

RESEARCH

Open Access



# Multi-tool copy number detection highlights common body size-associated variants in miniature pig breeds from different geographical regions

Jan Berghöfer<sup>1,2,3</sup>, Nadia Khaven<sup>1,3</sup>, Stefan Mundlos<sup>1,4,5</sup> and Julia Metzger<sup>1,3\*</sup>

## Abstract

**Background** Copy number variations (CNVs) represent a common and highly specific type of variation in the genome, potentially influencing genetic diversity and mammalian phenotypic development. Structural variants, such as deletions, duplications, and insertions, have frequently been highlighted as key factors influencing traits in high-production pigs. However, comprehensive CNV analyses in miniature pig breeds are limited despite their value in biomedical research.

**Results** This study performed whole-genome sequencing in 36 miniature pigs from nine breeds from America, Asia and Oceania, and Europe. By employing a multi-tool approach (CNVpytor, Delly, GATK gCNV, Smoove), the accuracy of CNV identification was improved. In total, 34 homozygous CNVs overlapped with exonic regions in all samples, suggesting a role in expressing specific phenotypes such as uniform growth patterns, fertility, or metabolic function. In addition, 386 copy number variation regions (CNVRs) shared by all breeds were detected, covering 33.6 Mb (1.48% of the autosomal genome). Further, 132 exclusive CNVRs were identified for American breeds, 47 for Asian and Oceanian breeds, and 114 for European breeds. Functional enrichment analysis revealed genes within the common CNVRs involved in body height determination and other growth-related parameters. Exclusive CNVRs were located in the region of genes enriched for lipid metabolism in American minipigs, reproductive traits in Asian and Oceanian breeds, and cardiovascular features and body height in European breeds. In the selected groups, quantitative trait loci associated with body size, meat quality, reproduction, and disease susceptibility were highlighted.

**Conclusion** This investigation of the CNV landscape of minipigs underlines the impact of selective breeding on structural variants and its role in the development of specific breed phenotypes across geographical areas. The multi-tool approach provides a valuable resource for future studies on the effects of artificial selection on livestock genomes.

**Keywords** Copy number variation, Pig genome, Next-generation sequencing, Miniature pigs, Quantitative trait loci, Functional enrichment, Regional adaptation, Genetic diversity, Biomedical research

\*Correspondence:  
Julia Metzger  
julia.metzger@tiho-hannover.de

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

## Background

Miniature pigs, extremely popular for their small stature and similarities to human physiology, have become indispensable models in biomedical research. However, the genetic background of their distinct phenotypic traits, including the role of copy number variations (CNVs), remains insufficiently characterised. Indeed, mammalian genomes harbour a remarkable diversity of genetic variants that underly breed- or population-specific phenotypes [1]. They range from single-nucleotide polymorphisms (SNPs) and small insertions or deletions (INDELs) to larger structural alterations [2–4]. Among these structural variants, CNVs are crucial for both natural and artificial selection [5]. CNVs can be duplicated or deleted elements, ranging from 50 base pairs (bp) to several megabase pairs [6–9]. They can influence gene expression by altering the copy number of genes and regulatory elements, affecting cellular function and phenotypic traits [10–12].

During evolution and subsequent domestication of animals, CNVs have been crucial in introducing genetic diversity and supporting the selection of desirable traits such as improved disease resistance, growth rate, and productivity [13–17]. The interplay between natural and artificial selection on specific CNVs has guided evolutionary trajectories in both wild and domesticated populations [18–20].

Various methods have been employed to detect CNVs, including SNP microarrays, array-based Comparative Genomic Hybridisation (aCGH), and Whole-Genome Sequencing (WGS) [21–23]. Though microarrays are widely used, WGS allows more accurate CNV detection due to their high resolution [24–26]. Recent advances in bioinformatics have facilitated the identification of small CNVs from short-read sequencing data despite difficulties arising from complex or repetitive genomic regions [27, 28]. In contrast, long-read sequencing technologies have been proposed to be advantageous for resolving complex structural variants. Their application in livestock genomics remains limited due to high costs and computational demands [27–29]. Consequently, short-read sequencing remains the primary method for large-scale CNV studies in domestic animal populations [26].

To enhance sensitivity and precision in short-read WGS-based CNV analysis, combining multiple CNV tools has been highlighted as an effective strategy [5, 16, 30]. This multi-tool approach captures a broader spectrum of variation by integrating signals from paired-end reads, split reads, or coverage depth [31, 32]. The annotation of resulting data remains limited but could be improved with more widespread use of WGS [5].

Recent livestock studies have taken advantage of these improved methods and demonstrated the significant impact of CNVs on various phenotypic traits [16,

33, 34]. For instance, a CNV in the *eukaryotic translation initiation factor 4A2* (*EIF4A2*) gene was linked to hip width, rump length, heart girth, and chest depth in several Chinese cattle breeds [35]. Furthermore, a CNV-based genome-wide association study (GWAS) in horses revealed three deleted CNV regions (CNVRs) associated with larger body sizes [16, 17, 36, 37]. In pigs, a comprehensive screening of CNVs across 18 diverse populations identified candidate copy number variable genes associated with complex traits [38]. Additionally, reproductive traits have been analysed for significant CNVs in pigs [39, 40], showing, for example, associations with litter size and teat number in Bama Xiang pigs [39]. Moreover, recent studies highlighted CNVs acting as modifiers for intramuscular fat content [41–43]. However, most CNV studies in pigs predominantly focused on traditional and commercial breeds commonly used in agriculture [17, 36, 37].

In addition, CNVs have been studied to identify the evolutionary genomic differences between domestic pigs and their wild ancestor [17, 44]. A comparison of the CNV-landscape between Anqingliubai pigs and Asian wild boars revealed genes related to growth (*CD36*), reproduction (*CIT*, *RLN*), detoxification (*CYP3A29*), and fatty acid metabolism (*ELOVL6*) [44]. Furthermore, genome sequence diversity has been put into context with phenotypic and production traits of Chinese and Western breeds, as well as with hypoxia and body size for Chinese native Tibetan, Dahe, and Wuzhishan pigs [45, 46]. Particular attention may be given to laboratory minipig herds to maintain genetic diversity and avoid the accumulation of fitness-reducing mutations [47].

Miniature pigs represent individuals with particularly small body sizes and distinct disease resistance profiles, making them suitable for biomedical research [48, 49]. Due to their physiological and anatomical similarities to humans, miniature pigs serve as valuable models for cardiovascular, diabetes, or obesity research and xenotransplantation [50–52]. Understanding copy number variability in miniature pigs could clarify the mechanisms underlying their unique phenotypic characteristics and reinforce their importance as large animal models in research [39, 53, 54].

The present study provides comprehensive insights into the CNV landscape and genetic diversity among miniature pig breeds, using a multi-tool approach for CNV detection. These findings improve our understanding of the genetic basis of phenotypic traits, enhance their utility as biomedical research models, and support more effective breeding strategies.

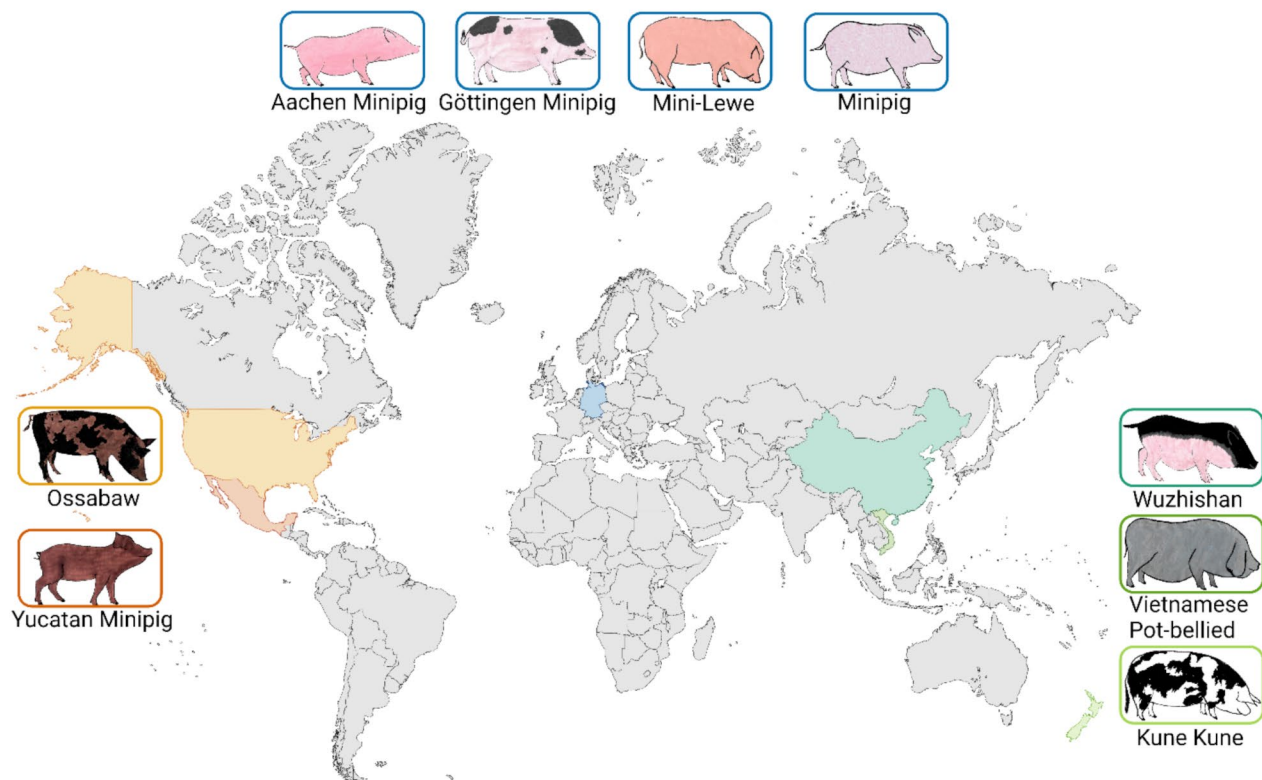
## Results

### CNVs discovery and distribution

WGS data from 36 miniature pigs representing nine breeds from different geographical origins worldwide were analysed to test the assumption that miniature pig breeds harbour CNVRs associated with their specific regional phenotype, as well as CNVRs common in all miniature pig breeds, regardless of their geographical origin (Fig. 1). Quality metrics from short-read WGS alignments revealed a mean alignment rate of 99.63% and mean genome coverage of 99.45%, reflecting a broad and consistent genome representation. The mean sequencing depth of 36.75X (minimum depth 14.06X) provided a resolution suitable for accurate detection of CNVs, including smaller variants (Additional file 1: Table S1). A summary of the number of CNVs identified by each tool (CNVpytor, Delly, GATK gCNV, and Smoove) based on CNV type (gain, loss or a mix of both) and the distribution of the CNV length per breed is presented in Additional file 2: Table S2. The distribution of CNV lengths showed most CNVs within a small range, with a sharp decline in frequency with increasing CNV size (Additional file 3: Figure S1). Among the tools, CNVpytor identified the highest number of CNVs; notably, all four

tools detected significantly more losses than gains. Based on the detected CNVs, multidimensional scaling (MDS) of the nine miniature pig breeds showed an apparent clustering of individuals by breed, with distinct groupings for European, Asian, and American breeds (Additional file 4: Figure S2). Similarly, the neighbour-joining (NJ) tree reflected these genetic separations, showing consistent clustering of individuals within breeds and highlighting their distinct genomic profiles and breeding history shaped by geographical origin and selective breeding practices.

To explore CNVs with potential functional impacts, filters were applied for homozygous CNVs and multi-copy regions (copy numbers > 4) overlapping with exons in all 36 minipig samples. 34 CNVs met these criteria in all samples, comprising 27 losses (deletions) and seven gains (duplications). These CNVs overlapped with and potentially affected 195 exonic regions corresponding to 64 genes (Additional file 5: Table S3), including *pregnancy-associated glycoprotein 6 (PAG6)*, *chromodomain helicase DNA binding protein 1 (CHD1L)*, *cadherin EGF LAG seven-pass G-type receptor 1 (CELSR1)*, *early endosome antigen 1 (EEA1)*, *cytochrome P450 family 4 subfamily A member 24 (CYP4A24)*, *collagen type VI alpha 1 chain*



**Fig. 1** Overview of the geographical origins and phenotypes. The phenotypes and geographical origins of the nine analysed miniature pig breeds from America (orange: Mexico and USA), Asia and Oceania (green: China, Vietnam, and New Zealand), and Europe (blue: Germany) are displayed. The figure was created with BioRender.com

(*COL6A1*), *phospholipase C delta 4 (PLCD4)* and *RPTOR independent companion of MTOR complex 2 (RICTOR)*. These loci might be involved in different biological processes, such as reproduction, metabolic regulation, and tissue development, supporting the relevance of these structural variants in shaping breed-specific traits.

The validation of seven selected CNVs overlapping with exonic regions by real-time quantitative PCR (qPCR) confirmed the losses or gains in the miniature pigs (Additional file 6: Table S4). Among five entirely randomly selected CNVs across different chromosomes, two regions with predicted losses and one region with a predicted gain could be validated, whereas two predicted gains could not be confirmed.

#### CNVs shared by all minipig populations and regionally grouped populations

After calling and filtering the CNVs, genome-wide CNVRs were assembled, and the chromosomal distribution of CNVRs and their types (gain, loss or both) were summarised (Additional file 7: Table S5). A total of 386 CNVRs were identified among CNVRs shared by at least two individuals from all nine miniature pig breeds distributed across different chromosomes (Fig. 2, see also Additional file 8: Figure S3 for a detailed view in an ideogram). These CNVRs included five copy number gains, 345 copy number losses, and 36 CNVRs with a mix of both types.

The sizes of the CNVRs ranged from 180 to 1,330,687 bp, with an average of 87,057.15 bp and



**Fig. 2** Distribution of shared minipig-CNVs across all nine global breeds and by regional breed groups. Grey bars represent all 18 pig autosomes. Green segments indicate deleted 'loss' CNVRs, while blue segments highlight duplicated 'gain' CNVRs. Orange segments show 'both' CNVRs, which contain regions with both 'loss' and 'gain'. The graph is divided into four panels: one for all minipig breeds combined (All Minipigs) and three for regional breed groups (America, Asia & Oceania, Europe). Note that only large CNVRs are visible in this graph



a median of 559 bp. The sum of CNVRs occupied 33,604,061 bp, representing 1.48% of the autosomal *Sus scrofa* 11.1 genome. Smaller CNVRs were more frequent, which aligns with prior observations that shorter structural variants are less likely to undergo purifying selection (Additional file 9: Figure S4). The number of CNVRs on each chromosome ranged from five on chromosome 18 to 34 on chromosome 2 (Additional file 10: Table S6) and was relative to the chromosomal length (Additional file 9: Figure S4). Taken together, these shared CNVRs underscore the presence of deletions across all minipig breeds and highlight the genome-wide prevalence of smaller variants with potential functional relevance.

### Comparison of CNVRs across regionally grouped populations

Subsequent analysis of the shared CNVRs resulted in 430 CNVRs in all ten individuals of the two miniature pig breeds from America (Ossabaw and Yucatan minipig, Additional file 7: Table S5). These CNVRs included nine copy number gains, 398 copy number losses, and 23 CNVRs with a mixture of both types (Fig. 2). The sizes of the CNVRs ranged from 194 to 1,330,687 bp, with a mean of 58,773.94 bp and a median of 597 bp. The CNVRs occupied 25,272,793 bp, representing 1.12% of the autosomal genome. Shorter CNVRs were more abundant than longer CNVRs in the genome (Additional file 9: Figure S4). The number of CNVRs on each chromosome ranged from three on chromosome 18 to 36 on chromosome 5, increasing with chromosome length (Additional file 10: Table S6; Additional file 9: Figure S4).

For the three miniature pig breeds from Asia and Oceania (Kune Kune, Vietnamese Potbellied Pig, and Wuzhishan Minipig), 382 CNVR were identified that are shared by all nine individuals (Additional file 7: Table S5). These CNVRs included seven copy number gains, 354 copy number losses, and 21 CNVRs with a mixture of both types (Fig. 2). The sizes of the CNVRs ranged from 191 to 1,330,687 bp, with an average of 63,869.72 bp and a median of 502 bp. The collective length of the CNVRs was 24,398,232 bp, representing 1.08% of the autosomal genome. The CNVRs identified on each chromosome ranged from five on chromosome 18 to 32 on chromosomes 2 and 8 (Additional file 10: Table S6). Shorter CNVRs were more prevalent than longer CNVRs in the genome, and the number of detected CNVRs increased with increasing chromosome length (Additional file 9: Figure S4).

Furthermore, for the four miniature pig breeds from Europe (Aachen Minipig, Göttingen Minipig, and Mini-Lewe Minipig; Additional file 7: Table S5), 388 CNVRs were called that are shared by all 17 individuals, including ten copy number gains, 360 copy number losses, and 18 CNVRs with a combination of both types. The sizes

of the CNVRs ranged from 191 to 1,330,687 bp, with an average of 69,129.91 bp and a median of 564.5 bp. The collective length of the CNVRs was 26,777,008 bp (1.18% of the autosomal genome). The number of CNVRs identified on each chromosome ranged from three on chromosome 18 to 34 on chromosome 4 (Additional file 10: Table S6). Consistent with the other regional groups, shorter CNVRs were more frequent, and their chromosomal distribution correlated with chromosome length. Altogether, these findings highlight that each regional group of minipigs exhibits a distinct but overlapping set of CNVRs.

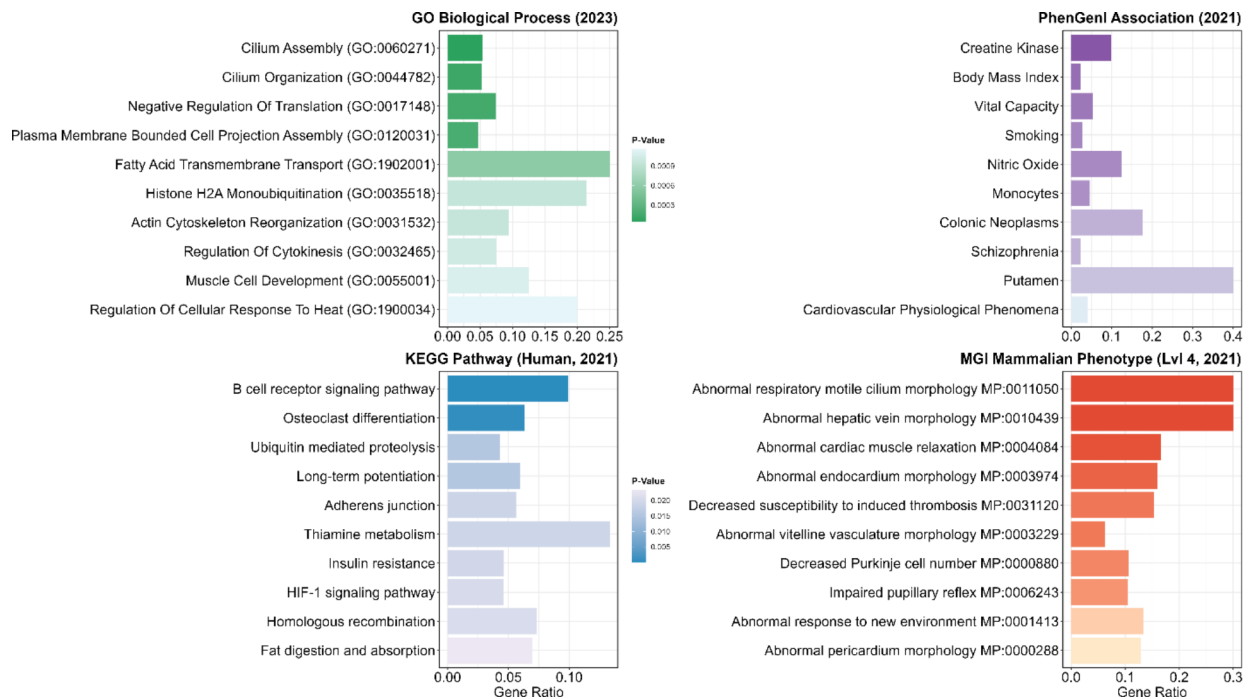
### Functional enrichment of genes overlapping with common CNVRs

For functional enrichment analysis, 359 genes (no duplicates) were extracted from 370 CNVRs shared by at least two individuals across all nine miniature pig breeds (see Additional file 7: Table S5). In addition, for the regional subsets, we identified 380 genes from 430 shared CNVRs (America), 357 genes from 382 shared CNVRs (Asia and Oceania), and 358 genes from 388 shared CNVRs (Europe; Additional file 7: Table S5).

Gene set enrichment analysis of 286 human orthologous genes from 359 pig genes in CNVRs shared by all minipig breeds revealed significantly enriched Gene Ontology (GO) biological processes, including cilium assembly (GO:0060271), fatty acid transmembrane transport (GO:1902001), actin cytoskeleton reorganisation (GO:0031532), and muscle cell development (GO:0055001) (Fig. 3; Additional file 11: Table S7). Pathways from the Kyoto Encyclopaedia of Genes and Genomes (KEGG), such as B-cell receptor signalling, osteoclast differentiation, insulin resistance, and fat digestion and absorption, were also enriched, along with terms like body mass index and monocytes from the PhenGenI association database. Among these, 46 significantly enriched genes, such as *CPNE8*, *CDH4*, and *INSR*, were linked to the PhenGenI term “body mass index”, whereas 34 genes, including *DPP10*, *CLSTN2*, and *ADAMTS17*, were associated with the term “height”. In addition, the MGI Mammalian Phenotypes database highlighted enriched terms related to respiratory motility (MP:0011050), hepatic vein morphology (MP:0010439), cardiac muscle relaxation (MP:0004084), and responses to novel environments (MP:0001413).

### Functional enrichment of genes overlapping with regionally grouped CNVRs

Analogous to the complete dataset, functional enrichment was also performed for CNVRs identified in pigs from distinct geographical regions (America, Asia & Oceania, Europe, see Fig. 4). These analyses included 315 human orthologous genes from American breeds (out of



**Fig. 3** Functional enrichment of genes in CNVRs detected in all nine minipig breeds. The illustration depicts the ten most significantly enriched GO terms from the databases “GO\_Biological\_Process\_2023”, “KEGG\_2021\_Human”, “MGI\_Mammalian\_Phenotype\_Level\_4\_2021”, and “PhenGenI\_Association\_2021”. The terms are ordered by *p*-value (colour intensity) and represent CNVR genes common to all nine minipig breeds. The X-axis indicates the gene ratio

380 pig CNVR-genes), 292 genes from Asian and Oceanian breeds (357 pig CNVR-genes) and 290 genes from European breeds (358 pig CNVRs-genes; Additional file 11: Table S7).

#### Functional enrichment in American miniature pig breeds

Genes in CNVRs shared by the two American breeds were significantly enriched for GO processes, such as negative regulation of Hippo signalling (GO:0035331), regulation of Hippo signalling (GO:0035330), glutamate receptor signalling pathway (GO:0007215), and negative regulation of GTPase activity (GO:0034260) (Fig. 4a). Enriched KEGG pathways included the B-cell receptor signalling pathway, ascorbate and aldarate metabolism, pentose and glucuronate interconversions, and osteoclast differentiation. In addition, the PhenGenI association highlighted significantly enriched terms such as creatine kinase and metabolome. The MGI Mammalian Phenotypes indicated traits like abnormal gait (MP:0001406) and abnormal hepatic vein morphology (MP:0010439).

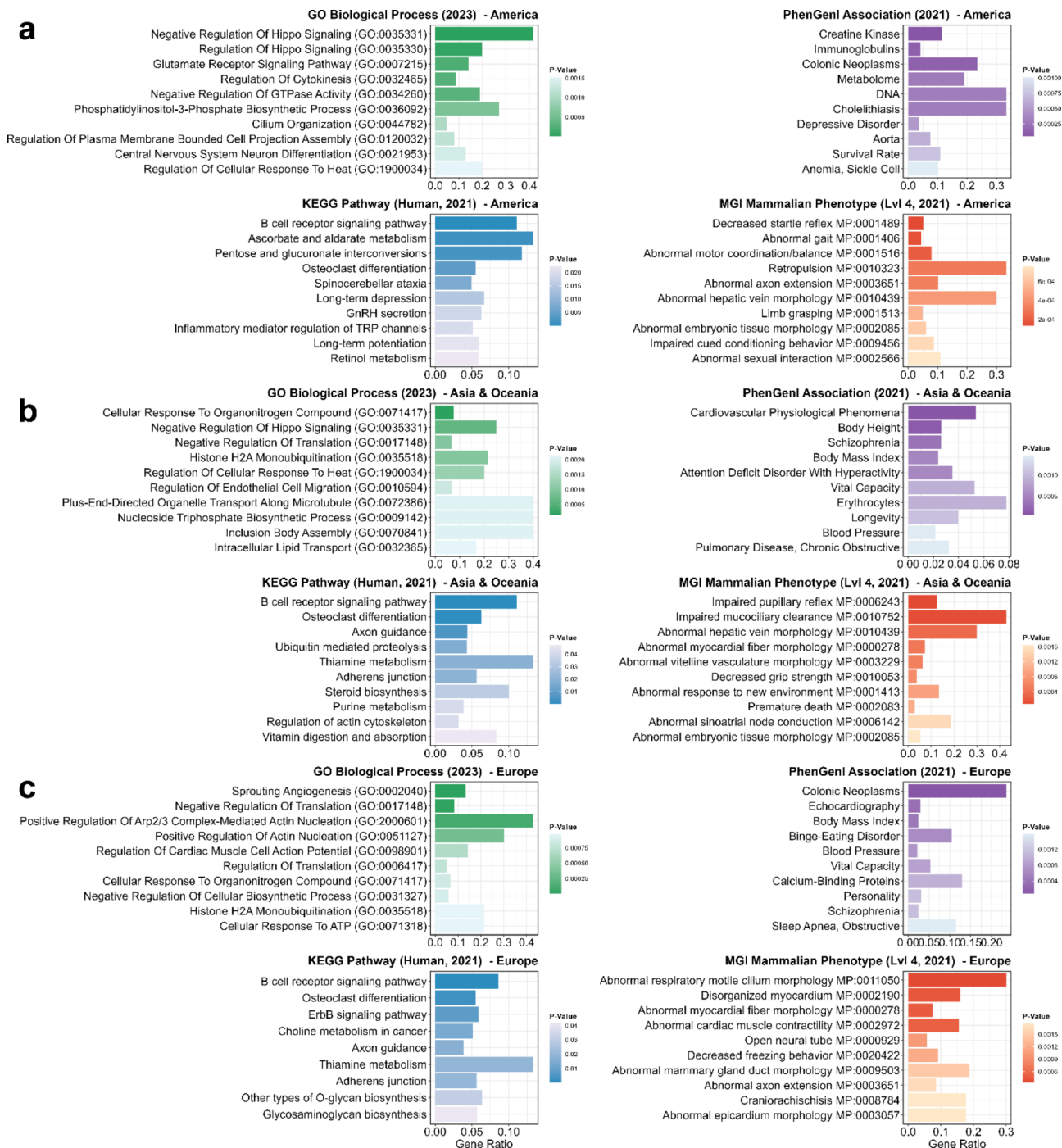
#### Functional enrichment in Asian and Oceanian miniature pig breeds

For genes in CNVRs shared by three breeds from Asia and Oceania, functional gene set enrichment analysis revealed significant GO terms such as cellular response to organonitrogen compound (GO:0071417), negative regulation of Hippo signalling (GO:0035331), negative

regulation of translation (GO:0017148), and histone H2A monoubiquitination (GO:0035518) (Fig. 4b). Significant KEGG pathways resembled those in American minipigs (e.g. B-cell receptor signalling; osteoclast differentiation), whereas PhenGenI associations featured cardiovascular traits, body height, body mass index, and vital capacity. Furthermore, the MGI Mammalian Phenotypes revealed enrichments related to sinoatrial node conduction (MP:0006142), hepatic vein morphology (MP:0010439), and myocardial fibre morphology (MP:0000278).

#### Functional enrichment in European miniature pig breeds

For genes in CNVRs shared by four breeds from Europe, several GO biological processes were enriched, including sprouting angiogenesis (GO:0002040), negative regulation of translation (GO:0017148), positive regulation of actin nucleation (GO:0051127), and regulation of cardiac muscle cell action potential (GO:0098901) (Fig. 4c). KEGG pathways included B-cell receptor signalling, osteoclast differentiation, ErbB signalling pathway, and choline metabolism in cancer. The PhenGenI association identified significantly enriched terms such as body mass index and vital capacity, and the MGI Mammalian Phenotypes showed significantly enriched terms, such as abnormal respiratory motile cilium morphology (MP:0011050), disorganised myocardium (MP:0002190), abnormal myocardial fibre morphology (MP:0000278), and abnormal cardiac muscle contractility (MP:0002972).

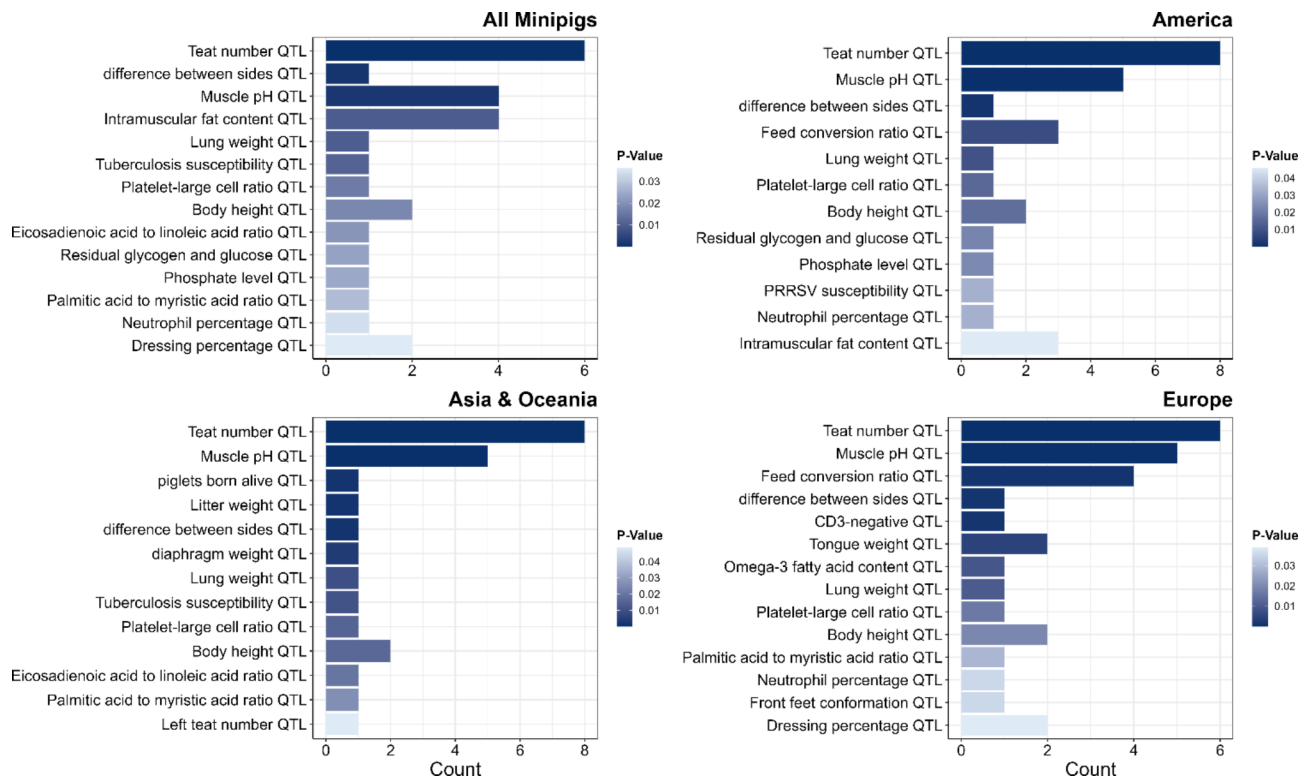


**Fig. 4** Functional enrichment analysis for genes in CNVRs called for regionally grouped minipig breeds. The figure illustrates the ten most significantly enriched terms for the databases “GO\_Biological\_Process\_2023”, “KEGG\_2021\_Human”, “MGI\_Mammalian\_Phenotype\_Level\_4\_2021”, and “PhenGenI\_Association\_2021”. The data set comprising CNVR genes common to all regionally grouped breeds originating from America (a), Asia and Oceania (b) or Europe (c) has been sorted by *p*-value (colour intensity). The X-axis indicates the gene ratio

### Enrichment of QTL overlapping with common CNVRs

As an additional functional evaluation of CNVRs, an overlap analysis was performed with quantitative trait loci (QTL) from the animal QTL database (Fig. 5; Additional file 12: Table S8). Among 31,312 QTL examined, 52 were associated with 36 phenotypic traits and overlapped

with CNVRs shared by all nine minipig breeds. QTL enrichment analysis revealed that approximately 39% of these traits were significantly overrepresented ( $p < 0.05$ ). These traits included body height, difference between two sides, muscle pH QTL, intramuscular fat content QTL, tuberculosis susceptibility QTL, and several



**Fig. 5** Enrichment analysis of QTL overlapping with CNVRs. Enrichment analysis was conducted for significant QTL found in the same regions overlapping with CNVRs shared by all nine miniature pig breeds or for regionally grouped minipig breeds (America, Asia & Oceania, or Europe). The figure shows all significantly enriched QTL in the CNVRs compared to the autosomal genome in total. The dataset comprising QTL common to all nine minipig breeds (a) or the common regionally grouped breeds originating from America (b), Asia and Oceania (c), and Europe (d) was sorted by *p*-values

haematological and metabolic QTL. Interestingly, the analysis revealed a significantly enriched height QTL within two common CNVRs shared by all nine minipig breeds. One overlapped with the *chromodomain helicase DNA binding protein 4* (*CHD4*) gene.

In addition, QTL overlap was tested for the regional breed groups. In the American breeds, 50 QTL were associated with 34 phenotypic traits, of which 12 (35%) showed significant overrepresentation. These traits were mainly related to the QTL body height, difference between sides, muscle pH QTL, intramuscular fat content QTL, PRRSV susceptibility, and haematological or metabolic parameters. In Asian and Oceanian breeds, 49 QTL representing 36 phenotypic traits overlapped with CNVRs, of which 13 (36%) were significantly overrepresented, suggesting an emphasis on body size, reproduction, and meat quality. In European minipigs, 54 QTL coincided with 37 phenotypic traits, of which 14 (38%) were significantly overrepresented, reflecting an association with body height, conformation, meat quality, disease susceptibility, and immunity. Overall, these region-specific QTL overlaps and the globally shared patterns indicate that CNVs are integral to phenotypic diversity in minipigs worldwide, mainly for growth, disease resistance, and production traits.

#### Exclusive CNVRs across regionally grouped populations

Finally, CNVRs uniquely detected in each regional group (named exclusive CNVRs) were investigated. Overall, 132 exclusive CNVRs were identified in two miniature pig breeds from America, harbouring one gain, 125 losses, and six CNVRs with a mixed gain-and-loss pattern (Additional file 13: Table S9). Their lengths ranged from 237 to 649,999 bp (average size: 23,651.70 bp, median: 1120 bp), encompassing 0.14% of the autosomal genome (3,122,025 bp). In the three miniature pig breeds from Asia and Oceania, 47 exclusive CNVRs were detected, comprising 45 losses and two mixed types with an average size of 21,302.47 bp (median: 661 bp). These regions covered 0.04% of the autosomal genome (1,001,216 bp). European minipigs (four breeds) exhibited 114 exclusive CNVRs with three copy number gains, 107 losses, and four mixed types. The lengths ranged from 217 to 585,179 bp (mean: 47,224.74 bp; median: 1629.5 bp), covering 0.24% of the autosomal genome (5,383,620 bp).

Finally, we compared biological themes to search for enriched functional profiles for the genes within the unique CNVRs. Genes were grouped according to their geographical origin: 125 genes from America, 47 from Asia and Oceania, and 109 from Europe (Additional file 14: Table S10). The analysis revealed several significantly



enriched functional clusters of GO biological processes and KEGG terms for each regional gene set that overlapped with exclusive CNVRs (Fig. 6). The American gene set showed many terms on metabolic regulation, including “nucleobase-containing small molecule biosynthetic process” (GO:0034404) and other biosynthetic pathways. In contrast, gene sets from Asia and Oceania highlighted “immune responses to bacterial and fungal pathogens”, suggesting specialisation in immune defence systems. The European gene set showed enrichment in ciliary function, organelle assembly, and lipid metabolism (e.g. glycosphingolipid biosynthesis), reflecting distinct local adaptations. Further analysis of KEGG pathways emphasised neurological and metabolic signalling in American minipigs, a more specialised immune response profile in Asia and Oceania, and cardiovascular and circadian regulatory adaptations in Europe. These region-specific enrichment patterns indicated exclusive CNVRs harboured genes under potential selection in local breeding histories.

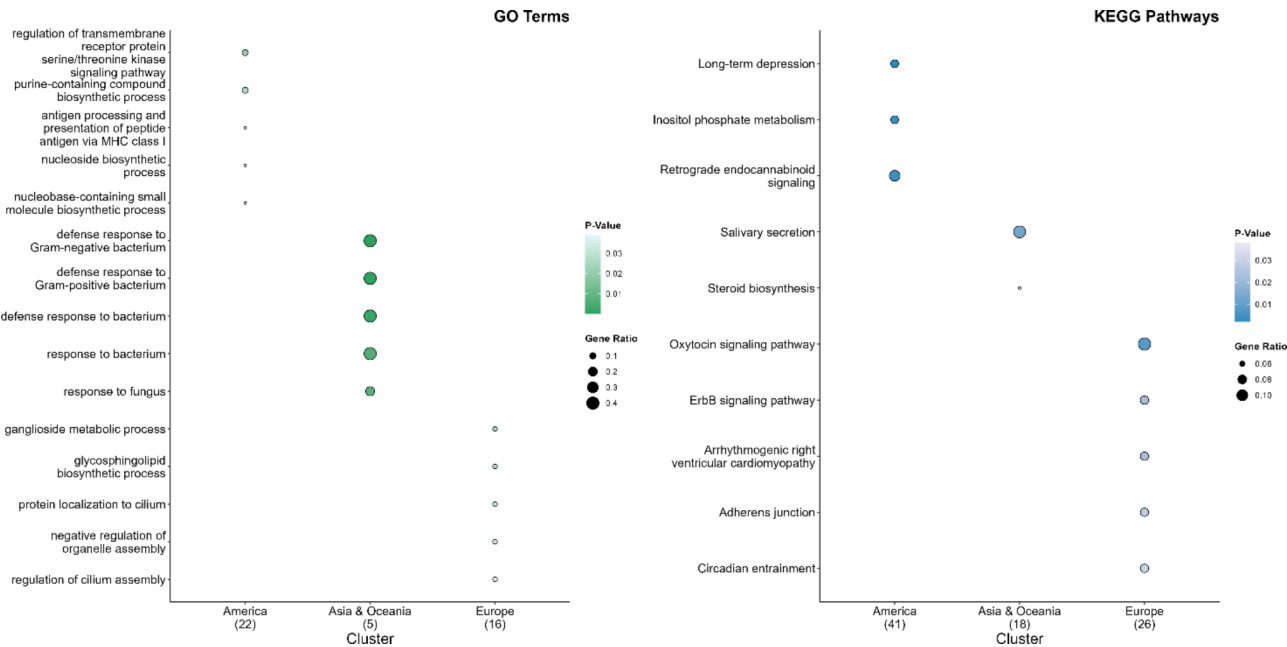
Distribution of QTL overlapping exclusive CNVRs

In addition, an overlap of exclusive CNVRs with QTL assessed for regionally grouped minipig breeds revealed distinct patterns in America, Asia and Oceania, and Europe (Table 1). In the American breeds, 15 QTL corresponding to 11 phenotypic traits overlapped with exclusive CNVRs, encompassing growth and developmental

characteristics such as teat number, feed conversion ratio, birth weight variability, and longissimus muscle area, in addition to meat and carcass quality traits (“muscle pH”, “intramuscular fat content”, “stearic acid content”), reproduction (“sperm abnormality” and “still-birth rates”), and a health-immunity QTL for “PRRSV susceptibility”. In Asia and Oceania, four QTL were identified, reflecting a narrower range of traits associated primarily with growth (“teat count”, “diaphragm weight”) and reproduction (“litter weight”, “live-born piglets”). By contrast, European minipig breeds exhibited a more extensive distribution, with 16 QTL linked to 15 phenotypic traits, including growth and development, meat and carcass quality, health and immunity, and reproduction. Noteworthy traits included “body height”, “shoulder weight”, “meat colour”, and immune-related markers such as “CD3-negative” and “CD8-negative” or “CD8-positive leukocyte percentage”. Altogether, these region-specific patterns underscore that exclusive CNVRs are closely tied to traits of regional importance, reflecting distinctive breeding histories and local adaptations.

Discussion

This study presents a comprehensive analysis of the genomic CNV landscape of nine miniature pig breeds from diverse geographical regions using a multi-tool approach. By integrating multiple CNV detection methods, both common and exclusive CNVRs were identified,



**Fig. 6** Biological theme comparison of genes overlapping with exclusive CNVRs in regionally grouped minipig breeds. Biological themes were compared for genes overlapping with exclusive CNVRs identified in all regionally grouped minipig breeds (America, Asia, Oceania, and Europe). The figure illustrates the five most significantly enriched Gene Ontology (GO, left) biological processes and KEGG term clusters (right). On the x-axis, each geographic region (America, Asia & Oceania, Europe) and the number of annotated genes are indicated (in parentheses). The *p*-value is displayed by colour intensity, and the dot size indicates the gene ratio for each cluster

**Table 1** Distribution of QTL overlapping exclusive CNVRs in regionally grouped minipig breeds. The table summarises all QTL overlapping with exclusive CNVRs identified for all regionally grouped minipig breeds (America, Asia & Oceania, and Europe) assigned to the functional categories as ‘growth and development’, ‘reproduction’, ‘meat and carcass quality’ and ‘health and immunity’ to provide insights into region-specific trait association

Category	QTL-Counts	America	Asia & Oceania	Europe
Growth and Development	Teat number	5	1	1
	Feed conversion ratio	1	0	2
	Birth weight variability	1	0	0
	Body height	0	0	1
	Diaphragm weight	0	1	0
	Longissimus muscle area	1	0	0
	Shoulder weight	0	0	1
	Subcutaneous fat thickness	1	0	0
	Tongue weight	0	0	1
	<b>Total QTL</b>	<b>9</b>	<b>2</b>	<b>6</b>
Health and Immunity	CD3-negative leukocyte percentage	0	0	1
	CD8-negative leukocyte percentage	0	0	1
	CD8-positive leukocyte percentage	0	0	1
	Front feet conformation	0	0	1
	PRRSV susceptibility	1	0	0
	<b>Total QTL</b>	<b>1</b>	<b>0</b>	<b>4</b>
Meat and Carcass Quality	Muscle pH	1	0	1
	Cis-11-Eicosenoic acid to oleic acid ratio	0	0	1
	Intramuscular fat content	1	0	0
	Meat colour	0	0	1
	Omega-3 fatty acid content	0	0	1
	Palmitoleic acid to palmitic acid ratio	0	0	1
	Stearic acid content	1	0	0
	<b>Total QTL</b>	<b>3</b>	<b>0</b>	<b>5</b>
Reproduction	Sperm abnormality rate	1	0	1
	Litter weight	0	1	0
	Number of stillborn	1	0	0
	Piglets born alive	0	1	0
	<b>Total QTL</b>	<b>2</b>	<b>2</b>	<b>1</b>

offering valuable insights into genetic diversity, regional adaptability, and possible implications for breeding strategies.

Employing a multi-tool approach likely increased the sensitivity and accuracy of CNV identification compared to traditional single-tool methods. In particular, the validation of those CNVs was conducted using a minimum of two tools, confirming the in-silico approach’s robustness. By integrating detection methods based on read depth, split reads, and paired-end reads, a broader range of CNVs could be captured, including smaller and more complex variants, which might otherwise remain undetected [5, 30]. This finding is consistent with a previous study, suggesting that the combination of tools represents a good compromise between specificity and sensitivity for calling CNVs of different types in short-read sequencing data [5]. Nevertheless, the detection efficiency is still highly dependent on the read length: longer reads tend to reduce mismatches and multiple alignments [55], whereas shorter reads (> 100 bp) can present challenges in resolving variant-dense regions [55, 56]. Although long-read sequencing technologies help overcome some complexities in repetitive genomic regions [57], their higher costs and computational demands continue to limit their application in livestock genomics.

In practice, both long- and short-read sequencing platforms offer distinct advantages and drawbacks: Long reads can more readily resolve CNVs in regions inaccessible to arrays or short-read data, whereas short-read approaches offer high resolution for smaller private CNV calls [58]. Moreover, both approaches are relatively independent of population-level CNV frequencies compared to SNP arrays. This is particularly advantageous for studies with small sample sizes or livestock breeds with lower genetic diversity, such as pigs [17, 59]. This advantage is enhanced by recent PCR-independent library preparations that deliver uniform signal intensities during sequencing, facilitating sample-specific depth normalisation without batch processing [60, 61]. Additional techniques, such as 10X Genomics linked-read sequencing [62], DNA template strand sequencing [57] and Hi-C [63], also show promise for increasing CNV detection accuracy, though their use remains limited so far.

The multi-tool analysis in this study allowed the detection of relatively large CNVRs of an average size of 87 kb. In contrast, a previous study in 16 wild and domestic high-production pigs - relying on a single (read-depth based) CNV caller - reported a smaller average CNVR size of 13 kb [17]. The detection of larger CNVRs here likely reflects the broader coverage afforded by a multi-tool method. Furthermore, it has been proposed that rare CNVs are usually longer than common CNVs due to purifying selection against large genomic alterations, particularly in variation-sensitive regions [31]. Consistent

with this, the present data show that shorter CNVRs occur more frequently than longer ones.

In addition, all four tools detected significantly more losses than gains, similar to what was observed in previous CNV studies [16, 64]. It was suggested that gains and losses might be differently balanced depending on the CNV-detection tool and subsequent detection method [65]. Apart from the technical context, losses are also proposed to be generally more frequent than gains due to their simpler mutational mechanisms, such as unequal crossover or replication slippage, and their lower likelihood of causing gene dosage imbalances [66, 67].

Interestingly, 195 exonic regions overlap with 34 homozygous deletions, duplications, or multi-copy CNVs, suggesting a potential functional impact of these structural variants in the analysed minipig breeds. Several genes appear to underpin key biological processes and might influence traits enhanced by breeding, including growth, reproductive success, and metabolic function [47, 68, 69]. Notably, genes such as *PAG6* involved in maternal recognition of pregnancy, immunomodulation and trophoblast invasion [70, 71], *CELSR1*, linked to teat number in pigs [72–74] and dosage-dependent effects on embryogenesis or craniofacial development [75, 76], as well as *EEA1*, more highly expressed in embryos of low fertility bulls [77], highlighted the potential role of CNVRs in embryonic development and survival of newborn piglets. However, CNVRs in the region of *EEA1* need to be considered with caution as it was previously suggested that the presence of minor CNV in the reference genome assembly, *Sus scrofa* 11.1, might lead to the predictions of an abnormally gained copy number on *EEA1*, as it was observed in multiple domesticated pig populations and wild boars [78].

Furthermore, members of the CYP4A family, including *CYP4A24* detected in our analysis, have been investigated in the context of liver metabolism and meat quality in pigs [79–81]. The intense selection pressure on performance traits such as meat quality and fat content observed in minipigs might be reflected by the CNVRs found in *COL6A1*, a major extracellular matrix component contributing to the structural integrity in connective tissues with a potential effect on muscle quality [82, 83], *CNOT9*, a factor for meat colour traits in commercial pig breeds [84, 85] or *PLCD4*, which was reported to harbour variants affecting feed conversion, intramuscular fat, and water exudation in pigs [86]. Moreover, it was interesting to note that also an exonic region of *RICTOR*, a component of the mTORC2 complex and a potential regulator of cytoskeletal organisation, metabolism and cell growth [87, 88], was highlighted in our results and might subsequently be associated with growth effects in miniature pigs. This outcome of the analysis of CNVs overlapping with exonic regions suggests that dosage effects might

control some key genes and are, therefore, potential candidates for future functional investigations.

In addition, 386 CNVRs common to all investigated miniature pig breeds were detected. As anticipated for miniature-sized breeds, our analysis revealed a significant enrichment of body height QTL within two common CNVRs, overlapping with *CDH4*, a major component of the nucleosome remodelling and deacetylase complex [89]. This gene plays an essential role in epigenetic transcriptional repression, highlighting the potential phenotypic impact of changes in the gene regulatory landscape, such as epigenetic mechanisms [90]. Furthermore, gene set enrichment analysis revealed 34 genes in “body height” and 46 genes in “body mass index”, including *CDH4* (PhenGenI Associations), further supporting the hypothesis that structural variants contribute to the small stature targeted in miniature pig breeding. Consequently, these findings align with previous findings from a genome-wide association study across 58 pig populations, highlighting various CNVs significantly associated with body length and weight [91].

Subsequent analysis found QTL for “body height” and “difference between sides” within CNVRs shared by the analysed minipig breeds, implying these as indicators for intense selection pressures for small body size [92], with miniature size preserved proportionally in all body parts [93]. Signatures of selection for this extreme phenotype were recognised in various breeds, including the Yucatan miniature pig, for which genomic regions under potential selection were associated with feed intake, weight regulation, cell cycle and cell proliferation [92, 94]. Genes related to body mass index, fat digestion, muscle cell development, and creatine kinase were also enriched, hinting at traits under selection associated with miniature body size or targeted breeding strategies for animal models in biomedical research [95]. For instance, fatty acid transport and insulin resistance support the metabolic adaptability of minipigs, making them suitable models for obesity research [95, 96].

Beyond these common CNVRs, exclusive regions in American, European, and Asian/Oceanian breeds were discovered. As emphasised by the phylogenetic tree, the minipigs showed a clustering pattern according to their geographical origins and breeding history, suggesting that CNVs serve as distinct targets for breeding strategies across different geographical contexts. A similar assumption was made in a recent study on minipig populations targeting demographic signatures of selection [94].

For example, we identified CNVRs in Ossabaw and Yucatan minipigs associated with lipid metabolism. This was highlighted by an overlap of genes in these regions enriched for ascorbate and aldarate metabolism and QTL linked to intramuscular fat content, stearic acid content, and subcutaneous fat thickness. This finding might hint

at the extremely increased ability of the Ossabaw minipig to convert food into body fat and its outstanding propensity to obesity and the development of metabolic syndrome [97, 98]. Similarly, the Yucatan minipig was studied as a model for obesity for the effects of a low-calorie/low-protein diet on weight and body composition [99]. This breed's predisposition to obesity might be further supported by our findings of genes enriched in CNVRs, overlapping with the Hippo signalling pathway, which is known to control the balanced regulation of adipocyte proliferation and differentiation [100] and was previously tied to meat quality traits [101].

In contrast to the American breeds, negative regulation of Hippo signalling was enriched in common CNVRs shared by the breeds from Asia and Oceania (Kune Kune, Vietnamese Potbellied, and Wuzhishan). This pathway might play a role in the specific phenotypic expression of the different breed types, mainly as Hippo signalling is relevant for cell proliferation, as mentioned above, and is also highly important for organ size control [102]. Furthermore, the identified QTL "litter weight" and "teat number" for this group of pigs, as well as the enrichment of genes in exclusive CNVRs for the oxytocin signalling pathway, likely reflect the efforts to increase breeding efficiency for conservation purposes [103]. In addition, early sexual maturity and good adaptability to harsh rearing conditions or poor nutrition were highlighted as unique characteristics common in these breeds [103–107].

Similar considerations apply to European minipig breeds (Aachen Minipig, Göttingen Minipig, Mini-LEWE), where QTL for traits such as "teat number" and "sperm abnormality rate" were found to overlap with exclusive CNVRs, indicating selection to refine reproductive attributes. We propose that these similarities regarding reproduction might be due to the breeding history of these minipigs, which resulted from the cross-breeding of Vietnamese potbellied pigs with commercial farm pigs in Europe to develop miniature-sized pigs for biomedical research [106, 108]. Additional exclusive CNVRs coincided with QTL "body height" and "shoulder weight," suggesting an association with breed-specific selection for reduced body size [108–110]. In addition, a large number of enriched genes were associated with the cardiovascular system, including "angiogenesis," "cardiac muscle cell action potential," and "abnormal myocardial fibre morphology," potentially reflecting the strong emphasis on cardiac research in European minipigs [111, 112]. It was proposed that the minipig's extensive similarities to the human heart in anatomy and physiology make it an attractive alternative to traditional non-rodent species for cardiac and respiratory safety studies [111, 113]. Studies for cardiovascular safety pharmacology have highlighted the significantly longer cardiac action potentials and QT intervals (time between the start of the Q

wave and the end of the T wave) of the (Göttingen) minipig compared to dogs at comparable cycle lengths as well as a relative resistance to arrhythmias [114]. Research on xenografts and immunological safety frequently involved Göttingen minipigs [115], which aligns with our findings of QTL for "CD3-negative" and "CD8-positive/negative leukocyte percentage" in CNVRs, presumably reflecting artificial selection for immunological traits [116, 117].

Overall, our CNVR findings in European minipigs and other geographically diverse breeds spotlight essential features tied to structural variation in desirable genes and genomic regions under selection in miniature pigs [93], including body size, conformation, meat quality, and disease susceptibility [68, 118]. Incorporating CNV data into breeding programmes can help maintain genetic diversity while improving desirable traits, as demonstrated in breeds such as the Aachen Minipig and Göttingen minipig [93, 108].

Nevertheless, gene set enrichment studies require cautious interpretation since they focus on coding regions and rely on existing gene annotations. In recent years, it has been highlighted that structural variants in non-coding DNA have modifying or pathogenic potential, affecting the spatial 3D organisation of the genome and leading to gene dosage effects [119, 120]. Future investigations should include CNVs in non-coding regions to provide a broader perspective on phenotypic impacts, particularly in growth control or disease susceptibility traits.

## Conclusions

This study provides a comprehensive view of CNVs in miniature pig breeds and advances our understanding of how geographic distribution impacts genetic diversity. The results reveal distinctive patterns shaped by targeted breeding strategies in pigs used for biomedical research purposes. This study sheds light on the complex CNV landscape as a valuable resource for future studies in pigs and improves the genetic understanding of breeding strategies and their impact on structural variation as footprints of artificial selection in livestock populations.

## Methods

### Sampling and whole genome sequencing

This study utilised WGS data from 36 miniature pig samples representing nine different miniature pig breeds from various locations worldwide, as detailed in Additional File 1: Table S1. In total, 20 samples were obtained from a previous study [121] (project identifier PRJNA795885) or downloaded directly from the National Center for Biotechnology Information (NCBI)-Sequence Read Archive (SRA). Ethical approval for these samples can be found in the respective publications.

In addition, further hair roots, ear cartilage (from ear tagging) or skin samples (from two piglets that died due



to perinatal mortality) were collected from 16 minipigs, including three Kune Kune, four Aachen Minipig, three Mini-Lewe, three Potbellied, and three other miniature pig breeds, and underwent WGS. Subsequently, DNA was extracted from non-degraded tissues to minimise any confounding impact on genomic integrity.

Sample preparation followed the protocol described in our previous study [121], adhering to all national and international guidelines. All samples were sequenced on an Illumina NovaSeq 6000 platform, generating 150 bp paired-end reads to ensure high accuracy and resolution. The animal experiments were approved by the animal ethics committee of the Lower Saxony State Veterinary Office (registered at 33.9-42502-05-17A217). All sampled pigs were owned by the Institute of Animal Genomics. Informed consent for using the samples was obtained from the pig owner. Due to the low burden of sampling, no anaesthesia was applied. This global data collection ensured the broad applicability of our findings to diverse genomic research and animal breeding contexts.

### WGS analysis

Raw sequencing reads were filtered using the fastp v. 0.23.4 with specific parameters to improve data integrity. These parameters included adapter trimming to remove sequencing adaptors, low complexity filtering to reduce artefacts (`--low_complexity_filter --complexity_threshold 1`), and quality trimming to ensure high-quality reads (`--cut_front --cut_front_window_size 1 --cut_front_mean_quality 20 and --cut_tail --cut_tail_window_size 1 --cut_tail_mean_quality 20`). Furthermore, bases were filtered based on a Phred score threshold (`--qualified_quality_phred 15`), with reads discarded if more than 70% of the bases were of low quality (`--unqualified_percent_limit 70`) or if the number of N bases exceeded 50 (`--n_base_limit 50`).

Filtered reads were aligned to the autosomal reference genome *Sus scrofa* 11.1 [36] accessed from Ensembl, using the Burrows-Wheeler alignment (BWA) tool v. 0.7.17-r1188 [122]. The parameters `-t 8` (for multithreading) and `-M` (to mark shorter split hits as secondary) were used to ensure proper handling of split alignments. Coverage and depth estimates were calculated using samtools v. 1.19.2 (command `samtools depth`) [123], averaging the read depth across all autosomal bases with mapped reads. Multiple CNV detection tools were employed to cross-validate CNVs, thereby reducing false positives in lower-depth regions and minimising coverage-related biases.

Duplicates were marked, and the processed data were indexed using Picard tools v. 3.1.0 [124]. Each sample's data quality was re-evaluated to ensure that low-depth or high-coverage regions would not disproportionately impact CNV detection.

### CNV detection and analysis

Following prior studies' recommendations, four tools were used to detect CNVs, providing a multi-tool approach to optimise CNV recall and precision [5, 59]. All tools were applied exclusively to autosomes to eliminate sex-related biases in CNV analysis.

Lumpy [125], integrated via Smoove v. 0.2.8, leveraged split-read signals [126]. Default parameters were applied, specifying a minimum variant size of 50 bp (`--min-event-size 50`) and genotypes were jointly called to obtain a population-wide CNV dataset. Furthermore, Delly v. 1.1.8 [127] was applied for a comprehensive analysis combining paired-end and split-read data with default settings, specifying `-z 50` as the minimum CNV size. Genome Analysis Toolkit (GATK) gCNV pipeline v. 4.5 relied on read-depth signals to detect CNVs, using a bin size of 1000 as recommended for 30x data to balance detection rates and false positive control [128]. Finally, CNVpytor v. 1.3.1 analysed CNVs based on read depth and allele imbalance [129], using a bin size of 500 bp following the recommendations for approximately 37x coverage and 150 bp reads [129]. All tools employed normalisation procedures (for GC content and local coverage variation) to reduce systematic biases. CNVs larger than 5 Mb were excluded to limit the number of false-positive rates due to repetitive regions or assembly uncertainties, as previously suggested [130].

Nine pig breeds were examined for differences in population structure using the CNVs pooled from the four CNV datasets. All CNVs were combined into a union dataset; duplicates were removed by merging CNVs of the same type using a minimum 50% reciprocal overlap criterion with GenomicRanges v. 1.56.1 [131] in R 4.3.3 [132]. A binary matrix was then constructed indicating the presence or absence of CNVs per sample. Principal component analysis for population structure analysis was performed using multidimensional scaling with the `cmdscale` function in R 4.3.3 [132] and visualised as a scatterplot using ggplot2 [133]. A neighbour-joining tree was constructed using a genetic distance matrix to assess phylogenetic relationships between the studied populations, which were visualised using MEGA v. 11 [134].

To further investigate variants with a higher likelihood of a functional impact, we extracted homozygous deletions (copy number state 0, potential for gene disruption) and homozygous duplications (copy number state  $\geq 4$ , prone to gene dosage imbalances) present in all 36 minipig samples from all CNV calling datasets for further analysis. These CNVs were merged using a minimum 50% reciprocal overlap criterion with GenomicRanges v. 1.56.1 [131] in R 4.3.3 [132]. Subsequently, the merged CNV dataset intersected with all exonic regions extracted from the Ensembl *Sus scrofa* 11.1 genome annotation (release 111). All CNVs with at least 50% overlap with

one or more exons were retained. Corresponding genes were assigned to these exonic regions, and gene annotations were performed using Ensembl biomaRt v. 2.60.1 [135, 136].

#### Real-time quantitative PCR

Seven CNVs within exonic regions (see analysis above) and a random selection of five CNVs across different chromosomes were validated using real-time quantitative PCR. In total, five gains (*EEA1*, *GLYAT*, *SUSD4*, *CYB5R1*, *TRMT6*) and seven losses (*CYP4A24*, *SCAND3*, *PLCD4*, *RICTOR*, *SPATC1*, *COL6A1*—two regions) were examined using isolated DNA from seven miniature pigs also used for sequencing and further two additional minipigs (Additional file 6: Table S4). All qPCR assays were designed following the TaqMan copy number assay guidelines (Thermo Fisher Scientific, Waltham, MA, USA). As a reference, the genomic region covering *CSK* (*C-terminal src kinase*) was used, which has not been reported to exhibit any CNVs in porcine studies so far but was found to be copy-number stable in the human genome [137, 138]. Reactions were prepared in triplicates, with each 20 µl multiplex reaction containing 20 ng DNA, 1X TaqPath ProAmp MasterMix (Thermo Fisher Scientific, Waltham, MA, USA), 1X assay (FAM dye), and 1X reference assay (VIC dye). qPCR was run on a QuantStudio 3 system (Applied Biosystems, Foster City, CA, USA) using a standard TaqMan copy number assay protocol: 95 °C for 10 min, followed by 40 cycles of 95 °C for 15 s and 60 °C for 60 s. The data were analysed using CopyCaller software v2.1 (Thermo Fisher Scientific, Waltham, MA, USA), and relative copy numbers were estimated based on the assumption of  $\Delta C_t = 0$  representing the diploid state.

#### CNV region detection and analysis

For each of the four CNV datasets, homozygous and heterozygous CNVs, including deletions (loss), duplications (gain), and any CNVs with higher copy numbers, were assembled into CNVRs. CNVs overlapping by at least one base pair were merged into CNVRs using the populationRanges function in CNVRanger v. 1.20.0 [139] in R 4.3.3 [132]. Genomic regions of low density (less than 10% of the total contributing individual calls within a combined region) were removed from the CNVRs to avoid false positives. The CNVRs were divided into deleted 'loss' and duplicated 'gain' regions, while overlapping 'gain' and 'loss' regions were combined into a single region categorised as 'both'.

Next, overlapping CNVRs of the same type across the four tool-specific datasets were combined if they met the mutual one bp overlap criterion and appeared in at least two methods. CNVRs in all nine miniature pig breeds were identified by retaining those found in at least two

individuals per breed. Regional CNVRs were defined by grouping the breeds (America, Asia & Oceania, and Europe) and retaining CNVRs present in every animal from that region. Singletons (CNVs present in only one individual) and doubletons (present in two individuals) were recorded but not included in the consensus CNVRs to ensure robustness. The distribution of CNVs across individuals and breeds is summarised in Additional File 2: Table S2.

The distribution of CNVRs was visualised as a bar plot using ggplot2 [133] and, additionally, with gene density, as an ideogram using RIdeogram v. 0.2.2 [140] in R 4.3.3 [132]. Gene density was calculated in 1-Mb windows based on the *Sus scrofa* 11.1 genome annotation. Chromosomal coordinates were retrieved from assembly metadata. CNVRs were plotted as overlays on gene density data to highlight regions of interest.

#### Enrichment of genes and quantitative trait loci overlapping CNVRs

Genome annotation was obtained from Ensembl (release 111), and pig QTL was retrieved from the animal QTL database [141] for the *Sus scrofa* 11.1 reference genome. An overlap of at least one bp between CNVRs and genomic features (genes or QTL) was used to identify potentially affected regions. A Fisher's exact test [142] was performed on all QTL overlapping with CNVRs to determine whether specific QTL were overrepresented compared to their overall genome distribution in R 4.3.3 [132]. The complete pig QTL dataset in the animal QTL database served as background references.

Functional enrichment analysis was performed by mapping pig genes to human orthologues using gprofiler2 v. 0.2.3 [143], then conducting enrichment analyses with enrichR v. 3.2 [144, 145] in R 4.3.3 [132], drawing upon "GO\_Biological\_Process\_2023" [146–148], "KEGG\_2021\_Human" [149–151], "MGI\_Mammalian\_PhenoType\_Level\_4\_2021" [152–154] and "PhenGenI\_Association\_2021" [155]. All terms with a  $p$ -value < 0.05 were considered significant. All Ensembl IDs were reviewed to remove duplicates from overlapping annotations in multi-copy CNVRs and avoid overrepresentation.

#### Analysis of exclusive CNVRs between different geographic regions

Non-overlapping, exclusive CNVRs specific to America, Asia & Oceania, or Europe were identified by comparing CNVRs from one region with those in other regions using GenomicRanges v. 1.56.1 [131] in R 4.3.3 [132]. Exclusive CNVRs did not overlap with CNVRs in other regions; for these region-specific CNVRs, associated genes and QTL were extracted, converting Ensembl to Entrez IDs with Ensembl biomaRt v. 2.60.1 [135, 136]. Subsequent functional annotation and biological theme comparisons

were performed using ClusterProfiler v. 4.12.3 [156, 157] with the *Sus scrofa* 11.1 reference genome in R 4.3.3 [132]. This analysis included GO biological processes and KEGG pathway enrichment to determine the biological themes represented by the genes overlapping with the exclusive CNVRs of each region. Finally, QTL linked to these exclusive CNVRs were summarised according to functional categories such as “growth and development”, “reproduction”, “meat and carcass quality”, and “health and immunity”, offering insights into region-specific trait–association.

#### Abbreviations

aCGH	Array-based Comparative Genomic Hybridisation
BWA	Burrows-Wheeler Alignment
CD36	CD36 molecule, thrombospondin receptor
CELSR1	Cadherin EGF LAG seven-pass G-type receptor 1
CHD1L	Chromodomain helicase DNA binding protein 1
CHD4	Chromodomain helicase DNA binding protein 4
CIT	Citron Rho-interacting serine/threonine kinase
CNOT9	CCR4-NOT transcription complex subunit 9
CNVRs	Copy number variant regions
CNVs	Copy number variants
COL6A1	Collagen type VI alpha 1 chain
CSK	C-terminal Src Kinase
CYP3A29	Cytochrome P450 family 3 subfamily A member 29
CYP4A24	Cytochrome P450 family 4 subfamily A member 24
EEA1	Early endosome antigen 1
EIF4A2	Eukaryotic translation initiation factor 4A2
ELOVL6	ELOVL fatty acid elongase 6
FISH	Fluorescence in situ hybridisation
GATK	Genome Analysis Toolkit
GO	Gene Ontology
GWAS	Genome-wide association study
INDEL	Insertion or deletion
KEGG	Kyoto Encyclopaedia of Genes and Genomes
MDS	Multidimensional scaling
MP	Mammalian Phenotype
NCBI	National Center for Biotechnology Information
NJ	Neighbour-joining
PAG6	Pregnancy-associated glycoprotein 6
PCA	Principal component analysis
PCR	Polymerase chain reaction
PLCD4	Phospholipase C delta 4
qPCR	Real-time quantitative PCR
QTL	Quantitative trait loci
RICTOR	RPTOR independent companion of MTOR complex 2
RLN	Relaxin family peptide hormone
SCAND3	SCAN domain containing 3
SNP	Single nucleotide polymorphism
SRA	Sequence Read Archive
SUSD4	Sushi domain containing 4
U6	U6 spliceosomal RNA
WGS	Whole genome sequencing

#### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-025-11446-8>.

Supplementary Material 1  
Supplementary Material 2  
Supplementary Material 3  
Supplementary Material 4  
Supplementary Material 5

Supplementary Material 6  
Supplementary Material 7  
Supplementary Material 8  
Supplementary Material 9  
Supplementary Material 10  
Supplementary Material 11  
Supplementary Material 12  
Supplementary Material 13  
Supplementary Material 14

#### Acknowledgements

We thank the MPI-MG Sequencing Core for help with sequencing and Mogens Kilian Drabert, Heike Klippert-Hasberg, and Jörn Wrede for technical assistance. Special thanks to Gerd Heinrichs for providing the Aachen minipigs investigated in this study. We also acknowledge the support of the North German Supercomputing Alliance (HLRN) for providing HPC resources. DeepL and Grammarly were used to enhance the clarity and readability of this manuscript. DeepL was employed for translation and improving grammar, expression, and spell-checking. Grammarly was used to improve grammar, phrasing, and spelling. Figure 1 was created using BioRender.com (agreement number HM278A8DHH).

#### Author contributions

J.B., J.M., and S.M. designed this study. J.B. and J.M. determined the methods used in this study. N.K. was responsible for sampling, library preparation and experimental validation. J.B. performed computational analyses. J.B. and J.M. prepared the manuscript draft. J.B., S.M., J.M., and N.K. reviewed and edited the manuscript. J.M. secured funding and supervised the project. All authors have read and approved the final version of the manuscript.

#### Funding

Open Access funding enabled and organized by Projekt DEAL. Stiftung Tierärztliche Hochschule Hannover (TIHO) (3134).

#### Data availability

All WGS data are available on the NCBI SRA (PRJNA635602, <https://www.ncbi.nlm.nih.gov/Traces/study/?acc=PRJNA635602>).

#### Declarations

##### Ethics approval and consent to participate

All animal experiments were conducted following national and international guidelines and were approved by the Animal Ethics Committee of the Lower Saxony State Veterinary Office Landesamt für Verbraucherschutz und Lebensmittelsicherheit, Oldenburg, Germany (registered under 33.9-42502-05-17A217). All sampled pigs were owned by the Institute of Animal Genomics. Informed consent for using the samples was obtained from the pig owner. Due to the low burden of sampling, no anaesthesia was applied.

##### Consent for publication

Not applicable.

##### Competing interests

The authors declare no competing interests.

##### Author details

- <sup>1</sup>Max Planck Institute for Molecular Genetics, Berlin, Germany
- <sup>2</sup>Institute of Chemistry and Biochemistry, Department of Biology, Chemistry and Pharmacy, Freie Universität Berlin, Berlin, Germany
- <sup>3</sup>Institute of Animal Genomics, University of Veterinary Medicine Hannover, Hanover, Germany
- <sup>4</sup>Institute for Medical and Human Genetics, Charité Universitätsmedizin Berlin, Berlin, Germany
- <sup>5</sup>Charité - Universitätsmedizin Berlin, BCRT - Berlin Institute of Health Centre for Regenerative Therapies, Berlin, Germany

Received: 30 October 2024 / Accepted: 5 March 2025

Published online: 22 March 2025

## References

1. Sindi S, Helman E, Bashir A, Raphael BJ. A geometric approach for classification and comparison of structural variants. *Bioinformatics*. 2009;25(12):i222–230.
2. Tang X, Zhu B, Ren R, Chen B, Li S, Gu J. Genome-wide copy number variation detection in a large cohort of diverse horse breeds by whole-genome sequencing. *Front Vet Sci*. 2023;10:1296213.
3. Cagan A, Blass T. Identification of genomic variants putatively targeted by selection during dog domestication. *BMC Evol Biol*. 2016;16:10.
4. Xia X, Zhang S, Zhang H, Zhang Z, Chen N, Li Z, Sun H, Liu X, Lyu S, Wang X, et al. Assessing genomic diversity and signatures of selection in Jiaxian red cattle using whole-genome sequencing data. *BMC Genomics*. 2021;22(1):43.
5. Coutelier M, Holtgrewe M, Jager M, Flottman R, Mensah MA, Spielmann M, Krawitz P, Horn D, Beule D, Mundlos S. Combining callers improves the detection of copy number variants from whole-genome sequencing. *Eur J Hum Genet*. 2022;30(2):178–86.
6. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shaperro MH, Carson AR, Chen W, et al. Global variation in copy number in the human genome. *Nature*. 2006;444(7118):444–54.
7. Sudmant PH, Rausch T, Gardner EJ, Handsaker RE, Abyzov A, Huddleston J, Zhang Y, Ye K, Jun G, Fritz MH, et al. An integrated map of structural variation in 2,504 human genomes. *Nature*. 2015;526(7571):75–81.
8. Zhang F, Gu W, Hurles ME, Lupski JR. Copy number variation in human health, disease, and evolution. *Annu Rev Genomics Hum Genet*. 2009;10:451–81.
9. MacDonald JR, Ziman R, Yuen RK, Feuk L, Scherer SW. The database of genomic variants: a curated collection of structural variation in the human genome. *Nucleic Acids Res*. 2014;42(Database issue):D986–992.
10. Henrichsen CN, Chaigat E, Raymond A. Copy number variants, diseases and gene expression. *Hum Mol Genet*. 2009;18(R1):R1–8.
11. Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, Thorne N, Redon R, Bird CP, de Grassi A, Lee C, et al. Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science*. 2007;315(5813):848–53.
12. Pos O, Radvanszky J, Buglyo G, Pos Z, Rusnakova D, Nagy B, Szemes T. DNA copy number variation: main characteristics, evolutionary significance, and pathological aspects. *Biomed J*. 2021;44(5):548–59.
13. Axelsson E, Ratnakumar A, Arendt ML, Maqbool K, Webster MT, Perloski M, Liberg O, Arnemo JM, Hedhammar A, Lindblad-Toh K. The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature*. 2013;495(7441):360–4.
14. Andersson L, Georges M. Domestic-animal genomics: deciphering the genetics of complex traits. *Nat Rev Genet*. 2004;5(3):202–12.
15. Estrada-Reyes ZM, Ogunade IM, Pech-Cervantes AA, Terrill TH. Copy number variant-based genome wide association study reveals immune-related genes associated with parasite resistance in a heritage sheep breed from the United States. *Parasite Immunol*. 2022;44(11):e12943.
16. Metzger J, Philipp U, Lopes MS, da Camara Machado A, Felicetti M, Silvestrelli M, Distl O. Analysis of copy number variants by three detection algorithms and their association with body size in horses. *BMC Genomics*. 2013;14:487.
17. Paudel Y, Madsen O, Megens HJ, Frantz LA, Bosse M, Bastiaansen JW, Crooijmans RP, Groenen MA. Evolutionary dynamics of copy number variation in pig genomes in the context of adaptation and domestication. *BMC Genomics*. 2013;14:449.
18. Keel BN, Lindholm-Perry AK, Snelling WM. Evolutionary and functional features of copy number variation in the cattle genome. *Front Genet*. 2016;7:207.
19. Wang J, Jiang J, Fu W, Jiang L, Ding X, Liu JF, Zhang Q. A genome-wide detection of copy number variations using SNP genotyping arrays in swine. *BMC Genomics*. 2012;13:273.
20. Revilla M, Puig-Oliveras A, Castello A, Crespo-Piazuelo D, Paludo E, Fernandez AI, Ballester M, Folch JM. A global analysis of CNVs in swine using whole genome sequence data and association analysis with fatty acid composition and growth traits. *PLoS ONE*. 2017;12(5):e0177014.
21. Al Abri MA, Holl HM, Kalla SE, Sutter NB, Brooks SA. Whole genome detection of sequence and structural polymorphism in six diverse horses. *PLoS ONE*. 2020;15(4):e0230899.
22. Li Y, Mei S, Zhang X, Peng X, Liu G, Tao H, Wu H, Jiang S, Xiong Y, Li F. Identification of genome-wide copy number variations among diverse pig breeds by array CGH. *BMC Genomics*. 2012;13:725.
23. Butty AM, Chud TCS, Miglior F, Schenkel FS, Kommadath A, Krivushin K, Grant JR, Hafliger IM, Drogemuller C, Canovas A, et al. High confidence copy number variants identified in Holstein dairy cattle from whole genome sequence and genotype array data. *Sci Rep*. 2020;10(1):8044.
24. Mills RE, Walter K, Stewart C, Handsaker RE, Chen K, Alkan C, Abyzov A, Yoon SC, Ye K, Cheetham RK, et al. Mapping copy number variation by population-scale genome sequencing. *Nature*. 2011;470(7332):59–65.
25. Teo SM, Pawitan Y, Ku CS, Chia KS, Salim A. Statistical challenges associated with detecting copy number variations with next-generation sequencing. *Bioinformatics*. 2012;28(21):2711–8.
26. Liu X, Chen W, Huang B, Wang X, Peng Y, Zhang X, Chai W, Khan MZ, Wang C. Advancements in copy number variation screening in herbivorous livestock genomes and their association with phenotypic traits. *Front Vet Sci*. 2023;10:1334434.
27. Mahmoud M, Gobet N, Cruz-Davalos DI, Mounier N, Dessimoz C, Sedlazeck FJ. Structural variant calling: the long and the short of it. *Genome Biol*. 2019;20(1):246.
28. Sedlazeck FJ, Lee H, Darby CA, Schatz MC. Piercing the dark matter: bioinformatics of long-range sequencing and mapping. *Nat Rev Genet*. 2018;19(6):329–46.
29. Logsdon GA, Vollger MR, Eichler EE. Long-read human genome sequencing and its applications. *Nat Rev Genet*. 2020;21(10):597–614.
30. Kosugi S, Momozawa Y, Liu X, Terao C, Kubo M, Kamatani Y. Comprehensive evaluation of structural variation detection algorithms for whole genome sequencing. *Genome Biol*. 2019;20(1):117.
31. Rafter P, Purfield DC, Berry DP, Parnell AC, Gormley IC, Kearney JF, Coffey MP, Carthy TR. Characterization of copy number variants in a large multibreed population of beef and dairy cattle using high-density single nucleotide polymorphism genotype data. *J Anim Sci*. 2018;96(10):4112–24.
32. Zhou J, Liu L, Reynolds E, Huang X, Garrick D, Shi Y. Discovering copy number variation in dual-purpose Xinjiang brown cattle. *Front Genet*. 2021;12:747431.
33. Hay EHA, Utsunomiya YT, Xu L, Zhou Y, Neves HHR, Carvalho R, Bickhart DM, Ma L, Garcia JF, Liu GE. Genomic predictions combining SNP markers and copy number variations in Nellore cattle. *BMC Genomics*. 2018;19(1):441.
34. Yang P, Wang G, Jiang S, Chen M, Zeng J, Pang Q, Du D, Zhou M. Comparative analysis of genome-wide copy number variations between Tibetan sheep and white Suffolk sheep. *Anim Biotechnol*. 2023;34(4):986–93.
35. Zhang Z, Peng M, Wen Y, Chai Y, Liang J, Yang P, Liu X, Li J, Huang Y, Li L, et al. Copy number variation of EIF4A2 loci related to phenotypic traits in Chinese cattle. *Vet Med Sci*. 2022;8(5):2147–56.
36. Warr A, Affara N, Aken B, Beiki H, Bickhart DM, Billis K, Chow W, Eory L, Finlayson HA, Fliecek P et al. An improved pig reference genome sequence to enable pig genetics and genomics research. *GigaScience*. 2020;9(6).
37. Bickhart DM, Liu GE. The challenges and importance of structural variation detection in livestock. *Front Genet*. 2014;5:37.
38. Chen C, Qiao R, Wei R, Guo Y, Ai H, Ma J, Ren J, Huang L. A comprehensive survey of copy number variation in 18 diverse pig populations and identification of candidate copy number variable genes associated with complex traits. *BMC Genomics*. 2012;13:733.
39. Mo J, Lu Y, Zhu S, Feng L, Qi W, Chen X, Xie B, Chen B, Lan G, Liang J. Genome-wide association studies, runs of homozygosity analysis, and copy number variation detection to identify reproduction-related genes in Bama Xiang pigs. *Front Vet Sci*. 2022;9:892815.
40. Wang Y, Zhang T, Wang C. Detection and analysis of genome-wide copy number variation in the pig genome using an 80 K SNP Beadchip. *J Anim Breed Genet*. 2020;137(2):166–76.
41. Wang L, Zhao L, Zhang L, Liu X, Hou X, Gao H, Yan H, Zhao F, Wang L. NTN1 affects porcine intramuscular fat content by affecting the expression of myogenic regulatory factors. *Anim (Basel)*. 2019;9(9).
42. Wei X, Shu Z, Wang L, Zhang T, Zhang L, Hou X, Yan H, Wang L. Copy number variations contribute to intramuscular fat content differences by affecting the expression of PELP1 alternative splices in pigs. *Anim (Basel)*. 2022;12(11).
43. Ramayo-Caldas Y, Crespo-Piazuelo D, Morata J, Gonzalez-Rodriguez O, Sebastia C, Castello A, Dalmau A, Ramos-Onsins S, Alexiou KG, Folch JM, et al. Copy number variation on ABC2-DNMBP loci affects the diversity and composition of the fecal microbiota in pigs. *Microbiol Spectr*. 2023;11(4):e0527122.
44. Fan S, Kong C, Chen Y, Zheng X, Zhou R, Zhang X, Wu X, Zhang W, Ding Y, Yin Z. Copy number variation analysis revealed the evolutionary difference



- between Chinese indigenous pigs and Asian wild boars. *Genes* (Basel). 2023;14(2).
45. Wang Y, Tang Z, Sun Y, Wang H, Wang C, Yu S, Liu J, Zhang Y, Fan B, Li K, et al. Analysis of genome-wide copy number variations in Chinese indigenous and western pig breeds by 60 K SNP genotyping arrays. *PLoS ONE*. 2014;9(9):e106780.
46. Dong K, Pu Y, Yao N, Shu G, Liu X, He X, Zhao Q, Guan W, Ma Y. Copy number variation detection using SNP genotyping arrays in three Chinese pig breeds. *Anim Genet*. 2015;46(2):101–9.
47. Shatokhin KS. Problems of mini-pig breeding. *Vavilovskii Zhurnal Genet Selektii*. 2021;25(3):284–91.
48. Prather RS, Lorton M, Ross JW, Whyte JJ, Walters E. Genetically engineered pig models for human diseases. *Annu Rev Anim Biosci*. 2013;1:203–19.
49. Lunney JK, Van Goor A, Walker KE, Hailstock T, Franklin J, Dai C. Importance of the pig as a human biomedical model. *Sci Transl Med*. 2021;13(621):eabd5758.
50. Morozov VA, Ludwig S, Ludwig B, Rotem A, Barkai U, Bornstein SR, Denner J. Islet cell transplantation from Gottingen minipigs to cynomolgus monkeys: analysis of virus safety. *Xenotransplantation*. 2016;23(4):320–7.
51. Dorado B, Ploen GG, Baretino A, Macias A, Gonzalo P, Andres-Manzano MJ, Gonzalez-Gomez C, Galan-Arriola C, Alfonso JM, Lobo M, et al. Generation and characterization of a novel knockin minipig model of Hutchinson-Gilford Progeria syndrome. *Cell Discov*. 2019;5:16.
52. Johansen T, Hansen HS, Richelsen B, Malmlof R. The obese Gottingen minipig as a model of the metabolic syndrome: dietary effects on obesity, insulin sensitivity, and growth hormone profile. *Comp Med*. 2001;51(2):150–5.
53. Liu X, Hu F, Wang W, Chen X, Niu X, Huang S, Wang Z, Wang J, Ran X. Genome-wide identification of copy number variations in wrinkled skin cases of Xiang pigs. *Sci Rep*. 2024;14(1):19695.
54. Le MT, Choi H, Lee H, Le VCQ, Ahn B, Ho CS, Hong K, Song H, Kim JH, Park C. SLA-1 genetic diversity in pigs: extensive analysis of copy number variation, heterozygosity, expression, and breed specificity. *Sci Rep*. 2020;10(1):743.
55. Duan J, Zhang JG, Deng HW, Wang YP. Comparative studies of copy number variation detection methods for next-generation sequencing technologies. *PLoS ONE*. 2013;8(3):e59128.
56. Cameron DL, Di Stefano L, Papenfuss AT. Comprehensive evaluation and characterisation of short read general-purpose structural variant calling software. *Nat Commun*. 2019;10(1):3240.
57. Chaisson MJ, Huddleston J, Dennis MY, Sudmant PH, Malig M, Hormozdiari F, Antonacci F, Surti U, Sandstrom R, Boitano M, et al. Resolving the complexity of the human genome using single-molecule sequencing. *Nature*. 2015;517(7536):608–11.
58. Lavrichenko K, Johansson S, Jonassen I. Comprehensive characterization of copy number variation (CNV) called from array, long- and short-read data. *BMC Genomics*. 2021;22(1):826.
59. Gabriellaite M, Torp MH, Rasmussen MS, Andreu-Sanchez S, Vieira FG, Pedersen CB, Kinalis S, Madsen MB, Kodama M, Demircan GS et al. A comparison of tools for copy-number variation detection in germline whole exome and whole genome sequencing data. *Cancers* (Basel). 2021;13(24).
60. Gross AM, Ajay SS, Rajan V, Brown C, Bluske K, Burns NJ, Chawla A, Coffey AJ, Malhotra A, Scocchia A, et al. Copy-number variants in clinical genome sequencing: deployment and interpretation for rare and undiagnosed disease. *Genet Med*. 2019;21(5):1121–30.
61. Roller E, Ivakhno S, Lee S, Royce T, Tanner S. Canvas: versatile and scalable detection of copy number variants. *Bioinformatics*. 2016;32(15):2375–7.
62. Mostovoy Y, Levy-Sakin M, Lam J, Lam ET, Hastie AR, Marks P, Lee J, Chu C, Lin C, Dzakula Z, et al. A hybrid approach for de novo human genome sequence assembly and phasing. *Nat Methods*. 2016;13(7):587–90.
63. Chakraborty A, Ay F. Identification of copy number variations and translocations in cancer cells from Hi-C data. *Bioinformatics*. 2018;34(2):338–45.
64. Wang H, Chai Z, Hu D, Ji Q, Xin J, Zhang C, Zhong J. A global analysis of CNVs in diverse Yak populations using whole-genome resequencing. *BMC Genomics*. 2019;20(1):61.
65. Liu G, Yang H, Yuan X. A shortest path-based approach for copy number variation detection from next-generation sequencing data. *Front Genet*. 2022;13:1084974.
66. Carvalho CM, Lupski JR. Mechanisms underlying structural variant formation in genomic disorders. *Nat Rev Genet*. 2016;17(4):224–38.
67. Feuk L, Carson AR, Scherer SW. Structural variation in the human genome. *Nat Rev Genet*. 2006;7(2):85–97.
68. Köhn F. History and development of miniature, micro- and minipigs. *Minipig Biomedical Res.*, CRC press, 2012:3–15., eBook ISBN 9780429149399
69. McNulty PA. The minipig in biomedical research. Boca Raton: CRC Press/Taylor & Francis; 2012.
70. Wang L, Zhang L, Li Y, Li W, Luo W, Cheng D, Yan H, Ma X, Liu X, Song X, et al. Data mining in networks of differentially expressed genes during sow pregnancy. *Int J Biol Sci*. 2012;8(4):548–60.
71. Klisch K, De Sousa NM, Beckers JF, Leiser R, Pich A. Pregnancy associated glycoprotein-1, -6, -7, and -17 are major products of bovine binucleate trophoblast giant cells at midpregnancy. *Mol Reprod Dev*. 2005;71(4):453–60.
72. Rohrer GA, Nonneman DJ. 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*. 2017;49(1):4.
73. Curtin JA, Quint E, Tsipouri V, Arkell RM, Cattanch B, Copp AJ, Henderson DJ, Spurr N, Stanier P, Fisher EM, et al. Mutation of *Celsr1* disrupts planar polarity of inner ear hair cells and causes severe neural tube defects in the mouse. *Curr Biol*. 2003;13(13):1129–33.
74. Glasco DM, Pike W, Qu Y, Reustle L, Misra K, Di Bonito M, Studer M, Fritzsche B, Goffinet AM, Tissir F, et al. The atypical Cadherin *Celsr1* functions non-cell autonomously to block rostral migration of facial branchiomotor neurons in mice. *Dev Biol*. 2016;417(1):40–9.
75. Shi DL. Wnt/planar cell polarity signaling controls morphogenetic movements of gastrulation and neural tube closure. *Cell Mol Life Sci*. 2022;79(12):586.
76. Robinson A, Escuin S, Doudney K, Vekemans M, Stevenson RE, Greene ND, Copp AJ, Stanier P. Mutations in the planar cell polarity genes *CELSR1* and *SCRIB* are associated with the severe neural tube defect craniorachischisis. *Hum Mutat*. 2012;33(2):440–7.
77. Kropp J, Carrillo JA, Namous H, Daniels A, Salih SM, Song J, Khatib H. Male fertility status is associated with DNA methylation signatures in sperm and transcriptomic profiles of bovine preimplantation embryos. *BMC Genomics*. 2017;18(1):280.
78. Jang J, Kim B, Jhang SY, Ahn B, Kang M, Park C, Cho ES, Kim YS, Park W, Kim H. Population differentiated copy number variation between Eurasian wild boar and domesticated pig populations. *Sci Rep*. 2023;13(1):1115.
79. Lundell K. Cloning and expression of two novel pig liver and kidney fatty acid hydroxylases [cytochrome P450 (CYP)4A24 and CYP4A25]. *Biochem J*. 2002;363(Pt 2):297–303.
80. Archer H, Soares RAN, Jafarikia M, Lillie BN, Schenkel F, Squires EJ. Effect of SNPs in the promoter region on the expression of cytochrome P450 2E1 (CYP2E1) in pig liver. *Anim (Basel)*. 2024;14(8).
81. Wang L, Zhang Y, Zhang B, Zhong H, Lu Y, Zhang H. Candidate gene screening for lipid deposition using combined transcriptomic and proteomic data from Nanyang black pigs. *BMC Genomics*. 2021;22(1):441.
82. Freiburg CD, Solomon-Degefa H, Freiburg P, Mörgelin M, Bolduc V, Schmitz S, Ala P, Muntoni F, Behrmann E, Bönnemann CG et al. The UCMD-causing COL6A1 (c.930 + 189C > T) intron mutation leads to the secretion and aggregation of single mutated collagen VI  $\alpha 1$  chains. *Hum Mutat*. 2023;2023.
83. Ropka-Molik K, Pawlina-Tyszko K, Zukowski K, Piorkowska K, Zak G, Gurgul A, Derebecka N, Wesoly J. Examining the genetic background of Porcine muscle growth and development based on transcriptome and MiRNAome data. *Int J Mol Sci*. 2018;19(4).
84. Heidartabar M, Huisman A, Krivushin K, Stothard P, Dervishi E, Charagu P, Bink M, Plastow GS. Imputation to whole-genome sequence and its use in genome-wide association studies for pork colour traits in crossbred and purebred pigs. *Front Genet*. 2022;13:1022681.
85. Garapaty S, Mahajan MA, Samuels HH. Components of the CCR4-NOT complex function as nuclear hormone receptor coactivators via association with the NRC-interacting factor NIF-1. *J Biol Chem*. 2008;283(11):6806–16.
86. Stuczynska A, Piorkowska K, Tyra M, Zukowski K. The effect of QTL-rich region polymorphisms identified by targeted DNA-seq on pig production traits. *Mol Biol Rep*. 2018;45(3):361–71.
87. Shiota C, Woo JT, Lindner J, Shelton KD, Magnuson MA. Multiallelic disruption of the rictor gene in mice reveals that mTOR complex 2 is essential for fetal growth and viability. *Dev Cell*. 2006;11(4):583–9.
88. Makkos A, Agg B, Varga ZV, Giricz Z, Gyongyosi M, Lukovic D, Schulz R, Bartekova M, Gorbe A, Ferdinandy P. Molecular network approach reveals rictor as a central target of cardiac ProtectomIRs. *Int J Mol Sci*. 2021;22(17).
89. Gomez-Del Arco P, Perdiguero E, Yunes-Leites PS, Acin-Perez R, Zeini M, Garcia-Gomez A, Sreenivasan K, Jimenez-Alcazar M, Segales J, Lopez-Maderuelo D, et al. The chromatin remodeling complex Chd4/NuRD controls striated muscle identity and metabolic homeostasis. *Cell Metab*. 2016;23(5):881–92.
90. Weiss K, Terhal PA, Cohen L, Brucoleri M, Irving M, Martinez AF, Rosenfeld JA, Machol K, Yang Y, Liu P, et al. De novo mutations in CHD4, an ATP-dependent

- chromatin remodeler gene, cause an intellectual disability syndrome with distinctive dysmorphisms. *Am J Hum Genet.* 2016;99(4):934–41.
91. Xu J, Fu Y, Hu Y, Yin L, Tang Z, Yin D, Zhu M, Yu M, Li X, Zhou Y, et al. Whole genome variants across 57 pig breeds enable comprehensive identification of genetic signatures that underlie breed features. *J Anim Sci Biotechnol.* 2020;11(1):115.
92. Kim H, Song KD, Kim HJ, Park W, Kim J, Lee T, Shin DH, Kwak W, Kwon YJ, Sung S, et al. Exploring the genetic signature of body size in Yucatan miniature pig. *PLoS ONE.* 2015;10(4):e0121732.
93. Simianer H, Kohn F. Genetic management of the Gottingen minipig population. *J Pharmacol Toxicol Methods.* 2010;62(3):221–6.
94. Kwon D, Ahn J, Kim H, Kim H, Kim J, Wy S, Ko Y, Kim J. Convergent dwarfism consequences of minipigs under independent artificial selections. *BMC Genomics.* 2024;25(1):761.
95. Swindle MM, Makin A, Herron AJ, Clubb FJ Jr, Frazier KS. Swine as models in biomedical research and toxicology testing. *Vet Pathol.* 2012;49(2):344–56.
96. Pedersen R, Ingerslev HC, Sturek M, Alloosh M, Cirera S, Christoffersen BO, Moesgaard SG, Larsen N, Boye M. Characterisation of gut microbiota in Ossabaw and Gottingen minipigs as models of obesity and metabolic syndrome. *PLoS ONE.* 2013;8(2):e56612.
97. Sturek M, Alloosh M, Sellke FW. Swine disease models for optimal vascular engineering. *Annu Rev Biomed Eng.* 2020;22:25–49.
98. Lloyd PG, Fang MY, Brisbin IL, Andersson L, Sturek M. AMP kinase gene mutation is consistent with a thrifty phenotype (metabolic syndrome) in a population of feral swine. *Faseb J.* 2006;20(4):A299–299.
99. Couvert A, Lacaze L, Touboulis S, Gautier S, Guerin S, Randuineau G, Rome V, Malbert CH, Val-Laillet D, Derbre F, et al. The Yucatan minipig model: A new preclinical model of malnutrition in obese patients with acute or chronic diseases. *Clin Nutr.* 2024;43(2):357–65.
100. Ardestani A, Lupse B, Maedler K. Hippo signaling: key emerging pathway in cellular and whole-body metabolism. *Trends Endocrin Met.* 2018;29(7):492–509.
101. Lee JB, Lim JH, Park HB. Genome-wide association studies to identify quantitative trait loci and positional candidate genes affecting meat quality-related traits in pigs. *J Anim Sci Technol.* 2023;65(6):1194–204.
102. Pan DJ. Hippo signaling in organ size control. *Gene Dev.* 2007;21(8):886–97.
103. Dang-Nguyen TQ, Tich NK, Nguyen BX, Ozawa M, Kikuchi K, Manabe N, Ratky J, Kanai Y, Nagai T. Introduction of various Vietnamese indigenous pig breeds and their conservation by using assisted reproductive techniques. *J Reprod Dev.* 2010;56(1):31–5.
104. Feng Shu-Tang CY-C. Studies on the biological characteristics, ex situ breeding and genetic diversity of Wuzhishan pig. *Biodiv Sci.* 1998;06(3):172–9.
105. Scudamore W. Small-scale outdoor pig breeding. *Crowood;* 2015.
106. Amalraj A, Matthijs A, Schoos A, Neiryck W, De Coensel E, Bernaerd E, Van Soom A, Maes D. Health and management of hobby pigs: a review. *Vlaams Diergen Tijds.* 2018;87(6):347–58.
107. Morel PCH, Lee TS, Moughan PJ. Effect of feeding level, live weight and genotype on the apparent faecal digestibility of energy and organic matter in the growing pig. *Anim Feed Sci Tech.* 2006;126(1–2):63–74.
108. Pawlowsky K, Ernst L, Steitz J, Stopinski T, Kogel B, Henger A, Tolba R. The Aachen minipig: phenotype, genotype, hematological and biochemical characterization, and comparison to the Gottingen minipig. *Eur Surg Res.* 2017;58(5–6):193–203.
109. Kohn F, Sharifi AR, Simianer H. Modeling the growth of the gottingen minipig. *J Anim Sci.* 2007;85(1):84–92.
110. Schachler K, Surie C, Minx J, Distl O, Metzger J. Estimation of the genetic diversity in the Mini-Lewe breeding population and influence of inbreeding on litter traits. *Zuchtungskunde.* 2019;91(3):227–45.
111. Laursen M, Olesen SP, Grunnet M, Mow T, Jespersen T. Characterization of cardiac repolarization in the Gottingen minipig. *J Pharmacol Toxicol Methods.* 2011;63(2):186–95.
112. Sedlarik K, Ebner E, Weidenbach H, Riedel E, Danz J. [Arterialization of the coronary veins as a therapeutic measure in acute myocardial ischemia. I. Experimental studies and morphology in miniswine]. *Z Exp Chir.* 1981;14(5):273–86.
113. Authier S, Gervais J, Fournier S, Gauvin D, Maghezzi S, Troncy E. Cardiovascular and respiratory safety pharmacology in Gottingen minipigs: pharmacological characterization. *J Pharmacol Toxicol Methods.* 2011;64(1):53–9.
114. Laursen M, Grunnet M, Olesen SP, Jespersen T, Mow T. Keeping the rhythm – pro-arrhythmic investigations in isolated Gottingen minipig hearts. *J Pharmacol Toxicol Methods.* 2011;64(2):134–44.
115. Descotes J, Allais L, Ancian P, Pedersen HD, Friry-Santini C, Iglesias A, Rubic-Schneider T, Skaggs H, Vestbjerg P. Nonclinical evaluation of immunological safety in Gottingen minipigs: the CONFIRM initiative. *Regul Toxicol Pharmacol.* 2018;94:271–5.
116. Flisikowska T, Egli J, Flisikowski K, Stumbaum M, Kung E, Ebeling M, Schmucki R, Georges G, Singer T, Kurome M, et al. A humanized minipig model for the toxicological testing of therapeutic recombinant antibodies. *Nat Biomed Eng.* 2022;6(11):1248–56.
117. Allais L, Perbet A, Condevaux F, Briffaux JP, Pallardy M. Immunosafety evaluation in juvenile Gottingen minipigs. *J Immunotoxicol.* 2022;19(1):41–52.
118. Wang SJ, Liu WJ, Yang LG, Sargent CA, Liu HB, Wang C, Liu XD, Zhao SH, Affara NA, Liang AX, et al. Effects of FUT1 gene mutation on resistance to infectious disease. *Mol Biol Rep.* 2012;39(3):2805–10.
119. Spielmann M, Mundlos S. Structural variations, the regulatory landscape of the genome and their alteration in human disease. *BioEssays.* 2013;35(6):533–43.
120. Allou L, Mundlos S. Disruption of regulatory domains and novel transcripts as disease-causing mechanisms. *BioEssays.* 2023;45(10):e2300010.
121. Berghöfer J, Khavesh N, Mundlos S, Metzger J. Simultaneous testing of rule- and model-based approaches for runs of homozygosity detection opens up a window into genomic footprints of selection in pigs. *BMC Genomics.* 2022;23(1):564.
122. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2010;26(5):589–95.
123. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. Genome project data processing S: the sequence alignment/map format and samtools. *Bioinformatics.* 2009;25(16):2078–9.
124. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010;20(9):1297–303.
125. Layer RM, Chiang C, Quinlan AR, Hall IM. LUMPY: a probabilistic framework for structural variant discovery. *Genome Biol.* 2014;15(6):R84.
126. Pedersen BS, Layer R, Quinlan AR. Smoove: structuralvariant calling and genotyping with existing tools. vers 0.2.8. Github; 2020. <https://github.com/brentp/smoove>
127. Rausch T, Zichner T, Schlattl A, Stutz AM, Benes V, Korbel JO. DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics.* 2012;28(18):i333–9.
128. Babadi M, Fu JM, Lee SK, Smirnov AN, Gauthier LD, Walker M, Benjamin DI, Zhao X, Karczewski KJ, Wong I, et al. GATK-gCNV enables the discovery of rare copy number variants from exome sequencing data. *Nat Genet.* 2023;55(9):1589–97.
129. Suvakov M, Panda A, Diesch C, Holmes I, Abyzov A. CNVpytor: a tool for copy number variation detection and analysis from read depth and allele imbalance in whole-genome sequencing. *Gigascience.* 2021;10(11).
130. Winberg J, Gustavsson P, Sahlin E, Larsson M, Ehren H, Fossum M, Wester T, Nordgren A, Nordenskjöld A. Pathogenic copy number variants are detected in a subset of patients with gastrointestinal malformations. *Mol Genet Genomic Med.* 2020;8(2):e1084.
131. Lawrence M, Huber W, Pages H, Aboyoun P, Carlson M, Gentleman R, Morgan MT, Carey VJ. Software for computing and annotating genomic ranges. *PLoS Comput Biol.* 2013;9(8):e1003118.
132. Core Team R. R: A language and environment for statistical computing. In, R 4.3.3 edn. Vienna, Austria: R Foundation for Statistical Computing; 2024.
133. Wickham H. ggplot2. Wiley interdisciplinary reviews: Comput Stat. 2011;3(2):180–5.
134. Tamura K, Stecher G, Kumar S. MEGA11: molecular evolutionary genetics analysis version 11. *Mol Biol Evol.* 2021;38(7):3022–7.
135. Durinck S, Moreau Y, Kasprzyk A, Davis S, De Moor B, Brazma A, Huber W. BioMart and bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics.* 2005;21(16):3439–40.
136. Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomart. *Nat Protoc.* 2009;4(8):1184–91.
137. Wang J, Wang H, Jiang J, Kang H, Feng X, Zhang Q, Liu JF. Identification of genome-wide copy number variations among diverse pig breeds using SNP genotyping arrays. *PLoS ONE.* 2013;8(7):e68683.
138. Johansson AC, Feuk L. Characterization of copy number-stable regions in the human genome. *Hum Mutat.* 2011;32(8):947–55.
139. da Silva V, Ramos M, Groenen M, Crooijmans R, Johansson A, Regitano L, Coutinho L, Zimmer R, Waldron L, Geistlinger L. CNVRanger: association analysis of CNVs with gene expression and quantitative phenotypes. *Bioinformatics.* 2020;36(3):972–3.

140. Hao Z, Lv D, Ge Y, Shi J, Weijers D, Yu G, Chen J. Rdeogram: drawing SVG graphics to visualize and map genome-wide data on the ideograms. *PeerJ Comput Sci.* 2020;6:e251.
141. Hu ZL, Park CA, Reecy JM. Bringing the animal QTLdb and CorrDB into the future: meeting new challenges and providing updated services. *Nucleic Acids Res.* 2022;50(D1):D956–61.
142. Fisher RA. On the interpretation of  $\chi^2$  from contingency tables, and the calculation of P. *J R Stat Soc.* 1922;85:87–94.
143. Kolberg L, Raudvere U, Kuzmin I, Vilo J, Peterson H. gprofiler2 -- an R package for gene list functional enrichment analysis and namespace conversion toolset g:Profiler. *F1000Res.* 2020;9.
144. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* 2016;44(W1):W90–7.
145. Xie Z, Bailey A, Kuleshov MV, Clarke DJB, Evangelista JE, Jenkins SL, Lachmann A, Wojciechowicz ML, Kropiwnicki E, Jagodnik KM, et al. Gene set knowledge discovery with enrichr. *Curr Protoc.* 2021;1(3):e90.
146. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. Gene ontology: tool for the unification of biology. The gene ontology consortium. *Nat Genet.* 2000;25(1):25–9.
147. Gene Ontology C. The gene ontology resource: enriching a gold mine. *Nucleic Acids Res.* 2021;49(D1):D325–34.
148. Gene Ontology C, Aleksander SA, Balhoff J, Carbon S, Cherry JM, Drabkin HJ, Ebert D, Feuermann M, Gaudet P, Harris NL et al. The gene ontology knowledgebase in 2023. *Genetics.* 2023;224(1).
149. Kanehisa M. Toward understanding the origin and evolution of cellular organisms. *Protein Sci.* 2019;28(11):1947–51.
150. Kanehisa M, Furumichi M, Sato Y, Ishiguro-Watanabe M, Tanabe M. KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res.* 2021;49(D1):D545–51.
151. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 2000;28(1):27–30.
152. Smith CL, Eppig JT. The mammalian phenotype ontology: enabling robust annotation and comparative analysis. *Wiley Interdiscip Rev Syst Biol Med.* 2009;1(3):390–9.
153. Smith CL, Eppig JT. Expanding the mammalian phenotype ontology to support automated exchange of high throughput mouse phenotyping data generated by large-scale mouse knockout screens. *J Biomed Semant.* 2015;6:11.
154. Smith CL, Eppig JT. The mammalian phenotype ontology as a unifying standard for experimental and high-throughput phenotyping data. *Mamm Genome.* 2012;23(9–10):653–68.
155. Ramos EM, Hoffman D, Junkins HA, Maglott D, Phan L, Sherry ST, Feolo M, Hindorff LA. Phenotype-genotype integrator (PheGenI): synthesizing genome-wide association study (GWAS) data with existing genomic resources. *Eur J Hum Genet.* 2014;22(1):144–7.
156. Xu S, Hu E, Cai Y, Xie Z, Luo X, Zhan L, Tang W, Wang Q, Liu B, Wang R et al. Using clusterprofiler to characterize multiomics data. *Nat Protoc.* 2024; 19:3292–320. <https://doi.org/10.1038/s41596-024-01020-z>
157. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, Feng T, Zhou L, Tang W, Zhan L, et al. ClusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innov (Camb).* 2021;2(3):100141.

## Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.