

# Genome-wide assessment characteristics of genes overlapping copy number variation regions in Duroc purebred population

**Zhipeng Wang**

NorthEast agricultural university <https://orcid.org/0000-0002-6041-8906>

**Yuanyuan Guo**

Northeast Agricultural University

**Tao Wang**

Northeast Agricultural University

**Chaoxin Zhang**

Northeast Agricultural University

**Shengwei Liu**

Northeast Agricultural University

**Bing Zhang**

Northeast Agricultural University

**Xue Li**

Northeast Agricultural University

**Li'ang Yang**

Northeast Agricultural University

**Qiusi Zhu**

Northeast Agricultural University

**Wenwen Wang**

Shandong Agricultural University

**Qin Zhang**

Shandong Agricultural University

**Xibiao Wang**

Northeast Agricultural University

**Qingli Meng** (✉ [475361814@qq.com](mailto:475361814@qq.com))

Northeast Agricultural University <https://orcid.org/0000-0003-4508-9953>

---

## Research article

**Keywords:** Copy number variations, genome-wide, Duroc purebred pig, CNV genes, dosage-sensitive genes, CNV-miRNAs

**Posted Date:** December 4th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-118996/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

# Abstract

**Background:** Copy number variations (CNVs) are important structural variations that can cause significant phenotypic diversity. Reliable CNVs mapping can be achieved by identification of CNVs from different genetic backgrounds. Investigations on the characteristics of overlapping CNV regions (CNVRs) between protein-coding genes (CNV genes) and miRNAs (CNV-miRNAs) can reveal the potential mechanisms of their regulation.

**Results:** In this study, we used 55K SNP arrays to detect CNVs in Duroc purebred pig. A total number of 211 CNVRs were detected with a total length of 118.48 Mb, accounting for 5.23% of the autosomal genome sequence. Of these CNVRs, 32 were gains, 175 losses, and 4 contained both types (loss and gain within the same region). The CNVRs we detected were non-randomly distributed in the swine genome and were significantly enriched in the segmental duplication and gene density region. Additionally, these CNVRs were overlapping with 1096 protein-coding genes, and 39 miRNAs, respectively. The CNV genes were enriched in terms of dosage-sensitive gene list. The expression of the CNV genes was significantly higher than that of the non-CNV genes in the adult Duroc liver and prostate. Of all detected CNV genes, 252 genes, which accounted for 22.99%, were tissue-specific (TSI > 0.9). Strong negative selection had been underway in the CNV genes as the ones that were located entirely within the loss CNVRs appeared to be evolving rapidly as determined by the median dN plus dS values. Non-CNV genes tended to be miRNA target than CNV genes. Furthermore, CNV-miRNAs tended to target more genes compared to non-CNV-miRNAs, and a combination of two CNV-miRNAs preferentially synergistically regulated the same target genes. We also focused our efforts on examining CNV genes and CNV-miRNAs, which were also involved in the lipid metabolism, including *DGAT1*, *DGAT2*, *MOGAT2*, miR143, miR335, and miRLET7.

**Conclusions:** Our analyses of CNV-genes and CNV-miRNAs provide new insights into the characteristics of CNVRs in Duroc purebred population. Further molecular experiments and independent large studies are needed to confirm our findings.

## Background

Based on the size of the segment modified, genome sequence variations can be divided into three main types: single nucleotide polymorphisms (SNPs), short insertions and deletions (Indels), and structural variations (SVs). The length of SVs is at least 50 bp (base pairs) (Mills et al., 2011). Generally, SVs contains two variation subtypes, including unbalanced (*e.g.*, deletions, duplications, and insertions) and balanced forms (*e.g.*, inversions and translocations) (Feuk et al., 2006). Unbalanced forms SVs induce copy-number variations. Recent findings have shown that structural DNA variations are widespread in animal genomes, such as those of rodents (Graubert et al 2007) and primates (Freeman et al., 2006). The copy number variation (CNV) has been considered a major type of structural variations, with a length ranging from one to several Mb (Feuk et al., 2006).

Currently, with recent advances in high-throughput sequencing technologies, various approaches can be applied to perform genome-wide CNV mapping, including DNA hybridization in BAC/PAC/oligonucleotide arrays, high-density SNP chips, and next-generation sequencing. Using genome-wide technologies of higher resolution, tremendous quantities of CNVs have been identified in many farm animal species, such as cattle (Liu et al., 2010; Mei et al., 2020), pig (Ramayo-Caldas et al. 2010; Jiang et al., 2014; Wang et al., 2015), sheep (Liu et al., 2013; Di Gerlando et al., 2019), and chicken (Griffin et al. 2008; Seol et al., 2019).

As in other domestic animals, reliable detection of CNVs in swine is still challenging, with a low concordance among different studies. Some evidence suggests that multiple populations should be surveyed to construct an accurate CNV map (Ramayo-Caldas et al., 2010; Liu et al., 2010). Identifying CNVs from different genetic backgrounds can validate the data on CNV regions discovered in various investigations and achieve reliable CNVs mapping that describes the genome-wide characteristics of various populations.

By molecular mechanisms, such as gene disruption, gene fusion, positive effect, and dosage effect, CNVs can cause Mendelian disease or traits, or be associated with complex disease or quantitative traits (Lupski and Stankiewicz, 2005). CNVs affect the phenotypic variation in domestic animal genomes. For example, Fliskowski *et al.* (2010) identified a 110-kb deletion of the *MIMT1* gene in the cattle genome, which was associated with abortions and stillbirth phenotype. Additionally, some chicken exterior traits, including feather growth, comb shape, dark brown plumage, and dermal hyperpigmentation, are caused by copy number variations of *PRLR*, *SOX5*, *SOX10*, and *EDN3* genes, respectively (Elferink et al. 2008; Wright et al. 2009; Gunnarsson et al., 2011; Dorshorst et al., 2011). The dominant white color of swine has been associated with a duplication of a 450-kb fragment encompassing the *KIT* gene (Giuffra et al. 2002; Seo et al., 2007). Recent studies found a high frequency in miRNA copy number abnormality. In this respect, Marcinkowska et al. (2011) detected miRNAs located in the human CNVR that also had potential functional variants. Moreover, Willemsen et al. (2011) reported a deletion of 1p21.3 containing MIR137, which induced miRNA downregulation and upregulation of targets in subjects with congenital abnormalities. The aforementioned examples reveal the association between copy number change and gene function, which leads to alteration of some phenotypes. Thus, characteristics of genes overlapping CNVRs are to be investigated, and the potential regulatory mechanisms of these genes are to be analyzed and established.

In this study, we performed genome-wide CNVR mapping in a Duroc swine population using a 55K SNP Chip. Our findings provide a useful addition to swine genomic structure variations and validate CNVs detected in previous investigations. Furthermore, we have presented the structure and characteristics of the CNV map obtained and had discussed in detail the impact of CNVRs on gene morphology and function.

## Results

### Genome-wide identification and characterization of CNVs

We identified a total number of 1371 CNVs within the autosome genome of Duroc populations in the (Table 1), whose sizes ranged from 8.37 kb to 2838.50 kb. The average and the median sizes were 386.30 kb and 270.05 kb, respectively. The copy number losses were 28.17 times more frequent than the copy number gains (1324 losses versus 47 gains). The size of the CN losses and CN gains ranged from 8.37 kb to 2838.50 kb and from 33.37 kb to 998.00 kb, respectively. The median and average sizes of the CN losses (270.05 kb and 390.39 kb) were slightly larger than those of the CN gains (223.71 kb and 271.03 kb). The distribution of CNVs size ranges are illustrated in Figure 1. In this Duroc swine population, 205 individuals had CNVs, whereas three individuals were without CNVs, with an average number of CNVs per individual genome of 6.59.

### Characteristics of porcine CNVRs

By aggregating the overlapping CNVs, a total number of 211 CNVRs across the autosomes were identified, which covered 118.48 Mb of the swine genome and corresponded to 5.23% of the length of the autosomal sequence. Of these CNVRs 175 were losses and 32 gains, whereas 4 contained both events which were within the following ranges 8.34–3882.48 kb, 19.98–1035.01 kb, and 596.40–2516.40 kb, respectively. The CNVRs of losses, gains, and both events had means or medians of 596.30 kb or 346.87 kb, 277.46 kb or 231.12kb, and 1312.96 kb or 1069.49 kb, correspondingly. The distribution of the CNVRs size ranges is depicted in Fig. 1. In this study, the loss events were approximately 5.47-fold more common than the gain events.

The numbers of CNVRs in each autosomal chromosome are presented in Table 1, and the location and characteristics of all CNVRs are displayed in Figure 2. Using regression analysis, we found a significant positive linear relationship between the chromosome sequence length and the number of CNVR located on it ( $P = 5.13\text{E-}4$ ) (Figure 3). Longer chromosomes had higher numbers of CNVRs located on them.

These CNVRs were unevenly distributed among different chromosomal regions. Certain chromosomal regions had a relatively high density of CNVs regions, such as each end of SSC1 and SSC2. We found a positive correlation between the number of CNVs and gene density; 77.82% of the total number of CNV regions were located on protein-coding genes, which were significantly more than non-CNV regions ( $P = 1.65\text{E-}4$ ).

In the genomes of many mammals, SD is a necessary condition and catalyst for CNVs formation. In this study, we found that 21.44% of the CNVR sequences (25.63 Mb/119.49 Mb) directly overlapped with porcine SD regions, as obtained from the results of Feng *et al.* (2014). Using Chi-squared test, we established significant enrichment of CNVRs in the SD region ( $P = 3.67\text{E-}9$ ). It is noteworthy that CNVs are known to co-occur with SDs, and SDs are more abundant in some locations of the porcine genome. Therefore, based on our findings, we suggest that porcine CNVRs are not uniformly distributed in the genome.

### Annotation analysis of swine CNVRs

Using the data obtained from the porcine QTL database, the CNVRs identified in this study were overlapping with 4472 QTLs, including 296 exterior, 643 health, 2715 meat and carcass, 499 production, and 319 reproduction traits.

Of the 211 CNVRs we identified, 30 CNVRs did not include genes as annotated in the ENSEMBL database. The loss CNVRs were overlapping with 1034 (26 miRNAs), the gain ones with 60 (10 miRNAs), and those involving both types and 41 (3 miRNAs) genes. The total number of genes was 1135, including 1096 protein-coding genes (also called CNV-gene) (such as the olfactory receptor gene family, taste receptor gene family, *DGAT1*, *PPARA*), and 39 miRNA genes, also called CNV-miRNA, including MIR143, MIR335, and MIRLET7.

Of those 1096 CNV genes 145 were dosage-sensitive genes (86 monoallelic expression, 22 imprinting expression, and 37 other type dosage-sensitive genes) and 77 transcription factor genes belonging to 70 TF gene families. Using Fisher's exact test, we found dosage-sensitive gene ( $P = 3.50\text{E-}3$ ), imprinted genes ( $P = 7.64\text{E-}4$ ) and monoallelically expressed genes  $P = 4.67\text{E-}2$  enriched in CNV genes, respectively, whereas TF genes were neither enriched nor underrepresented in CNV genes ( $P > 0.05$ ). In addition, 602 CNV genes (54.93%) were regulated by miRNAs, while 7872 non-CNV genes (58.13%) were target genes of miRNAs. Therefore, target genes encompassed more non-CNV than CNV genes ( $P = 4.18\text{E-}2$ ).

Here, we compared the expression profiles of CNV and non-CNV genes in 27 adult Duroc pig tissue types. We found that the expression of CNV genes was significant higher than that of non-CNV genes in the liver ( $P = 3.99\text{E-}3$ ) and the prostate ( $P = 2.62\text{E-}5$ ). Of those CNV genes, 252 or

22.99% were tissue-specific (TSI > 0.9). However, 24.94% of the non-CNV genes (3390) were tissue-specific genes, which were significantly more than CNV genes ( $P = 4.99E-2$ ).

To evaluate the functional annotation of these CNVRs, we conducted gene ontology (GO) and KEGG enrichment analyses using 1096 CNV genes. We found 34 GO terms and 7 KEGG pathways that were statistically significant ( $P < 0.05$ ) (Table 2). Of the major GO terms 14 were associated with biological processes, 14 with cellular components, and 6 with molecular function categories, such as skeletal muscle cell differentiation (GO:0035914), glucose metabolic processes regulation (GO:0010906), and negative regulation of oligodendrocyte differentiation (GO:0048715). Significant KEGG pathways included the Hippo signaling pathway (ssc04390), Wnt signaling pathway (ssc04310), taste transduction (ssc04742), and glycerolipid metabolism (ssc00561).

### Classification and characteristics of CNV-gene based on structural relationship

Based on the structural relationships between protein-coding genes and CNVRs, we classified the genes into three types, as previously suggested by Woodward and Bateman (2011). Type I CNV gene was contained entirely within the CNV. Type II CNV gene partially overlapped the CNV, which were often disrupted and even with fusion genes formed. Type III genes were those that contained the CNV within the gene. In this study, we identified are 862 type I, 206 type II, and 28 type III CNV genes (Table 3).

To better understand the biology of the aforementioned three types of CNV genes, we investigated their basic characteristics, selective pressures, and functional annotation. On average, type I CNV genes were shorter (43.86 kb) than type II ones (201.33 kb); type III genes were the longest (461.52 kb). Obviously, type I genes were included within CNVs, whereas type III genes contained CNVs.

The dN and dS values of the pig/human ortholog were obtained from Ensembl Compare database using PAML. Then, the Kolmogorov-Smirnov test results showed that the mean or median dN plus dS values of type I genes were higher than those of non-CNV genes (Table 3). Therefore, type I genes tend to be rapidly evolving and to have increased mutation rates. Based on the dN + dS values, we also established that type II and III genes mutated more slowly than non-CNV genes ( $P = 4.24E-3$ ,  $P = 5.64E-3$ , respectively). Additionally, the median or mean dN + dS values of the genes overlapping the loss type CNVRs were higher than those of non-CNV genes ( $P = 2.2E-16$ ). Our findings evidence that type I genes overlapping loss CNVRs were rapidly evolving. The very low dN/dS values of all CNV genes would show that strong negative selection is acting on them. Those genes were remained due to genetic drift or difficulties to remove on genome.

Type I genes tend to be the involved in the regulation of glucose metabolic processes (GO:0010906), skeletal muscle cell differentiation (GO:0035914), glycerolipid metabolism (ssc00561), Wnt signaling pathway (ssc04310), and taste transduction (ssc04742). On the other hand, type II genes were associated with vesicle-mediated transport (GO:0016192), intracellular protein transport (GO:0006886), and Hippo signaling pathway (ssc04390).

### Characteristics of the target genes participating in CNV-miRNAs regulation

A total number of 39 miRNAs were located in the Duroc CNVRs, which were called CNV-miRNAs. The remaining miRNAs were referred to as non-CNV miRNAs. Using Wilcox rank-sum test, we found that the number of CNV-miRNA target genes and binding sites were significantly higher than those of non-CNV-miRNA ( $p_{\#TargetGene} = 0.0192$ ,  $p_{\#BindingSites} = 0.0243$ ), respectively. Thus, CNV-miRNAs appeared to regulate more target genes than non-CNV-miRNAs.

To characterize CNV-miRNA target genes, we classified miRNA target genes into three groups, as described earlier Wu *et al.* (2012). The first target genes group had 368 genes regulated exclusively by CNV-miRNAs. Of these, 361 target genes were regulated by one CNV-miRNA, whereas the remaining target genes were regulated by two CNV-miRNAs. The second target gene group included 38 genes regulated by a combination of non-CNV miRNAs and more than one CNV-miRNA. The third group contained 528 target genes regulated only by non-CNV miRNAs. Of these, 504 target genes were regulated by one non-CNV miRNA.

To investigate the target-recognition preference of CNV-miRNAs, we used a sampling simulation strategy to identify whether the observed number of target genes for each regulatory type could be expected from random sampling. These simulations provided clues for identifying the regulatory patterns of CNV-miRNAs. We found the number of target genes regulated only by two CNV-miRNAs was significantly higher than the expected after the application of random simulations ( $P = 3.57E-2$ ). In this study, we found seven target genes that were regulated exclusively by two CNV-miRNAs, such as *CRK* gene regulated by miR-4331 and miR9817. That is, some genes are preferentially targeted by combination of some CNV-miRNAs. Obviously, the copy number alterations of one miRNA influences that of other miRNAs if their binding sites are co-located in the same UTRs. The dosage of miRNAs should be balanced to synergistically regulate the same genes.

In this study, we performed gene ontology (GO) and KEGG enrichment analyses using three groups of target genes regulated by the CNV-miRNAs and non-CNV miRNAs. ErbB signaling pathway (ssc04012) was enriched in the first group of target genes, whereas KEGG pathways

analysis in the third group of target genes showed significantly enriched sphingolipid signaling pathway (ssc04071), NF-kappa B signaling pathway (ssc04064), and Wnt signaling pathway (ssc04310).

## Discussion

### Characteristics of the CNVRs distribution on Duroc genome

Recently, accumulating evidence has indicated the widespread distribution of CNVs in the genome. Furthermore, their involvement in genetic variation, phenotypic diversity, and evolutionary adaptation has been acknowledged as a major contribution (Yim et al, 2010). At least 10% of the human genome is considered to be covered by CNVs (Redon et al. 2006; Wong et al. 2007), and speculations exist that human CNVRs may cover up to 13% of the genome sequences (Stankiewicz & Lupski 2010). Additionally, Cutler *et al.* (2007) discovered that CNVs constituted up to 10% of the mouse genome. In this study, 211 CNVRs were identified, which accounted for 5.23% of the autosomal sequences. Nevertheless, this figure might be conservative, because some CNVs could not be detected, including small (< 10 kb) and large CNVs, which is possibly due to the small sample size and low homology probes, and as well as to limitations of current reference genomes such as sequence gaps.

Here, the abundance of loss CNVR events was approximately 5.47-fold higher than that of gain CNVR events, which is consistent with findings obtained in previous studies on cattle (Bae et al., 2010; Hou et al., 2011), goat (Fontanesi et al., 2010) and sheep (Fontanesi et al., 2011; Liu et al., 2013; Hou et al 2015). This result might be explained by action of biological factors, as suggested by Fadista *et al.* (2010). Initially, non-allelic homologous recombination (NAHR) seemed to be one of the main mechanisms responsible for CNVs formation (Zhang et al., 2009). For example, Locke *et al.* (2006) suggested losses were under stronger selection than gains. In this respect, Turner *et al.* (2008) showed that NAHR tended to generate more loss than gain. Furthermore, our results also confirm that type I CNV genes overlapping loss CNVRs appear to be rapidly evolving.

We observed that CNVRs tended to have a non-uniform distribution in the porcine genome and were enriched in the gene density and segmental duplications regions. In the human genome, CNVRs were found to genome examinations genome examinations be more frequently located in some regions in the genome and chromosomes such as the pericentromeric and the subtelomeric regions (Zarrei et al 2015). Studies have shown that the non-uniform distribution of CNVRs may arise from nearby repetitive sequences. Moreover, human CNVs were significantly overrepresented in simple tandem repeat sequences (Stankiewicz and Lupski 2002; Hurles 2005; Lupski and Stankiewicz 2005). Nguyen et al. (2006). In primate genome, CNVs were discovered to occur together with SDs (Kim *et al.*, 2008). Therefore, SDs may promote CNV formation (Dumas et al., 2007; Lee et al., 2008). In addition, human CNVRs were observed to be unusually enriched in protein-coding genes. The elevated gene density of CNVs might have been caused by the retention of duplicated sequences that were of adaptive benefit (Nguyen et al., 2006).

### Comparison of our findings with those of previous studies on porcine CNVR

Twenty studies were focused on genome-wide identification of porcine CNVs. SNP genotyping platforms, array-based comparative genomic hybridization (aCGH), and next-generation sequencing were applied in these investigations, and a total number of 16,396 CNVRs were detected, with a total length of CNVR ranging from 9.66 Mb to 560.30 Mb in the different studies (Table 4). The CNVR distributions established in each of the genome examinations are presented in Table 4. Of the 211 CNVR detected in this research, 191 had been also previously detected in earlier studies. These results indicate that approximately 90% of the CNVRs identified here can be validated by previous investigations, whereas 10% of our findings are original, first detected herein.

The most overlapped CNVR counts (98) were consistent with those reported by Stafuzza *et al.* (2019), who detected 3520 CNVR events based on the SNP chip data of 3520 Duroc pigs. Additionally, Chen *et al.* (2012), Jiang *et al.* (2014), Wang *et al.* (2015a), Long *et al.* (2016), and Keel *et al.* (2019) used SNP chip or NGS platforms to identify CNVRs in Duroc pig populations. A total number of 54, 96, 88, 46, and 75 CNVR were identified in these studies that were overlapping, respectively. These results implied that these overlapping CNVRs contained some Duroc breed genome-specific CNVRs.

Remained reports have lower proportion CNVRs overlapped with our study. The issue of low overlapping rates between different reports was also occurred in CNV studies of other studies. We deem that the following reasons could have contributed to the observed differences. First, the study populations of different breeds have various genetic backgrounds. Many previous studies have also shown the presence of breed/line-specific CNVRs in the genome. Chen *et al.* (2012) surveyed CNVs in 18 diverse pig populations and discovered that only 20 CNVRs of the 565 CNVRs were available in more than nine pig populations, whereas most CNVRs (72.9%) were limited to only one pig population. Second, there are differences in the sampling methods and genetic drift events among studies on the same breed. Third, different detected platforms have been used, CGH arrays, SNP genotyping, or NGS. Finally, many structural variations in the genome might have remained undiscovered.

### Duroc CNV genes morphology and functions

The varying copy number of CNV genes changes gene expression due to altered gene dosage and disruption effects by gene structural variations. If CNV is located in the coding region, it alters the protein function, whereas its location in the regulatory region changes the gene expression level. Dosage sensitivity of the included genes is the most popular hypothesis that attempts to explain pathogenic CNVs. We discovered that dosage genes, imprinted genes, and monoallelically expressed genes were enriched in the CNV regions in the genome of Duroc pigs. The CNV regions of the human genome may be its most dosage-sensitive regions, in which CNVs are likely to be associated with disease development (Zarrei et al., 2015). However, it is worth emphasizing that changes in gene copy number do not always lead to differences in gene expression. Many factors, such as lack of regulatory elements in duplication event, the chromatin environment, and dosage compensation, might maintain stable mRNA levels.

Structural variation and miRNA are two genetic elements which affect gene expression and regulation. Here, we predicted the potential number of miRNA targets of various genes that were located either in CNV or non-CNV regions. In the genome of a Duroc pig population, we found that miRNAs regulated less CNV than non-CNV genes, but the mean number of miRNA per CNV-gene is similar to that per non-CNV gene. Felekakis *et al.* (2011) and Jovelin (2015) demonstrated that the genes located in the CNV regions of the human genome were targeted by more miRNA molecules, and CNV genes had more miRNA-binding sites than non-CNV genes. Similarly to the human genome, miRNA regulates more the CNV genes in the fruit fly genome than non-CNV genes. However, Jovelin (2015) argued that this principle was not universal. In this previous study, worm and zebrafish showed the opposite pattern and had significantly more miRNAs and target sites per non-CNV genes. Therefore, structure variations such as duplication and deletion do not necessarily lead to increased miRNA target sites for CNV genes. The distinct results among species could result from functional differences between CNV and non-CNV genes, differential abundance of CNV types, and the accuracy of CNV annotations. The evolutionary interaction between miRNAs and CNVs could have been obscured by interspecies differences.

Previous reports have addressed the impact of CNVs on the phenotypic variation of domestic animals species. For instance, Clop *et al.* (2012) supposed that bridging the gap between CNV genotypes and complex phenotypes will be the next genetic challenge. In addition, Fontanesi *et al.* (2011) showed that duplication of the *ASIP* (agouti-signaling protein) locus was associated with a grey coat in the Massese sheep. The majority of CNVRs identified in this study overlapped with pig QTLs. Earlier, Paudel *et al.* (2015) hypothesized that copy number variations provided the means for rapid adaptation to different environments during speciation/diversification. Here, we also deem that some genes with CNVs have had a possibly prominent role in the ongoing speciation, and might have impacted certain phenotypes through gene dosage alteration or *via* a positional effect, in which the structural variant might have altered the genomic landscape of the regulatory elements modulating the expression of these genes.

According to the enrichment analysis results and the already known basic gene function, genes related to fat metabolic were identified in this study, including *DGAT1*, *DGAT2*, *MOGAT2*, *AGPAT2*, *FABP1*, *PPARA*, *ANGPTL3*, and *NPC2*. The basic functions of these genes are described below (Table 5).

*DGAT1* and *DGAT2* control the synthesis of triglycerides and transmembrane proteins localized in the endoplasmic reticulum, which affects fat metabolism and lipid deposition in tissues. They also participate in the regulation of energy synthesis and catabolism. *MOGAT2* is critically involved in the uptake of dietary fat by the human small intestine. Additionally, the *AGPAT2* gene has been associated with congenital generalized lipodystrophy, characterized by severe insulin resistance and near-absence of adipose tissue. The roles of *FABP1* are associated with fatty acid uptake, transport, and metabolism, related to the lipid metabolism regulation by the PPAR signaling pathway. *ANGPTL3* acts in part as a hepatokine and is involved in the regulation of lipid and glucose metabolism (Koishi et al., 2002; Tikka et al., 2014). The *NPC2* gene has important functions in the transfer of cholesterol from the human lysosome (Infante et al., 2008).

### Functions and regulation of Duroc CNV-miRNAs

Previous studies on human CNV discovery have reported the presence of copy number variable miRNA genes (Wong et al 2007, Lin et al 2008). For example, Marcinkowska *et al.* (2011) found that approximately 30% of genome miRNAs were located in the human CNV regions. Additionally, Ha et al. (2009) discovered that miRNAs had an equilibrating role in genomic dosage phenomena. The results of numerous studies have clearly evidenced the feasibility of using the dysregulation of CNV-ncRNAs as a biological marker for disease screening. In this study, we detected 39 miRNA genes that overlapped with CNVRs, including some miRNAs involved in precursor adipocyte differentiation and lipid deposition, such as *MIR143*, *MIR335*, *MIR378*, and *MIRLET7* (Table 5). An earlier study by An et al. (2016) revealed that *MIR143* was promoted the adipogenic differentiation of porcine bone marrow-derived mesenchymal stem cells. In another investigation, Li *et al.* (2016) evaluated differentially expressed liver miRNAs between Tibetan and Yorkshire pigs and identified differentially expressed miRNAs (*MIR335* and *MIR378*) that participated in the glucose and lipid metabolism. It is noteworthy that Li *et al.* (2011) adopted a deep sequencing approach to determine the identity and abundance of miRNAs in swine adipose tissue development and found that *MIR143* and *MIRLET7* were the miRNAs with the highest expression.

Our present analysis results indicate that in the porcine genome CNV-miRNAs tend to target a higher number of genes than non-CNV-miRNAs with a pattern similar to that in the human genome, earlier established by Wu *et al.* (2012). These scientists also found that this regulation

model might play important roles in the prevention of CNV-miRNA purification. From an evolutionary viewpoint, certain CNV-miRNAs seem to have beneficial effects on biological processes in organisms. Our further analysis revealed that genes targeted by CNV-miRNAs participate in a wide range of biological responses to environmental factors. Obviously, CNV-miRNAs provide a possibility of increasing regulatory complexity using a strategy that increases the number target genes.

## Conclusions

In this study, we constructed a CNVR map for the Duroc pig population and identified 211 CNVRs, accounting for 5.23% of the autosomal genome. These CNVRs were non-randomly distributed in the porcine genome and were significantly enriched in the segmental duplication and gene density regions. These CNVRs overlapped with 1096 genes, 39 miRNA, and 4472 QTLs. Of these protein-coding genes, we established that dosage sensitivity and imprinted and monoallelically expressed genes were enriched in CNV genes. Strong negative selection was acting on CNV genes, and the genes contained entirely within the loss CNVRs appeared to be rapidly evolving. CNV-miRNAs tended to target more genes than non-CNV-miRNAs, and a combination of two CNV-miRNAs was found to preferentially synergistically regulate the same genes. Some CNV genes and CNV-miRNAs involved in the lipid metabolism were assessed in this study, including *DGAT1*, *DGAT2*, *MOGAT2*, miR143, miR335, and miRLET7. Nevertheless, further molecular experiments and independent large studies are needed to validate our findings.

## Methods

### Animal population

A total number of 208 Duroc pigs (10 males and 198 females) were used in this study, which were obtained from the whole foundation herd of the Beijing Breeding Swine Center. The pigs were located on the same farm, under similar environmental conditions and an identical standard feeding schedule. All animals were inspected for the presence of open wounds, any illness, or abnormal behavior. All pigs are alive and without genetic modification.

### SNP genotyping and quality control

We used the phenol-chloroform method to extract genomic DNA from blood. Genotyping of a total number of 50,703 SNPs across the whole genome was performed using the GeneSeek Porcine 50K SNP Chip (Neogen, Lincoln, NE, USA). We performed the following quality control through PLINK (V1.90) software (Purcell et al., 2007) and determined the numbers of SNPs in the following categories: (1) SNPs with minor allele frequencies (MAF) 5% and (2) SNPs and individual call rates 95%. Only autosomal SNPs, with a total number of 40,070 SNPs, were considered for subsequent analyses. The genotyping module of BeadStudio tool (Illumina, Inc., San Diego, CA, USA) was used to determine the genotypes signal intensity of the individuals, including log R ratio (LRR) and B-allele frequency (BAF).

### Identification of swine CNVs and CNVRs

In the present study, the PennCNV (Wang *et al.*, 2007) algorithm was used to identify porcine CNVs. Based on the hidden Markov model (HMM), this algorithm can detect CNVs from high-density SNPs genotyping data, which includes abundant information including the signal intensity and the population frequency (PFB) at each SNP marker, and the distance between SNPs, based on the *Sus scrofa* (Sscrofa11.1) genome assembly. To salvage the sample affected by genomic wave, a porcine GC-model file was created by calculating the GC content of the 1-Mb region surrounding each SNP and the -gcmodel option in PennCNV was used for adjustment. After detection of CNVs, PennCNV quality filters were used with the following cutoff values: (1) Standard deviation of LRR < 0.30; (2) BAF drift < 0.01; and (3) Waviness factor value within  $\pm 0.05$ . To reduce the false positive rate, we acquired a CNV containing three or more consecutive SNPs. Referring to the criteria of Redon *et al.* (2006), CNV regions (CNVRs) were determined by aggregating overlapping CNVs identified across all samples, which had to be present in at least two individuals in each breed. We divided the CNVRs into three types, including gains, losses, and both types (including gain and loss events).

Using regression analysis, we assessed the relationship between the numbers of CNVRs and the length of each chromosome. From the results of Feng *et al.* (2017), we obtained the segmental duplication (SD) regions of the swine genome, and analyzed the relationship between CNVRs and SD using Chi-squared test.

To date, only 20 studies have been focused on genome-wide CNV identification in pigs. Of them, 2 studies employed Sscrofa9.2, and 17 utilized Sscrofa10.2 genome, respectively. To increase the accuracy of the comparisons among studies, CNVRs located on the Sscrofa9.2 and Sscrofa10.2 assembly were converted into the Sscrofa11.1 genome using NCBI Remap tools.

### Function annotation and analysis of CNVRs

Swine transcripts and annotations were downloaded from the Ensembl database. According to the position of the CNVRs and genes, we identified the protein-coding genes and miRNA partially or completely overlapping with the CNVRs. The DAVID Bioinformatics Resources was used for function analysis, including Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG). We downloaded porcine QTLs from the QTL database.

Next, we curated dosage-sensitivity gene list, including the imprinted genes, monoallelically genes, and other type dosage sensitivity genes. These genes were taken from the Database of Chromosomal Imbalance and Phenotype in Humans using Ensembl Resources (DECIPHER, <http://decipher.sanger.ac.uk/index>) (Firth et al., 2009), the International Standards for Cytogenomic Arrays (ISCA, <http://www.iscaconsortium.org>) (Riggs et al., 2012), the Catalogue of Parent of Origin Effects database (Morison et al., 2001; Gimelbrant et al., 2007), and the Geneimprint database ([www.geneimprint.com](http://www.geneimprint.com)) (Chen et al., 2020). The swine genome contains 21 and 369 imprinted and monoallelically expression genes, correspondingly. Based on data from the Ensembl Genome Compare database, we selected the porcine ortholog gene with human dosage-sensitivity genes. Overall, we established a total number of 1542 dosage-sensitive genes in swine genome, including 166 imprinted genes, 1043 monoallelically expressed genes.

The sequenced RNA-seq raw data of 27 adult Duroc tissue types from the retina, pancreas, gut, brain, gall bladder, lung, liver, testes, salivary gland, longissimus dorsi, spinal cord, thyroid, lymph, urinary bladder, spleen, prostate, kidney, adrenal gland, esophagus, stomach, heart, nasopharynx, fat, ovary, breast, placenta, and uterus) were downloaded from NCBI SRA (Sequence Read Archive) database with the BioProject number PRJNA392949 (Zhao *et al.*, 2018). After the QC step conducted using FASTQC, Trimmomatic tools (v3.6), RNA-seq clean data were mapped to the *Sus scrofa* 11.1 genome release version with Hisat2. To obtain expression levels of all genes in the samples of each of the tissue types, fragments per kilobase of exon model per million mapped reads (FPKM) and counts were calculated using StringTie 1.3.4 and FeatureCounts1.6.0 tools, respectively.

We used the tissue specificity index ( $\tau$ ) (Itai *et al.*, 2005) to grade the scalar measurements of the expression specificity, which ranged from 0 for housekeeping genes to 1 for tissue-specific genes. The index  $\tau$  is defined as  $\tau = \frac{1}{N} \sum_{i=1}^N \frac{x_i}{\max_j x_j}$ , where N is the number of tissues, and  $x_i$  is the expression normalized by the maximal expression value.

We used the miRanda tool (Doron *et al.* 2008) to predict the target gene regulated by miRNAs. The miRanda algorithm integrated biological knowledge on target rules of mammalian microRNAs. In this study, Tot Score and Tot Energy values set 140 and -20, respectively.

To identify the target-recognition preference of miRNAs overlapped CNVR, we employed a random sampling method, based on the procedure proposed by Wu *et al.* (2012). The simulation process included two steps: (a) CNV-miRNAs were randomly selected from all miRNAs in the porcine genome, called pseudo-CNV-miRNAs; (b) Based on the relationship between miRNA and the target genes predicted by miRanda, we marked the relationships target genes and pseudo-CNV-miRNAs or pseudo-non-CNV miRNAs, respectively; (c) For each regulatory type, we re-recorded the number of target genes. Steps (a)–(c) were repeated 10,000 times.

In this study, all statistical analyses, including regression analysis, Kolmogorov-Smirnov test, Wilcoxon rank-sum test, Fisher's exact test, and Chi-squared test were performed using R4.0.0 software.

## Abbreviations

CNVs: Copy number variations; CNVRs: CNV regions; SNPs: single nucleotide polymorphisms; Indels: short insertions and deletions; SVs: structural variations; MAF: minor allele frequencies; LRR: log R ratio; BAF: B-allele frequency; HMM: hidden Markov model; PFB: population frequency; SD: segmental duplication; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; NAHR: non-allelic homologous recombination; aCGH: array-based comparative genomic hybridization.

## Declarations

### Acknowledgements

The authors would like to thank the members of the Beijing Breeding Swine Center for managing the pigs and collecting the data.

### Authors' contributions

ZPW and QLM conceived the study. WWW, QZ and XBW participated in its design. ZPW and YYG performed all data analysis. ZPW, YYG, TW and QLM drafted the manuscript. CXZ, SWL, BZ, XL, LAY, QSZ were involved in the acquisition of data, and contributed to the writing and editing. All authors have read and approved the manuscript.

### Funding



This study was supported by Natural Science Foundation of China (No. 32070571), the Academic Backbone Project of Northeast Agricultural University (No.15XG14), NEAU Research Founding for Excellent Young Teachers (2010RCB29). The funding bodies did not influence the design of the study, data collection, analysis, interpretation of data, and in writing the manuscript.

### **Availability of data and materials**

The datasets used and/or analysed during the current study available from the corresponding author on reasonable request.

### **Ethics approval and consent to participate**

All animal sample collections and experimental procedures carefully followed the guidelines for the Care and Use of Experimental Animals published by the Ministry of Science and Technology of the People's Republic of China (Approval number: 2006–398). This study was approved by the Beijing Breeding Swine Center (Beijing, China) and the Northeast Agricultural University (Harbin, China) Animal Care and Treatment Committee (NEAU-EC20). Written informed consent was obtained from the owners for the participation of their animals in this study.

### **Consent for publication**

Not applicable.

### **Competing interests**

The authors declare that they have no competing interests.

### **Author details**

- 1.College of Animal Science and Technology, Northeast Agricultural University, Harbin, China
- 2.Beijing Breeding Swine Center, Beijing 100194, China.
3. Bioinformatics Center, Northeast Agricultural University, Harbin, China
- 4.College of Animal Science and Technology, Shandong Agricultural University, Tai'an, China

## **References**

- An X, Ma K, Zhang Z, Zhao T, Zhang X, Tang B, Li Z. miR-17, miR-21, and miR-143 Enhance Adipogenic Differentiation from Porcine Bone Marrow-Derived Mesenchymal Stem Cells. *DNA Cell Biol.* 2016;35(8):410-416. <https://doi.org/10.1089/dna.2015.3182>.
- Bae JS, Cheong HS, Kim LH, NamGung S, Park TJ, Chun JY, Kim JY, Pasaje CF, Lee JS, Shin HD. Identification of copy number variations and common deletion polymorphisms in cattle. *BMC Genomics.* 2010;11:232. <https://doi.org/10.1186/1471-2164-11-232>.
- 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. <https://doi.org/10.1186/1471-2164-13-733>.
- Chen C, Begcy K. Genome-Wide Identification of Allele-Specific Gene Expression in a Parent-of-Origin Specific Manner. *Methods Mol Biol.* 2020;2072:129-139. [https://doi.org/10.1007/978-1-4939-9865-4\\_11](https://doi.org/10.1007/978-1-4939-9865-4_11).
- Clop A, Vidal O, Amills M. Copy number variation in the genomes of domestic animals. *Anim Genet.* 2012;43(5):503-517. <https://doi.org/10.1111/j.1365-2052.2012.02317.x>.
- Cutler G, Marshall LA, Chin N, Baribault H, Kassner PD. Significant gene content variation characterizes the genomes of inbred mouse strains. *Genome Res.* 2007;17(12):1743-1754. <https://doi.org/10.1101/gr.6754607>.
- Di Gerlando R, Suter AM, Mastrangelo S, Tolone M, Portolano B, Sottile G, Bagnato A, Strillacci MG, Sardina MT. Genome-wide association study between CNVs and milk production traits in Valle del Belice sheep. *PLoS One.* 2019;14(4):e0215204. <https://doi.org/10.1371/journal.pone.0215204>.
- 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):101-109. <https://doi.org/10.1111/age.12247>.

Doron Betel, Manda Wilson, Aaron Gabow, Debora S Marks, Chris Sander. The microRNA.org resource: targets and expression. *Nucleic Acids Res.* 2008;36(suppl 1):D149-53.

Dorshorst B, Molin AM, Rubin CJ, Johansson AM, Strömstedt L, Pham MH, Chen CF, Hallböök F, Ashwell C, Andersson L. A complex genomic rearrangement involving the endothelin 3 locus causes dermal hyperpigmentation in the chicken. *PLoS Genet.* 2011;7(12):e1002412. <https://doi.org/10.1371/journal.pgen.1002412>.

Dumas L, Kim YH, Karimpour-Fard A, Cox M, Hopkins J, Pollack JR, Sikela JM. Gene copy number variation spanning 60 million years of human and primate evolution. *Genome Res.* 2007;17(9):1266-1277. <https://doi.org/10.1101/gr.6557307>.

Elferink MG, Vallée AA, Jungerius AP, Crooijmans RP, Groenen MA. Partial duplication of the PRLR and SPEF2 genes at the late feathering locus in chicken. *BMC Genomics.* 2008;9:391. <https://doi.org/10.1186/1471-2164-9-391>.

Fadista J, Thomsen B, Holm LE, Bendixen C. Copy number variation in the bovine genome. *BMC Genomics.* 2010;11:284. <https://doi.org/10.1186/1471-2164-11-284>.

Feng X, Jiang J, Padhi A, Ning C, Fu J, Wang A, Mrode R, Liu JF. Characterization of genome-wide segmental duplications reveals a common genomic feature of association with immunity among domestic animals. *BMC Genomics.* 2017;18(1):293. <https://doi.org/10.1186/s12864-017-3690-x>.

Fernández AI, Barragán C, Fernández A, Rodríguez MC, Villanueva B. Copy number variants in a highly inbred Iberian porcine strain. *Anim Genet.* 2014;45(3):357-66. <https://doi.org/10.1111/age.12137>.

Feuk L, Carson AR, Scherer SW. Structural variation in the human genome. *Nat Rev Genet.* 2006;7(2):85-97. <https://doi.org/10.1038/nrg1767>.

Firth HV, Richards SM, Bevan AP, Clayton S, Corpas M, Rajan D, Van Vooren S, Moreau Y, Pettett RM, Carter NP. DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *Am J Hum Genet.* 2009;84(4):524-533. <https://doi.org/10.1016/j.ajhg.2009.03.010>.

Fontanesi L, Dall'Olio S, Beretti F, Portolano B, Russo V. Coat colours in the Massese sheep breed are associated with mutations in the agouti signalling protein (ASIP) and melanocortin 1 receptor (MC1R) genes. *Animal.* 2011;5(1):8-17. <https://doi.org/10.1017/S1751731110001382>.

Fontanesi L, Martelli PL, Beretti F, Riggio V, Dall'Olio S, Colombo M, Casadio R, Russo V, Portolano B. An initial comparative map of copy number variations in the goat (*Capra hircus*) genome. *BMC Genomics.* 2010;11:639. <https://doi.org/10.1186/1471-2164-11-639>.

Flisikowski K, Venhoranta H, Nowacka-Woszek J, McKay SD, Flyckt A, Taponen J, Schnabel R, Schwarzenbacher H, Szczerbal I, Lohi H, Fries R, Taylor JF, Switonski M, Andersson M. A novel mutation in the maternally imprinted PEG3 domain results in a loss of MIMT1 expression and causes abortions and stillbirths in cattle (*Bos taurus*). *PLoS One.* 2010;5(11):e15116. <https://doi.org/10.1371/journal.pone.0015116>.

Freeman JL, Perry GH, Feuk L, Redon R, McCarroll SA, Altshuler DM, Aburatani H, Jones KW, Tyler-Smith C, Hurles ME, Carter NP, Scherer SW, Lee C. Copy number variation: new insights in genome diversity. *Genome Res.* 2006;16(8):949-961. <https://doi.org/10.1101/gr.3677206>.

Felekis K, Voskarides K, Dweep H, Sticht C, Gretz N, Deltas C. Increased number of microRNA target sites in genes encoded in CNV regions. Evidence for an evolutionary genomic interaction. *Mol Biol Evol.* 2011;28(9):2421-2424. <https://doi.org/10.1093/molbev/msr078>.

Gimelbrant A, Hutchinson JN, Thompson BR, Chess A. Widespread monoallelic expression on human autosomes. *Science.* 2007;318(5853):1136-1140. <https://doi.org/10.1126/science.1148910>.

Giuffra E, Törnsten A, Marklund S, Bongcam-Rudloff E, Chardon P, Kijas JM, Anderson SI, Archibald AL, Andersson L. A large duplication associated with dominant white color in pigs originated by homologous recombination between LINE elements flanking KIT. *Mamm Genome.* 2002;13(10):569-577. <https://doi.org/10.1007/s00335-002-2184-5>.

Graubert TA, Cahan P, Edwin D, Selzer RR, Richmond TA, Eis PS, Shannon WD, Li X, McLeod HL, Cheverud JM, Ley TJ. A high-resolution map of segmental DNA copy number variation in the mouse genome. *PLoS Genet.* 2007;3(1):e3. <https://doi.org/10.1371/journal.pgen.0030003>.

Griffin DK, Robertson LB, Tempest HG, Vignal A, Fillon V, Crooijmans RP, Groenen MA, Deryusheva S, Gaginskaya E, Carré W, Waddington D, Talbot R, Völker M, Masabanda JS, Burt DW. Whole genome comparative studies between chicken and turkey and their implications for avian genome evolution. *BMC Genomics.* 2008;9:168. <https://doi.org/10.1186/1471-2164-9-168>.

Gunnarsson U, Kerje S, Bed'hom B, Sahlqvist AS, Ekwall O, Tixier-Boichard M, Kämpe O, Andersson L. The Dark brown plumage color in chickens is caused by an 8.3-kb deletion upstream of SOX10. *Pigment Cell Melanoma Res.* 2011;24(2):268-274. <https://doi.org/10.1111/j.1755-148X.2011.00825.x>.

Ha M, Lu J, Tian L, Ramachandran V, Kasschau KD, Chapman EJ, Carrington JC, Chen X, Wang XJ, Chen ZJ. Small RNAs serve as a genetic buffer against genomic shock in Arabidopsis interspecific hybrids and allopolyploids. *Proc Natl Acad Sci U S A.* 2009;106(42):17835-17840. <https://doi.org/10.1073/pnas.0907003106>.

Hou Y, Liu GE, Bickhart DM, Cardone MF, Wang K, Kim ES, Matukumalli LK, Ventura M, Song J, VanRaden PM, Sonstegard TS, Van Tassell CP. Genomic characteristics of cattle copy number variations. *BMC Genomics.* 2011;12:127. <https://doi.org/10.1186/1471-2164-12-127>.

Hou CL, Meng FH, Wang W, Wang SY, Xing YP, Cao JW, et al. Genome-wide analysis of copy number variations in Chinese sheep using array comparative genomic hybridization. *Small Ruminant Res.* 2015; 128:19–26. <https://doi.org/10.1016/j.smallrumres.2015.04.014>

Hurles M. How homologous recombination generates a mutable genome. *Hum Genomics.* 2005;2(3):179-186. <https://doi.org/10.1186/1479-7364-2-3-179>.

Infante RE, Wang ML, Radhakrishnan A, Kwon HJ, Brown MS, Goldstein JL. NPC2 facilitates bidirectional transfer of cholesterol between NPC1 and lipid bilayers, a step in cholesterol egress from lysosomes. *Proc Natl Acad Sci U S A.* 2008;105(40):15287-15292. <https://doi.org/10.1073/pnas.0807328105>.

Itai Yanai, Hila Benjamin, Michael Shmoish, Vered Chalifa-Caspi, Maxim Shklar, Ron Ophir, Arren Bar-Even, Shirley Horn-Saban, Marilyn Safran, Eytan Domany, Doron Lancet, Orit Shmueli. Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics.* 2005;21(5):650-659. <https://doi.org/10.1093/bioinformatics/bti042>.

Jiang J, Wang J, Wang H, Zhang Y, Kang H, Feng X, Wang J, Yin Z, Bao W, Zhang Q, Liu JF. Global copy number analyses by next generation sequencing provide insight into pig genome variation. *BMC Genomics.* 2014;15(1):593. <https://doi.org/10.1186/1471-2164-15-593>.

Jovelín R. No Evidence that MicroRNAs Coevolve with Genes Located in Copy Number Regions. *Mol Biol Evol.* 2015;32(7):1890-1894. <https://doi.org/10.1093/molbev/msv073>.

Keel BN, Nonneman DJ, Lindholm-Perry AK, Oliver WT, Rohrer GA. A Survey of Copy Number Variation in the Porcine Genome Detected From Whole-Genome Sequence. *Front Genet.* 2019;10:737. <https://doi.org/10.3389/fgene.2019.00737>.

Kim PM, Lam HY, Urban AE, Korbel JO, Affourtit J, Grubert F, Chen X, Weissman S, Snyder M, Gerstein MB. Analysis of copy number variants and segmental duplications in the human genome: Evidence for a change in the process of formation in recent evolutionary history. *Genome Res.* 2008;18(12):1865-1874. <https://doi.org/10.1101/gr.081422.108>.

Koishi R, Ando Y, Ono M, Shimamura M, Yasumo H, Fujiwara T, Horikoshi H, Furukawa H. Angptl3 regulates lipid metabolism in mice. *Nat Genet.* 2002;30(2):151-157. <https://doi.org/10.1038/ng814>.

Li G, Li Y, Li X, Ning X, Li M, Yang G. MicroRNA identity and abundance in developing swine adipose tissue as determined by Solexa sequencing. *J Cell Biochem.* 2011;112(5):1318-1328. <https://doi.org/10.1002/jcb.23045>.

Li Y, Li X, Sun WK, Cheng C, Chen YH, Zeng K, Chen X, Gu Y, Gao R, Liu R, Lv X. Comparison of liver microRNA transcriptomes of Tibetan and Yorkshire pigs by deep sequencing. *Gene.* 2016;577(2):244-250. <https://doi.org/10.1016/j.gene.2015.12.003>.

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. <https://doi.org/10.1186/1471-2164-13-725>.

Lin CH, Li LH, Ho SF, Chuang TP, Wu JY, Chen YT, Fann CS. A large-scale survey of genetic copy number variations among Han Chinese residing in Taiwan. *BMC Genet.* 2008;9:92. <https://doi.org/10.1186/1471-2156-9-92>.

Liu GE, Hou Y, Zhu B, Cardone MF, Jiang L, Cellamare A, Mitra A, Alexander LJ, Coutinho LL, Dell'Aquila ME, Gasbarre LC, Lacalandra G, Li RW, Matukumalli LK, Nonneman D, Regitano LC, Smith TP, Song J, Sonstegard TS, Van Tassell CP, Ventura M, Eichler EE, McDanel TG, Keele JW. Analysis of copy number variations among diverse cattle breeds. *Genome Res.* 2010;20(5):693-703. <https://doi.org/10.1101/gr.105403.110>.

Liu J, Zhang L, Xu L, Ren H, Lu J, Zhang X, Zhang S, Zhou X, Wei C, Zhao F, Du L. Analysis of copy number variations in the sheep genome using 50K SNP BeadChip array. *BMC Genomics.* 2013;14:229. <https://doi.org/10.1186/1471-2164-14-229>.

- Locke DP, Sharp AJ, McCarroll SA, McGrath SD, Newman TL, Cheng Z, Schwartz S, Albertson DG, Pinkel D, Altshuler DM, Eichler EE. Linkage disequilibrium and heritability of copy-number polymorphisms within duplicated regions of the human genome. *Am J Hum Genet*. 2006;79(2):275-290. <https://doi.org/10.1086/505653>.
- Long Y, Su Y, Ai H, Zhang Z, Yang B, Ruan G, Xiao S, Liao X, Ren J, Huang L, Ding N. A genome-wide association study of copy number variations with umbilical hernia in swine. *Anim Genet*. 2016;47(3):298-305. <https://doi.org/10.1111/age.12402>.
- Lupski JR, Stankiewicz P. Genomic disorders: molecular mechanisms for rearrangements and conveyed phenotypes. *PLoS Genet*. 2005;1(6):e49. <https://doi.org/10.1371/journal.pgen.0010049>
- Lee AS, Gutiérrez-Arcelus M, Perry GH, Vallender EJ, Johnson WE, Miller GM, Korbel JO, Lee C. Analysis of copy number variation in the rhesus macaque genome identifies candidate loci for evolutionary and human disease studies. *Hum Mol Genet*. 2008;17(8):1127-1136. <https://doi.org/10.1093/hmg/ddn002>.
- Morison IM, Paton CJ, Cleverley SD. The imprinted gene and parent-of-origin effect database. *Nucleic Acids Res*. 2001;29(1):275-276. <https://doi.org/10.1093/nar/29.1.275>.
- Marcinkowska M, Szymanski M, Krzyzosiak WJ, Kozłowski P. Copy number variation of microRNA genes in the human genome. *BMC Genomics*. 2011;12:183. <https://doi.org/10.1186/1471-2164-12-183>.
- Mei C, Junjvlieke Z, Raza SHA, Wang H, Cheng G, Zhao C, Zhu W, Zan L. Copy number variation detection in Chinese indigenous cattle by whole genome sequencing. *Genomics*. 2020;112(1):831-836. <https://doi.org/10.1016/j.ygeno.2019.05.023>.
- Mills RE, Walter K, Stewart C, Handsaker RE, Chen K, Alkan C, Abyzov A, Yoon SC, Ye K, Cheetham RK, Chinwalla A, Conrad DF, Fu Y, Grubert F, Hajirasouliha I, Hormozdiari F, Iakoucheva LM, Iqbal Z, Kang S, Kidd JM, Konkel MK, Korn J, Khurana E, Kural D, Lam HY, Leng J, Li R, Li Y, Lin CY, Luo R, Mu XJ, Nemesh J, Peckham HE, Rausch T, Scally A, Shi X, Stromberg MP, Stütz AM, Urban AE, Walker JA, Wu J, Zhang Y, Zhang ZD, Batzer MA, Ding L, Marth GT, McVean G, Sebat J, Snyder M, Wang J, Ye K, Eichler EE, Gerstein MB, Hurles ME, Lee C, McCarroll SA, Korbel JO. Mapping copy number variation by population-scale genome sequencing. *Nature*. 2011;470(7332):59-65. <https://doi.org/10.1038/nature09708>.
- Nguyen DQ, Webber C, Ponting CP. Bias of selection on human copy-number variants. *PLoS Genet*. 2006;2(2):e20. <https://doi.org/10.1371/journal.pgen.0020020>.
- Paudel Y, Madsen O, Megens HJ, Frantz LA, Bosse M, Crooijmans RP, Groenen MA. Copy number variation in the speciation of pigs: a possible prominent role for olfactory receptors. *BMC Genomics*. 2015;16(1):330. <https://doi.org/10.1186/s12864-015-1449-9>.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559-575. <https://doi.org/10.1086/519795>.
- Ramayo-Caldas Y, Castelló A, Pena RN, Alves E, Mercadé A, Souza CA, Fernández AI, Perez-Enciso M, Folch JM. Copy number variation in the porcine genome inferred from a 60 k SNP BeadChip. *BMC Genomics*. 2010;11:593. <https://doi.org/10.1186/1471-2164-11-593>.
- Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W, Cho EK, Dallaire S, Freeman JL, González JR, Gratacòs M, Huang J, Kalaitzopoulos D, Komura D, MacDonald JR, Marshall CR, Mei R, Montgomery L, Nishimura K, Okamura K, Shen F, Somerville MJ, Tchinda J, Valsesia A, Woodwark C, Yang F, Zhang J, Zerjal T, Zhang J, Armengol L, Conrad DF, Estivill X, Tyler-Smith C, Carter NP, Aburatani H, Lee C, Jones KW, Scherer SW, Hurles ME. Global variation in copy number in the human genome. *Nature*. 2006;444(7118):444-454. <https://doi.org/10.1038/nature05329>.
- Revay T, Quach AT, Maignel L, Sullivan B, King WA. Copy number variations in high and low fertility breeding boars. *BMC Genomics*. 2015;16(1):280. <https://doi.org/10.1186/s12864-015-1473-9>.
- Revilla M, Puig-Oliveras A, Castelló A, Crespo-Piazuelo D, Paludo E, Fernández 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. <https://doi.org/10.1371/journal.pone.0177014>.
- Riggs ER, Church DM, Hanson K, Horner VL, Kaminsky EB, Kuhn RM, Wain KE, Williams ES, Aradhya S, Kearney HM, Ledbetter DH, South ST, Thorland EC, Martin CL. Towards an evidence-based process for the clinical interpretation of copy number variation. *Clin Genet*. 2012;81(5):403-412. <https://doi.org/10.1111/j.1399-0004.2011.01818.x>.
- Schiavo G, Dolezal MA, Scotti E, Bertolini F, Calò DG, Galimberti G, Russo V, Fontanesi L. Copy number variants in Italian Large White pigs detected using high-density single nucleotide polymorphisms and their association with back fat thickness. *Anim Genet*. 2014;45(5):745-749.

<https://doi.org/10.1111/age.12180>.

Seo BY, Park EW, Ahn SJ, Lee SH, Kim JH, Im HT, Lee JH, Cho IC, Kong IK, Jeon JT. An accurate method for quantifying and analyzing copy number variation in porcine KIT by an oligonucleotide ligation assay. *BMC Genet.* 2007;8:81. <https://doi.org/10.1186/1471-2156-8-81>.

Seol D, Ko BJ, Kim B, Chai HH, Lim D, Kim H. Identification of Copy Number Variation in Domestic Chicken Using Whole-Genome Sequencing Reveals Evidence of Selection in the Genome. *Animals (Basel).* 2019;9(10):809. <https://doi.org/10.3390/ani9100809>.

Stafuzza NB, Silva RMO, Fragomeni BO, Masuda Y, Huang Y, Gray K, Lourenco DAL. A genome-wide single nucleotide polymorphism and copy number variation analysis for number of piglets born alive. *BMC Genomics.* 2019;20(1):321. <https://doi.org/10.1186/s12864-019-5687-0>.

Stankiewicz P, Lupski JR. Genome architecture, rearrangements and genomic disorders. *Trends Genet.* 2002;18(2):74-82. [https://doi.org/10.1016/s0168-9525\(02\)02592-1](https://doi.org/10.1016/s0168-9525(02)02592-1).

Stankiewicz P, Lupski JR. Structural variation in the human genome and its role in disease. *Annu Rev Med.* 2010;61:437-455. <https://doi.org/10.1146/annurev-med-100708-204735>.

Tikka A, Soronen J, Laurila PP, Metso J, Ehnholm C, Jauhiainen M. Silencing of ANGPTL 3 (angiopoietin-like protein 3) in human hepatocytes results in decreased expression of gluconeogenic genes and reduced triacylglycerol-rich VLDL secretion upon insulin stimulation. *Biosci Rep.* 2014;34(6):e00160. <https://doi.org/10.1042/BSR20140115>.

Turner DJ, Miretti M, Rajan D, Fiegler H, Carter NP, Blayney ML, Beck S, Hurles ME. Germline rates of de novo meiotic deletions and duplications causing several genomic disorders. *Nat Genet.* 2008;40(1):90-95. <https://doi.org/10.1038/ng.2007.40>.

Wang H, Wang C, Yang K, Liu J, Zhang Y, Wang Y, Xu X, Michal JJ, Jiang Z, Liu B. Genome Wide Distributions and Functional Characterization of Copy Number Variations between Chinese and Western Pigs. *PLoS One.* 2015;10(7):e0131522. <https://doi.org/10.1371/journal.pone.0131522>.

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. <https://doi.org/10.1186/1471-2164-13-273>.

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.* 2013a;8(7):e68683. <https://doi.org/10.1371/journal.pone.0068683>.

Wang J, Jiang J, Wang H, Kang H, Zhang Q, Liu JF. Enhancing genome-wide copy number variation identification by high density array CGH using diverse resources of pig breeds. *PLoS One.* 2014;9(1):e87571. <https://doi.org/10.1371/journal.pone.0087571>

Wang J, Jiang J, Wang H, Kang H, Zhang Q, Liu JF. Improved Detection and Characterization of Copy Number Variations Among Diverse Pig Breeds by Array CGH. *G3 (Bethesda).* 2015a;5(6):1253-1261. <https://doi.org/10.1534/g3.115.018473>.

Wang K, Li MY, Hadley D, Liu R, Glessner J, Grant SFA, Hakonarson H, Bucan M. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res.* 2007;17(11):1665-1674. <https://doi.org/10.1101/gr.6861907>.

Wang L, Liu X, Zhang L, Yan H, Luo W, Liang J, Cheng D, Chen S, Ma X, Song X, Zhao K, Wang L, et al. Genome-wide copy number variations inferred from SNP genotyping arrays using a Large White and Minzhu intercross population. *PLoS One.* 2013;8(10):e74879. <https://doi.org/10.1371/journal.pone.0074879>.

Wang Y, Tang Z, Sun Y, Wang H, Wang C, Yu S, Liu J, Zhang Y, Fan B, Li K, Liu B. Analysis of genome-wide copy number variations in Chinese indigenous and western pig breeds by 60 K SNP genotyping arrays. *PLoS One.* 2014a; 9(9):e106780. <https://doi.org/10.1371/journal.pone.0106780>.

Wiedmann RT, Nonneman DJ, Rohrer GA. Genome-Wide Copy Number Variations Using SNP Genotyping in a Mixed Breed Swine Population. *PLoS One.* 2015;10(7):e0133529. <https://doi.org/10.1371/journal.pone.0133529>.

Willemsen MH, Vallès A, Kirkels LA, Mastebroek M, Olde Loohuis N, Kos A, Wissink-Lindhout WM, de Brouwer AP, Nillesen WM, Pfundt R, Holder-Espinasse M, Vallée L, Andrieux J, Coppens-Hofman MC, Rensen H, Hamel BC, van Bokhoven H, Aschrafi A, Kleefstra T. Chromosome 1p21.3 microdeletions comprising DPYD and MIR137 are associated with intellectual disability. *J Med Genet.* 2011; 48(12):810-818. <https://doi.org/10.1136/jmedgenet-2011-100294>.

- Wong KK, deLeeuw RJ, Dosanjh NS, Kimm LR, Cheng Z, Horsman DE, MacAulay C, Ng RT, Brown CJ, Eichler EE, Lam WL. A comprehensive analysis of common copy-number variations in the human genome. *Am J Hum Genet.* 2007; 80(1):91-104. <https://doi:10.1086/510560>
- Woodward C, Bateman A. The characterisation of three types of genes that overlie copy number variable regions. *PLoS One.* 2011; 6(5):e14814. <https://doi:10.1371/journal.pone.0014814>.
- Wright D, Boije H, Meadows JR, Bed'hom B, Gourichon D, Vieaud A, Tixier-Boichard M, Rubin CJ, Imsland F, Hallböök F, Andersson L. Copy number variation in intron 1 of SOX5 causes the Pea-comb phenotype in chickens. *PLoS Genet.* 2009; 5(6):e1000512. <https://doi:10.1371/journal.pgen.1000512>.
- Wu X, Zhang D, Li G. Insights into the regulation of human CNV-miRNAs from the view of their target genes. *BMC Genomics.* 2012; 13:707. <https://doi:10.1186/1471-2164-13-707>.
- Yim SH, Kim TM, Hu HJ, Kim JH, Kim BJ, Lee JY, Han BG, Shin SH, Jung SH, Chung YJ. Copy number variations in East-Asian population and their evolutionary and functional implications. *Hum Mol Genet.* 2010; 19(6):1001-1008. <https://doi:10.1093/hmg/ddp564>.
- Zhang F, Gu W, Hurles ME, Lupski JR. Copy number variation in human health, disease, and evolution. *Annu Rev Genomics Hum Genet.* 2009; 10:451-481. <https://doi:10.1146/annurev.genom.9.081307.164217>.
- Zarrei M, MacDonald JR, Merico D, Scherer SW. A copy number variation map of the human genome. *Nat Rev Genet.* 2015; 16(3):172-183. <https://doi:10.1038/nrg3871>.
- Zhao P, Zheng X, Feng W, Wang H, Kang H, Ning C, Du H, Yu Y, Li B, Zhao Y, Liu JF. Profiling long noncoding RNA of multi-tissue transcriptome enhances porcine noncoding genome annotation. *Epigenomics.* 2018; 10(3):301-320. <https://doi:10.2217/epi-2017-0149>

## Tables

**Table 1.** CNVR distributions in the each chromosome of Duroc purebred population

SSC	Length (Mb)	CNV			CNVR			
		Total	Gain	Loss	Total	Gain	Loss	Both
1	274.33	74	0	74	14	0	14	0
2	151.94	108	4	104	17	4	13	0
3	132.85	190	4	186	9	2	6	1
4	130.91	81	2	79	8	1	7	0
5	104.53	49	0	49	13	0	13	0
6	170.84	179	1	178	14	1	13	0
7	121.84	75	2	73	15	1	13	1
8	138.97	86	1	85	11	1	10	0
9	139.51	24	5	19	14	4	9	1
10	69.36	27	1	26	6	1	5	0
11	79.17	28	13	15	14	5	9	0
12	61.6	82	0	82	9	0	9	0
13	208.34	117	1	116	9	0	9	0
14	141.76	95	1	94	13	1	12	0
15	140.41	75	3	72	18	2	15	1
16	79.94	40	5	35	12	5	7	0
17	63.49	29	2	27	8	2	6	0
18	55.98	12	2	10	7	2	5	0
Total	2265.77	1371	47	1324	211	32	175	4
Average*	-	6.59	0.23	6.37	1.01	0.15	0.84	0.02

\*: At sample level, each sample has 6.59 (1371/208) CNVs for Duroc.

**Table 2.** GO and KEGG pathway analyses of genes in the identified CNVRs

Category	Term	GO or KEGG Name	P Value
Biological Process	GO:0048715	negative regulation of oligodendrocyte differentiation	0.0039
	GO:0016192	vesicle-mediated transport	0.0059
	GO:0007596	blood coagulation	0.0067
	GO:0010906	regulation of glucose metabolic process	0.0159
	GO:0051897	positive regulation of protein kinase B signaling	0.0167
	GO:0007030	Golgi organization	0.0190
	GO:0006898	receptor-mediated endocytosis	0.0199
	GO:0051726	regulation of cell cycle	0.0249
	GO:0032456	endocytic recycling	0.0256
	GO:0003151	outflow tract morphogenesis	0.0314
	GO:0046488	phosphatidylinositol metabolic process	0.0338
	GO:0006886	intracellular protein transport	0.0387
	GO:0035914	skeletal muscle cell differentiation	0.0408
	GO:0060412	ventricular septum morphogenesis	0.0448
Cellular Component	GO:0005654	nucleoplasm	2.60E-04
	GO:0035102	PRC1 complex	0.0010
	GO:0005769	early endosome	0.0032
	GO:0030659	cytoplasmic vesicle membrane	0.0070
	GO:0043235	receptor complex	0.0071
	GO:0005604	basement membrane	0.0135
	GO:0005581	collagen trimer	0.0194
	GO:0005737	cytoplasm	0.0209
	GO:0000813	ESCRT I complex	0.0226
	GO:0000791	euchromatin	0.0226
	GO:0005829	cytosol	0.0247
	GO:0016020	membrane	0.0282
	GO:0000777	condensed chromosome kinetochore	0.0328
	GO:0005925	focal adhesion	0.0438
Molecular Function	GO:0030246	carbohydrate binding	7.88E-04
	GO:0005096	GTPase activator activity	0.0081
	GO:0008270	zinc ion binding	0.0124
	GO:0005089	Rho guanyl-nucleotide exchange factor activity	0.0237
	GO:0003727	single-stranded RNA binding	0.0239
	GO:0005509	calcium ion binding	0.0378
KEGG Pathway	ssc04390	Hippo signaling pathway	0.0041
	ssc04310	Wnt signaling pathway	0.0131
	ssc04142	Lysosome	0.0187
	ssc04742	Taste transduction	0.0425
	ssc04520	Adherens junction	0.0443



ssc00561	Glycerolipid metabolism	0.0463
ssc00051	Fructose and mannose metabolism	0.0473

**Table 3.** The mean and median values of length of genes, dN/dS, and dN+dS for the three types of CNV genes

Gene type	# Genes	Mean(Median) length (kb)	dN/dS Mean(Median)	K-S test (P value)	dN+dS Mean(Median)	K-S test (P value)
Type I	862	43.86 (20.32)	0.1540(0.1149)	7.30E-4	0.8385(0.6717)	2.20E-16
Type II	206	201.33 (123.19)	0.1390(0.1077)	3.30E-2	0.5199(0.4132)	4.24E-3
Type III	28	461.52 (371.50)	0.1056(0.0599)	1.42E-3	0.4080(0.3146)	5.64E-3
Non CNV	13542	71.11(30.39)	0.1676(0.1268)	NA	0.5910(0.4725)	NA

Note: \* probability values for K-S test are given for the comparison of CNV gene type and non CNV genes.

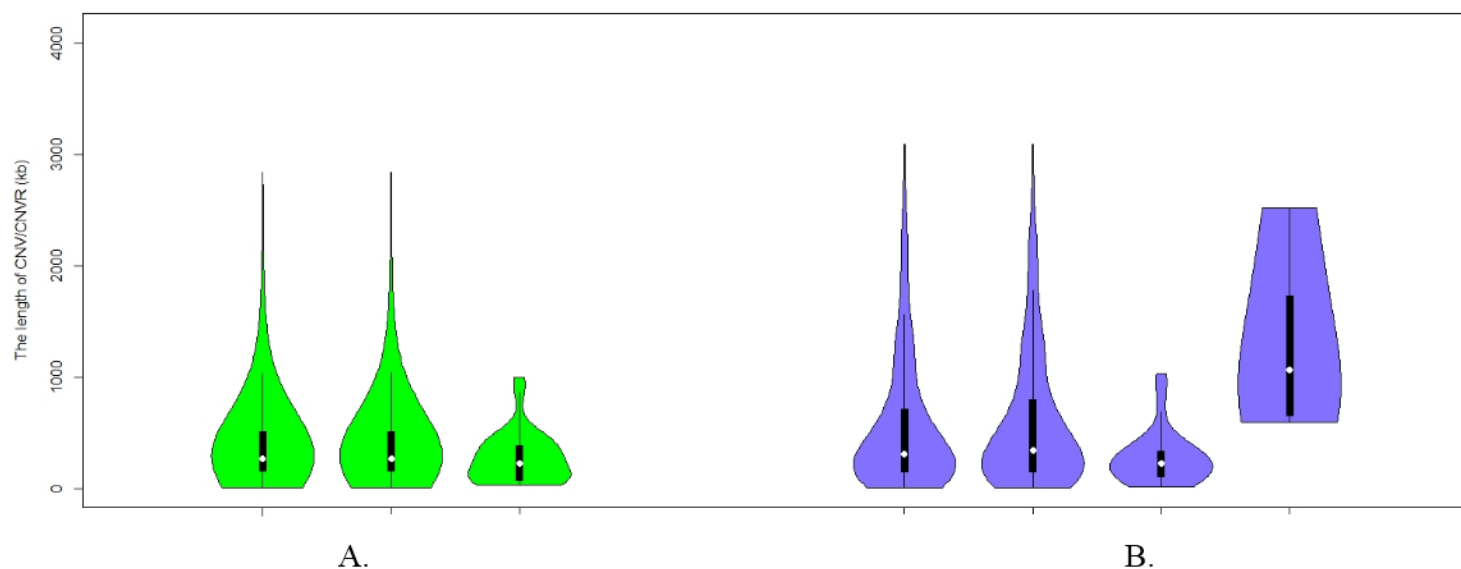
**Table 4.** Comparison of CNVRs identified in this study with previous studies

Study	Platform	Sample	CNVR	Total length (Mb)	Average length (kb)	Range (kb)	Gain	Loss	Both	Genomic	Concordant number
Ramayo-Caldas et al (2010)	SNP Chip	55	49	36.97/1.51%	754.59	44.70-10700.00	19	8	22	9.2	–
Wang et al (2012)	SNP Chip	474	382	95.76/4.23%	250.70	5.03-2702.70	34	296	52	10.2	53
Chen et al (2012)	SNP Chip	1693	565	143.03/5.84%	247.55	50.39-8100.00	225	261	79	10.2	83
Li et al (2012)	aCGH	12	259	16.85/0.74%	65.07	2.30-1550.00	93	140	26	10.2	18
Wang et al (2013a)	SNP Chip	14	63	9.98/0.36%	158.37	3.20-827.21	26	36	1	10.2	6
Wang et al (2013)	SNP Chip	585	249	560.30/26.22%	2305.77	29.20-27290.00	70	43	136	9.2	–
Wang et al (2014)	aCGH	12	1344	47.79/1.70%	35.56	3.37-1319.00	557	760	27	10.2	90
Wang et al (2014a)	SNP Chip	302	348	150.49/6.14%	443.24	4.93-12410.00	88	243	17	10.2	57
Schiavo et al (2014)	SNP Chip	305	170	72.33/2.95%	425.47	25.20-1700.00	7	161	2	10.2	28
Fernandez et al (2014)	SNP Chip	223	65	9.68/0.33%	148.99	3.06-1070.00	32	21	12	10.2	11
Jiang et al (2014)	NGS	13	3131	102.80/4.20%	32.80	10.00-555.10	1702	1366	63	10.2	147
Wiedmann et al (2015)	SNP Chip	1802	502	495.29/19.1%	986.63	0.93-31727.39	–	–	–	10.2	105
Wang et al (2015a)	NGS	49	3131	42.10/1.72%	13.40	1.00-88.80	745	2364	22	10.2	142
Wang et al (2015)	aCGH	12	758	47.43/1.69%	62.58	7.02-2635.29	189	472	28	10.2	44
Revay et al (2015)	SNP Chip	38	35	36.50/1.30%	1043.73	7.47-3755.29	5	28	2	10.2	15
Dong et al (2015)	SNP Chip	96	105	16.71/0.68%	159.10	0.31-2751.85	50	45	10	10.2	12
Long et al (2016)	SNP Chip	905	737	93.70/3.82%	126.23	0.31-2989.80	475	255	7	10.2	73
Revilla et al (2017)	NGS	32	540	9.66/0.39%	17.88	3.21-1106.44	231	305	4	10.2	34
Stafuzza et al (2019)	SNP Chip	3520	425	197.00/7.01%	463.62	2.50-9718.40	19	342	64	10.2	126
Keel et al (2019)	NGS	240	3538	22.90/0.94%	6.80	0.23-398.90	144	3372	22	11.1	—

**Table 5.** Some candidate genes overlapped with CNVRs involved fatness metabolic and development.

Gene symbol	Location(Mb)	Full name	Major function of involving in fatness metabolic and development
<i>MIR143</i>	SSC2:157.34-157.34	microRNA mir143	Promote the adipogenic differentiation. The most abundant expression in developing swine adipose tissue.
<i>MIR335</i>	SSC18:19.34-19.34	microRNA mir335	Participate in the metabolism of glucose and lipid.
<i>MIR378</i>	SSC2:157.64-157.64	microRNA mir378	Participate in the metabolism of glucose and lipid.
<i>MIRLET7</i>	–	microRNA let7 family	The most abundant expression in developing swine adipose tissue.
<i>DGAT1</i>	SSC4:0.60-0.61	Diacylglycerol O-acyltransferase 1	Affect fat metabolism and lipid deposition in tissues, and participate in the regulation of energy synthesis and catabolism.
<i>DGAT2</i>	SSC9:11.16-11.18	Diacylglycerol O-acyltransferase 2	Affect fat metabolism and lipid deposition in tissues, and participate in the regulation of energy synthesis and catabolism.
<i>MOGAT2</i>	SSC9:11.12-11.13	Monoacylglycerol O-acyltransferase 2	Take part in some pathway related to fat digestion and absorption and metabolism.
<i>AGPAT2</i>	SSC1:313.74-313.74	1-acylglycerol-3-phosphate O-acyltransferase 2	Associate with congenital generalized lipodystrophy, or Berardinelli-Seip syndrome.
<i>FABP1</i>	SSC3:60.62-60.63	Fatty acid binding protein 1	Role include fatty acid uptake, transport, and metabolism.
<i>PPARA</i>	SSC5:0.47-0.49	Peroxisome proliferator-activated receptor alpha	A key regulator of lipid metabolism.
<i>ANGPTL3</i>	SSC1:313.74-313.74	Angiopoietin like 3	Involve in regulation of lipid and glucose metabolism. Inhibit endothelial lipase, causing increased plasma levels of HDL cholesterol and phospholipids.
<i>NPC2</i>	SSC7:103.57-103.58	NPC intracellular cholesterol transporter 2	Plays an important role in the egress of cholesterol from the lysosomal compartment

## Figures



**Figure 1**

Violin plots of the total CNVs (CNVRs) lengths, gain or loss CNVs (CNVRs) lengths, and both CNVRs lengths in Duroc purebred population.

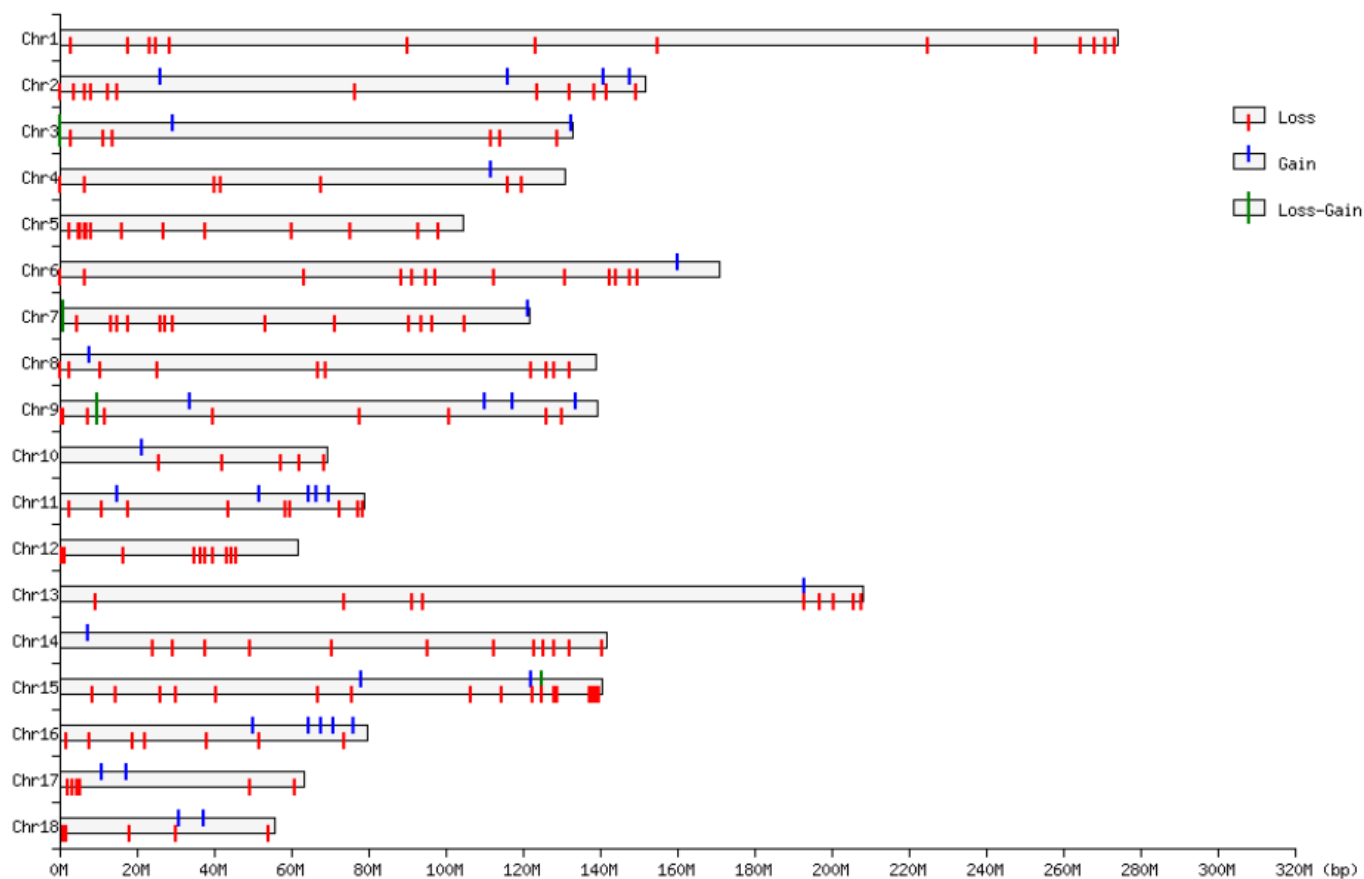


Figure 2

CNVR distributions in the genome of Duroc purebred population

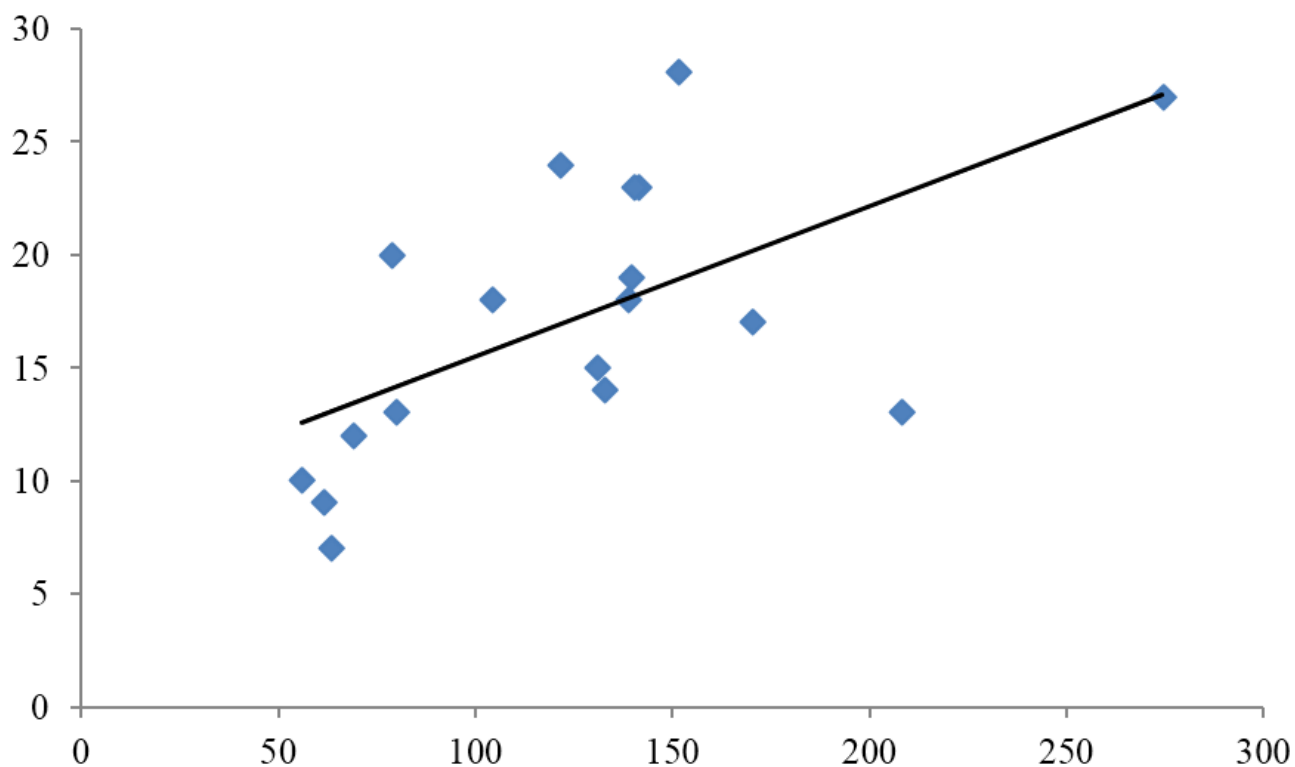


Figure 3

Correlation CNVR numbers and chromosome length.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [GICSD2001585AuthorChecklistFull.pdf](#)