships (Barlow & Bartolomei, 2014). Indeed, the second imprinted gene that we had found to be highly differentiated between the Fra and Ger populations (Lorenc et al., 2014), Peg13, has also turned out to be part of a regulatory network that organizes the social brain (Keshavarz & Tautz, 2021). Hence, the primary functional role of the tandem repeat clusters of SNORD115 and SNORD116 genes may be in modulating the expression of a suite of genes that contribute to individual phenotypic variation, including morphology, as shown here, but possibly also behavioural traits.

ACKNOWLEDGEMENTS

We thank Chen Xie for his valuable advice with transcriptome analysis, Stefan Dennenmoser and Cemalettin Bekpen for their advice on SNORDs copy number analysis, Sven Künzel for sequencing, Kristian Ullrich for retrieving reads from the genome sequence data, Ellen McConnell, Nicole Thomsen, and Cornelia Burghardt for their great lab-related support, Elke Blohm-Sievers for her help in CT scanning, the entire mouse team at the MPI for Evolutionary Biology, in particular Christine Pfeifle, Susanne Holz and Camilo Medina for taking care of the mice in this study. The work was funded by institutional funds of the MPG to DT. Open Access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

AUTHOR CONTRIBUTIONS

Maryam Keshavarz and Diethard Tautz designed the research, Maryam Keshavarz, Peter Refki, R. Guy Reeves performed the research, Maryam Keshavarz, Yoland Savriama, R. Guy Reeves, Diethard Tautz analysed the data, and Maryam Keshavarz and Diethard Tautz wrote the manuscript.

OPEN RESEARCH BADGES

This article has earned an Open Data badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available at https://zenodo.org/record/47503 71#.YKIIhhJS9aR and https://doi.org/10.5281/zenodo.4750371.

DATA AVAILABILITY STATEMENT

RNA sequence reads have been deposited at the European Nucleotide Archive under ENA accession number PRJEB29473.

CT data are deposited at Zenodo and are available at https://zenodo.org/record/4750371#.YKIIhhJS9aR and https://doi.org/10.5281/ zenodo.4750371.

ORCID

Maryam Keshavarz D https://orcid.org/0000-0002-9029-4318 Yoland Savriama D https://orcid.org/0000-0002-9980-584X R. Guy Reeves D https://orcid.org/0000-0001-9454-6175 Diethard Tautz D https://orcid.org/0000-0002-0460-5344

REFERENCES

- Adams, D. C., & Otarola-Castillo, E. (2013). geomorph: an r package for the collection and analysis of geometric morphometric shape data. *Methods in Ecology and Evolution*, 4(4), 393–399. https://doi. org/10.1111/2041-210x.12035
- Ashley, C. T., & Warren, S. T. (1995). Trinucleotide repeat expansion and human disease. *Annual Review of Genetics*, *29*, 703–728. https://doi. org/10.1146/annurev.ge.29.120195.003415
- Barbaric, I., Perry, M. J., Dear, T. N., Da Costa, A. R., Salopek, D., Marusic, A., Hough, T., Sara Wells, A., Hunter, J., & Cheeseman, M. & Brown, S. D. M. (2008). An ENU-induced mutation in the Ankrd11 gene results in an osteopenia-like phenotype in the mouse mutant Yoda. *Physiological Genomics*, 32(3), 311–321. https://doi.org/10.1152/physiolgenomics.00116.2007
- Barlow, D. P., & Bartolomei, M. S. (2014). Genomic imprinting in mammals. Cold Spring Harbor Perspectives in Biology, 6, a018382. https:// doi.org/10.1101/cshperspect.a018382
- Bazeley, P. S., Shepelev, V., Talebizadeh, Z., Butler, M. G., Fedorova, L., Filatov, V., & Fedorov, A. (2008). snoTARGET shows that

WILFY-MOLECULAR ECOLOGY

human orphan snoRNA targets locate close to alternative splice junctions. *Gene*, 408(1–2), 172–179. https://doi.org/10.1016/j. gene.2007.10.037

- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114– 2120. https://doi.org/10.1093/bioinformatics/btu170
- Bughio, F., & Maggert, K. A. (2019). The peculiar genetics of the ribosomal DNA blurs the boundaries of transgenerational epigenetic inheritance. *Chromosome Research*, 27(1-2), 19-30. https://doi. org/10.1007/s10577-018-9591-2
- Burns, C. M., Chu, H., Rueter, S. M., Hutchinson, L. K., Canton, H., SandersBush, E., & Emeson, R. B. (1997). Regulation of serotonin-2C receptor G-protein coupling by RNA editing. *Nature*, 387(6630), 303–308. https://doi.org/10.1038/387303a0
- Cassidy, S. B., Schwartz, S., Miller, J. L., & Driscoll, D. J. (2012). Prader-Willi syndrome. *Genetics in Medicine*, 14(1), 10–26. https://doi. org/10.1038/gim.0b013e31822bead0
- Cavaille, J. (2017). Box C/D small nucleolar RNA genes and the Prader-Willi syndrome: A complex interplay. *Wiley Interdisciplinary Reviews-Rna*, 8(4), e1417. https://doi.org/10.1002/wrna.1417
- Chang, C. C., Chow, C. C., Tellier, L., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience*, 4, s13742–015–0047–8. https:// doi.org/10.1186/s13742-015-0047-8
- Chebib, J., Jackson, B. C., Lopez-Cortegano, E., Tautz, D., & Keightley, P. D. (2021). Inbred lab mice are not isogenic: Genetic variation within inbred strains used to infer the mutation rate per nucleotide site. *Heredity*, 126(1), 107–116. https://doi.org/10.1038/s41437-020-00361-1
- Didion, J. P., Morgan, A. P., Clayshulte, A.-F., Mcmullan, R. C., Yadgary, L., Petkov, P. M., Bell, T. A., Gatti, D. M., Crowley, J. J., Hua, K., Aylor, D. L., Bai, L., Calaway, M., Chesler, E. J., French, J. E., Geiger, T. R., Gooch, T. J., Garland, T., Harrill, A. H., ... Pardo-Manuel de Villena, F. (2015). A multi-megabase copy number gain causes maternal transmission ratio distortion on mouse chromosome 2. *Plos Genetics*, 11(2), e1004850. https://doi.org/10.1371/journ al.pgen.1004850
- Ellegren, H. (2004). Microsatellites: Simple sequences with complex evolution. Nature Reviews Genetics, 5(6), 435-445. https://doi. org/10.1038/nrg1348
- Falaleeva, M., Welden, J. R., Duncan, M. J., & Stamm, S. (2017). C/D-box snoRNAs form methylating and non-methylating ribonucleoprotein complexes: Old dogs show new tricks. *BioEssays*, *39*(6), 1600264. https://doi.org/10.1002/bies.201600264
- Fedorov, A., Beichel, R., Kalpathy-Cramer, J., Finet, J., Fillion-Robin, J.-C., Pujol, S., Bauer, C., Jennings, D., Fennessy, F., Sonka, M., Buatti, J., Aylward, S., Miller, J. V., Pieper, S., & Kikinis, R. (2012). 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magnetic Resonance Imaging*, 30(9), 1323–1341. https:// doi.org/10.1016/j.mri.2012.05.001
- Fondon, J. W., & Garner, H. R. (2004). Molecular origins of rapid and continuous morphological evolution. Proceedings of the National Academy of Sciences of the United States of America, 101(52), 18058– 18063. https://doi.org/10.1073/pnas.0408118101
- Fondon, J. W., & Garner, H. R. (2007). Detection of length-dependent effects of tandem repeat alleles by 3-D geometric decomposition of craniofacial variation. *Development Genes and Evolution*, 217(1), 79–85. https://doi.org/10.1007/s00427-006-0113-4
- Gallagher, D., Voronova, A., Zander, M. A., Cancino, G. I., Bramall, A., Krause, M. P., Abad, C., Tekin, M., Neilsen, P. M., Callen, D. F., Scherer, S. W., Keller, G. M., Kaplan, D. R., Walz, K., & Miller, F. D. (2015). Ankrd11 Is a chromatin regulator involved in autism that is essential for neural development. *Developmental Cell*, 32(1), 31–42. https://doi.org/10.1016/j.devcel.2014.11.031
- Gauthier, I., & Nuss, P. (2015). Anxiety disorders and GABA neurotransmission: A disturbance of modulation. *Neuropsychiatric Disease and Treatment*, 11, 165–175. https://doi.org/10.2147/ndt.s58841

- Gemayel, R., Vinces, M. D., Legendre, M., & Verstrepen, K. J. (2010). Variable tandem repeats accelerate evolution of coding and regulatory sequences. Annual Review of Genetics, 44(44), 445–477. https://doi.org/10.1146/annurev-genet-072610-155046
- Hannan, A. J. (2018). Tandem repeats mediating genetic plasticity in health and disease. *Nature Reviews Genetics*, 19(5), 286–298. https://doi.org/10.1038/nrg.2017.115
- Harr, B., Karakoc, E., Neme, R., Teschke, M., Pfeifle, C., Pezer, Ž., Babiker, H., Linnenbrink, M., Montero, I., Scavetta, R., Abai, M. R., Molins, M. P., Schlegel, M., Ulrich, R. G., Altmüller, J., Franitza, M., Büntge, A., Künzel, S., & Tautz, D. (2016). Genomic resources for wild populations of the house mouse, Mus musculus and its close relative Mus spretus. *Scientific Data*, *3*, 160075. https://doi.org/10.1038/ sdata.2016.75
- Hebras, J., Marty, V., Personnaz, J., Mercier, P., Krogh, N., Nielsen, H., Aguirrebengoa, M., Seitz, H., Pradere, J.-P., Guiard, B. P., & Cavaille, J. (2020). Reassessment of the involvement of Snord115 in the serotonin 2c receptor pathway in a genetically relevant mouse model. *Elife*, 9, e60862. https://doi.org/10.7554/eLife.60862
- Ka, M. H., & Kim, W. Y. (2018). ANKRD11 associated with intellectual disability and autism regulates dendrite differentiation via the BDNF/
 TrkB signaling pathway. Neurobiology of Disease, 111, 138–152. https://doi.org/10.1016/j.nbd.2017.12.008
- Katz, D. C., Aponte, J. D., Liu, W., Green, R. M., Mayeux, J. M., Pollard, K. M., Pomp, D., Munger, S. C., Murray, S. A., Roseman, C. C., Percival, C. J., Cheverud, J., Marcucio, R. S., & Hallgrímsson, B. (2020). Facial shape and allometry quantitative trait locus intervals in the Diversity Outbred mouse are enriched for known skeletal and facial development genes. *PLoS One*, *15*(6), e0233377. https://doi. org/10.1371/journal.pone.0233377
- Keshavarz, M., & Tautz, D. (2021). The imprinted IncRNA Peg13 regulates sexual preference and the sex-specific brain transcriptome in mice. Proceedings of the National Academy of Sciences of the United States of America, 118(10), e2022172118. https://doi.org/10.1073/ pnas.2022172118
- Kim, D., Landmead, B., & Salzberg, S. L. (2015). HISAT: A fast spliced aligner with low memory requirements. *Nature Methods*, 12(4), 357-360. https://doi.org/10.1038/nmeth.3317
- Kishore, S., Khanna, A., Zhang, Z., Hui, J., Balwierz, P. J., Stefan, M., Beach, C., Nicholls, R. D., Zavolan, M., & Stamm, S. (2010). The snoRNA MBII-52 (SNORD 115) is processed into smaller RNAs and regulates alternative splicing. *Human Molecular Genetics*, 19(7), 1153–1164. https://doi.org/10.1093/hmg/ddp585
- Kishore, S., & Stamm, S. (2006). The snoRNA HBII-52 regulates alternative splicing of the serotonin receptor 2C. *Science*, 311(5758), 230– 232. https://doi.org/10.1126/science.1118265
- Kiss, T. (2001). Small nucleolar RNA-guided post-transcriptional modification of cellular RNAs. *Embo Journal*, 20(14), 3617–3622. https:// doi.org/10.1093/emboj/20.14.3617
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., & Homer, N., Marth, G., Abecasis, G., Durbin, R., & Genome Project Data, P (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), 2078–2079. https://doi.org/10.1093/bioin formatics/btp352.
- Linnenbrink, M., Teschke, M., Montero, I., Vallier, M., & Tautz, D. (2018). Meta-populational demes constitute a reservoir for large MHC allele diversity in wild house mice (*Mus musculus*). *Frontiers in Zoology*, 15, 15. https://doi.org/10.1186/s12983-018-0266-9
- Linnenbrink, M., Ullrich, K. K., McConnell, E., & Tautz, D. (2020). The amylase gene cluster in house mice (*Mus musculus*) was subject to repeated introgression including the rescue of a pseudogene. *Bmc Evolutionary Biology*, 20(1), 56. https://doi.org/10.1186/s12862-020-01624-5
- Lorenc, A., Linnenbrink, M., Montero, I., Schilhabel, M. B., & Tautz, D. (2014). Genetic differentiation of hypothalamus parentally biased transcripts in populations of the house mouse implicate the

prader-willi syndrome imprinted region as a possible source of behavioral divergence. *Molecular Biology and Evolution*, 31(12), 3240– 3249. https://doi.org/10.1093/molbev/msu257

- Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology, 15(12), 550. https://doi.org/10.1186/s13059-014-0550-8
- Mardia, K. V., Bookstein, F. L., & Moreton, I. J. (2000). Statistical assessment of bilateral symmetry of shapes. *Biometrika*, 87(2), 285–300. https://doi.org/10.1093/biomet/87.2.285
- Moretti, C., Blanco, M., Ialy-Radio, C., Serrentino, M.-E., Gobé, C., Friedman, R., Battail, C., Leduc, M., Ward, M. A., Vaiman, D., Tores, F., & Cocquet, J. (2020). Battle of the sex chromosomes: Competition between X and Y chromosome-encoded proteins for partner interaction and chromatin occupancy drives multicopy gene expression and evolution in muroid rodents. *Molecular Biology* and Evolution, 37(12), 3453–3468. https://doi.org/10.1093/molbe v/msaa175
- North, H. L., Caminade, P., Severac, D., Belkhir, K., & Smadja, C. M. (2020). The role of copy-number variation in the reinforcement of sexual isolation between the two European subspecies of the house mouse. *Philosophical Transactions of the Royal Society B-Biological Sciences*, 375(1806), 20190540. https://doi.org/10.1098/ rstb.2019.0540
- Paigen, K., & Petkov, P. (2010). Mammalian recombination hot spots: Properties, control and evolution. *Nature Reviews Genetics*, 11(3), 221-233. https://doi.org/10.1038/nrg2712
- Pajic, P., Pavlidis, P., Dean, K., Neznanova, L., Romano, R.-A., Garneau, D., Daugherity, E., Globig, A., Ruhl, S., & Gokcumen, O. (2019). Independent amylase gene copy number bursts correlate with dietary preferences in mammals. *Elife*, *8*, e44628. https://doi. org/10.7554/eLife.44628
- Pallares, L. F., Carbonetto, P., Gopalakrishnan, S., Parker, C. C., Ackert-Bicknell, C. L., Palmer, A. A., & Tautz, D. (2015). Mapping of craniofacial traits in outbred mice identifies major developmental genes involved in shape determination. *Plos Genetics*, 11(11), e1005607. https://doi.org/10.1371/journal.pgen.1005607
- Pallares, L. F., Harr, B., Turner, L. M., & Tautz, D. (2014). Use of a natural hybrid zone for genomewide association mapping of craniofacial traits in the house mouse. *Molecular Ecology*, 23(23), 5756–5770. https://doi.org/10.1111/mec.12968
- Perry, G. H., Dominy, N. J., Claw, K. G., Lee, A. S., Fiegler, H., Redon, R., Werner, J., Villanea, F. A., Mountain, J. L., Misra, R., Carter, N. P., Lee, C., & Stone, A. C. (2007). Diet and the evolution of human amylase gene copy number variation. *Nature Genetics*, 39(10), 1256– 1260. https://doi.org/10.1038/ng2123
- Pezer, Z., Chung, A. G., Karn, R. C., & Laukaitis, C. M. (2017). Analysis of copy number variation in the Abp gene regions of two house mouse subspecies suggests divergence during the gene family expansions. *Genome Biology and Evolution*, 9(6), 1393–1405. https:// doi.org/10.1093/gbe/evx099
- Pezer, Z., Harr, B., Teschke, M., Babiker, H., & Tautz, D. (2015). Divergence patterns of genic copy number variation in natural populations of the house mouse (*Mus musculus domesticus*) reveal three conserved genes with major population-specific expansions. *Genome Research*, 25(8), 1114–1124. https://doi.org/10.1101/ gr.187187.114
- Pinheiro, L. B., Coleman, V. A., Hindson, C. M., Herrmann, J., Hindson, B. J., Bhat, S., & Emslie, K. R. (2012). Evaluation of a droplet digital polymerase chain reaction format for DNA Copy number quantification. *Analytical Chemistry*, 84(2), 1003–1011. https://doi. org/10.1021/ac202578x
- Pointer, M. A., Kamilar, J. M., Warmuth, V., Chester, S. G. B., Delsuc, F., Mundy, N. I., Asher, R. J., & Bradley, B. J. (2012). RUNX2 tandem repeats and the evolution of facial length in

placental mammals. Bmc Evolutionary Biology, 12, 103. https://doi.org/10.1186/1471-2148-12-103

- Quan, P. L., Sauzade, M., & Brouzes, E. (2018). dPCR: A technology review. Sensors, 18(4), 1271. https://doi.org/10.3390/s18041271
- Raabe, C. A., Voss, R., Kummerfeld, D.-M., Brosius, J., Galiveti, C. R., Wolters, A., Seggewiss, J., Huge, A., Skryabin, B. V., & Rozhdestvensky, T. S. (2019). Ectopic expression of Snord115 in choroid plexus interferes with editing but not splicing of 5-Ht2c receptor pre-mRNA in mice. *Scientific Reports*, *9*, 4300. https://doi. org/10.1038/s41598-019-39940-6
- Ritzman, T. B., Banovich, N., Buss, K. P., Guida, J., Rubel, M. A., Pinney, J., Khang, B., Ravosa, M. J., & Stone, A. C. (2017). Facing the facts: The Runx2 gene is associated with variation in facial morphology in primates. *Journal of Human Evolution*, 111, 139–151. https://doi. org/10.1016/j.jhevol.2017.06.014
- Runte, M., Huttenhofer, A., Gross, S., Kiefmann, M., Horsthemke, B., & Buiting, K. (2001). The IC-SNURF-SNRPN transcript serves as a host for multiple small nucleolar RNA species and as an antisense RNA for UBE3A. *Human Molecular Genetics*, 10(23), 2687–2700. https://doi.org/10.1093/hmg/10.23.2687
- Savriama, Y., & Klingenberg, C. P. (2011). Beyond bilateral symmetry: Geometric morphometric methods for any type of symmetry. *Bmc Evolutionary Biology*, 11, 280. https://doi. org/10.1186/1471-2148-11-280
- Schlager, S. (2017). Morpho and Rvcg Shape Analysis in R. In G. Zheng, S. Li, & G. Szekely (Eds.), *Statistical shape and deformation analysis* (pp. 217–256). Academic Press.
- Schlotterer, C., & Tautz, D. (1994). Chromosomal homogeneity of Drosophila ribosomal DNA arrays suggests intrachromosomal exchanges drive concerted evolution. *Current Biology*, 4(9), 777–783. https://doi.org/10.1016/s0960-9822(00)00175-5
- Simons, A. M. (2011). Modes of response to environmental change and the elusive empirical evidence for bet hedging. Proceedings of the Royal Society B-Biological Sciences, 278(1712), 1601–1609. https:// doi.org/10.1098/rspb.2011.0176
- Sirmaci, A., Spiliopoulos, M., Brancati, F., Powell, E., Duman, D., Abrams, A., Bademci, G., Agolini, E., Guo, S., Konuk, B., Kavaz, A., Blanton, S., Digilio, M. C., Dallapiccola, B., Young, J., Zuchner, S., & Tekin, M. (2011). Mutations in ANKRD11 cause KBG syndrome, characterized by intellectual disability, skeletal malformations, and macrodontia. American Journal of Human Genetics, 89(2), 289–294. https://doi.org/10.1016/j.ajhg.2011.06.007
- Stamm, S., Gruber, S. B., Rabchevsky, A. G., & Emeson, R. B. (2017). The activity of the serotonin receptor 2C is regulated by alternative splicing. *Human Genetics*, 136(9), 1079–1091. https://doi. org/10.1007/s00439-017-1826-3
- Staubach, F., Lorenc, A., Messer, P. W., Tang, K., Petrov, D. A., & Tautz, D. (2012). Genome patterns of selection and introgression of haplotypes in natural populations of the house mouse (*Mus musculus*). *Plos Genetics*, 8(8), e1002891. https://doi.org/10.1371/journ al.pgen.1002891
- Sutherland, G. R., & Richards, R. I. (1995). Simple tandem DNA repeats and human genetic disease. Proceedings of the National Academy of Sciences, 92(9), 3636–3641. https://doi.org/10.1073/pnas.92.9.3636
- Taft, R. A., Davisson, M., & Wiles, M. V. (2006). Know thy mouse. *Trends in Genetics*, 22(12), 649–653. https://doi.org/10.1016/j.tig.2006.09.010
- Tautz, D. (1989). Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Research, 17(16), 6463–6471. https://doi.org/10.1093/nar/17.16.6463
- Tautz, D., & Renz, M. (1984). Simple sequences are uibiquitous repetitive components of eukaryotic genomes. *Nucleic Acids Research*, 12(10), 4127–4138. https://doi.org/10.1093/nar/12.10.4127
- Walsh, B., & Lynch, M. (2018). Evolution and selection of quantitative traits. Oxford University Press.

4721

WILEY-<mark>MOLECULAR ECOLOGY</mark>

- Zhang, Y.-J., Yang, J.-H., Shi, Q.-S., Zheng, L.-L., Liu, J., Zhou, H., Zhang, H., & Qu, L.-H. (2014). Rapid birth-and-death evolution of imprinted snoRNAs in the prader-willi syndrome locus: Implications for neural development in euarchontoglires. *PLoS One*, 9(6), e0100329. https://doi.org/10.1371/journal.pone.0100329
- Zhao, Y., Xia, Q. Y., Yin, Y. P., & Wang, Z. K. (2016). Comparison of droplet digital PCR and quantitative PCR assays for quantitative detection of *Xanthomonas citri* subsp citri. *PLoS One*, 11(7), e0159004. https:// doi.org/10.1371/journal.pone.0159004
- Zurita, E., Chagoyen, M., Cantero, M., Alonso, R., González-Neira, A., López-Jiménez, A., López-Moreno, J. A., Landel, C. P., Benítez, J., Pazos, F., & Montoliu, L. (2011). Genetic polymorphisms among C57BL/6 mouse inbred strains. *Transgenic Research*, 20(3), 481– 489. https://doi.org/10.1007/s11248-010-9403-8

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Keshavarz, M., Savriama, Y., Refki, P., Reeves, R. G., & Tautz, D. (2021). Natural copy number variation of tandemly repeated regulatory SNORD RNAs leads to individual phenotypic differences in mice. *Molecular Ecology*, 30, 4708–4722. https://doi.org/10.1111/mec.16076