ANNALS OF ANIMAL SCIENCE
ISSN: 2300-8733, https://sciendo.com/journal/AOAS

ACCEPTED AUTHOR VERSION OF THE MANUSCRIPT:

Trios-based inquiry into *de novo* copy number variants in the swine genome

DOI: 10.2478/aoas-2025-0120

Magdalena Frąszczak, Błażej Nowak, Martyna Kaźmierczak, Magda Mielczarek♦

Department of Genetics, Wrocław University of Environmental and Life Sciences, Kożuchowska 7, 51-631 Wrocław, Poland

♦Corresponding author: magda.mielczarek@upwr.edu.pl

Received date: 18 February 2025
Accepted date: 14 October 2025

To cite this article: (2026). Frąszczak M., Nowak B., Kaźmierczak M., Mielczarek M. (2026). Trios-based inquiry into *de novo* copy number variants in the swine genome, Annals of Animal Science, DOI: 10.2478/aoas-2025-0120

Trios-based inquiry into *de novo* copy number variants in the swine genome

Magdalena Frąszczak, Błażej Nowak, Martyna Kaźmierczak, Magda Mielczarek♦

Department of Genetics, Wrocław University of Environmental and Life Sciences, Kożuchowska 7, 51-631 Wrocław, Poland

♦Corresponding author: magda.mielczarek@upwr.edu.pl

DOI: 10.2478/aoas-2025-0120
Abbreviated title: **Trios-Based Inquiry into *de novo* CNV in the Swine Genome**

* This work was supported by the Wrocław University of Environmental and Life Sciences (Poland) as part of the Ph.D. research program "MISTRZ", No N090/0005/21. The APC/BPC is financed by Wroclaw University of Environmental and Life Sciences.

Abstract

Copy number variation (CNV) plays an important role in determining multiple phenotypic traits due to these large genomic segments having the potential to disrupt functionally important genomic regions. The current study investigated the prevalence, distribution, and functional impacts of *de novo* CNVs in swine genomes and compared them with those that were inherited. CNV patterns showed high similarity within nuclear families, indicating that most CNVs are inherited, yet 9% arose *de novo* and contributed to individual variation. The underlying systematic pattern indicated that *de novo* deletions are shorter than inherited ones, which has essential meaning, especially for functionally important genomic locations. Moreover, *de novo* deletions were not uniformly distributed along the genome, forming clusters. Despite the potentially more severe impact of deletions, more *de novo* deletions than duplications were present in offspring genomes. The highest CNV-QTL (quantitative trait loci) overlap was found for teat number, reflecting strong selection for this trait. Overlaps also occurred in key reproductive and production traits. Multiple CNVs in the same gene may suggest regions prone to structural variation, as seen in testis-expressed 14 (*TEX14*), linked to pig infertility, and as parkin RBR E3 ubiquitin protein ligase (*PRKN*), a known CNV hotspot in humans. Deletions were primarily located in introns, with no significant enrichment of the Gene Ontology (GO) terms or Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. *De novo* duplications mainly affected genes related to environmental interactions, such as the immune response and olfactory receptors.

Key words: copy number variants, *de novo* CNV, inheritance, nuclear families, structural variants

The pig is one of the most economically important livestock species worldwide (Jiang et al., 2014), serving as an excellent model organism for research focused on human health and disease, contributing to the understanding of human phenotypes and disease. Indeed, the anatomy, physiology, and immunology of pigs are comparable to those of humans. Furthermore, the pig genome shows a higher similarity to the human genome than the mouse, which is considered a standard model organism (Pabst, 2020; Walters and Prather, 2013).

Genomes contain many types of deoxyribonucleic acid (DNA) polymorphisms. Such genetic variation manifests in different genomic dimensions, from single-nucleotide polymorphisms (SNPs) to large-scale structural variants (SVs), and often involves variations in the number of copies of long genomic segments. CNVs are a particular subtype of SVs caused by deletions or duplications that range from approximately 50 base pairs (bp) to several megabase pairs (Mb). In recent years, CNVs were most commonly detected in porcine genomes using array-based comparative genomic hybridisation (aCGH; see Fadista et al., 2008; Li et al., 2012; Wang et al., 2014), SNP arrays (Ramayo-Caldas et al., 2023; Wang et al., 2012; 2013; Xu et al., 2023), or whole genome sequence (WGS) data (Esteve-Codina et al., 2013; Liu et al., 2017; Paudel et al., 2013; Revilla et al., 2017; Wang et al., 2017; Zheng et al., 2020; Qian et al., 2023; Jang et al., 2023).

Since CNVs encompass more nucleotide sequences than SNPs, they have greater potential to impact phenotypic variation and disease susceptibility by altering genes and gene promoter sequences, for example. In pigs, genes related to disease (Long et al., 2016), olfaction, neurological processes (Paudel et al., 2013; 2015) coat colour (Giuffra et al., 2002), fatty acid composition (Revilla et al., 2017), and production performance (Jiang et al., 2014) are reportedly enriched in CNVs. Furthermore, Ramayo-Caldas et al. (2023) found associations between porcine CNV and the diversity and composition of pig faecal microbiota.

All polymorphisms, including CNVs, are inherited or can arise *de novo*. Hehir-Kwa et al. (2011) analysed one of the largest cohorts (3,443 individuals) available for studying *de novo* CNVs using SNP microarrays and observed a significant association between paternal age and *de novo* CNV mutation rate. Interestingly, a recent large-scale aCGH analysis of 2,323 individuals by Wadhawan et al. (2020) found a significant association between maternal age and *de novo* CNV mutation rate. In humans, *de novo* CNVs are linked to multiple neurological diseases such as schizophrenia (Kirov et al., 2012), intellectual disability (Gilissen et al., 2014), autism spectrum disorder (Sanders et al., 2015), and neurodevelopmental disorders (Hamanaka et al., 2022). Arias et al. (2023) studied the *de novo* formation of CNVs in pigs using a sample of 478 parent-offspring trios, though SNP array genotyping only provided a low-accuracy estimation of CNV breakpoint positions. Meanwhile, Steensma et al. (2023) examined *de novo* SVs in commercial pig lines based on the WGS data of 37 trios, but did not focus on SV distribution. Instead, they aimed to highlight the potential of livestock breeding programmes to provide a suitable population structure for *de novo* SV identification and characterisation using ear, hair, and semen samples. However, the Steensma et al. (2023) data are not publicly available, so it is not possible to compare CNVs detected using such a vast dataset with those found in the current study.

Our study aimed to determine the prevalence of *de novo* formed CNVs in the porcine genome based on WGS data, and characterise the distribution of such polymorphisms across

the genome. Moreover, we compared the length and number of *de novo* CNVs to inherited CNVs and annotated them in a functional context.

Material and methods

Sequenced individuals

The data set consisted of WGSs, obtained using the Illumina HiSeq 2000, of twelve individuals representing the Polish Large White breed. The sequenced individuals represented two unrelated nuclear families comprising sire - dam - single offspring trios (two trios) or sire - dam – two full sibs (four trios) (Figure 1). All animals were housed in one closed piggery, divided into sectors, with standard environmental, microclimatic, and nutritional conditions. The temperature and humidity in the piggery ranged from 20 - 22°C and 70 - 80%, respectively. The animals were fed a commercial diet and had constant access to water. The datasets generated and analysed during the study are available in the National Center for Biotechnology Information database (Bioproject ID: PRJNA1172736) of the Sequence Read Archive (SRA) repository.

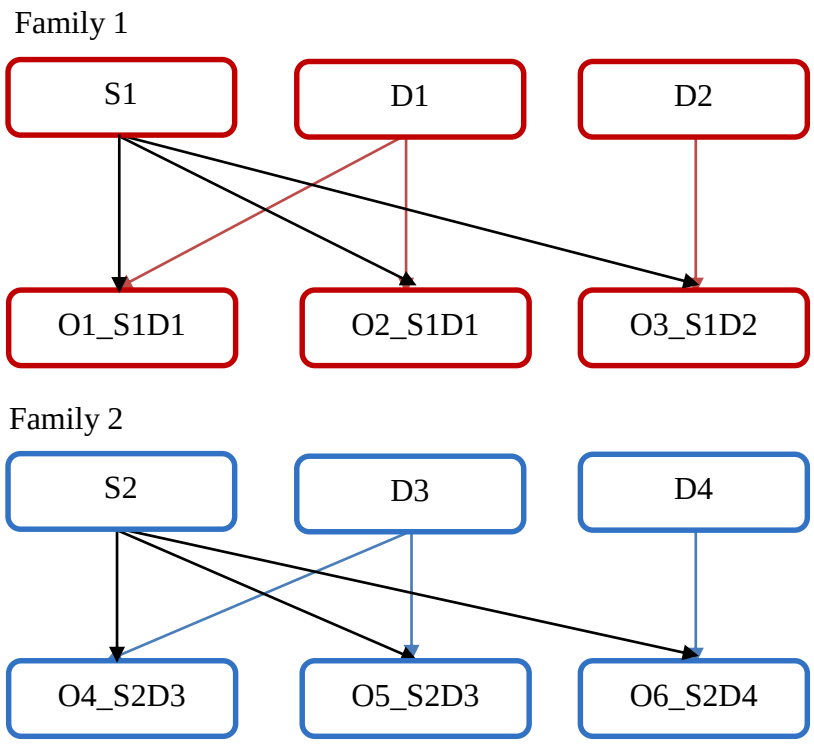


Figure 1. The whole genome sequencing experimental design. A black arrow indicates the paternal contribution, while a red/blue arrow indicates the maternal. S corresponds to sire, D to dam, and O to offspring. Family 1 is represented by red, while Family 2 is represented by blue

Bioinformatic pipeline

CNV identification from WGS data involved (1) raw data quality control using FastQC (Andrews, 2010) and MultiQC (Ewels et al., 2016), (2) quality-based read trimming using

Trimmomatic (Bolger et al., 2014), (3) alignment to the Sscrofa11.1 reference genome by the BWA-MEM software (Li and Durbin, 2009), (4) post-alignment processing with the Picard (http://broadinstitute.github.io/picard) and SAMtools (Li et al., 2009) packages, (5) detection of CNV deletions and duplications with CNVnator (Abyzov et al., 2011) and Pindel (Ye et al., 2009), (6) CNV filtration, and (7) functional analysis with Variant Effect Predictor (McLaren et al., 2016), ShinyGO (Ge et al., 2020), and custom scripts. (1) Default parameters were used for FastQC, which included assessments of per-base sequence quality, average sequence quality, sequence duplication levels, and adapter content, among others. MultiQC was used with default settings to combine FastQC reports into a single HTML report. (2) Raw data cleaning involved trimming low-quality reads by applying a four-base sliding window and cutting the read once the average quality in the window dropped below 20 (SLIDINGWINDOW:4:20). Any reads shorter than 60 bp after trimming were removed (MINLEN:60). (3) Alignment to the Sscrofa11.1 reference genome (the assembly version: https://ftp.ensembl.org/pub/release-113/fasta/sus_scrofa/dna/) was performed with default parameters, including a seed length of 19, a mismatch penalty of four, and a gap open penalty of six. The read group ID was attached to every read in the output. (4) BAM files from multiple lanes were first merged using SAMtools merge (i), then sorted by genomic coordinates (ii) with SAMtools sort. Mate-pair information was corrected (iii) using SAMtools fixmate-m, which ensures proper pairing information for downstream analysis. Genome-wide depth of coverage was calculated (iv) using SAMtools depth-a, which reports coverage at all genomic positions, including those with zero coverage. A custom script was then used to compute the average genome coverage (v) from the SAMtools depth output. Additionally, alignment summary statistics (vi), such as total reads, mapped reads, and properly paired reads, were obtained using SAMtools flagstat. (5) Both CNV detection programmes were used with default parameters, except for CNVnator, for which a window size of 200 bp was chosen. As Abyzov et al. (2011) suggest, the larger bin size should be used for the lower genome coverage, which we determined using the CNVnator eval option. (6) Identified raw CNVs were post-processed by (i) removing variants shorter than 50 bp or longer than 1,000,000 bp, (ii) those detected with Pindel software supported by less than three reads, and (iii) those overlapping with gaps in the reference genome. (iv) The final set of CNVs was based on the consensus output from CNVnator and Pindel. In summary, the CNVnator output was used as a baseline, while the validated variants comprised CNVs that had at least 90% overlap with CNVs detected by Pindel, considering the +/- 100 bp breakpoint accuracy. All steps from this section were described in detail by Mielczarek et al. (2023). In this study, only CNVs located on autosomes were further considered. CNVs detected in offspring that were not present in the parents or any other animals were considered *de novo*. (7) To demonstrate the potential relationships between d*e novo* CNVs and known pig QTL, the overlap between them was also analysed. Swine QTLs from the Sscrofa 11.1 genome were downloaded from the Animal QTL database (http://www.animalgenome.org/cgi-bin/QTLdb/SS/index; release 56; Apr 24, 2025) (Hu et al., 2022), which includes 57,041 known QTL representing 406 different base traits and 1,088 variants of these traits. The custom script was used to find the overlap at the bp level. The Variant Effect Predictor tool was used to annotate *de novo* CNVs by identifying their locations within or near genes. To explore potential functional implications of gene-level overlaps, we retrieved all affected genes to determine whether any were impacted by more than one large-

scale variant. Canonical transcripts overlapping with *de novo* CNVs were selected for the functional enrichment analysis. The ShinyGO tool was used for exploring enrichment in Gene Ontology (GO) terms representing the molecular function category (Ashburner et al., 2000; The Gene Ontology Consortium et al., 2023) and in pathways defined by the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa et al., 2017). A p-value cut-off (false discovery rate [FDR]) of 0.05 was used after implementation in ShinyGO.

CNV determined in offspring only, common for full siblings or full siblings and their half siblings, were considered as putative *de novo* germline mutations occurring in one of the parents and were thus subjected to functional annotation. This subset was not included in the gene-level overlap, QTL, or formal statistical analyses, as our study did not focus on *de novo* events occurring in the parents. However, we considered this an interesting observation. Given the potential biological relevance of these CNVs, we decided to perform a functional analysis of them. It is worth keeping in mind that without CNV detections performed on the single-cell or at least single-tissue resolution, it is difficult to differentiate which variants are inherited as germline mutations of the parent and which arise *de novo* in offspring. The systematic differentiation between these two sources of *de novo* variation would be an interesting follow-up to our study.

Statistical modelling of *de novo* copy number variants

All tests were performed separately on autosomal deletions and duplications. First, to define the subsequent statistical handling of the data (parametric or non-parametric methods), the null hypothesis of a normal distribution of CNV lengths and counts for both inherited and *de novo* variants was examined using the Kolmogorov-Lilliefors test for normality. The test implies neither a predefined expected value nor the variance of the normal distribution, and the test statistic is expressed by the maximal absolute difference between the empirical and theoretical normal cumulative distribution functions. The p-value was approximated using the Dallal-Wilkinson formula (Dallal and Wilkinson, 1986). Due to a lack of normality, non-parametric methods were applied. For the assessment of similarities in CNV distribution between pairs of individuals, a genome was divided into 1,000 bp regions, following the approach proposed by Jang et al. (2023). Each region was then classified as containing a CNV(s) when a CNV overlapped with the region, or as CNV-free when there was no overlap. Based on this classification, the Jaccard similarity coefficients were calculated for each pair of individuals:

$$J = \frac{s_{11}}{s_{11} + s_{10} + s_{01}},$$

where s_{11} denotes the number of regions containing a CNV in both individuals, and s_{01} and s_{10} denote the number of regions where only one individual contained a CNV. Furthermore, Kruskal multidimensional scaling (MDS) was performed, based on the values of the Jaccard coefficient calculated for all possible pairs of individuals, to visualise the similarities between all animals. The next step determined CNV regions (CNVRs) by merging CNVs of the same type that wholly or partly overlapped between individuals, as applied in Revilla et al. (2017). Moreover, a CNV unique for an individual was also considered a CNVR. CNVRs were then used to construct UpSet plots to present the number of regions common for related animals, as well as the *de novo* variants. Permutation tests evaluated the null hypothesis of no differences

in the number of *de novo* and inherited CNVs using test statistics from the Wilcoxon signed rank test. The uniform distribution of *de novo* CNVs along the genome was tested using a Pearson goodness-of-fit test:

$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - np)^2}{np} \sim \chi^2_{k-1}$$

where n_i denotes the distances between consecutive *de novo* CNVs, $n = 2{,}265{,}774{,}640$ is the total genome length (only autosomal chromosomes), and $p = \frac{1}{d}$, where d is the number of *de novo* deletions/duplications.

The null hypothesis of equal lengths of *de novo* and inherited CNVs was tested using the Mann-Whitney U test:

$$Z = \frac{U - \frac{km}{2}}{\sqrt{\frac{km(k+m+1)}{12}}} \sim N(0,1),$$

where $U = \sum_{i=1}^{k} R_i - \frac{k(k-1)}{2}$, R_i denotes the rank of the length of the i-th *de novo* variant in the vector of lengths of all CNVs, while k and m are the number of *de novo* and inherited CNVs, respectively. All calculations were performed using the corresponding functions implemented in the R package (ade4, ComplexHeatmap, dplyr, ggplot2, numbers, plotify, UpSetR, tidyverse; R Core Team, 2022).

Results

A comprehensive overview of all copy number variants

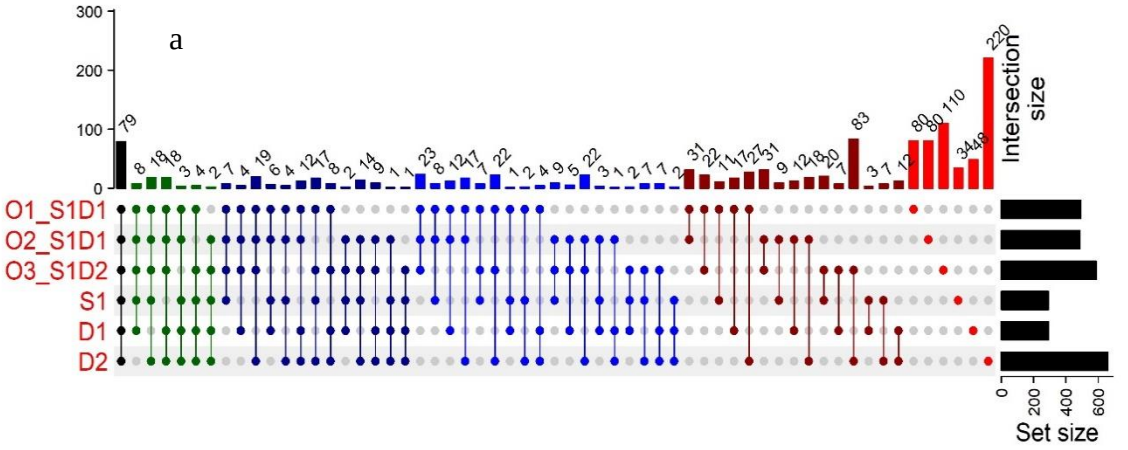
Most sequenced reads (between 98.11% and 98.60%, depending on the individual) were aligned to the reference genome, while the percentage of properly paired reads ranged from 95.31% to 96.00%. The resulting genome average coverage ranged from 10x to 19x (Additional file 1). The total number of deletions per individual ranged from 325 to 753. For almost all animals, the highest number of deletions (34 - 81) was located on pig chromosome SSC2, although it has almost half the length (151,935,994 bp) of SSC1 (274,330,532 bp). As expected, the lowest number of deletions for each individual (0 - 8) was always on the shortest (55,982,971 bp) autosome (SSC18) (Additional file 2). The number of duplications per individual ranged from 282 to 444, with the highest number (38 - 80) located on SSC7, while SSC18 contained the lowest number of duplications (0 - 3) (Additional file 3). The length of CNVs varied between 600 bp and 195,000 bp (6,662 ± 13,002 bp) for deletions, and between 1,200 bp and 561,400 bp (10,036 ± 33,438 bp) for duplications. Depending on the individual, this comprised 0.11% to 0.21% of the total autosomes length being deleted, and 0.26% to 0.35% of the autosomal genome being duplicated. The lowest total number of CNVs was detected on SSC18 for deletions (48) and duplications (11), while the highest number of deletions (667) was found on SSC2 overlapping this chromosome by 0.75%. However, longer deletions occurred on SSC9, resulting in the highest percentage coverage (0.96%) for this chromosome. Most duplications were identified on SSC7, covering 1.75% of its length (Additional file 4), with the total number and length of CNVs depending on the chromosome.

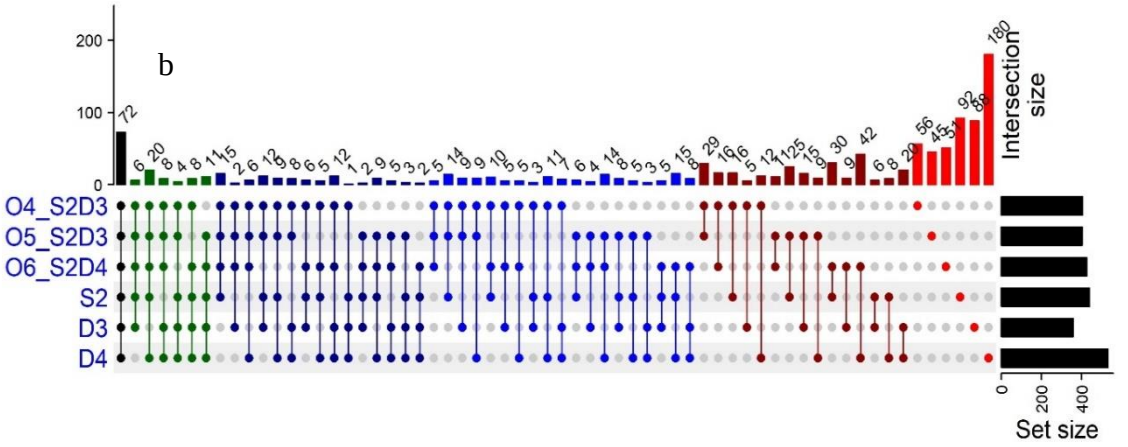
The CNV patterns within nuclear families exhibited a high degree of similarity, suggesting that a considerable proportion of CNVs are inherited. Full siblings demonstrated greater genomic similarity than unrelated individuals in terms of deletions and duplications (Figures 2 and 3). Moreover, the offspring exhibited a higher degree of similarity to the mother than to the father when considering duplications. However, some variation within the family was still observed and was caused by *de novo* CNVs.

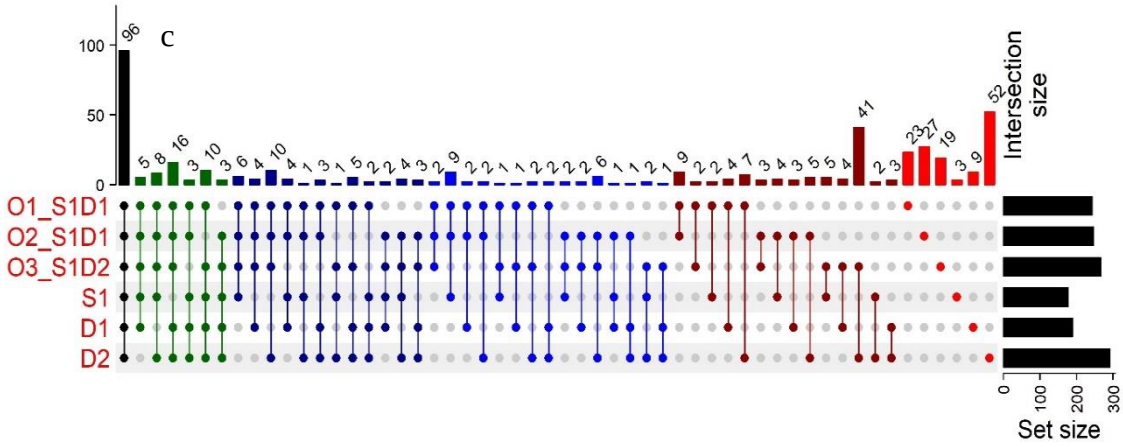
Figure 2. The similarity between animals in overlapping deletions. S corresponds to sire, D to dam, and O to offspring. Family 1 is represented by red, while Family 2 is represented by blue

Figure 3. The similarity between animals in overlapping duplications. S corresponds to sire, D to dam, and O to offspring. Family 1 is represented by red, while Family 2 is represented by blue

All individuals had 42 deletions and 66 duplications in common, compared to the reference genome. The highest fractions of shared deletions and duplications were determined among full siblings and varied from 50% to 52% and 70% to 78%, respectively. The frequency of CNVs shared with family depended on the animal, and ranged between 84% - 89% for deletions and 90% - 96% for duplications. *De novo* variants constituted only 2% - 7% of all identified duplications, and between 9% and 13% of all deletions. The number of CNVRs within nuclear families, as well as the number of *de novo* CNVs, is presented in Figure 4 (a-d). It is worth noting that there were 29 deletions and seven duplications detected in full-sibling genomes, as well as six deletions and two duplications in the genomes of full-siblings and their half-siblings. Since these polymorphisms were not detected in parental genomes, but were defined in siblings, they could still be considered putative *de novo* germline mutations occurring in one of the parents.


a
Intersection size
O1_S1D1
O2_S1D1
O3_S1D2
S1
D1
D2
Set size


b
Intersection size
O4_S2D3
O5_S2D3
O6_S2D4
S2
D3
D4
Set size


c
Intersection size
O1_S1D1
O2_S1D1
O3_S1D2
S1
D1
D2
Set size

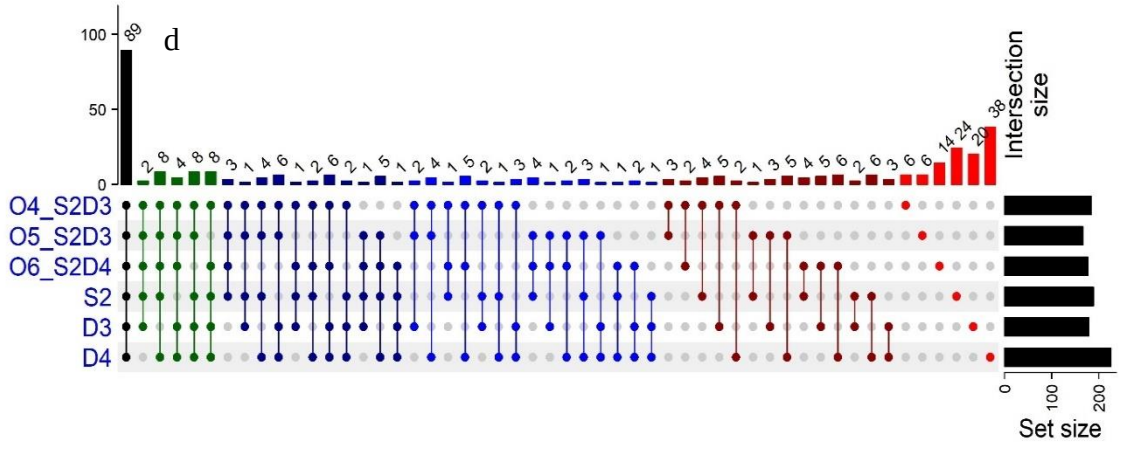

d
Intersection size
O4_S2D3
O5_S2D3
O6_S2D4
S2
D3
D4
Set size

Figure 4. Copy number variant (CNV) overlap within families. The left black column shows the number of CNV regions (CNVRs) detected in all animals. The green/navy blue/blue/navy red columns give the number of CNVRs found in five/four/three/two animals, and the dots below the columns denote the animals with these variants. The red columns correspond to CNVs unique to each individual. The first three red columns always correspond to the number of *de novo* CNVRs present in offspring. The horizontal columns are the total CNVRs identified in each animal. The first two graphs (a and b) show deletion regions, and the latter (c and d) refer to duplication regions

De novo copy number variant survey

Considering that *de novo* CNVs make up only 9% of all CNVs, most (14%) were located on SSC1. Significantly fewer CNVs were *de novo* than inherited ($p = 0.008$ for deletions and 0.006 for duplications). The frequency of *de novo* deletions was the highest on chromosomes three (0.236) and 18 (0.32), while the highest frequency of *de novo* duplications occurred on chromosomes 16 (0.167) and 18 (0.333). The lowest frequency of *de novo* deletions occurred on SSC11 (0.069), with the lowest duplication frequency occurring on SSC5 (0.005). Despite the potentially higher detrimental impact of deletions over duplications, there were more *de novo* deletions than *de novo* duplications, regardless of the chromosome (Figure 5 and Additional file 5).

The distribution of *de novo* CNVs along the genome was non-uniform (p approximated zero for deletions and duplications). The distance between neighbouring *de novo* deletions varied from 600 bp to 67,171,600 bp, and duplications varied between 1,800 bp and 84,799,836 bp. The longest distances were observed on SSC8 for both deletions and duplications. In addition, we observed clusters of deletion variants, with only a few separated single deletions (see Figure 6 for deletions and Figure 7 for duplications).

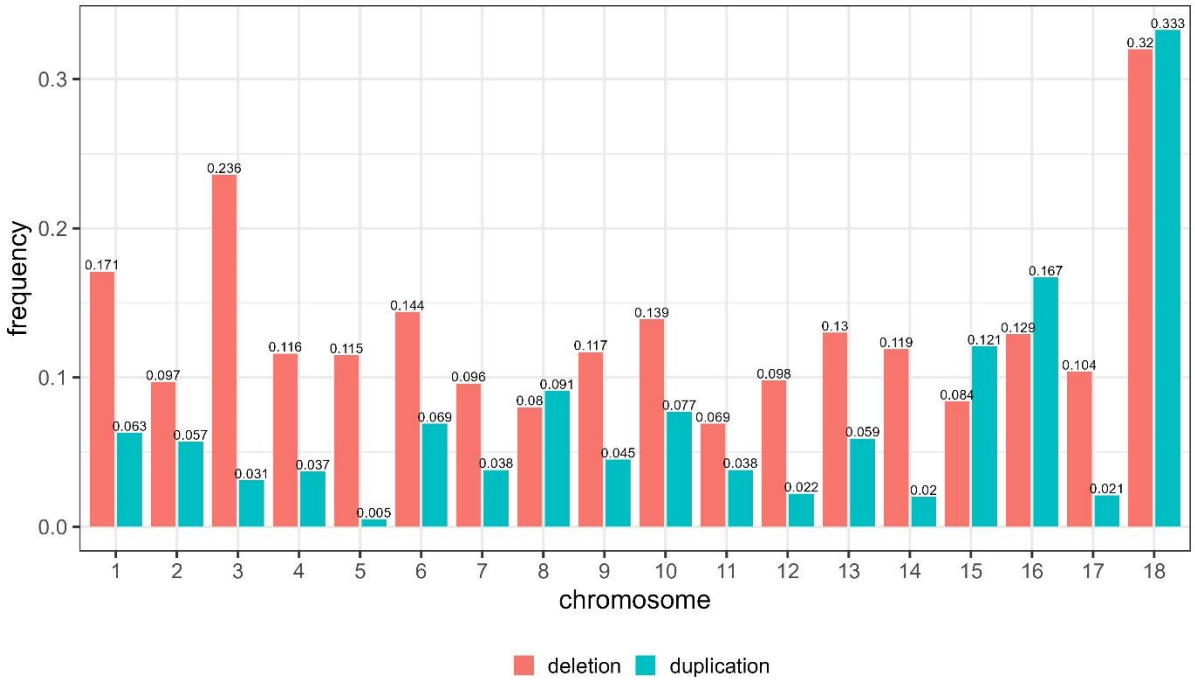


Figure 5. The frequency of *de novo* copy number variants (CNVs) to all CNVs

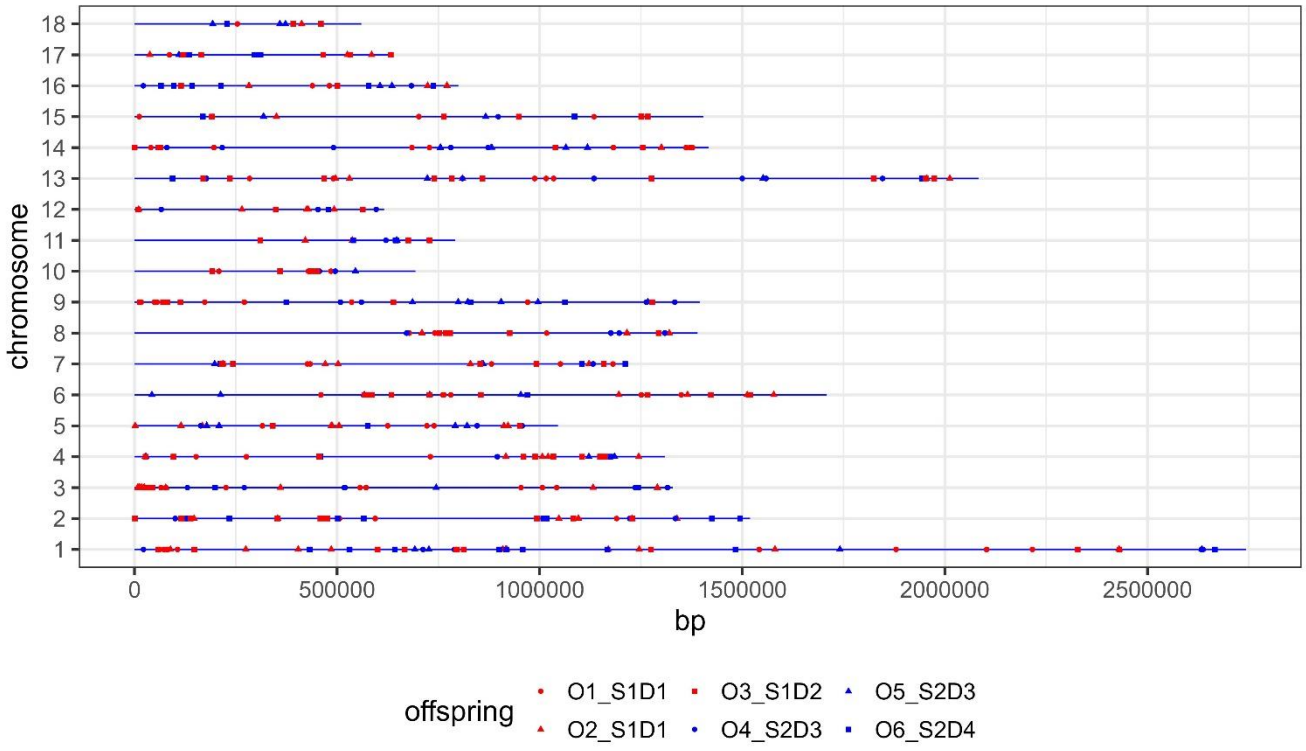


Figure 6. Genomic distribution of *de novo* deletions identified in the six offspring

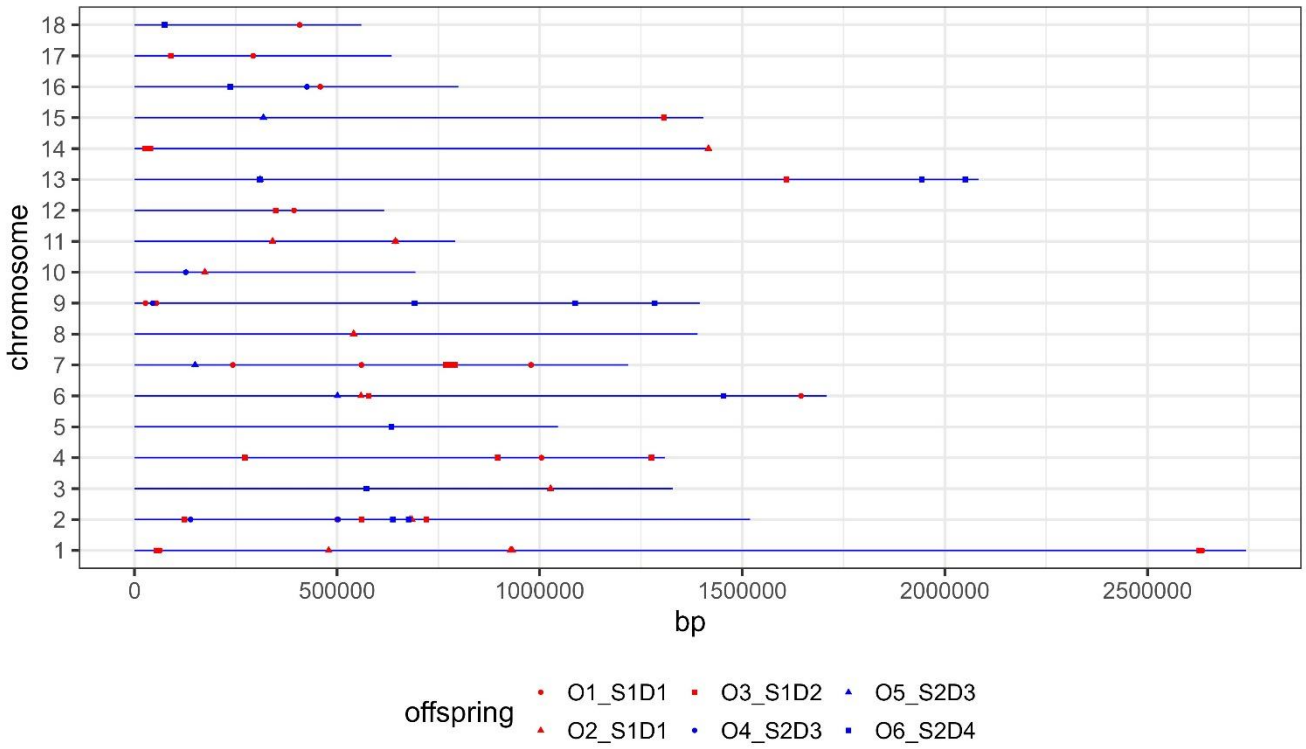


Figure 7. Genomic distribution of *de novo* duplications identified in the six offspring

In all individuals, *de novo* deletions were shorter than inherited deletions (Additional file 6), with length varying considerably among animals (600 bp to 70,000 bp), and the median length of *de novo* deletions was one-third of the inherited ones. On the other hand, no significant differences were found in the length of *de novo* and inherited duplications. Tables containing basic statistics for the length of *de novo* and inherited CNVs can be found in Additional file 7 (divided by individuals) and Additional File 5 (divided by chromosomes).

Functional perspective on *de novo* copy number variants

In the case of deletions, overlapping CNVs with QTLs were identified on 10 chromosomes for five reproduction traits, including sperm abnormality rate (SSC1 and 12), number of functional sperm (SSC1), teat number (SSC4, 7, 9, 10, 14, and 15), sperm progressive motility (SSC7), and offspring number (SSC13). They were also identified in the longissimus muscle depth (SSC5 and 13), number of ribs (SSC7), and subcutaneous fat thickness (SSC13) production traits, as well as one immune-related trait (interleukin-8 level; SSC15).

In the case of duplications, a relationship was found on nine chromosomes for 14 different traits, including five reproductive traits (number of functional sperm [SSC1 and 4], semen odour [SSC1], boar sexuality score [SSC1], sperm progressive motility [SSC4], and teat number [SSC7, 17, and 18]), six production traits (longissimus muscle area [SSC1], body condition score [SSC1], average daily gain [SSC4], body length [SSC5], number of ribs [SSC7], and meat colour [SSC14]), and three physiological traits (interleukin-12 [SSC13], blood lipase and, immunoglobulin G levels [both SSC15]). Gene-level overlap between *de novo* CNVs was found for parkin RBR E3 ubiquitin protein ligase (*PRKN*) (ENSSSCG00000004032;

chr1:5465312–6730872) and testis-expressed 14 (*TEX14*) (ENSSSCG00000017645; chr12:34815476–34912766), which were affected by deletions and duplications. However, the deletions and duplications did not overlap within the gene in either case. Specifically, for ENSSSCG00000004032, three CNVs were identified within the gene boundaries (one deletion [chr1:5908601–5911400; 0.22% for the gene length] and two duplications [chr1:6264201 6273000; 0.69%, and chr1:5514201–5516400; 0.17%]). For ENSSSCG00000017645, one deletion (chr12:34899601–34900400; 0.82%) and one duplication (chr12:34888201–34899600; 11.71%) were located in adjacent but non-overlapping gene segments. A visualisation of all regions in the genomic context is provided in Additional file 8.

Most (127) of the 184 canonical transcripts that overlapped with *de novo* deletions were located in introns, meaning they had no direct effects on proteins. However, more severe consequences were also identified, including feature truncation (34), transcript ablation (11), and stop lost (11). It is worth noting that not all affected transcripts are well-characterised since they represent novel genes. Nonetheless, severe consequences of *de novo* deletions were also determined for known genes, including stop codon loss in LOC100515185 acyl-coenzyme A amino acid N-acyltransferase 2 (ENSSSCG00000038171), *HGSNAT* heparan-alpha-glucosaminide N-acetyltransferase (ENSSSCG00000038960), *ELFN1* extracellular leucine-rich repeat and fibronectin type III domain containing 1 (ENSSSCG00000030485), *ACTL8* actin-like 8 (ENSSSCG00000024062), and LOC106504900 olfactory receptor 8B3-like (ENSSSCG00000058524). However, no enrichment of GO terms or KEGG pathways was found for *de novo* deletions.

For *de novo* duplications affecting 55 canonical transcripts, transcript amplification (24), intron (13), coding sequence variant (12), non-coding transcript exon variant (4), and feature elongation (2) sequence ontologies were identified. The enriched GOs incorporated immunoglobulin receptor binding (GO:0034987; $p = 1.7 \cdot 10^{-2}$), immune receptor activity (GO:0140375; $p = 4.3 \cdot 10^{-2}$), antigen binding (GO:0003823; $p = 4.3 \cdot 10^{-2}$), transmembrane signalling receptor activity (GO:0004888; $p = 4.3 \cdot 10^{-2}$), and molecular transducer activity (GO:0060089; $p = 4.3 \cdot 10^{-2}$). Additional file 9 shows GOs with fold enrichment and the number of genes corresponding to the GO term. Furthermore, the KEGG pathway for olfactory transduction was determined (ssc04740; $p = 9.8 \cdot 10^{-3}$).

Among CNVs that indicate *de novo* mutations in the parental germline tract (present in siblings and absent in their parents' genomes), 10 deletions and nine duplications overlapped with canonical transcripts. Deletions were mainly located in introns (9). However, one was reported as a feature truncation variant of the LOC100515852 polymeric immunoglobulin receptor-like gene (ENSSSCG00000017235). Considering duplications, transcript amplification (5), intron variants (2), and coding sequence variants (2) were determined according to sequence ontology. Some of the transcripts affected by these duplications correspond to novel genes that are not well characterised, except *FCN1* ficolin (collagen/fibrinogen domain containing) 1 (ENSSSCG00000029414; amplification), *FCN2* ficolin (collagen/fibrinogen domain containing lectin) 2 (ENSSSCG00000023333; amplification), and *CYP4A24* cytochrome P450 family 4 subfamily A member 24 (ENSSSCG00000062158; coding sequence variant).

Discussion

In recent years, extensive research on CNVs has highlighted their important role in population diversity, disease development, and evolution (Pös et al., 2021). It is well-known that CNVs strongly affect phenotypes by changing gene structure, dosage, and regulation (1000 Genomes Project et al., 2011; Geistlinger et al., 2018). The high impact of CNVs is caused by their dimensions, spanning from 50 bp to several Mb, making a single CNV capable of encompassing several genes (Du et al., 2022; Pös et al., 2021). Indeed, there is a corresponding change in gene expression when CNVs occur, with 85% to 95% of CNVs in humans and mice associated with expression changes in the affected genes (Henrichsen et al., 2009; Tang and Amon, 2013).

De novo mutations are a major cause of severe genetic disorders (Acuna-Hidalgo et al., 2016), which explains why the *de novo* CNVs identified in this study were shorter and less common than inherited ones. Shorter CNVs span smaller regions of the genome, meaning they may have a lower impact on phenotypes or genome stability. Larger inherited deletions may have been retained in the population due to their neutral or beneficial effects, whereas *de novo* mutations are typically subject to immediate selection pressures due to their potential deleterious impacts (Acuna-Hidalgo et al., 2016). These findings are consistent with our results, with significantly fewer CNVs arising *de novo* compared to inherited CNVs, and in line with current knowledge (McCarroll et al., 2008; Wen et al., 2022) showing that up to 99% of CNVs are inherited (van Ommen, 2005). The location of mutations is not random across the genome and is determined by multiple factors, including sequence composition and its functional role (Acuna-Hidalgo et al., 2016).

The highest overlap between CNVs and QTLs (for both deletions and duplications) was found for teat number, which is consistent with research showing this trait to be among the 10 with the highest overlap (Reavy et al., 2015; Keel et al., 2019). This may be related to the large number (n = 2,936) of known QTLs for this trait, which results from the fact that selection for increased teat number has been conducted for a long time, and is also necessary because litter size typically exceeds the number of available nipples (Rohrer and Nonneman, 2017; Yang et al., 2023). Although CNVs are primarily the result of replication and recombination events, there are indications that artificial reproductive conditions may indirectly influence their development or perpetuation. In pigs, this may be related to artificial insemination, the most important tool in modern pig breeding, enabling intensive boar selection based on semen quality (Gao et al., 2019; Zhuang et al., 2023). Additionally, Large White breeds are widely used for crossbreeding due to their high reproductive performance and excellent meat production (Zhang et al., 2022). Among these breeds, the *FANCM* (Fanconi anaemia complementation group M) gene is involved in germ cell development, and mutations can cause male reproductive disorders due to sperm deformation and reduced sperm number and motility (Yin et al., 2019). Our study confirmed this by demonstrating overlap between the CNVs and QTLs related to sperm morphology and physiology. Production traits such as backfat thickness, meat colour, and rib number play a key role in intensive selection in pig breeding. To date, CNVs related to meat colour and backfat thickness have been identified in the pig genome, involving the *TGFBR3* (TGF-beta receptor type III) gene (Wang et al., 2015; Zhang et al., 2024). However, rib number appears to be particularly important, especially for pig producers, as a higher

number correlates with greater carcass length. The QTL region affecting rib number is located on SSC7, encompassing the vertebrae development-associated (*VRTN*) gene, which corresponds to our results. This gene plays a key role in spine development, and a specific intron insertion (e.g., g.20311_20312ins291) significantly affects rib count and carcass length (Borchers et al., 2004).

Undoubtedly, the distribution of *de novo* CNVs along the genome is non-uniform, which was also demonstrated in our study. We observed clusters of variants, with only a few separated single deletions, especially for deletions. Mutational clusters have been identified, and they correspond to multiple *de novo* mutations in very close vicinity in a single individual (Acuna-Hidalgo et al., 2016; Chan and Gordenin, 2015). Interestingly, the presence of deletions and duplications within the same gene suggests that the gene may be located in a genomic region prone to structural variation. Such regions are often enriched for repetitive elements and segmental duplications, which serve as hotspots for recurrent structural alterations mediated by mechanisms such as non-allelic homologous recombination (NAHR) and related processes (Höps et al., 2024; Lin and Gokcumen, 2019; Paudel et al., 2013; Soto et al., 2023).

Gene-level overlap between *de novo* CNVs was found for *TEX14* and *PRKN*. Sironen et al. (2011) described the role of *TEX14* in spermatogenesis in Yorkshire pigs and highlighted the importance of specific genomic remodelling events as causes for inherited defects. A specific male infertility in Yorkshire pigs, characterised by early meiotic spermatogenic arrest, was linked to a 2 Mb region on SSC12. Sequencing of the candidate gene (*TEX14*) revealed a 51 bp insertion leading to a premature stop codon. The insertion was likely the result of an original duplication event, followed by recombination and repositioning. This explanation corresponds well with the way structural variations like insertions and duplications arise and evolve in genomes, often involving duplication followed by recombination-mediated rearrangements. Evidence of hotspots in the pig *PRKN* gene has not been described; however, it is a well-established hotspot for CNVs in humans. The *PRKN* gene is located in one of several genomic regions of very high deletion frequency ('hotspots'), where rare deletions are found at frequencies of up to 100-fold higher than the average for the genome as a whole (Toft and Ross, 2010).

CNVs, whether they affect a QTL, a single gene, or the entire chromosome, have been identified as causes of not only diseases and developmental abnormalities, but also as sources of adaptive potential (Tang and Amon, 2013). The latter determines whether an organism can compete for resources and survive changing environmental conditions (Pös et al., 2021). The contribution of environmental factors to the origin of CNVs is still poorly understood. However, research shows that CNVs are enriched for genes associated with environmental factors, i.e., genes that are not critical for the organism's development, but rather facilitate its response to and interaction with an ever-changing environment. This includes, among others, enrichment for immune and inflammatory response genes. According to Tizaoui (2018), infections, chronic inflammation, cellular stress, and free radicals generated by inflammation favour *de novo* CNV formation. This is in line with our findings for duplications that significantly enriched genes related to immunoglobulin receptor binding, immune receptor activity, and antigen binding. Another example in which *de novo* duplicated genes tend to encode proteins that interface with the external environment includes those related to olfactory receptors (The Bovine Genome Sequencing and Analysis Consortium et al., 2009). Odours and chemosensory stimuli are

detected and identified by olfactory receptors that are crucial for finding food, detecting mates and offspring, recognising territories, and avoiding danger. Olfactory receptor genes are duplicated very widely within mammalian genomes (Chen et al., 2012; Groenen et al., 2012; Moreno-Estrada et al., 2007), suggesting they may be under strong selection (Groenen et al., 2012). Interestingly, pigs have the largest repertoire of functional olfactory receptor genes, indicating their importance for scavenging (Groenen et al., 2012). This is consistent with our findings, such as enrichment of the KEGG olfactory transduction pathway, signalling receptors, and molecular transducer activity. The latter transmits the signal from one side of the membrane to the other to initiate a change in cell activity or state as part of signal transduction, which captures, among others, olfactory receptor activity. No enrichment was determined for *de novo* deleted genes, though this may be explained by the fact that deletions can have severe consequences, such as the loss of specific genes or regulatory elements. If the deleted region includes important genes, it can alter normal cellular functions and result in developmental abnormalities (Acuna-Hidalgo et al., 2016; Liu and Bickhart, 2012).

It is important to acknowledge the limitations of this study. We analysed a small sample of 12 individuals, which may limit the detection of rare *de novo* CNVs and complicate the assessment of population-level relevance. However, it is important to note that our study was based on trio data, which is crucial for the accurate detection of *de novo* CNVs. The trio design allows for distinguishing true *de novo* variants from inherited ones, significantly reducing false positives and improving accuracy (Boonin et al., 2025). Moreover, studies investigating *de novo* CNVs based on WGS data in livestock species are relatively rare, and this study represents a valuable early step in applying such analyses to domestic animals. Considering the environmental effects, unlike studies where environmental variability (e.g., diet, stress, exposure to mutagens) may confound the interpretation of CNV formation, our use of standardized and controlled conditions minimizes such effects, allowing for a clearer interpretation of CNVs in the context of genetic rather than environmental factors (Arlt et al., 2012; Feuk et al., 2006). This study relied on blood DNA sequences; therefore, tissue-specific or somatic mosaic CNVs may have been missed. Nevertheless, blood is one of the most accessible and minimally invasive tissues to obtain, making it one of the most widely used biological materials (Svärd et al., 2025). Blood is also widely accepted for identifying germline *de novo* CNVs, which are expected to be present across all somatic tissues (Krepischi et al., 2012; Pereira et al., 2024; Stadler et al., 2012). Our approach provides a solid foundation for detecting germline CNVs in livestock. Although no experimental validation was performed, the use of appropriate experimental design, well-established CNV detection tools, and stringent filtering criteria increases confidence in the computational results. Moreover, excluding sex chromosomes from the CNV analysis, which may be seen as another limitation of the study, helped avoid technical and biological issues. Detection of CNV on the pig sex chromosomes (SSCX and SSCY) is challenging. The high number of repetitive elements on both chromosomes reduces short-read alignment accuracy. These highly repetitive elements cause ambiguity in short-read mapping because reads originating from repeats can map to multiple locations, reducing mapping quality and increasing false positives or negatives in CNV detection (Bickhart and Liu, 2014; Skinner et al., 2016). Moreover, hemizygosity of SSCX in males and the haploid nature of SSCY introduce complexities in ploidy normalisation and read depth interpretation, which are problematic for CNV detection tools (Keel et al., 2019).

Multiple CNV detection methods, especially those based on read depth, are further limited by reduced precision in defining CNV boundaries on sex chromosomes. These limitations may lead to the exclusion of SSCY from CNV analyses. Despite the aforementioned constraints, our study demonstrates the feasibility of detecting *de novo* CNVs using trio-based designs in livestock and provides valuable insights that can guide future, larger-scale studies.

To conclude, CNV patterns showed a high degree of similarity within nuclear families, indicating that a significant proportion of CNVs are inherited. However, 9% of all CNVs are due to *de novo* events and contribute to individual variation. No significant difference in the length of *de novo* and inherited duplications was recorded. However, *de novo* deletions were shorter than inherited deletions, which has implications for functionally important genomic locations. Despite the potentially greater detrimental impact of deletions compared to duplications, more *de novo* deletions were retained in the offspring genomes, and their distribution was non-uniform across the genome. The highest CNV-QTL overlap was found for teat number, reflecting strong and long-term selection for this trait. CNV-QTL overlaps were also associated with key reproductive and production traits, suggesting that artificial selection may influence CNV patterns in pigs. In terms of the functional impact on genes, they were primarily located in introns. The presence of multiple CNVs within the same gene suggests it may lie in a genomic region prone to structural variation, often associated with repetitive sequences and recombination hotspots. Notably, such overlap was observed in *TEX14* and *PRKN*, with *TEX14* linked to male infertility in pigs and *PRKN* known as a CNV hotspot in the human genome. Despite structural gene changes, no significant enrichment of GO terms or KEGG pathways was identified for them. However, *de novo* gene duplications occurred predominantly in genes involved in environmental interactions, particularly those associated with immune responses and olfactory receptor mechanisms.

Acknowledgements

We would like to express our sincere gratitude to Professor Joanna Szyda for her valuable comments on this research. This work was supported by the Wrocław University of Environmental and Life Sciences (Poland) as part of the Ph.D. research program “MISTRZ”, No N090/0005/21. Computations were carried out at Poznan Supercomputing and Networking Centre.

Authors' contributions

MF generated the idea for the study, performed statistical analysis, and co-wrote the manuscript draft. BN collected DNA samples and contributed to manuscript writing. MK performed some statistical analysis and contributed to manuscript writing. MM implemented the bioinformatic pipeline, performed the functional analysis, and co-wrote the manuscript draft. All authors read and approved the manuscript.

References

1000 Genomes Project, Mills R.E., Walter K., Stewart C., Handsaker R.E., Chen K., Alkan C., Abyzov A., Yoon S.C., Ye K., Cheetham R.K., Chinwalla A., Conrad D.F., Fu Y., Grubert F., Hajirasouliha I., Hormozdiari F., Iakoucheva L.M., Iqbal Z., Kang S., Kidd J.M., Konkel M.K., Korn J., Khurana E., Kural D., Lam H.Y.K., Leng J., Li R., Li Y.,

Lin C.-Y., Luo R., Mu X.J., Nemesh J., Peckham H.E., Rausch T., Scally A., Shi X., Stromberg M.P., Stütz A.M., Urban A.E., Walker J.A., Wu J., Zhang Y., Zhang Z.D., Batzer M.A., Ding L., Marth G.T., McVean G., Sebat J., Snyder M., Wang J., Ye K., Eichler E.E., Gerstein M.B., Hurles M.E., Lee C., McCarroll S.A., Korbel J.O. (2011). Mapping copy number variation by population-scale genome sequencing. Nature, 470: 59–65.

Abyzov A., Urban A.E., Snyder M., Gerstein M. (2011). CNVnator: An approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. Genome Res., 21: 974–984.

Acuna-Hidalgo R., Veltman J.A., Hoischen A. (2016). New insights into the generation and role of de novo mutations in health and disease. Genome Biol., 17: 241.

Andrews S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data.

Arias K.D., Pablo Gutiérrez J., Fernandez I., Menéndez-Arias N.A., Álvarez I., Goyache F. (2023). Segregation patterns and inheritance rate of copy number variations regions assessed in a Gochu Asturcelta pig pedigree. Gene, 854: 147111.

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

Ashburner M., Ball C.A., Blake J.A., Botstein D., Butler H., Cherry J.M., Davis A.P., Dolinski K., Dwight S.S., Eppig J.T., Harris M.A., Hill D.P., Issel-Tarver L., Kasarskis A., Lewis S., Matese J.C., Richardson J.E., Ringwald M., Rubin G.M., Sherlock G. (2000). Gene Ontology: tool for the unification of biology. Nat. Genet., 25: 25–29.

Bickhart D.M., Liu G.E. (2014). The challenges and importance of structural variation detection in livestock. Front. Genet., 5: 37.

Bolger A.M., Lohse M., Usadel B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics, 30: 2114–2120.

Boonin P., Klumsathian S., Iemwimangsa N., Sensorn I., Charoenyingwatana A., Chantratita W., Chareonsirisuthigul T. (2025). Detection of Genetic Variants in Thai Population by Trio-Based Whole-Genome Sequencing Study. Biology, 14: 301.

Borchers N., Reinsch N., Kalm E. (2004). The number of ribs and vertebrae in a Piétrain cross: variation, heritability and effects on performance traits. J. Anim. Breed. Genet., 121: 392–403.

Chan K., Gordenin D.A. (2015). Clusters of Multiple Mutations: Incidence and Molecular Mechanisms. Annu. Rev. Genet., 49: 243–267.

Chen R., Irwin D.M., Zhang Y.-P. (2012). Differences in Selection Drive Olfactory Receptor Genes in Different Directions in Dogs and Wolf. Mol. Biol. Evol., 29: 3475–3484.

Dallal G.E., Wilkinson L. (1986). An Analytic Approximation to the Distribution of Lilliefors's Test Statistic for Normality. Am. Stat., 40: 294–296.

Du H., Jolly A., Grochowski C.M., Yuan B., Dawood M., Jhangiani S.N., Li H., Muzny D., Fatih J.M., Coban-Akdemir Z., Carlin M.E., Scheuerle A.E., Witzl K., Posey J.E., Pendleton M., Harrington E., Juul S., Hastings P.J., Bi W., Gibbs R.A., Sedlazeck F.J., Lupski J. R., Carvalho C.M.B., Liu P. (2022). The multiple de novo copy number variant (MdnCNV) phenomenon presents with peri-zygotic DNA mutational signatures and multilocus pathogenic variation. Genome Med., 14: 122.

Esteve-Codina A., Paudel Y., Ferretti L., Raineri E., Megens H.-J., Silió L., Rodríguez M.C., Groenen M.A., Ramos-Onsins S.E., Pérez-Enciso M. (2013). Dissecting structural and nucleotide genome-wide variation in inbred Iberian pigs. BMC Genomics, 14: 148.

Ewels P., Magnusson M., Lundin S., Käller M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics, 32: 3047–3048.

Fadista J., Nygaard M., Holm L.-E., Thomsen B., Bendixen C. (2008). A Snapshot of CNVs in the Pig Genome. PLoS ONE, 3: e3916.

Feuk L., Carson A.R., Scherer S.W. (2006). Structural variation in the human genome. Nat. Rev. Genet., 7: 85–97.

Gao N., Chen Y., Liu X., Zhao Y., Zhu L., Liu A., Jiang W., Peng X., Zhang C., Tang Z., Li X. Chen Y. (2019). Weighted single-step GWAS identified candidate genes associated with semen traits in a Duroc boar population. BMC Genomics, 20: 797.

Ge S.X., Jung D., Yao R. (2020). ShinyGO: a graphical gene-set enrichment tool for animals and plants. Bioinforma. Oxf. Engl., 36: 2628–2629.

Geistlinger L., Da Silva V.H., Cesar A.S.M., Tizioto P.C., Waldron L., Zimmer R., Regitano L.C.D.A., Coutinho L.L. (2018). Widespread modulation of gene expression by copy number variation in skeletal muscle. Sci. Rep., 8: 1399.

Gilissen C., Hehir-Kwa J.Y., Thung D.T., Van De Vorst M., Van Bon B.W.M., Willemsen M.H., Kwint M., Janssen I.M., Hoischen A., Schenck A., Leach R., Klein R., Tearle R., Bo T., Pfundt R., Yntema H.G., De Vries B.B.A., Kleefstra T., Brunner H.G., Vissers L.E.L.M., Veltman J.A. (2014). Genome sequencing identifies major causes of severe intellectual disability. Nature, 511: 344–347.

Giuffra E., Törnsten A., Marklund S., Bongcam-Rudloff E., Chardon P., Kijas J.M.H., Anderson S.I., Archibald A.L., Andersson L. (2002). A large duplication associated with dominant white color in pigs originated by homologous recombination between LINE elements flanking KIT. Mamm. Genome, 13: 569–577.

Groenen M.A.M., Archibald A.L., Uenishi H., Tuggle C.K., Takeuchi Y., Rothschild M.F., Rogel-Gaillard C., Park C., Milan D., Megens H.-J., Li S., Larkin D.M., Kim H., Frantz L.A.F., Caccamo M., Ahn H., Aken B.L., Anselmo A., Anthon C., Auvil L., Badaoui B., Beattie C.W., Bendixen C., Berman D., Blecha F., Blomberg J., Bolund L., Bosse M., Botti S., Bujie Z., Bystrom M., Capitanu B., Carvalho-Silva D., Chardon P., Chen C., Cheng R., Choi S.-H., Chow W., Clark R.C., Clee C., Crooijmans R.P.M.A., Dawson H.D., Dehais P., De Sapio F., Dibbits B., Drou N., Du Z.-Q., Eversole K., Fadista J., Fairley S., Faraut T., Faulkner G.J., Fowler K.E., Fredholm M., Fritz E., Gilbert J.G.R., Giuffra E., Gorodkin J., Griffin D.K., Harrow J.L., Hayward A., Howe K., Hu Z.-L., Humphray S.J., Hunt T., Hornshøj H., Jeon J.-T., Jern P., Jones M., Jurka J., Kanamori H., Kapetanovic R., Kim J., Kim J.-H., Kim K.-W., Kim T.-H., Larson G., Lee K., Lee K.-T., Leggett R., Lewin H.A., Li Y., Liu W., Loveland J.E., Lu Y., Lunney J.K., Ma J., Madsen O., Mann K., Matthews L., McLaren S., Morozumi T., Murtaugh M.P., Narayan J., Truong Nguyen D., Ni P., Oh S.-J., Onteru S., Panitz F., Park E.-W., Park H.-S., Pascal G., Paudel Y., Perez-Enciso M., Ramirez-Gonzalez R., Reecy J.M., Rodriguez-Zas S., Rohrer G.A., Rund L., Sang Y., Schachtschneider K., Schraiber J.G., Schwartz J., Scobie L., Scott C., Searle S., Servin B., Southey B.R., Sperber G., Stadler P., Sweedler J.V., Tafer H., Thomsen B., Wali R., Wang J., Wang J., White S., Xu X.,

Yerle M., Zhang G., Zhang J., Zhang J., Zhao S., Rogers J., Churcher C., Schook L.B. (2012). Analyses of pig genomes provide insight into porcine demography and evolution. Nature, 491: 393–398.

Hamanaka K., Miyake N., Mizuguchi T., Miyatake S., Uchiyama Y., Tsuchida N., Sekiguchi F., Mitsuhashi S., Tsurusaki Y., Nakashima M., Saitsu H., Yamada K., Sakamoto M., Fukuda H., Ohori S., Saida K., Itai T., Azuma Y., Koshimizu E., Fujita A., Erturk B., Hiraki Y., Ch'ng G.-S., Kato M., Okamoto N., Takata A., Matsumoto N. (2022). Large-scale discovery of novel neurodevelopmental disorder-related genes through a unified analysis of single-nucleotide and copy number variants. Genome Med., 14: 40.

Hehir-Kwa J.Y., Rodriguez-Santiago B., Vissers L.E., De Leeuw N., Pfundt R., Buitelaar J.K., Perez-Jurado L.A., Veltman J.A. (2011). De novo copy number variants associated with intellectual disability have a paternal origin and age bias. J. Med. Genet., 48: 776–778.

Henrichsen C.N., Vinckenbosch N., Zöllner S., Chaignat E., Pradervand S., Schütz F., Ruedi M., Kaessmann H., Reymond A. (2009). Segmental copy number variation shapes tissue transcriptomes. Nat. Genet., 41: 424–429.

Höps W., Rausch T., Jendrusch M., Human Genome Structural Variation Consortium (HGSVC), Ashraf H., Audano P.A., Austine O., Basile A.O., Beck C.R., Jan Bonder M., Byrska-Bishop M., Chaisson M.J.P., Chong Z., Corvelo A., Devine S.E., Ebert P., Ebler J., Eichler E.E., Gerstein M.B., Hallast P., Harvey W.T., Hasenfeld P., Hastie A.R., Henglin M., Hoekzema K., Hsieh P., Hunt S., Konkel M.K., Kordosky J., Lansdorp P.M., Lee C., Lee W.-P., Lewis A.P., Li C., Lin J., Loftus M., Logsdon G.A., Marschall T., Mills R.E., Mostovoy Y., Munson K.M., Narzisi G., Pang A., Porubsky D., Prodanov T., Rodriguez-Martin B., Shi X., Surapaneni L., Talkowski M.E., Yilmaz F., Yoo D., Zhou W., Zody M.C., Korbel J.O., Sedlazeck F.J. (2024). Impact and characterization of serial structural variations across humans and great apes. Nat. Commun., 15: 8007.

Hu Z.-L., Park C.A., Reecy J.M. (2022). Bringing the Animal QTLdb and CorrDB into the future: meeting new challenges and providing updated services. Nucleic Acids Res., 50: 956–961.

Jang J., Kim B., Jhang S.Y., Ahn B., Kang M., Park C., Cho E.S., Kim Y.-S., Park W., Kim H. (2023). Population differentiated copy number variation between Eurasian wild boar and domesticated pig populations. Sci. Rep., 13: 1115.

Jiang J., Wang J., Wang H., Zhang Y., Kang H., Feng X., Wang J., Yin Z., Bao W., Zhang Q., Liu J.-F. (2014). Global copy number analyses by next generation sequencing provide insight into pig genome variation. BMC Genomics, 15: 593.

Kanehisa M., Furumichi M., Tanabe M., Sato Y., Morishima K. (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res., 45: D353–D361.

Keel B.N., Nonneman D.J., Lindholm-Perry A.K., Oliver W.T., Rohrer G.A. (2019). A Survey of Copy Number Variation in the Porcine Genome Detected From Whole-Genome Sequence. Front. Genet., 10: 737.

Kirov G., Pocklington A.J., Holmans P., Ivanov D., Ikeda M., Ruderfer D., Moran J., Chambert K., Toncheva D., Georgieva L., Grozeva D., Fjodorova M., Wollerton R., Rees E., Nikolov I., Van De Lagemaat L.N., Bayés À., Fernandez E., Olason P.I., Böttcher Y.,

Komiyama N.H., Collins M.O., Choudhary J., Stefansson K., Stefansson H., Grant S.G.N., Purcell S., Sklar P., O'Donovan M.C., Owen M.J. (2012). De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. Mol. Psychiatry, 17: 142–153.

Krepischi A.C., Achatz M.I.W., Santos E.M., Costa S.S., Lisboa B.C., Brentani H., Santos T.M., Gonçalves A., Nóbrega A.F., Pearson P.L., Vianna-Morgante A.M., Carraro D.M., Brentani R.R., Rosenberg C. (2012). Germline DNA copy number variation in familial and early-onset breast cancer. Breast Cancer Res., 14: R24.

Li H., Durbin R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics, 25: 1754–1760.

Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R., 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. Bioinforma. Oxf. Engl., 25: 2078–2079.

Li Y., Mei S., Zhang X., Peng X., Liu G., Tao H., Wu H., Jiang S., Xiong Y., Li F. (2012). Identification of genome-wide copy number variations among diverse pig breeds by array CGH. BMC Genomics, 13: 725.

Lin, Y.-L., Gokcumen O. (2019). Fine-Scale Characterization of Genomic Structural Variation in the Human Genome Reveals Adaptive and Biomedically Relevant Hotspots. Genome Biol. Evol., 11: 1136–1151.

Liu G.E., Bickhart D.M. (2012). Copy number variation in the cattle genome. Funct. Integr. Genomics, 12: 609–624.

Liu Y., Yang X., Jing X., He X., Wang L., Liu Y., Liu D. (2017). Transcriptomics Analysis on Excellent Meat Quality Traits of Skeletal Muscles of the Chinese Indigenous Min Pig Compared with the Large White Breed. Int. J. Mol. Sci., 19: 21.

Long Y., Su Y., Ai H., Zhang Z., Yang B., Ruan G., Xiao S., Liao X., Ren J., Huang L., Ding N. (2016). A genome-wide association study of copy number variations with umbilical hernia in swine. Anim. Genet., 47: 298–305.

McCarroll S.A., Kuruvilla F.G., Korn J.M., Cawley S., Nemesh J., Wysoker A., Shapero M.H., de Bakker P.I.W., Maller J.B., Kirby A., Elliott A.L., Parkin M., Hubbell E., Webster T., Mei R., Veitch J., Collins P.J., Handsaker R., Lincoln S., Nizzari M., Blume J., Jones K.W., Rava R., Daly M.J., Gabriel S.B., Altshuler D. (2008). Integrated detection and population-genetic analysis of SNPs and copy number variation. Nat. Genet., 40: 1166–1174.

McLaren W., Gil L., Hunt S.E., Riat H.S., Ritchie G.R.S., Thormann A., Flicek P., Cunningham F. (2016). The Ensembl Variant Effect Predictor. Genome Biol., 17: 122.

Mielczarek M., Frąszczak M., Zielak-Steciwko A.E., Nowak B., Hofman B., Pierścińska J., Kruszyński W., Szyda J. (2023). An effect of large-scale deletions and duplications on transcript expression. Funct. Integr. Genomics, 23: 19.

Moreno-Estrada A., Casals F., Ramirez-Soriano A., Oliva B., Calafell F., Bertranpetit J., Bosch E. (2007). Signatures of Selection in the Human Olfactory Receptor OR5I1 Gene. Mol. Biol. Evol., 25: 144–154.

van Ommen G.-J.B. (2005). Frequency of new copy number variation in humans. Nat. Genet., 37: 333–334.

Pabst R. (2020). The pig as a model for immunology research. Cell Tissue Res., 380: 287–304.

Paudel Y., Madsen O., Megens H.-J., Frantz L.A., Bosse M., Bastiaansen J.W., Crooijmans R.P., Groenen M.A. (2013). Evolutionary dynamics of copy number variation in pig genomes in the context of adaptation and domestication. BMC Genomics, 14: 449.

Paudel Y., Madsen O., Megens H.-J., Frantz L.A.F., Bosse M., Crooijmans R.P.M.A., Groenen M.A.M. (2015). Copy number variation in the speciation of pigs: a possible prominent role for olfactory receptors. BMC Genomics, 16: 330.

Pereira S.S.S., Pinto I.P., Santos V.C.D.P., Silva R.C., Costa E.O.A., Cruz A.S.D., Cruz A.D.D., Silva C.C.D., Minasi L.B. (2024). Analysis of parental origin of de novo pathogenic CNVs in patients with intellectual disability. Genet. Mol. Biol., 47: e20230313.

Pös O., Radvanszky J., Buglyó G., Pös Z., Rusnakova D., Nagy B., Szemes T. (2021). DNA copy number variation: Main characteristics, evolutionary significance, and pathological aspects. Biomed. J., 44: 548–559.

Qian R., Xie F., Zhang W., Kong J., Zhou X., Wang C., Li X. (2023). Genome-wide detection of CNV regions between Anqing six-end-white and Duroc pigs. Mol. Cytogenet., 16: 12.

R Core Team (2022). R: A language and environment for statistical computing.

Ramayo-Caldas Y., Crespo-Piazuelo D., Morata J., González-Rodríguez O., Sebastià C., Castello A., Dalmau A., Ramos-Onsins S., Alexiou K.G., Folch J.M., Quintanilla R., Ballester M. (2023). Copy Number Variation on ABCC2-DNMBP Loci Affects the Diversity and Composition of the Fecal Microbiota in Pigs. Microbiol. Spectr., 11: e05271-22.

Revay T., Quach A.T., Maignel L., Sullivan B., King W.A. (2015). Copy number variations in high and low fertility breeding boars. BMC Genomics, 16: 280.

Revilla M., Puig-Oliveras A., Castelló A., Crespo-Piazuelo D., Paludo E., Fernández A.I., Ballester M., Folch J.M. (2017). A global analysis of CNVs in swine using whole genome sequence data and association analysis with fatty acid composition and growth traits. PLOS ONE, 12: e0177014.

Rohrer G.A., Nonneman D.J. (2017). Genetic analysis of teat number in pigs reveals some developmental pathways independent of vertebra number and several loci which only affect a specific side. Genet. Sel. Evol., 49: 4.

Sanders S.J., He X., Willsey A.J., Ercan-Sencicek A.G., Samocha K.E., Cicek A.E., Murtha M.T., Bal V.H., Bishop S.L., Dong S., Goldberg A.P., Jinlu C., Keaney J.F., Klei L., Mandell J.D., Moreno-De-Luca D., Poultney C.S., Robinson E.B., Smith L., Solli-Nowlan T., Su M.Y., Teran N.A., Walker M.F., Werling D.M., Beaudet A.L., Cantor R.M., Fombonne E., Geschwind D.H., Grice D.E., Lord C., Lowe J.K., Mane S.M., Martin D.M., Morrow E.M., Talkowski M.E., Sutcliffe J.S., Walsh C.A., Yu T.W., Ledbetter D.H., Martin C.L., Cook E.H., Buxbaum J.D., Daly M.J., Devlin B., Roeder K., State M.W. (2015). Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. Neuron, 87: 1215–1233.

Sironen A., Uimari P., Venhoranta H., Andersson M., Vilkki J. (2011). An exonic insertion within Tex14 gene causes spermatogenic arrest in pigs. BMC Genomics, 12: 591.

Skinner B.M., Sargent C.A., Churcher C., Hunt T., Herrero J., Loveland J.E., Dunn M., Louzada S., Fu B., Chow W., Gilbert J., Austin-Guest S., Beal K., Carvalho-Silva D.,

Cheng W., Gordon D., Grafham D., Hardy M., Harley J., Hauser H., Howden P., Howe K., Lachani K., Ellis P.J.I., Kelly D., Kerry G., Kerwin J., Ng B.L., Threadgold G., Wileman T., Wood J.M.D., Yang F., Harrow J., Affara N.A., Tyler-Smith C. (2016). The pig X and Y Chromosomes: structure, sequence, and evolution. Genome Res., 26: 130–139.

Soto D.C., Uribe-Salazar J.M., Shew C.J., Sekar A., McGinty S.P., Dennis M.Y. (2023). Genomic structural variation: A complex but important driver of human evolution. Am. J. Biol. Anthropol., 181: 118–144.

Stadler Z.K., Esposito D., Shah S., Vijai J., Yamrom B., Levy D., Lee Y., Kendall J., Leotta A., Ronemus M., Hansen N., Sarrel K., Rau-Murthy R., Schrader K., Kauff N., Klein R.J., Lipkin S.M., Murali R., Robson M., Sheinfeld J., Feldman D., Bosl G., Norton L., Wigler M., Offit K. (2012). Rare De Novo Germline Copy-Number Variation in Testicular Cancer. Am. J. Hum. Genet., 91: 379–383.

Steensma M.J., Lee Y.L., Bouwman A.C., Pita Barros C., Derks M.F.L., Bink M.C.A.M., Harlizius B., Huisman A.E., Crooijmans R.P.M.A., Groenen M.A.M., Mulder H.A., Rochus C.M. (2023). Identification and characterisation of de novo germline structural variants in two commercial pig lines using trio-based whole genome sequencing. BMC Genomics, 24: 208.

Svärd A.A., Viberg E., Von Platen I., Jönsson I., Lundgren M., on behalf of the DiPiS study group, Ramelius A., Andersson C., Bennet R., Ask M., Bremer J., Brundin C., Cilio C., Elding Larsson H., Hansson C., Hansson G., Ivarsson S., Jonsdottir B., Jonsson I., Lindberg B., Lernmark B., Lernmark Å., Melin J., Lundgren M., Carlsson A., Cedervall E., Jönsson B., Larsson K., Neiderud J. (2025). Feasibility of extracting usable DNA from blood samples stored up to 21 years in the DiPiS study. Sci. Rep., 15: 25637.

Tang Y.-C., Amon A. (2013). Gene copy-number alterations: a cost-benefit analysis. Cell, 152: 394–405.

The Bovine Genome Sequencing and Analysis Consortium, Elsik C.G., Tellam R.L., Worley K.C., Gibbs R.A., Muzny D.M., Weinstock G.M., Adelson D.L., Eichler E.E., Elnitski L., Guigó R., Hamernik D.L., Kappes S.M., Lewin H.A., Lynn D.J., Nicholas F.W., Reymond A., Rijnkels M., Skow L.C., Zdobnov E.M., Schook L., Womack J., Alioto T., Antonarakis S.E., Astashyn A., Chapple C.E., Chen H.-C., Chrast J., Câmara F., Ermolaeva O., Henrichsen C.N., Hlavina W., Kapustin Y., Kiryutin B., Kitts P., Kokocinski F., Landrum M., Maglott D., Pruitt K., Sapojnikov V., Searle S.M., Solovyev V., Souvorov A., Ucla C., Wyss C., Anzola J.M., Gerlach D., Elhaik E., Graur D., Reese J.T., Edgar R.C., McEwan J.C., Payne G.M., Raison J.M., Junier T., Kriventseva E.V., Eyras E., Plass M., Donthu R., Larkin D.M., Reecy J., Yang M.Q., Chen L., Cheng Z., Chitko-McKown C.G., Liu G.E., Matukumalli L.K., Song J., Zhu B., Bradley D.G., Brinkman F.S.L., Lau L.P.L., Whiteside M.D., Walker A., Wheeler T.T., Casey T., German J.B., Lemay D.G., Maqbool N.J., Molenaar A.J., Seo S., Stothard P., Baldwin C.L., Baxter R., Brinkmeyer-Langford C.L., Brown W.C., Childers C.P., Connelley T., Ellis S.A., Fritz K., Glass E.J., Herzig C.T.A., Iivanainen A., Lahmers K.K., Bennett A.K., Dickens C.M., Gilbert J.G.R., Hagen D.E., Salih H., Aerts J., Caetano A.R., Dalrymple B., Garcia J.F., Gill C.A., Hiendleder S.G., Memili E., Spurlock D., Williams J.L., Alexander L., Brownstein M.J., Guan L., Holt R.A.,

Jones S.J.M., Marra M.A., Moore R., Moore S.S., Roberts A., Taniguchi M., Waterman R.C., Chacko J., Chandrabose M.M., Cree A., Dao M.D., Dinh H.H., Gabisi R.A., Hines S., Hume J., Jhangiani S.N., Joshi V., Kovar C.L., Lewis L.R., Liu Y., Lopez J., Morgan M.B., Nguyen N.B., Okwuonu G.O., Ruiz S.J., Santibanez J., Wright R.A., Buhay C., Ding Y., Dugan-Rocha S., Herdandez J., Holder M., Sabo A., Egan A., Goodell J., Wilczek-Boney K., Fowler G.R., Hitchens M.E., Lozado R.J., Moen C., Steffen D., Warren J.T., Zhang J., Chiu R., Schein J.E., Durbin K.J., Havlak P., Jiang H., Liu Y., Qin X., Ren Y., Shen Y., Song H., Bell S.N., Davis C., Johnson A.J., Lee S., Nazareth L.V., Patel B.M., Pu L.-L., Vattathil S., Williams R.L., Curry S., Hamilton C., Sodergren E., Wheeler D.A., Barris W., Bennett G.L., Eggen A., Green R.D., Harhay G.P., Hobbs M., Jann O., Keele J.W., Kent M.P., Lien S., McKay S.D., McWilliam S., Ratnakumar A., Schnabel R.D., Smith T., Snelling W.M., Sonstegard T.S., Stone R.T., Sugimoto Y., Takasuga A., Taylor J.F., Van Tassell C.P., MacNeil M.D., Abatepaulo A.R.R., Abbey C.A., Ahola V., Almeida I.G., Amadio A.F., Anatriello E., Bahadue S.M., Biase F.H., Boldt C.R., Carroll J.A., Carvalho W.A., Cervelatti E.P., Chacko E., Chapin J.E., Cheng Y., Choi J., Colley A.J., De Campos T.A., De Donato M., Santos I.K.F.D.M., De Oliveira C.J.F., Deobald H., Devinoy E., Donohue K.E., Dovc P., Eberlein A., Fitzsimmons C.J., Franzin A.M., Garcia G.R., Genini S., Gladney C.J., Grant J.R., Greaser M.L., Green J.A., Hadsell D.L., Hakimov H.A., Halgren R., Harrow J.L., Hart E.A., Hastings N., Hernandez M., Hu Z.-L., Ingham A., Iso-Touru T., Jamis C., Jensen K., Kapetis D., Kerr T., Khalil S.S., Khatib H., Kolbehdari D., Kumar C.G., Kumar D., Leach R., Lee J.C.-M., Li C., Logan K.M., Malinverni R., Marques E., Martin W.F., Martins N.F., Maruyama S.R., Mazza R., McLean K.L., Medrano J.F., Moreno B.T., Moré D.D., Muntean C.T., Nandakumar H.P., Nogueira M.F.G., Olsaker I., Pant S.D., Panzitta F., Pastor R.C.P., Poli M.A., Poslusny N., Rachagani S., Ranganathan S., Razpet A., Riggs P K., Rincon G., Rodriguez-Osorio N., Rodriguez-Zas S.L., Romero N.E., Rosenwald A., Sando L., Schmutz S.M., Shen L., Sherman L., Southey B.R., Lutzow Y.S., Sweedler J.V., Tammen I., Telugu B.P.V.L., Urbanski J.M., Utsunomiya Y.T., Verschoor C.P., Waardenberg A.J., Wang Z., Ward R., Weikard R., Welsh T.H., White S. N., Wilming L.G., Wunderlich K.R., Yang J., Zhao F.-Q. (2009). The Genome Sequence of Taurine Cattle: A Window to Ruminant Biology and Evolution. Science, 324: 522–528.

The Gene Ontology Consortium, Aleksander S.A., Balhoff J., Carbon S., Cherry J.M., Drabkin H.J., Ebert D., Feuermann M., Gaudet P., Harris N.L., Hill D.P., Lee R., Mi H., Moxon S., Mungall C.J., Muruganugan A., Mushayahama T., Sternberg P.W., Thomas P.D., Van Auken K., Ramsey J., Siegele D.A., Chisholm R.L., Fey P., Aspromonte M.C., Nugnes M.V., Quaglia F., Tosatto S., Giglio M., Nadendla S., Antonazzo G., Attrill H., Dos Santos G., Marygold S., Strelets V., Tabone C.J., Thurmond J., Zhou P., Ahmed S.H., Asanitthong P., Luna Buitrago D., Erdol M.N., Gage M.C., Ali Kadhum M., Li K.Y.C., Long M., Michalak A., Pesala A., Pritazahra A., Saverimuttu S.C.C., Su R., Thurlow K.E., Lovering R.C., Logie C., Oliferenko S., Blake J., Christie K., Corbani L., Dolan M.E., Drabkin H.J., Hill D.P., Ni L., Sitnikov D., Smith C., Cuzick A., Seager J., Cooper L., Elser J., Jaiswal P., Gupta P., Jaiswal P., Naithani S., Lera-Ramirez M., Rutherford K., Wood V., De Pons J.L., Dwinell M.R., Hayman G.T., Kaldunski M.L.,

Kwitek A.E., Lauledеrkind S.J.F., Tutaj M.A., Vedi M., Wang S.-J., D'Eustachio P., Aimo L., Axelsen K., Bridge A., Hyka-Nouspikel N., Morgat A., Aleksander S.A., Cherry J.M., Engel S.R., Karra K., Miyasato S.R., Nash R.S., Skrzypek M.S., Weng S., Wong E.D., Bakker E., Berardini T.Z., Reiser L., Auchincloss A., Axelsen K., Argoud-Puy G., Blatter M.-C., Boutet E., Breuza L., Bridge A., Casals-Casas C., Coudert E., Estreicher A., Livia Famiglietti M., Feuermann M., Gos A., Gruaz-Gumowski N., Hulo C., Hyka-Nouspikel N., Jungo F., Le Mercier P., Lieberherr D., Masson P., Morgat A., Pedruzzi I., Pourcel L., Poux S., Rivoire C., Sundaram S., Bateman A., Bowler-Barnett E., Bye-A-Jee H., Denny P., Ignatchenko A., Ishtiaq R., Lock A., Lussi Y., Magrane M., Martin M.J., Orchard S., Raposo P., Speretta E., Tyagi N., Warner K., Zaru R., Diehl A. D., Lee R., Chan J., Diamantakis S., Raciti D., Zarowiecki M., Fisher M., James-Zorn C., Ponferrada V., Zorn A., Ramachandran S., Ruzicka L., Westerfield M., (2023). The Gene Ontology knowledgebase in 2023. Genetics, 224: iyad031.

Tizaoui K. (2018). De novo vs. inherited copy number variations in multiple sclerosis susceptibility. Cell. Mol. Immunol., 15: 812–814.

Toft M., Ross O. A. (2010). Copy number variation in Parkinson's disease. Genome Med., 2: 62.

Wadhawan I., Hai Y., Foyouzi Yousefi N., Guo X., Graham J.M., Rosenfeld J.A. (2020). De novo copy number variants and parental age: Is there an association? Eur. J. Med. Genet., 63: 103829.

Walters E.M., Prather R.S. (2013). Advancing swine models for human health and diseases. Mo. Med., 110: 212–215.

Wang J., Jiang J., Fu W., Jiang L., Ding X., Liu J.-F., Zhang Q. (2012). A genome-wide detection of copy number variations using SNP genotyping arrays in swine. BMC Genomics, 13: 273.

Wang J., Jiang J., Wang H., Kang H., Zhang Q., Liu J.-F. (2014). Enhancing Genome-Wide Copy Number Variation Identification by High Density Array CGH Using Diverse Resources of Pig Breeds. PLoS ONE, 9: e87571.

Wang J., Wang H., Jiang J., Kang H., Feng X., Zhang Q., Liu, J.-F. (2013). Identification of Genome-Wide Copy Number Variations among Diverse Pig Breeds Using SNP Genotyping Arrays. PLoS ONE, 8: e68683.

Wang L., Xu L., Liu X., Zhang T., Li N., Hay E.H., Zhang Y., Yan H., Zhao K., Liu G.E., Zhang L., Wang, L. (2015). Copy number variation-based genome wide association study reveals additional variants contributing to meat quality in Swine. Sci. Rep., 5: 12535.

Wang Z., Chen Q., Liao R., Zhang Z., Zhang X., Liu X., Zhu M., Zhang W., Xue M., Yang H., Zheng Y., Wang Q., Pan Y. (2017). Genome-wide genetic variation discovery in Chinese Taihu pig breeds using next generation sequencing. Anim. Genet., 48: 38–47.

Wen Q., Wang X., Zhang H., Liu X., Xu Z. (2022). Distribution and transmission of copy number variations of uncertain significance in 105 trios. Mol. Genet. Genomic Med., 10: e2030.

Xu C., Zhang W., Jiang Y., Zhou M., Liu L., Su S., Li X., Wang C. (2023). Genome-Wide Detection and Analysis of Copy Number Variation in Anhui Indigenous and Western Commercial Pig Breeds Using Porcine 80K SNP BeadChip. Genes, 14: 654.

Yang L., Li X., Zhuang Z., Zhou S., Wu J., Xu C., Ruan D., Qiu Y., Zhao H., Zheng E., Cai G., Wu Z., Yang J. (2023). Genome-Wide Association Study Identifies the Crucial Candidate Genes for Teat Number in Crossbred Commercial Pigs. Animals, 13: 1880.

Ye K., Schulz M.H., Long Q., Apweiler R., Ning Z. (2009). Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics, 25: 2865–2871.

Yin H., Ma H., Hussain S., Zhang H., Xie X., Jiang L., Jiang X., Iqbal F., Bukhari I., Jiang H., Ali A., Zhong L., Li T., Fan S., Zhang B., Gao J., Li Y., Nazish J., Khan T., Khan M., Zubair M., Hao Q., Fang H., Huang J., Huleihel M., Sha J., Pandita T.K., Zhang Y., Shi Q. (2019). A homozygous FANCM frameshift pathogenic variant causes male infertility. Genet. Med., 21: 62–70.

Zhang C., Yang H., Xu Q., Liu M., Chao X., Chen J., Zhou B. (2024). Genome-Wide Analysis Reveals Copy Number Variant Gene TGFBR3 Regulates Pig Back Fat Deposition. Animals, 14: 2657.

Zhang C., Zhao J., Guo Y., Xu Q., Liu M., Cheng M., Chao X., Schinckel A.P., Zhou B. (2022). Genome-Wide Detection of Copy Number Variations and Evaluation of Candidate Copy Number Polymorphism Genes Associated With Complex Traits of Pigs. Front. Vet. Sci., 9: 909039.

Zheng X., Zhao P., Yang K., Ning C., Wang H., Zhou L., Liu J. (2020). CNV analysis of Meishan pig by next-generation sequencing and effects of AHR gene CNV on pig reproductive traits. J. Anim. Sci. Biotechnol., 11: 42.

Zhuang Z., Li K., Yang K., Gao G., Li Z., Zhu X., Zhao Y. (2024). Genome-Wide Association Study Reveals Novel Candidate Genes Influencing Semen Traits in Landrace Pigs. Animals, 14: 1839.

Received: 18 II 2025

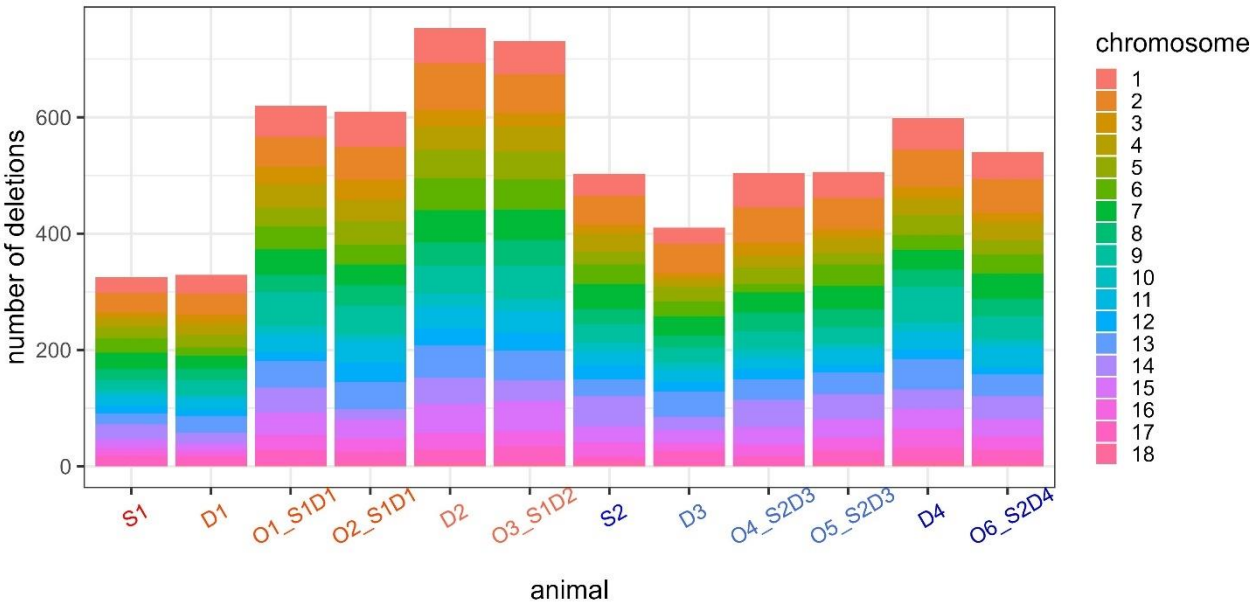
Accepted: 14 X 2025

Additional files

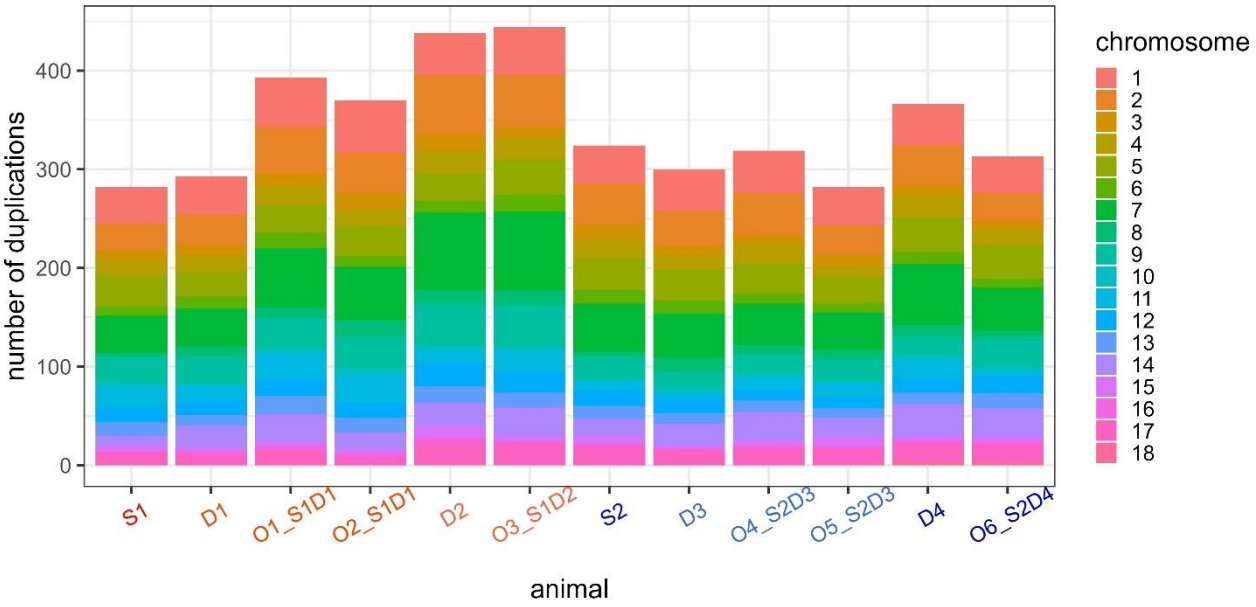
Sample name	Sample ID	% of mapped reads	% of properly mapped reads	Average genome coverage after alignment
Family 1				
Sire 1	S1	98.42	95.36	13
Dam 1	D1	98.53	96	11
Dam 2	D2	98.11	95.37	15
Offspring 1	O1_S1D1	98.41	95.58	16
Offspring 2	O2_S1D1	98.36	95.36	15

Offspring 3	O3_S1D2	98.3	95.31	14
Family 2				
Sire 2	S2	98.48	95.47	10
Dam 3	D3	98.6	95.82	12
Dam 4	D4	98.57	95.8	19
Offspring 4	O4_S2D3	98.42	95.54	12
Offspring 5	O5_S2D3	98.55	95.53	12
Offspring 6	O6_S2D4	98.37	95.42	18

Additional file 1. A summary of basic alignment statistics, including the percentage of aligned reads and properly paired reads, as well as the average genome coverage after alignment.



Additional file 2. Number of deletions identified in each animal.



Additional file 3. Number of duplications identified in each animal.

chr	chr length	cnv counts	cnv length	coverage (%)	max size	mean length	min size
deletions							
1	274330532	560	824000	0.300	41000	3951.07	600
2	151935994	667	1141400	0.751	70000	7324.74	600
3	132848913	238	404000	0.304	57200	4168.07	600
4	130910915	344	638400	0.488	114600	10021.51	600
5	104526007	364	478200	0.457	34200	4053.30	600
6	170843587	387	1076600	0.630	119600	8284.24	600
7	121844099	466	668200	0.548	99600	6597.85	600
8	138966237	355	667200	0.48	56400	6371.27	600
9	139512083	478	1343600	0.963	183400	8183.26	600
10	69359453	157	271800	0.392	65000	3526.11	600
11	79169978	340	449600	0.568	59200	5214.12	600
12	61602749	233	547200	0.888	195000	11570.82	600
13	208334590	481	1159200	0.556	140200	9968.40	600
14	141755446	423	718600	0.507	100200	6244.44	600

15	140412725	377	609400	0.434	53000	5449.87	600
16	79944280	262	329400	0.412	53400	3223.66	600
17	63494081	248	417600	0.658	189600	6933.06	600
18	55982971	49	131800	0.235	44800	5779.59	600

chr	chr length	cnv counts	cnv length	coverage (%)	max size	mean length	min size
duplications							
1	274330532	510	1423600	0.519	163600	13505.88	2200
2	151935994	482	1779800	1.171	174200	23009.96	1400
3	132848913	138	421400	0.317	52000	15779.71	2000
4	130910915	217	694400	0.530	193800	16544.70	1200
5	104526007	367	1103400	1.056	217200	20102.45	2600
6	170843587	144	582200	0.341	80200	17461.11	2800
7	121844099	629	2127800	1.746	198400	19948.17	2000
8	138966237	125	1001800	0.721	561400	64956.8	2000
9	139512083	333	1355600	0.972	239000	22518.32	1800
10	69359453	49	129600	0.187	27400	14934.69	2800
11	79169978	189	494400	0.624	69600	11987.30	2200
12	61602749	190	538600	0.874	49400	16998.95	1600
13	208334590	163	713000	0.342	157600	30195.09	2000
14	141755446	279	732600	0.517	54000	13179.93	2200
15	140412725	70	382400	0.272	91400	25354.29	2000
16	79944280	32	199800	0.250	52800	13981.25	2800
17	63494081	196	633600	0.998	117200	18442.86	1600
18	55982971	11	92200	0.165	54200	11054.55	4600

Additional file 4. Basic statistics for length and counts of all copy number variants (CNVs), depending on the chromosome.

chr	chr length	cnv counts	cnv length	coverage (%)	max size	mean length	min size
***de novo* deletions**							
1	274330532	55	222000	0.081	36000	4036.36	600
2	151935994	34	186000	0.122	70000	5470.58	600
3	132848913	33	74600	0.056	13400	2260.60	600
4	130910915	23	63400	0.048	11800	2756.52	600
5	104526007	22	79400	0.075	24600	3609.09	600
6	170843587	30	97200	0.056	17600	3240.00	600
7	121844099	24	60800	0.049	11000	2533.33	600
8	138966237	16	71000	0.051	36400	4437.50	600
9	139512083	31	179600	0.128	41000	5793.54	600
10	69359453	11	18400	0.026	3800	1672.72	600
11	79169978	13	37200	0.046	9200	2861.53	600
12	61602749	12	16800	0.027	4200	1400.00	600
13	208334590	33	117800	0.056	28200	3569.69	600
14	141755446	27	61600	0.043	12400	2281.48	600
15	140412725	18	78800	0.056	29000	4377.77	600
16	79944280	18	38600	0.048	10200	2144.44	600
17	63494081	14	61600	0.097	28000	4400.00	600
18	55982971	8	23200	0.041	13000	2900.00	600

chr	chr length	cnv counts	cnv length	coverage (%)	max size	mean length	min size
***de novo* duplications**							
1	274330532	17	233200	0.085	39200	13717.65	2200

2	151935994	14	118400	0.078	27800	8457.14	1400
3	132848913	2	10800	0.008	6200	5400.00	4600
4	130910915	4	48200	0.037	33800	12050.00	3000
5	104526007	1	14800	0.014	14800	14800.00	14800
6	170843587	5	55600	0.033	29000	11120.00	3200
7	121844099	12	193800	0.159	70000	16150.00	3200
8	138966237	6	52000	0.037	15000	8666.67	3800
9	139512083	8	60800	0.044	16400	7600.00	4400
10	69359453	2	31000	0.045	17600	15500.00	13400
11	79169978	4	40400	0.051	14000	10100.00	5000
12	61602749	2	19200	0.031	11400	9600.00	7800
13	208334590	5	47400	0.023	15600	9480.00	5400
14	141755446	3	26200	0.018	16000	8733.33	4000
15	140412725	4	31600	0.023	11600	7900.00	3200
16	79944280	3	43000	0.054	19600	14333.33	10000
17	63494081	2	40400	0.064	32200	20200.00	8200
18	55982971	2	14600	0.026	8800	7300.00	5800

Additional file 5. Basic statistics for length and counts of *de novo* copy number variants (CNVs), depending on the chromosome.

	O1_S1D1	**O2_S1D1**	**O3_S1D2**	**O4_S2D3**	**O5_S2D3**	**O6_S2D4**
P -value	$3.76 \cdot 10^{-5}$	$1.26 \cdot 10^{-6}$	$6.49 \cdot 10^{-7}$	$0.13 \cdot 10^{-1}$	$0.17 \cdot 10^{-3}$	$8.18 \cdot 10^{-7}$

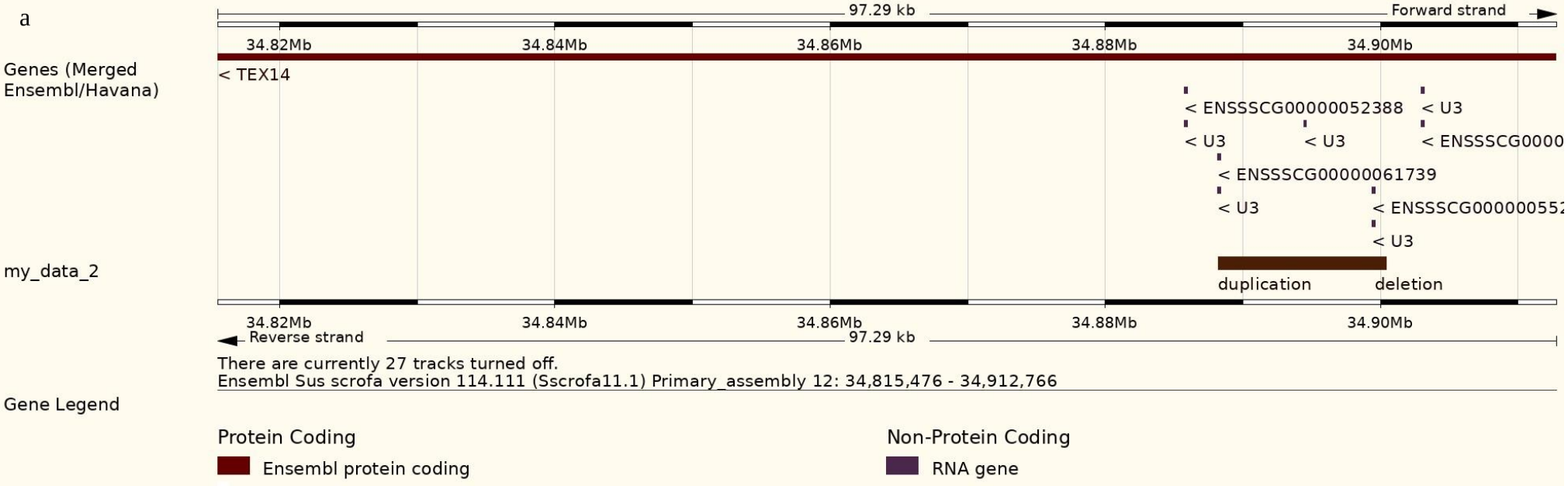
Additional file 6. Mann-Whitney U test showing that *de novo* deletions are shorter than inherited.

min	**1st qu.**	**median**	**mean**	**3rd qu**	**max**	**sd**	**number**
de novo deletions							
600	800	1400	2835	2850	23000	3890.368	80
600	800	1400	3472.5	3450	36000	5684.466	80

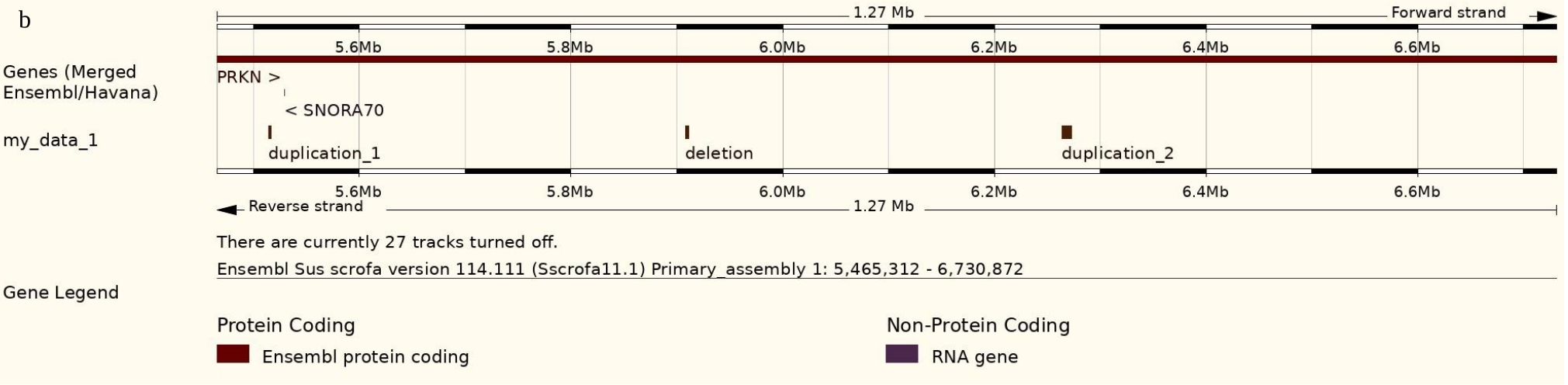
600	800	2600	3935.714	4050	36400	5695.179	56
600	800	1200	4515.556	3800	29000	7038.012	45
600	800	1600	2972.549	3000	28000	4424.662	51
600	800	1800	3710.909	2800	70000	7991.858	110
inherited deletions							
600	1800	4200	11637.64	9600	194600	23621.46	457
600	1600	3800	11657.99	9550	195000	25709.8	438
600	1800	4200	10456.46	9600	187800	18723.54	379
600	2000	4400	10830.46	9200	195000	22087.34	394
600	2200	4600	12348.79	11350	194600	24186.72	414
600	1600	3500	10816.85	8600	195000	24060.52	534
de novo duplications							
1800	5200	7600	13147.83	17100	44200	12181.09	23
3200	7000	8200	11711.11	13900	39200	7987.169	27
5400	7150	10500	10866.67	14000	17600	4865.662	6
5400	8200	8800	8800	9850	11600	2097.618	6
2600	4900	7800	8200	10600	16400	4170.224	14
1400	3500	6400	11500	12750	70000	15423.77	20
inherited duplications							
2200	12000	23600	41910.22	50400	563200	57319.37	225
2200	12000	23200	42234.06	49600	563200	60354.32	229
1200	13700	26000	46173.26	54500	563200	63677.16	187
2200	13800	27000	46524.85	55400	563200	63077.09	165
1200	13850	26200	47055.17	53300	563200	65834.85	174
2200	11600	23000	39696.85	44950	563200	56621.46	254

Additional file 7. Basic statistics for the length of *de novo* and inherited copy number variants (CNVs) in offspring.

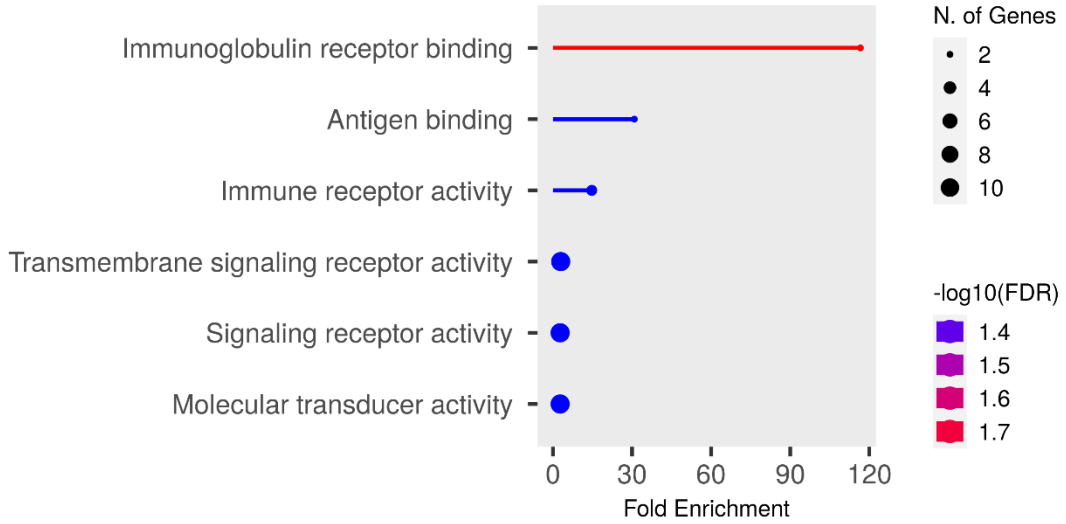
a


97.29 kb
Forward strand
34.82Mb
34.84Mb
34.86Mb
34.88Mb
34.90Mb
Genes (Merged Ensembl/Havana)
< TEX14
< ENSSSCG00000052388
< U3
< U3
< U3
< ENSSSCG0000
< ENSSSCG00000061739
< U3
< ENSSSCG000000552
< U3
my_data_2
duplication
deletion
34.82Mb
34.84Mb
34.86Mb
34.88Mb
34.90Mb
Reverse strand
97.29 kb
There are currently 27 tracks turned off.
Ensembl Sus scrofa version 114.111 (Sscrofa11.1) Primary_assembly 12: 34,815,476 - 34,912,766
Gene Legend
Protein Coding
Ensembl protein coding
Non-Protein Coding
RNA gene

b


1.27 Mb
Forward strand
5.6Mb
5.8Mb
6.0Mb
6.2Mb
6.4Mb
6.6Mb
Genes (Merged Ensembl/Havana)
PRKN >
< SNORA70
my_data_1
duplication_1
deletion
duplication_2
5.6Mb
5.8Mb
6.0Mb
6.2Mb
6.4Mb
6.6Mb
Reverse strand
1.27 Mb
There are currently 27 tracks turned off.
Ensembl Sus scrofa version 114.111 (Sscrofa11.1) Primary_assembly 1: 5,465,312 - 6,730,872
Gene Legend
Protein Coding
Ensembl protein coding
Non-Protein Coding
RNA gene

Additional file 8. Gene-level overlap between the *de novo* copy number variants (CNVs). (a) Visualisation of the ENSSSCG00000017645 (*TEX14*; chr12:34815476–34912766). Two CNVs (my_data_2) were identified within the gene boundaries: chr12:34888201-34899600 (duplication), chr12:34899601-34900400 (deletion). (b) Visualisation of the ENSSSCG00000004032 gene location (*PRKN*; chr1:5465312–6730872). Three CNVs (my_data_1) were identified within the gene boundaries: chr1:5514201-5516400 (duplication_1), chr1:5908601-5911400 (deletion), chr1:6264201-6273000 (duplication_2).



Additional file 9. Gene ontologies (GOs) with fold enrichment and the number of genes corresponding to the GO term. The false discovery rate (FDR) reflects the statistical significance of the enrichment (adjusted p-values for multiple testing to control the proportion of type I errors). Fold Enrichment measures the enrichment magnitude. Higher values indicate stronger enrichment and are an important metric of effect size. N. of Genes is the number of genes in the ontology that overlap with a gene list provided by a user.