



## OPEN ACCESS

## EDITED BY

Giovanni Malerba,  
University of Verona, Italy

## REVIEWED BY

Alsamman M. Alsamman,  
International Center for Agriculture  
Research in the Dry Areas (ICARDA),  
Egypt  
Paolo Zambonelli,  
University of Bologna, Italy

## \*CORRESPONDENCE

Maciej Szydlowski,  
✉ maciej.szydlowski@up.poznan.pl

RECEIVED 06 June 2023

ACCEPTED 11 September 2023

PUBLISHED 20 September 2023

## CITATION

Antkowiak M and Szydlowski M (2023),  
Uncovering structural variants associated  
with body weight and obesity risk in  
labrador retrievers: a genome-  
wide study.  
*Front. Genet.* 14:1235821.  
doi: 10.3389/fgene.2023.1235821

## COPYRIGHT

© 2023 Antkowiak and Szydlowski. This is  
an open-access article distributed under  
the terms of the [Creative Commons  
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,  
distribution or reproduction in other  
forums is permitted, provided the original  
author(s) and the copyright owner(s) are  
credited and that the original publication  
in this journal is cited, in accordance with  
accepted academic practice. No use,  
distribution or reproduction is permitted  
which does not comply with these terms.

# Uncovering structural variants associated with body weight and obesity risk in labrador retrievers: a genome-wide study

Michał Antkowiak and Maciej Szydlowski\*

Department of Genetics and Animal Breeding, Poznań University of Life Sciences, Poznań, Poland

Although obesity in the domestic dog (*Canis lupus familiaris*) is known to decrease well-being and shorten lifespan, the genetic risk variants associated with canine obesity remain largely unknown. In our study, which focused on the obesity-prone Labrador Retriever breed, we conducted a genome-wide analysis to identify structural variants linked to body weight and obesity. Obesity status was based on a 5-point body condition score (BCS) and the obese dog group included all dogs with a BCS of 5, along with dogs with the highest body weight within the BCS 4 group. Data from whole-gene sequencing of fifty dogs, including 28 obese dogs, were bioinformatically analyzed to identify potential structural variants that varied in frequency between obese and healthy dogs. The seven most promising variants were further analyzed by droplet digital PCR in a group of 110 dogs, including 63 obese. Our statistical evidence suggests that common structural mutations in or near six genes, specifically *ALPL*, *KCTD8*, *SGSM1*, *SLC12A6*, *RYR3*, and *VPS26C*, may contribute to the variability observed in body weight and body condition scores among Labrador Retriever dogs. These findings emphasize the need for additional research to validate the associations and explore the specific functions of these genes in relation to canine obesity.

## KEYWORDS

labrador retriever, obesity, *ALPL*, *KCTD8*, *SGSM1*, *SLC12A6*, *RYR3*, *VPS26C*

## Introduction

Canine obesity is a significant problem because of its high prevalence and its association with unfavorable physiological conditions and diseases. Obesity is a common phenomenon in countries where dogs are kept as pets, being perhaps as common as human obesity. For example, in Dutch populations, overweight affects forty percent of dogs and obesity another twenty percent, whereas over forty percent of dogs in Beijing are affected by obesity (Courcier et al., 2010; Mao et al., 2013). Obesity is associated with obesity-related metabolic dysfunction (Tvarijonavičute et al., 2012), orthopaedic disease, breathing difficulties and urinary incontinence, which reduce wellbeing and shorten life expectancy.

As in humans, common canine obesity is polygenic with contribution from a large number of genes, most with small effect size, and a significant environmental influence. In dogs, the contribution of polygenes to the overall variability, traditionally measured by the heritability coefficient, is difficult to estimate due to population fragmentation (both between and within breeds), incomplete pedigree records, and the lack of systematic data collection. Various characteristics of obesity, however, show similarity across mammals, and reliable heritability estimates exist for humans, mice, and pigs. From this point of view, it is believed

that obesity is moderately inherited in dogs (Switonski and Mankowska, 2013). A study of Strandberg et al. (2023) presented a comprehensive investigation into the heritability of body weight - a trait influenced by both body size and body fat accumulation. The analysis across 19 breeds estimated the average heritability of body weight to be 51%, with a range of 35%–70%. Moreover, the fact that certain breeds—such as Labrador retriever, Golden retriever, Pug, and Beagle—are especially prone to obesity indicates that some polygenes are fixed or very common in these breeds. The observation that Labrador dogs with the lowest birth weights are more likely to become overweight in adulthood can be explained by a genetic correlation between the two traits and indirectly suggests the existence of significant genetic variability related to obesity (Mugnier et al., 2020).

In contrast to human obesity, we know relatively little about the effects of individual genes and markers of obesity (Wallis and Raffan, 2020). From studies on humans and model animals, it appears that the genes implicated in the monogenic type of obesity may also carry mutations that stand out from the polygenic background (*LEP*, *PCSK1*, *LEPR*, *POMC*, *MC4R*, *SIM1*, *NTRK2*, *BDNF*, *SH2B1*, *MRAP2*, *KSR2*, *ADCY3*). This applies primarily to genes in the melanocortin pathway. A few effects of individual canine genes have been identified, including *POMC*, *TNF*, *MC4R*, *GPR120* and *ADCY3* (Zeng et al., 2014; Miyabe et al., 2015; Mankowska et al., 2016; Raffan et al., 2016; Mankowska et al., 2017; Sypniewski and Szydlowski, 2023). The outcomes of genome-wide association studies (GWAS) pertaining to body mass—an attribute intricately correlated with body fat—imply in a roundabout manner that genetic variations within *ACSL4*, a gene responsible for regulating back fat thickness in pigs, might likewise play a role in the susceptibility to canine obesity (Plassais et al., 2017). In contrast to investigations concerning human obesity, association analyses conducted in canines exhibit diminished statistical power, and the outcomes are frequently lacking validation within separate populations (except *POMC*). Consequently, our capacity is limited to identifying solely those variants that yield substantial influence over obesity-related traits.

The modest headway in pinpointing short variants linked to canine obesity underscores the potential advantages of incorporating an exploration into structural variants (SVs) within obesity research, given their anticipated capacity for exerting more pronounced individual effects. This move is facilitated by progress in computer-aided identification of structural variants based on genome-wide sequencing (Ionita-Laza et al., 2009; Teo et al., 2012). Arguments for the role of structural variants in obesity traits can be found in the discussions on the relationship between copy number variation (CNV) in the *AMY* gene and process of domestication (Axelsson et al., 2013). The first clue that structural variants in genes from the *AMY* family are associated with obesity was found in humans, where some data show correlation between the copy number of the *AMY1* gene, coding for pancreatic amylase, and body mass in children (Mejía-Benítez et al., 2015; León-Mimila et al., 2018), as well as Type 2 diabetes (Liu et al., 2020).

In this study, we search the dog genome for structural variants that may affect body weight and the predisposition of Labrador retriever dogs to obesity. We analyze genome-wide sequencing data from fifty dogs, including obese and nonobese individuals, and use

**TABLE 1** Summary of the sample of 110 Labrador retriever dogs used in this study. Presented values are medians (and range). The records were collected from 2014 to 2017.

| Dogs             | N   | Age (month)   | Body weight (kg)  |
|------------------|-----|---------------|-------------------|
| All              | 110 | 48 (12–120)   | 36.0 (25.0–50.0)  |
| Males            | 46  | 60 (12–120)   | 38.0 (31.0–50.0)  |
| Females          | 64  | 48 (12–120)   | 34.0 (25.0–46.0)  |
| Neutered males   | 7   | 60 (36–96)    | 35.0 (32.0–44.0)  |
| Neutered females | 16  | 72 (36–120)   | 31.1 (26.0–44.0)  |
| BCS = 3          | 32  | 36 (12–108)   | 30.5 (25.0–43.0)  |
| BCS = 4          | 52  | 48 (12–120)   | 36.1 (30.0–48.1)  |
| BCS = 5          | 26  | 76.5 (24–108) | 41.75 (35.0–50.0) |
| Nonobese         | 47  | 36 (12–120)   | 32.0 (25.0–40.0)  |
| Obese            | 63  | 60 (12–120)   | 39.0 (32.4–50.0)  |

bioinformatic methods to select polymorphic sites that may be significant for obesity-related traits. We validate these sites using the dd-PCR technique in a sample of 110 dogs. We find statistical evidence of a link between six structural variants in or near six protein-coding genes and obesity-related traits. We then justify the need to verify these preliminary associations in a broader study.

## Materials and methods

### Samples and whole-genome sequencing data

We used data on 110 Labrador retriever dogs, collected at four veterinary clinics during routine examinations from 2014 to 2017. All the dog owners gave their consent for participation in the research. The following data were collected: sex, age, body weight, neuter status, and body condition score on a five-point scale. Pedigree data were not collected.

To facilitate statistical analysis, the population was categorized into two groups: nonobese dogs and obese dogs. The nonobese dog group comprised all dogs with a body condition score (BCS) of 3, as well as dogs with the lowest body weight within the BCS 4 group (<33 kg). Conversely, the obese dog group included all dogs with a BCS of 5, along with dogs with the highest body weight within the BCS 4 group (>33 kg). The mean age and body weight were 56.7 months (SD = 33.8) and 36.2 kg (SD = 5.4 kg), respectively. The sample is summarized in Table 1.

Of these 110 dogs, 28 obese and 22 nonobese dogs were subjected to whole-genome sequencing as part of our previous study (Szydlowski and Antkowiak, 2022). FASTQ files are publicly available in the European Nucleotide Archive (PRJEB47658). Detailed information on mapping sequence reads on a reference genome is given in our previous report. Briefly, raw FASTQ files for each sample were aligned with the canine reference genome CanFam3.1 using BWA software (Li and Durbin, 2010). The aligned reads in SAM format were coordinate-sorted and converted to BAM format using SAMTOOLS (version 0.1.18; Li

et al., 2009). In this study, we reused the BAM files to search for structural variants that are potentially associated with obesity.

## Bioinformatic detection of structural variants

In order to detect SVs in short paired-end sequencing data, we used two bioinformatic methods: DELLY (Rausch et al., 2012) and TARDIS (Soylev et al., 2019). DELLY has the ability to detect deletions, inversions, duplications, and interchromosomal translocations with its integrated analysis of split reads. TARDIS is a more recent method with the ability to detect more complex events, such as a tandem duplication.

## Principal components analysis

Top principal components were included as covariates in the association analysis to account for population stratification. Principal components analysis (PCA) was conducted using plink software (ver. 2.0) on a sample of 50 sequenced dogs. We utilized 6.5 million short variants with a minor allele frequency (MAF) greater than 0.05.

## Selecting variants for droplet digital PCR

Due to the separate analysis of each genome by the two algorithms, the predicted localizations of structural variants (SVs) may vary across samples. To identify chromosomal regions potentially associated with obesity, we utilized the CoNVaQ algorithm (Larsen et al., 2018). In CoNVaQ, a new region is initiated whenever an SV from any dog starts or ends. This tool enables the comparison of two sets of SV segments, searching for genomic regions where the occurrence of SVs is significantly associated with the phenotype. In our study, we compared 28 obese dogs with 22 nonobese dogs.

We considered regions with a CoNVaQ raw *p*-value below 0.05 located within or in close proximity (<10 kb apart) to one or more protein-coding genes. CoNVaQ utilizes a straightforward statistical model that incorporates Fisher's exact test, thereby excluding the consideration of other potential risk factors, such as sex, age, castration, and genotype at the *POMC* locus. The CoNVaQ output, however, includes the "Gain/Loss" status for each individual, facilitating the verification of associations in a statistical model that incorporates known risk factors.

The SV calls (Gain/Loss) obtained from CoNVaQ were used in the general linear model for body weight and the logistic regression model for obesity. These models also incorporated age (as a covariate), sex (male/female), neuter status (yes/no), the top three principal components, and the presence or absence of the *POMC* deletion (14-bp Del variant, as described by Raffan et al., 2016). With regard to the *POMC* gene, the homozygous deletion was observed in only two dogs, therefore the Del/Del dogs and heterozygotes were classified as a single group.

In order to narrow down the chromosome regions identified in the previous procedure and select regions for ddPCR assays, we compiled a

list of genes associated with BMI in humans. To do this, we utilized the GWAS Catalog ([www.ebi.ac.uk/gwas](http://www.ebi.ac.uk/gwas), accessed on 5 February 2022), which provided us with single-nucleotide polymorphisms (SNPs) reported in genome-wide association studies (GWAS) that demonstrated a significant association ( $p < 1.0 \times 10^{-9}$ ) with BMI, along with their corresponding mapped genes. This comprehensive process resulted in a list of 1966 unique gene names.

## Droplet digital PCR for CNV analysis

DNA was extracted from frozen blood samples with MasterPure DNA Purification Kit for Blood Version II (Epicenter). The DNA quality was tested using a Qubit 2.0 Fluorometer (Invitrogen). Primers and assays for droplet digital PCR (ddPCR) of the fragments were designed by the tool on the BioRad website and by the Primer3Plus program. The finished primers and assays were obtained from BioRad or the Institute of Biochemistry and Biophysics of the Polish Academy of Sciences. A T100TM thermal cycler, a QX200TM droplet generator, a PX1TM PCR plate sealer, and a QX200TM droplet reader were obtained from BioRad. The reaction PCR was prepared using ddPCR Supermix for Probes (BioRad). The restriction enzymes (*Hae*III, *Eco*RI, *Msp*I, *Xba*I) were added to the reaction mix to improve the distribution of DNA across the droplets.

## Association analysis

To further explore the association between structural mutations identified by ddPCR and obesity-related traits in the sample of 110 dogs, we employed the general linear model (Type III SS) with body weight and logistic regression model with obesity status as the outcomes. These models incorporated age (as a covariate), sex, neuter status, presence of the *POMC* deletion (presence of 14-bp Del variant), and copy number determined by ddPCR (as a covariate). It is important to note that since principal components were not available for the unsequenced dogs, the association analysis was not adjusted for population stratification. Regression analyzes were performed using R statistical software version 4.2.2.

We adopted a significance level of  $\alpha = 0.05$ , and the resulting *p*-values were left unadjusted. Regardless of whether multiple-test correction is implemented or not, a cohort consisting of 110 individuals inherently lacks the capacity to offer unequivocal statistical substantiation. Within a limited population, the correction of *p*-values for variants pre-selected with a relatively high *a priori* likelihood substantially diminishes the likelihood of identifying prospective associations.

## Results

### Bioinformatic detection of structural variants

Analysis of the paired short-read sequencing data allowed for initial identification of almost 400,000 structural variants within 50 dog genomes. It should be noted, however, that the percentage of

**TABLE 2** Summary of structural variants detected by bioinformatic analysis of paired short-read sequencing data. Ten percent of the largest variants from each method were excluded.

| Type               | DELLY |             |                  | TARDIS |            |                  |
|--------------------|-------|-------------|------------------|--------|------------|------------------|
|                    | N/dog | Size range  | Median size (bp) | N/dog  | Size range | Median size (bp) |
| Deletion           | 7,054 | 33bp–207 kb | 389              | 318    | 31bp–62 kb | 461              |
| Insertion          | 180   | 38bp–86bp   | 58               | 77     | 31bp–60 kb | 93               |
| Inversion          | 715   | 50bp–206 kb | 2,561            | 2      | 40bp–60 kb | 1,013            |
| Tandem duplication | —     | —           | —                | 102    | 33bp–59 kb | 3,205            |
| All                | 7,950 | 33bp–207 kb | 402              | 500    | 31bp–62 kb | 512              |

false results for short-read-based methods may be extremely high. To ameliorate this problem, we discarded the 10% longest variants from each method and excluded all variants shorter than 30bp. The two algorithms differed significantly in the number and size of the variants they found: DELLY identified an average of 7,950 variants per dog (with a median size of 402bp) while TARDIS detected only 500 variants per dog (with a median size of 512bp). A summary of the results is presented in Table 2. Because most false positive variants are represented by a single individual, such variants are unlikely to be statistically associated with obesity.

We proceeded to use the CoNVaQ method to identify chromosomal regions that exhibited differences in structural variation (SV) frequency between obese and nonobese dogs. A total of 1,291 regions demonstrated a raw  $p$  below 0.05 (ranging from  $7.7 \times 10^{-7}$  to  $4.7 \times 10^{-2}$ ), with 241 of these regions overlapping with one or two protein-coding genes (Supplementary Figure S1). The SV calls (gain/loss) obtained from CoNVaQ were utilized in regression models for body weight and obesity, which included age, sex, neuter status, the top three principal components, and *POMC* deletion. The score plot for the first two principal components is shown on Supplementary Figure S2. The graph does not show an unambiguous picture of population stratification. Principal components and the effect of deletion in the *POMC* gene were not statistically significant, but were included in the models. Following the exclusion of regions with a  $p$  exceeding 0.05, a total of 56 regions remained (raw  $p$  ranging from  $7.6 \times 10^{-5}$  to  $4.8 \times 10^{-2}$ ).

Among the 56 regions that were analyzed, only one region corresponded to the list of GWAS signals associated with human BMI. To determine which SV variants should be further examined, we took into account both the results of the initial association analysis (considering low  $p$ ) and the proximity of the variants to protein coding genes. After careful deliberation, we chose the following seven genes for additional molecular analysis: *ALPL*, *KCTD8*, *RYR3*, *SGM1*, *SLC12A6*, *VPS26C*, and *MME* (with corresponding  $p$  of  $2.8 \times 10^{-3}$ ,  $3.1 \times 10^{-3}$ ,  $4.8 \times 10^{-4}$ ,  $2.8 \times 10^{-3}$ ,  $4.8 \times 10^{-4}$ ,  $2.9 \times 10^{-3}$ , and  $1.4 \times 10^{-4}$ , respectively). It is worth noting that initially, four other genes (*RECQL4*, *RBPI*, *XRNI*, and *GTF2A16*) were also considered. However, designing assays for them proved to be impossible due to the presence of highly similar fragments in different parts of the reference genome. *RYR3* was selected due to its paralog, *RYR1*, which has been shown to influence the distribution of adipose tissue in gilts (Fujii et al., 1991; Sadkowski et al., 2015).

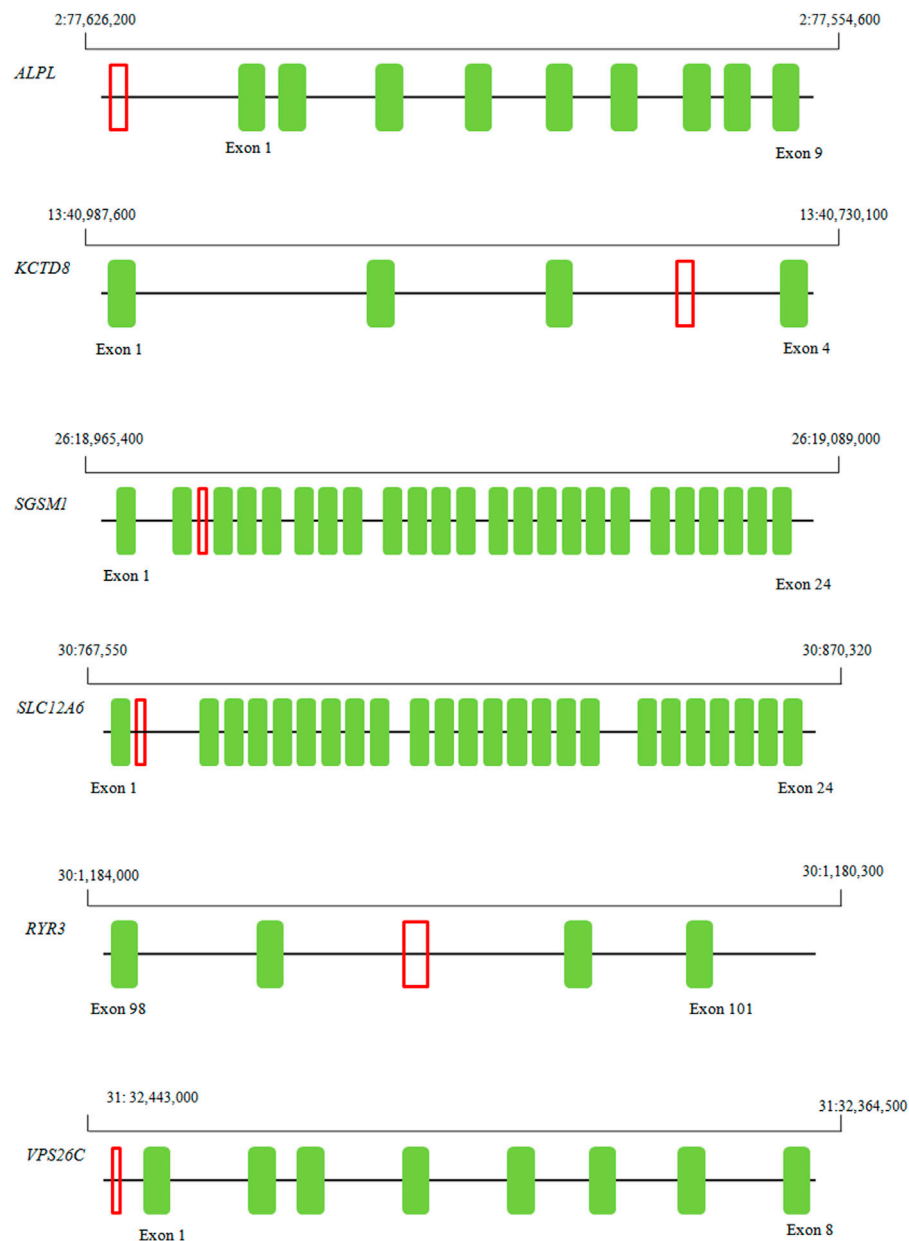
## Results of ddPCR and association study

The assays were successfully designed for seven fragments. We confirmed the presence of structural polymorphism in six genes: *ALPL*, *KCTD8*, *SGM1*, *SLC12A6*, *RYR3*, and *VPS26C*, though ddPCR showed no signal of structural variants for the *MME* gene. The chromosomal location of the six amplified fragments are presented on Figure 1.

We performed an association analysis between ddPCR results and obesity characteristics observed in the group of 110 Labrador. Again, the effect of deletions in the *POMC* gene was not statistically significant, but was included in the models. The structural variants observed in six genes exhibited associations with obesity, with four out of the six genes showing associations with both body weight and obesity (unadjusted  $p = 0.03$ , as shown in Table 3 and Figure 2; Figure 3; Figure 4). The effects of the four statistically significant mutations on body weight ranged from 34% to 54% SD. Notably, the results remain consistent for both traits, indicating that variants associated with increased body weight also elevate the risk of developing obesity.

The findings for the *RYR3* and *SLC12A6* genes are particularly intriguing, as these genes are closely located on chromosome 30, and their genotype distributions displayed remarkable similarity, suggesting a strong linkage disequilibrium. In the case of the *RYR3* gene, we amplified a 124-bp fragment located in intron 99 (CanFam3.1). The number of copies of this mutation ranged from zero to two, with a mean of 1.6 for dogs with a body condition score (BCS) of 3, and only 1.0 for those with a BCS of 5. This suggests that dogs carrying a deletion at the *RYR3* gene are more predisposed to excessive fat accumulation, while the risk of obesity decreases with each copy of the reference variant. Similarly, the deletion within the *SLC12A6* gene increases the risk of obesity. Collectively, the results pertaining to the *RYR3* and *SLC12A6* genes provide evidence of the potential importance of the chromosome 30 region in canine obesity, warranting further investigation.

The aforementioned associations were not adjusted for population stratification. In the preceding step, which encompassed fifty sequenced dogs, we computed the top twenty principal components. The eigenvalues for these components ranged from 2.6 to 1.0. The score plot for the first two principal components indicated genetic homogeneity within the population (Supplementary Figure S1). We believe that the procedure used here



**FIGURE 1**

Localization of six ddPCR-amplified fragments exhibiting overlap with structural variants that hold potential implications for body weight and obesity susceptibility in Labrador.

would reveal hidden stratification of the population, including that resulting from two types of Labrador (American and English).

## Discussion

The genetics of canine obesity is in its infancy compared to the genetics of human obesity, but for this reason, human genetics can offer a number of valid hypotheses, the verification of which in the canine population has recently brought significant progress. A number of candidate gene studies in relatively small cohorts have identified short mutations in the dog genome that shape traits associated with obesity in various breeds. A 14-bp intronic

deletion in the *POMC* gene greatly increases the body mass of Labrador retriever dogs, while altering the function of  $\beta$ -MSH and  $\beta$ -endorphin, two peptide hormones associated with food motivation and adiposity (Raffan et al., 2016; Mankowska et al., 2017). Moreover, two SNPs in the noncoding parts of *TNF* gene have been shown to be important for polygenic obesity in Labrador retriever dogs (Mankowska et al., 2016), whereas two exonic SNPs in the *MC4R* gene have been linked to body weight in Beagle dogs (Zeng et al., 2014). These findings indicate that the contribution of single genes to polygenic obesity may be greater in dogs than in humans, and suggests the possibility of identifying key markers that would be useful in risk prediction and treatment of obesity without the need for expensive genome-wide profiling.



**TABLE 3 Results of association study between body weight and obesity and structural variants in six protein-coding genes.**

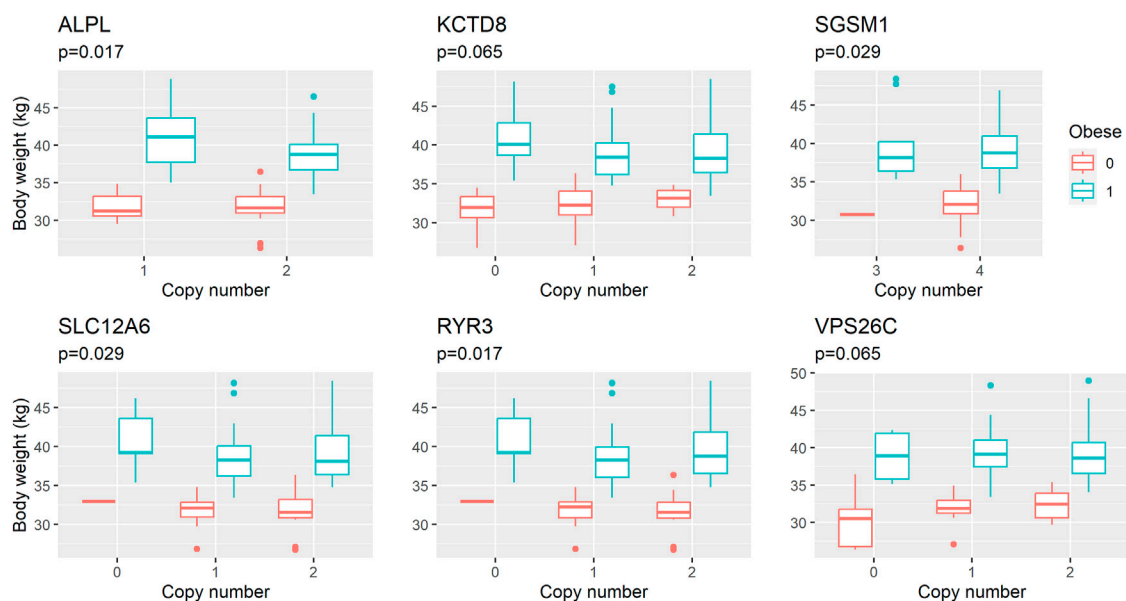
| Localization <sup>a</sup>  | Range of ddPCR output <sup>b</sup> | Association                                       |   |
|--|------------------------------------|---|---|
|  |                                    | Body weight (kg) <sup>c</sup> (per copy $\pm$ SE) | Obesity <sup>d</sup> (odds ratio)               |
| 2:77,622,662–77,622,741 deletion, intergenic region between <i>ALPL</i> and <i>ECE1</i>      | 0–2.5                              | $-2.88 \pm 1.17$ $p = \mathbf{0.017}$             | OR = 0.18 95%CI: 0.04–0.58 $p = \mathbf{0.008}$ |
| 13:40,904,429–40,904,512 deletion, intron 3 of <i>KCTD8</i>                                  | 0–2.1                              | $1.55 \pm 0.83$ $p = 0.065$                       | OR = 2.36 95%CI: 1.14–5.23 $p = \mathbf{0.025}$ |
| 26:18,996,505–18,996,597 duplication, intron 1 or 2 of <i>SGSM1</i>                          | 2.8–5.8                            | $-2.68 \pm 1.20$ $p = \mathbf{0.029}$             | OR = 0.16 95%CI: 0.03–0.58 $p = \mathbf{0.013}$ |
| 30:774,403–774,537 deletion/duplication, intron 1 of <i>SLC12A6</i>                          | 0–3.3                              | $-1.84 \pm 0.82$ $p = \mathbf{0.029}$             | OR = 0.41 95%CI: 0.18–0.84 $p = \mathbf{0.020}$ |
| 30:1,182,396–1,182,519 deletion, intron 99 of <i>RYR3</i>                                    | 0–2.1                              | $-2.17 \pm 0.88$ $p = \mathbf{0.017}$             | OR = 0.34 95%CI: 0.14–0.76 $p = \mathbf{0.012}$ |
| 31:32,435,989–32,436,088 deletion, intergenic region between <i>VPS26C</i> and <i>DYRK1A</i> | 0–2.2                              | $1.64 \pm 0.87$ $p = 0.065$                       | OR = 2.57 95%CI: 1.20–5.87 $p = \mathbf{0.019}$ |

<sup>a</sup>ddPCR, amplified fragment (CanFam 3.1). The true chromosomal location of the polymorphism is unknown, but it covers the amplified fragment.

<sup>b</sup>Values obtained from ddPCR, are real numbers and approximate the number of copies of the variant in the genome.

<sup>c</sup>Estimated effect in kg per copy and the corresponding standard error as estimated in general linear model.

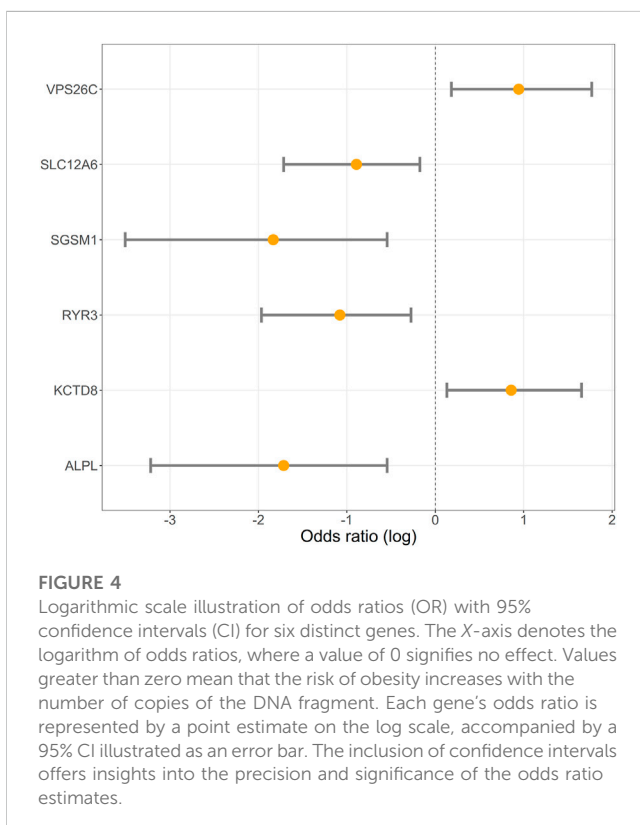
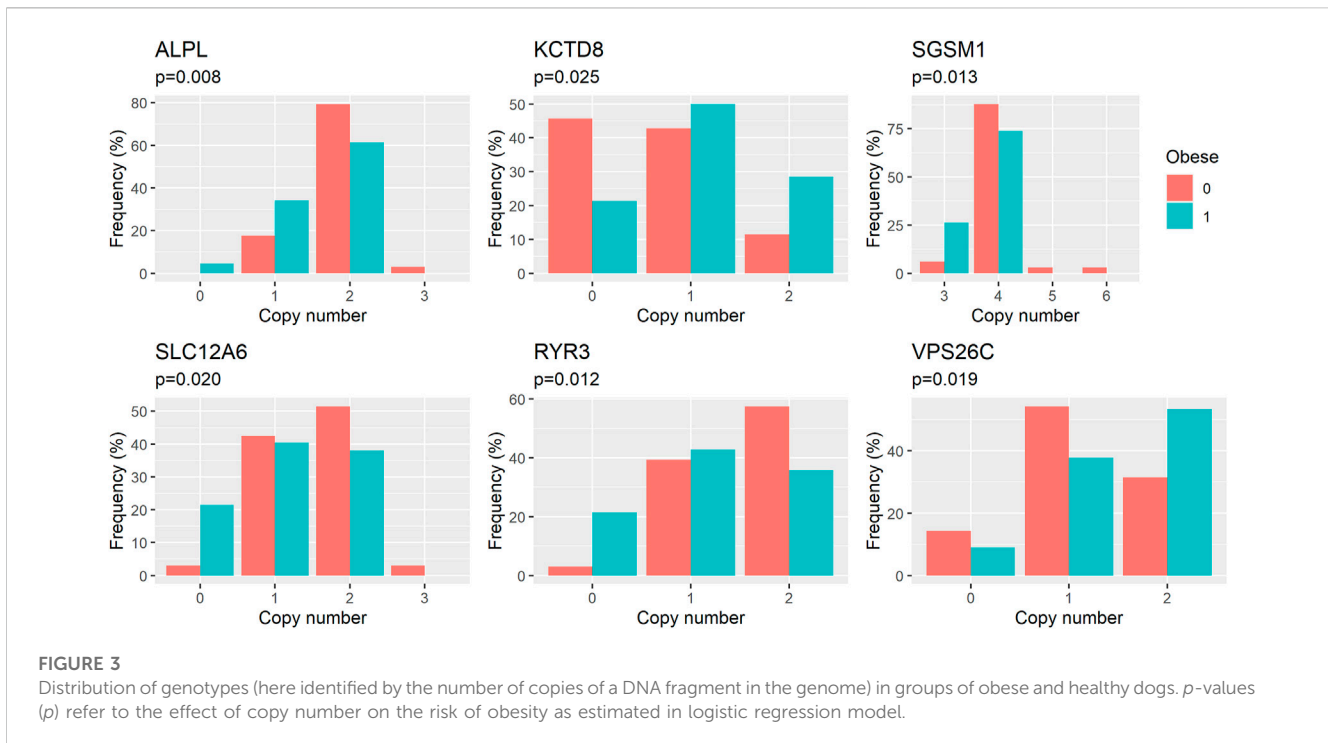
<sup>d</sup>Odds ratio (OR) and the corresponding 95% confidence interval (95%CI) as estimated in logistic regression for binary trait. Odds ratio = 1 indicates no difference between the groups. Values <0.05 are highlighted in bold.

**FIGURE 2**

Composite boxplot illustrating the relationship between body weight and copy number across six distinct genes. Each boxplot within the composition showcases the distribution of body weight values associated with varying copy number states for a specific gene. The median body weight within each gene's copy number category is indicated by the central line within the box, with the top and bottom edges of the box representing the third and first quartiles, respectively. Whiskers extend to 1.5 times the interquartile range (IQR) from the box boundaries, encompassing the majority of data points. Outliers, depicted as individual data points, are situated beyond the whiskers. The Y-axis represents body weight (grams), while the X-axis designates the six different genes and their respective copy number states. The Y-axis represents body weight (in kilograms), corrected by age (as a covariate), sex, neuter status and presence of the *POMC* deletion (presence of 14-bp Del variant). Owing to the limited size of certain groups, they are omitted from the plot for clarity.  $p$ -values ( $p$ ) refer to the effects of polymorphisms estimated in linear regression model.

In this study, our focus was on structural polymorphisms in the dog genome. Generally, structural polymorphisms are less prevalent than short variants, but their potential impacts can be anticipated to be relatively significant and easier to detect in

small populations. While copy number variations (CNVs) have not been as extensively investigated as short variants, there is compelling evidence linking certain CNVs to BMI in humans. For instance, such variants have been identified near the



*NEGR1*, *GPRC5B*, and *PPYR1* genes (Willer et al., 2009; Speliotes et al., 2010; Jarick et al., 2011). Additionally, the 1p21.1 multi-allele CNV encompasses *AMY1A*, which encodes salivary  $\alpha$ -amylase.

We did not find any evidence linking dog obesity traits to structural variants in or near genes that are considered crucial for obesity development, such as genes involved in the leptin-melanocortin signalling pathway and genes important for neuronal circuitry development (e.g., *BDNF*, *NTRK2*, *SEMA3A-G*). In human and animal models, these genes often harbour rare mutations that are responsible for severe and early-onset obesity. Furthermore, it is increasingly recognized that variants in or near these genes influence the risk of common polygenic obesity in humans. Despite the dominance of studies on these genes in the field of obesity genetics, it appears unlikely that these genes alone account for a significant portion of the genetic variation in obesity. In human studies, genome-wide association studies (GWAS) for obesity-related traits have identified thousands of loci, the majority of which have unknown underlying mechanisms. Our research, limited to common structural variants, suggests that these variants may be involved in mechanisms that have yet to be understood.

We obtained preliminary evidence suggesting that variants in six genes warrant further investigation. In our study, a structural deletion near the *ALPL* (Alkaline Phosphatase, Biomineralization Associated) gene was associated with reduced body weight and a higher risk of obesity. In humans, mutations in this gene have been linked to hypophosphatasia, an inherited condition that disrupts bone and teeth mineralization. Moreover, it is intriguing that this gene has also been associated with adaptive thermogenesis, which has garnered significant attention for its role in increasing systemic energy expenditure and combating obesity and diabetes. In mice, the targeted removal of *ALPL* specifically in adipocytes leads to decreased overall energy expenditure and the onset of obesity, without affecting movement or feeding behavior (Sun et al., 2021). These

findings highlight the potential importance of *ALPL* in regulating energy balance and adiposity.

We have obtained statistical evidence suggesting that a structural deletion in or near the *KCTD8* (Potassium Channel Tetramerization Domain Containing 8) gene could potentially increase body condition score. Additionally, it has been reported that an SNP in the Arabic population may be linked to diabetes (Musambil and Siddiqui, 2019). While clear associations between this gene and obesity traits are lacking, it is of interest due to its role in bitter and sweet taste signaling pathways. The gene is particularly highly expressed in the brain, but lower expression levels have also been observed in adipose tissue.

Our data also revealed that a duplication in the first intron of the *SGSM1* (Small G Protein Signaling Modulator 1) gene may lead to reduced body weight and a lower risk of obesity. The *SGSM1* gene is involved in GTPase activator activity and small GTPase binding activity, and it exhibits high expression levels in the brain. Interestingly, an intron variant in this gene has been associated with insulin measurement (Julienne et al., 2020).

We have demonstrated that deletions in or near the *SLC12A6* (Solute Carrier Family 12 Member 6) and *RYR3* (Ryanodine Receptor 3) genes are statistically associated with decreased body weight and a lower risk of obesity. These two genes are located in close proximity to each other on chromosome 30, and their effects on the two obesity traits are similar. In humans, mutations in the *SLC12A6* gene are generally associated with neuropathy (Bogdanova-Mihaylova et al., 2021), but there is no known connection between mutations in this gene and obesity-related traits, except for one genome-wide association study signal for BMI (Dong et al., 2018). Interestingly, the *SLC12A6* gene has been studied in both humans and dogs, revealing that truncating variants are responsible for movement disorders, although the phenotypes differ between species (Van Poucke et al., 2019).

The ryanodine receptor, encoded by the *RYR3* gene, plays a crucial role in releasing calcium from intracellular storage for various cellular processes, particularly in skeletal muscle contraction. It has been found that *RYR3* is under strong selection in sport-hunting genes (Kim et al., 2018). In humans, a single nucleotide polymorphism of this gene has been correlated with diabetes and hypertension, diseases associated with obesity (Gong et al., 2018). Genome-wide linkage analysis studies have also demonstrated a connection between polymorphism in the *RYR3* gene and the function of adiponectin in a Chinese population (Chang et al., 2016). In pigs, a short mutation in the *RYR1* gene, which is part of the intracellular calcium ion channel family like *RYR3*, is known to affect meat quality and fatness traits (Sadowski et al., 2015). Additionally, an intronic variant in the human *RYR3* gene has been associated with hippocampal volume (Alliey-Rodriguez et al., 2019). Studies on rats have shown that obesity is associated with abnormal neuronal differentiation in the hippocampus (Liškiewicz et al., 2021).

We also observed an association between a deletion near or in the *VPS26C* (VPS26 Endosomal Protein Sorting Factor C) gene and an increased risk of obesity. However, we are not aware of any studies that link this gene specifically to obesity.

In our studies, we did not observe any impact of the known deletion in the *POMC* gene on body weight or obesity. The reason behind our failure to detect the effect of the *POMC* gene may be

attributed to the rare occurrence of the deleterious variant in the homozygous genotype. It is plausible to assume that this deletion, which significantly disrupts gene function, follows a recessive pattern, thereby explaining our findings. We included the effect of *POMC* in all of our statistical models as a precautionary measure. In populations of limited size that lack substantial statistical power to detect significant sources of variation, incorporating a factor based solely on statistical significance can potentially lead to inaccuracies. Consequently, the inclusion of the *POMC* gene in the model was guided by existing knowledge that establishes its significance as a source of variation in Labrador.

Our association study is predicated upon a comparatively modest population, wherein a majority of individuals have not undergone sequencing. The extensive repertoire of association studies conducted within the human population has established that, owing to the vastness of the genome and the inherent role of chance, investigations involving small populations are particularly susceptible to yielding false positive outcomes. Although the mitigation of spurious associations is feasible through adjustments for both multiple testing and population stratification, these safeguards were not applicable in the context of our study. As alluded to earlier, a significant portion of individuals lacked complete genome screening, thus impeding a comprehensive accounting for potential population stratification. Given the constraints of the small sample size and the inherent limitations of a singular test, no provisions were made for multiple testing adjustments. Consequently, our findings should be regarded as moderately indicative statistical evidence that does not definitively resolve the matter of association but holds potential to inform and direct subsequent investigations.

Our research did not include a functional analysis, nor did we attempt to explain the mechanisms underlying the statistical associations described here. We acknowledge that elucidating such mechanisms will be challenging, particularly because the genes identified as important for Labrador obesity traits in our research have not been studied in the context of human obesity thus far. However, irrespective of the underlying mechanism, confirming the statistical relationship between the identified structural variants and obesity can serve as an indirect objective in the breeding of Labrador.

In conclusion, we have provided statistical evidence suggesting that common structural mutations in or near six genes, namely, *ALPL*, *KCTD8*, *SGSM1*, *SLC12A6*, *RYR3*, and *VPS26C*, may play a role in the variability of body weight and body condition score in Labrador retriever dogs. These findings emphasize the need for additional research to validate the associations.

## Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: European Nucleotide Archive (PRJEB47658).

## Ethics statement

The animal studies were approved by Bioethical Commission for Animal Care and Use in Poznań, Poland (31/2013). The studies were conducted in accordance with the local legislation and institutional



requirements. Written informed consent was obtained from the owners for the participation of their animals in this study.

## Author contributions

Formal analysis: MA; Investigation, MA; Writing—original draft, MA; Writing—review and editing, MS; Supervision, MS; Project administration, MS; Funding acquisition, MS. All authors contributed to the article and approved the submitted version.

## Funding

This work was supported by the National Science Centre, Poland (project no. 2016/23/B/NZ2/01762).

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## References

- Alliey-Rodriguez, N., Grey, T. A., Shafee, R., Asif, H., Lutz, O., Bolo, N. R., et al. (2019). NRXN1 is associated with enlargement of the temporal horns of the lateral ventricles in psychosis. *Transl. Psychiatry* 17, 230. doi:10.1038/s41398-019-0564-9
- Axelsson, E., Ratnakumar, A., Arendt, M., Maqbool, K., Webster, M. T., Perolski, M., et al. (2013). The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature* 495, 360–364. doi:10.1038/nature11837
- Bogdanova-Mihaylova, P., McNamara, P., Burton-Jones, S., and Murphy, S. M. (2021). Expanding the phenotype of SLC12A6-associated sensorimotor neuropathy. *BMJ Case Rep.* 27, e244641. doi:10.1136/bcr-2021-244641
- Chang, Y., Chiu, Y., He, Ch., Sheu, W. H., Lin, M., Seto, T. B., et al. (2016). Genome-wide linkage analysis and regional fine mapping identified variants in the RYR3 gene as a novel quantitative trait locus for circulating adiponectin in Chinese population. *Med. Baltim.* 95, e5174. doi:10.1097/MD.00000000000005174
- Courcier, E. A., Thomson, R. M., Mellor, D. J., and Yam, P. S. (2010). An epidemiological study of environmental factors associated with canine obesity. *J. Small Anim. Pract.* 51, 362–367. doi:10.1111/j.1748-5827.2010.00933.x
- Dong, J., Levine, D. M., Buas, M. F., Zhang, R., Onstad, L., Fitzgerald, R. C., et al. (2018). Interactions between genetic variants and environmental factors affect risk of esophageal Adenocarcinoma and Barrett's esophagus. *Clin. Gastroenterol. Hepatol.* 16, 1598–1606.e4. doi:10.1016/j.cgh.2018.03.007
- Fujii, J., Otsu, K., Zorzato, F., de Leon, S., Khanna, V. K., Weiler, J. E., et al. (1991). Identification of a mutation in porcine ryanodine receptor associated with malignant hyperthermia. *Science* 253 (5018), 448–451. doi:10.1126/science.1862346
- Gong, S., Bin Su, B., Tovar, H., Mao, Ch., Gonzalez, V., Liu, Y., et al. (2018). Polymorphisms within RYR3 gene are associated with risk and age at onset of hypertension, diabetes, and Alzheimer's disease. *Am. J. Hypertens.* 31, 818–826. doi:10.1093/ajh/hpy046
- Ionita-Laza, I., Rogers, A. J., Lange, C., Raby, B. A., and Lee, C. (2009). Genetic association analysis of copy-number variation (CNV) in human disease pathogenesis. *Genomics* 93, 22–26. doi:10.1016/j.ygeno.2008.08.012
- Jarick, I., Vogel, C. I., Scherag, S., Schäfer, H., Hebebrand, J., Hinney, A., et al. (2011). Novel common copy number variation for early onset extreme obesity on chromosome 11q11 identified by a genome-wide analysis. *Hum. Mol. Genet.* 20, 840–852. doi:10.1093/hmg/ddq518
- Julienne, H., Lechat, P., Guillemot, V., Lasry, C., Yao, C., Araud, R., et al. (2020). Jass: command line and web interface for the joint analysis of GWAS results. *Nar. Genom. Bioinform* 2, lqaa003. doi:10.1093/nargab/lqaa003
- Kim, J., Williams, F. J., Dreger, D. L., Plassais, J., Davis, B. W., Parker, H. G., et al. (2018). Genetic selection of athletic success in sport-hunting dogs. *Proc. Natl. Acad. Sci. U. S. A.* 115, E7212–E7221. doi:10.1073/pnas.1800455115
- Larsen, S. J., do Canto, L. M., Rogatto, S. R., and Baumbach, J. (2018). CoNVaQ: A web tool for copy number variation-based association studies. *BMC Genomics* 19, 369. doi:10.1186/s12864-018-4732-8
- León-Mimila, P., Villamil-Ramírez, H., López-Contreras, B., Morán-Ramos, S., Macías-Kauffer, L. R., Acuña-Alonzo, V., et al. (2018). Low salivary amylase gene (AMY1) copy number is associated with obesity and gut prevotella abundance in Mexican children and adults. *Nutrients* 10, 1607. doi:10.3390/nu10111607
- Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* 26, 589–595. doi:10.1093/bioinformatics/btp698
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi:10.1093/bioinformatics/btp352
- Liśkiewicz, A. D., Liśkiewicz, D., Marczak, Ł., Przybyła, M., Grabowska, K., Student, S., et al. (2021). Obesity-associated deterioration of the hippocampus is partially restored after weight loss. *Brain Behav. Immun.* 96, 212–226. doi:10.1016/j.bbi.2021.05.030
- Liu, Y., Smith, C. E., Parnell, L. D., Lee, Y., An, P., Straka, R. J., et al. (2020). Salivary AMY1 copy number variation modifies age-related type 2 diabetes risk. *Clin. Chem.* 66, 718–726. doi:10.1093/clinchem/hvaa072
- Mankowska, M., Stachowiak, M., Graczyk, A., Ciazynska, P., Gogulski, M., Nizanski, W., et al. (2016). Sequence analysis of three canine adipokine genes revealed an association between TNF polymorphisms and obesity in Labrador dogs. *Anim. Genet.* 47, 245–249. doi:10.1111/age.12390
- Mankowska, M., Krzeminska, P., Graczyk, M., and Switonski, M. (2017). Confirmation that a deletion in the POMC gene is associated with body weight of Labrador retriever dogs. *Res. Vet. Sci.* 112, 116–118. doi:10.1016/j.rvsc.2017.02.014
- Mao, J., Xia, Z., Chen, J., and Yu, J. (2013). Prevalence and risk factors for canine obesity surveyed in veterinary practices in Beijing, China. *Prev. Vet. Med.* 112, 438–442. doi:10.1016/j.prevetmed.2013.08.012
- Mejía-Benítez, M. A., Bonnefond, A., Yengo, L., Huyvaert, M., Dechaume, A., Peralta-Romero, J., et al. (2015). Beneficial effect of a high number of copies of salivary amylase AMY1 gene on obesity risk in Mexican children. *Diabetologia* 58, 290–294. doi:10.1007/s00125-014-3441-3
- Miyabe, M., Gin, A., Onozawa, E., Daimon, M., Yamada, H., Yamada, H., et al. (2015). Genetic variants of the unsaturated fatty acid receptor GPR120 relating to obesity in dogs. *J. Vet. Med. Sci.* 77, 1201–1206. doi:10.1292/jvms.15-0031
- Mugnier, A., Morin, A., Cellard, F., Devaux, L., Delmas, M., Adib-Lesaux, A., et al. (2020). Association between birth weight and risk of overweight at adulthood in Labrador dogs. *PLoS One* 15, e0243820. doi:10.1371/journal.pone.0243820

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2023.1235821/full#supplementary-material>

### SUPPLEMENTARY FIGURE S1

Manhattan plot for chromosomal regions potentially associated with obesity in Labrador retrievers. P-values are derived from the comparison of the frequency of structural variants between the groups of obese and healthy dogs using the CoNVaQ tool. Only 1291 regions with p-values > 0.05 are shown. Chromosome 39 represents the X chromosome.

### SUPPLEMENTARY FIGURE S2

The score plot for the first two principal components for the sample of 50 dogs.

- Musambil, M., and Siddiqui, K. (2019). Genetics and genomics studies in type 2 diabetes: A brief review of the current scenario in the arab region. *Diabetes Metab. Syndr.* 13, 1629–1632. doi:10.1016/j.dsx.2019.03.017
- Plassais, J., Rimbault, M., Williams, F. J., Davis, B. W., Schoenebeck, J. J., and Ostrander, E. A. (2017). Analysis of large versus small dogs reveals three genes on the canine X chromosome associated with body weight, muscling and back fat thickness. *PLoS Genet.* 13, e1006661. doi:10.1371/journal.pgen.1006661
- Raffan, E., Dennis, R. J., O'Donovan, C. J., Becker, J. M., Scott, R. A., Smith, S. P., et al. (2016). A deletion in the canine POMC gene is associated with weight and appetite in obesity-prone Labrador retriever dogs. *Cell. Metab.* 23, 893–900. doi:10.1016/j.cmet.2016.04.012
- Rausch, T., Zichner, T., Schlattl, A., Stutz, A. M., Benes, V., and Korbel, J. O. (2012). Delly: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* 28, i333–i339. doi:10.1093/bioinformatics/bts378
- Sadkowski, S., Molińska-Glura, M., Moliński, K., Szczepankiewicz, D., Świtoński, M., and Szydlowski, M. (2015). A well-known mutation in RYR1 alters distribution of adipose tissue in gilts. *Animal Sci. Pap. Rep.* 33, 147–154.
- Soylev, A., Le, T., Amini, H., Alkan, C., and Hormozdiari, F. (2019). Discovery of tandem and interspersed segmental duplications using high throughput sequencing. *Bioinformatics* 35, 3923–3930. doi:10.1093/bioinformatics/btz237
- Speliotes, E., Willer, C., Berndt, S., Monda, K. L., Thorleifsson, G., Jackson, A. U., et al. (2010). Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat. Genet.* 42, 937–948. doi:10.1038/ng.686
- Strandberg, E., Andersson, L., Emanuelson, U., Bjørnvad, C. R., Ringmark, S., Hedhammar, Å., et al. (2023). Heritability and genetic trend of body weight in dogs of different breeds in Sweden. *J. Anim. Sci.* 101, skad173. doi:10.1093/jas/skad173
- Sun, Y., Rahbani, J. F., Jedrychowski, M. P., Riley, C. L., Vidoni, S., Bogoslavski, D., et al. (2021). Mitochondrial TNAP controls thermogenesis by hydrolysis of phosphocreatine. *Nature* 593, 580–585. doi:10.1038/s41586-021-03533-z
- Switonski, M., and Mankowska, M. (2013). Dog obesity: the need for identifying predisposing genetic markers. *Res. Vet. Sci.* 95, 831–836. doi:10.1016/j.rvsc.2013.08.015
- Sypniewski, M., and Szydlowski, M. (2023). A study of 41 canine orthologues of human genes involved in monogenic obesity reveals marker in the ADCY3 for body weight in Labrador Retrievers. *Vet. Sci.* 10, 390. doi:10.3390/vetsci10060390
- Szydlowski, M., and Antkowiak, M. (2022). No evidence that long runs of homozygosity tend to harbor risk variants for polygenic obesity in Labrador retriever dogs. *J. Appl. Genet.* 63, 557–561. doi:10.1007/s13353-022-00693-0
- Teo, S. M., Pawitan, Y., Ku, C. S., and Salim, A. (2012). Statistical challenges associated with detecting copy number variations with next-generation sequencing. *Bioinformatics* 28, 2711–2718. doi:10.1093/bioinformatics/bts535
- Tvarijonavičiute, A., Ceron, J. J., Holden, S. L., Cuthbertson, D. J., Biourge, V., Morris, P. J., et al. (2012). Obesity-related metabolic dysfunction in dogs: A comparison with human metabolic syndrome. *BMC Vet. Res.* 8, 147. doi:10.1186/1746-6148-8-147
- Van Poucke, M., Stee, K., Sonck, L., Stock, E., Bosseler, L., Van Dorpe, J., et al. (2019). Truncating SLC12A6 variants cause different clinical phenotypes in humans and dogs. *Eur. J. Hum. Genet.* 27, 1561–1568. doi:10.1038/s41431-019-0432-3
- Wallis, N., and Raffan, E. (2020). The genetic basis of obesity and related metabolic diseases in humans and companion animals. *Genes. (Basel)* 20, 1378. doi:10.3390/genes11111378
- Willer, C. J., Speliotes, E. K., Loos, R. J., Li, S., Lindgren, C. M., Heid, I. M., et al. (2009). Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nat. Genet.* 41, 25–34. doi:10.1038/ng.287
- Zeng, R., Zhang, Y., and Du, P. (2014). SNPs of melanocortin 4 receptor (MC4R) associated with body weight in Beagle dogs. *Exp. Anim.* 63, 73–78. doi:10.1538/expanim.63.73