Genome-wide detection of CNV regions and their potential association with growth and fatness traits in Duroc pigs

Yibin Qiu
South China Agricultural University College of Animal Science
Rongrong Ding
Guangdong Wens Breeding Swine Technology Co., Ltd.
Zhanwei Zhuang
South China Agricultural University College of Animal Science
Jie Wu
South China Agricultural University College of Animal Science
Ming Yang
Guangdong Wens Breeding Swine Technology Co., Ltd.
Shenping Zhou
South China Agricultural University College of Animal Science
Yong Ye
South China Agricultural University College of Animal Science
Qian Geng
South China Agricultural University College of Animal Science
Zheng Xu
South China Agricultural University College of Animal Science
Sixiu Huang
South China Agricultural University College of Animal Science
Gengyuan Cai
South China Agricultural University College of Animal Science
Jie Yang (jieyang2012@hotmail.com)
South China Agricultural University https://orcid.org/0000-0002-7031-2160
Zhenfang Wu
South China Agricultural University College of Animal Science

Research

Keywords: Copy number variation, CNVR-based association, Growth, Fatness, Duroc pigs

Posted Date: October 8th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-86722/v1

Version of Record: A version of this preprint was published at BMC Genomics on May 8th, 2021. See the published version at https://doi.org/10.1186/s12864-021-07654-7.

Abstract

Background: In the process of pig breeding, the average daily gain (ADG), days to 100 kg (AGE), and backfat thickness (BFT) are directly related to growth rate and fatness. However, the genetic mechanisms involved are not well understood. As an essential source of genetic diversity, copy number variation (CNV) that can affect a variety of complex traits and diseases, has gradually been thrust into the limelight. In this study, we reported the genome-wide CNVs of Duroc pigs by SNP genotyping data of 6,627 Duroc pigs. Moreover, we also performed CNV region (CNVR) based association analysis for growth and fatness traits in two Duroc populations.

Results: Our study identified a total of 835 nonredundant CNVRs in American and Canadian Duroc pigs, which covered 226.73 Mb (~ 10.00%) of the pig autosomal genome. Among them, we identified 682 CNVRs in the American Duroc pigs and 424 CNVRs in the Canadian Duroc pigs, and 271 CNVRs were detected in both populations. Experimentally, 77.8% of randomly selected CNVRs were validated by quantitative PCR (qPCR). We also identified 24 significant CNVRs for growth and fatness traits though CNVR-based association analysis. Among these, six and three CNVRs in American and Canadian Duroc pigs were found to be associated with both ADG and AGE traits, respectively. Notably, four CNVRs showed both significant in ADG, AGE, and BFT, indicating that these CNVRs may play a pleiotropic role in regulating pig growth and fat deposition. Besides, further bioinformatics analysis determined a subset of potential candidate genes, such as *PDGFRB*, *INSIG1*, *GPER1*, *PDGFA*, and *LPCAT1*.

Conclusions: The present study provides a necessary supplement to the CNV map of the Duroc genome through a large-scale population genotyping. Also, the CNVR-based association results provide a meaningful way to elucidate the genetic mechanisms of complex traits. The identified CNVRs can be used as molecular markers for genetic improvement in molecular-guided breeding of modern commercial pigs

Background

In human and animal genomes, variation occurs in many forms, including single nucleotide polymorphisms (SNPs), insertion/deletion (INDEL) of small fragments, and copy number variations (CNVs). Among these genetic variations, CNVs are a particular subtype of genomic structural variation that roughly ranges from 50 bp to several Mbs and is mainly represented by deletions and duplications [1–4]. Adjacent copy number variation areas with overlapping regions can be combined into a large genome segment, known as the copy number variation region (CNVR) [5]. In terms of total bases involved, CNVs encompass more nucleotide sequences and arise more frequently than SNPs [6]. Therefore, they have higher mutation probability and more significant potential impacts [7], such as changing gene structure and altering gene dosage and thus dramatically affect gene expression and adaptive phenotypes [8]. Additionally, some of these CNVs were even associated with several complex diseases [9–11]. These observations lead us to predict that CNVs are a primary contributor to phenotypic variation and disease susceptibility.

Indeed, there are multiple studies have suggested that CNVs play an essential role in affecting some complex traits and causing disease. In humans, Aitman et al. [12] demonstrated that copy number polymorphism in *Fcgr3* genes is a determinant of susceptibility to immunologically mediated renal disease, and a recent study identified that copy number variation in *NPY4R* might be related to the pathogenesis of obesity [13]. Similarly, phenotypic variations and diseases caused by CNVs are also widespread in domesticated species. In the pigs-the focus of this study, a typical example of trait affected by CNVs is an increase in the copy number (CN) of the *KIT* gene that causes the dominant white phenotype [14, 15]. In reproductive performance, the CNV of the *MTHFSD* gene was correlated with litter size traits of Xiang pigs [16]. Zheng et al. [17] also showed that a higher CN of the *AHR* gene had a positive effect on litter size. In productive performance, Revilla et al. [18] discovered a CNVR containing the *GPAT2* gene, which might be associated with several growth-related traits. Thus, analyzing CNVs and identifying their potential association with complex traits have gradually become an essential part of genetic studies.

Growth rate and fatness are vital objectives in the process of pig breeding, which are directly related to economic advantages. The growth rate measured at different stages mainly included average daily gain (ADG) during the test period as well as age (AGE) that adjusted to a certain weight [19]. Fat deposition is also a critical biological process, generally measured by backfat thickness (BFT) or intermuscular fat content. Until now, Considerable association analysis has been focused on finding single-site variants, quantitative trait loci (QTLs), and related candidate functional genes that might influence growth and fatness traits [20–22]. However, the systematic association studies for complex quantitative traits based on CNVs were rarely explored [18, 23], and the full relevance of CNVs to the genetic basis involved is yet to be clarified. Besides, the genetic architectures of these traits are usually complex and commonly controlled by multiple genes [19], considering that the majority of association studies for growth traits and fatness in pigs have only used a small sample of genotyped animals, which has limited the statistical power of the association analysis [24]. It is necessary to conduct CNVs association analysis in a population with sufficient sample size.

In this study, we performed genome-wide CNV detection on the large-scale population of Duroc pigs from the American and Canadian origin. Moreover, CNVR-based association analysis on growth and fatness traits were applied in two experimental populations. Our results include the identified CNVR and candidate genes that can provide more information on the molecular mechanisms of important economic traits and promote the rapid development of molecular breeding in pigs.

Methods

Ethics statement

The animals and experimental methods used in this study are following the guidelines of the Ministry of Agriculture of China and Use Committee of South China Agricultural University (SCAU). The ethics committee of SCAU (Guangzhou, China) approved all of the animal experiments.

Samples and phenotype data

Our sample includes American (n=3,770) and Canadian (n=2,857) origin Duroc pigs collected from the Wens Foodstuff Group Co., Ltd. (Guangdong, China). In addition, data such as ADG (average daily gain at 100 kg), AGE (days to 100 kg), and BTF (backfat thickness at 100kg) were obtained from each population. During the experiment, all the pigs were raised under normal management conditions and the same feeding standards. The ADG and AGE of each pig were measured from 30 to 115 kg and then adjusted to 100 kg. The adjusted formula of AGE are as follows:

$$AGE\ adjusted\ to\ 100kg = Measured\ age - \left(\frac{Measured\ weight - 100kg}{Correction\ factor\ one}\right)$$

where correction factors one is different between sire and dam, and the formulas are as follows:

$$Sire: Correction\ factor\ one = \frac{Measured\ weight}{Measured\ age} \times 1.826040$$

$$Dam: Correction\ factor\ one = \frac{Measured\ weight}{Measured\ age} \times 1.714615$$

The following formula for adjusted ADG:

$$ADG\ adjusted\ to\ 100kg = \frac{100kg}{AGE\ adjusted\ to\ 100kg}$$

The phenotype of BTF was measured from the 10th-rib to 11th-rib at approximately 100 kg using an Aloka 500V SSD B ultrasound (Corometrics Medical Systems, USA) [25], and the adjusted formula of BFT is as follows:

$$BF\ adjusted\ to\ 100k = Measured\ backfat\ thickness \times Correction\ factor\ two$$

where $Correction\ factor\ two = \frac{A}{A+[B\times(Measured\ Weight-100)]}$, A = 13.468 and B = 0.111528 in sires, while A = 15.654 and B = 0.156646 in dams. Before the association analysis, outliers outside the mean ± 3 standard deviations were removed.

SNP genotyping and quality control

Genomic DNA was extracted from ear tissue with the traditional phenol/chloroform method, the DNA quality of samples was assessed by light absorption ratio ($A_{260/280}$ and $A_{260/230}$) and gel electrophoresis, the concentration was controlled to be 50 ng/μL [26]. Samples were genotyped using Illumina GeneSeek 50K SNP array (Neogen, Lincoln, NE, United States) with 50,703 SNP markers across the entire genome. The quality control (QC) were performed with PLINK v1.90 software [27], the parameters of QC include: animal call rates > 0.9, SNP call rates > 0.9, minor allele frequencies > 0.005 and $P > 10^{-6}$ for Hardy-Weinberg equilibrium test. The SNPs located in sex chromosomes or without positional information were also discarded. Finally, a set of 39,600 informative SNPs from 3,770 American origin Duroc pigs and 36,700 informative SNPs from 2,857 Canadian origin Duroc pigs were used for CNVs detection, respectively.

CNVs detection

PennCNV software [28] was used to identify individual-baed CNV by combining the SNP signal data of log R Ratio (LRR) and B allele frequency (BAF) as well as the population frequency of B allele (PFB). Among them, the LRR and BAF values were computed by GenomeStudio software v2.0 (Illumina, Inc., USA) for each SNP. The perl comppile_pfb.pl command was used to calculate the PFB files based on the BAF of each SNP. Moreover, the wave adjustment procedure was conducted using the -gcmodel option in PennCNV software to reduce the impact of genomic waves [29]. We calculated the GC content of the 500kb genomic region around each SNP according to the Sscrofa 11.1 version of the pig reference genome (http://ensemble.org/Sus_scrofa/Info/Index). The final CNVs were generated by retaining high-quality samples according to the following criteria: LRR < 0.3, BAF drift < 0.01, and GC wave factor of LRR < 0.05. Meanwhile, to further decrease false-positive CNVs, CNVs with consecutive SNPs ≥ 3 and CNV length ≥ 10 kb were remained. We also used the bedtools software v2.26.0 [30, 31] to merge CNVs with at least 1bp overlap in all samples to define CNV region (CNVR) [17], and CNVRuler software [32] was used to define three types of CNVR: loss, gain and mixed (gain and loss occur in the same region). Besides, we matched the CNVs of respective populations with the corresponding CNVR to obtain the CNVRs of each Duroc pig population. In other words, the CNVRs with full coverage CNV sequences were considered as the population-based CNVRs. Finally, a set of 682 CNVRs mapped in American Duroc pigs and 424 CNVRs mapped in Canadian Duroc pigs were used for subsequent analyses.

Quantitative PCR validation

We chose Quantitative PCR (qPCR) to validate the CNVRs detected by PennCNV software. A total of 9 CNVRs was randomly selected according to the CNVR type (loss, gain, and mixed) and frequency in the population. Due to the uncertain boundary of the identified CNVRs, we used Oligo 7 software [33] to design primers for the specific region of *ADGRA1*, *PUSL1*, *MAPRE2*, *SGMS2*, *PCID2*, *DSCAM*, *GATD3A*, *ELFN1* and *LIFR* genes (see Additional file 4: Table S4). We also selected the *GCG* gene as the reference locus since this gene was highly conserved among pigs and existed as a single copy in animals [17, 34, 35]. Normal samples identified with no copy number change in the test region were used as

a reference. qPCR was conducted on a LightCycler 480 Real-Time PCR System (Roche, Basel, Switzerland). The PCR reaction was performed with a total volume of 10 μl consisting of the following reagents: 1 μl DNA (50 ng/μl), 0.3 μl of both forward primer and reverse primer, 5 μl Blue-SYBR-Green mix (2×) and 3.4 μl water. The PCR conditions were as follows: 10 min at 95 ℃ followed by 40 cycles for 10 s, 60 ℃ for 15 s and 72 ℃ for 20 s. All reactions were carried out on 384-well clear reaction plates and each of sample was performed in triplicate, with average C_t values calculated for further copy number determination. The relative copy number difference of the test region was determined by $2 \times 2^{-\Delta\Delta Ct}$. Where $\Delta\Delta C_t$ = [(mean C_t of the target gene in test sample) - (mean C_t of *GCG* in test sample)] - [(mean C_t of the target gene in reference sample) - (mean C_t of *GCG* in reference sample)] [36]. The values around 2 were considered normal. A value of 3 or more and a value of 1 or less represent the status of gain and loss, respectively.

CNVR genotyping and association analysis

In order to provide the required input for association analysis, a specific genotyping was necessary. We genotyped CNVR type according to the CN value. For instance, A/A for copy deletion (CN = 0,1), A/G for normal copy (CN=2) and G/G for copy increase (CN=3,4).

In this study, a linear model was used to conduct association analysis between CNVR with frequencies large than 0.5% and single trait in each population, including fixed effects (overall mean, covariate) and random effects (residual effect). The analysis was completed by GEMMA software [37]. The linear model used was as follows:

$$Y_{ijkl} = \mu + S_i + P_j + W_k + G_l + e_{ijkl}$$

where Y_{ijkl} is the corrected phenotypic value for each population; μ is the overall mean; S_i is the fixed effects of gender; P_j is the fixed effects of parity; W_k is the fixed effects of birth weight; G_l is the CNVR genotypic effect; e_{ijkl} is the random residuals corresponding to the observed values of the trait. We applied the false discovery rate (FDR) to determine the significance threshold according to the previous studies [38, 39]. *P*-value was defined as follows:

$$P = FDR \times N/M$$

where FDR was set as 0.05, N represents the CNVRs number of $P < 0.05$ in the association results, and M is the total number of CNVRs.

CNVR overlapped with relevant QTLs

All QTL data based on Sscrofa 11.1 version were downloaded from the pig QTL database [40, 41]. We retrieved autosomal QTL regions associated with the ADG, AGE, and BFT traits from pigQTLdb, respectively. Significant CNVRs assessed in the corresponding traits were used to map QTLs.

Functional Genes annotation

The physical position information was obtained from the Sscrofa 11.1 version of the pig reference genome. Genes that overlapped or adjacent to the significant CNVRs were selected for the KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways and GO (Gene Ontology) analysis using KOBAS v3.0 [42]. The statistical method used in the enrichment analysis was Fisher's exact test with a significant threshold of $P < 0.05$. Also, GeneCards (http://www.genecards.org/) and Ensembl (www.ensembl.org/biomart/martview) were used to query gene functions.

Results

Genome-wide CNVs detection in two populations

We performed CNVs detection with PennCNV software on 18 autosomes in Duroc pigs of Canadian and American origin. As a result, a total of 16,142 CNVs (3,693 loss and 12,449 gain) with 5,061 individuals were obtained successfully. Among them, 11,100 CNVs were from 2,945 Duroc pigs of American, and 5,042 CNVs were from 2,116 Duroc pigs of Canadian. All of these CNVs were used to merge into CNVRs (Additional file 1: Table S1). As a result, a total of 835 CNVRs were generated in two populations with 342 gain, 337 loss, and 156 mixed. Table 1 and CNVR map (Fig. 1) summarizes the distribution of total CNVRs on different autosomes. CNVRs in chromosome 4 (SSC4) had the highest coverage (23.86%), while the lowest in SSC1 (5.96%). The number of CNVRs varied from 18 (SSC18) to 74 (SSC4), and the total size of CNVRs detected in this study was 226.73 Mb, accounting for ~ 10.00% of the pig autosomal genome.

Table 1
Chromosome distribution of all 835 CNVRs in the pig autosomes

Chr	Chr length (kb)	CNVR counts	Length of CNVR (kb)	Coverage (%)	Max size (kb)	Average size (kb)	Min size (kb)
1	274,330.50	61	16,345.13	5.96	1,855.10	128.76	11.81
2	151,936	54	19,672.21	12.95	4,253.11	168.98	22.45
3	132,848.90	44	11,265.54	8.48	1,864.38	153.35	18.74
4	130,910.90	74	31,241.02	23.86	1,791.54	244.18	24.44
5	104,526	40	9,314.40	8.91	1,247.57	110.66	14.21
6	170,843.60	61	17,164.90	10.05	1,404.71	184.41	24.07
7	121,844.10	55	11,426.88	9.38	899.03	160.76	18.22
8	138,966.20	47	11,450.49	8.24	1,981.86	122.7	32.22
9	139,512.10	49	10,938.51	7.84	1,329.81	146.11	26.12
10	69,359.45	21	4,729.65	6.82	676.99	200.79	40.11
11	79,169.98	41	9,569.71	12.09	1,635.59	136.85	43.2
12	61,602.75	39	10,251.95	16.64	2,157.61	186.27	21.53
13	208,334.60	61	15,956.15	7.66	1,662.75	186.86	24.59
14	141,755.50	62	15,592.69	11	2,234.96	126.92	14.93
15	140,412.70	46	12,263.88	8.73	2,643.96	119.46	33.1
16	79,944.28	30	7,369.11	9.22	946.64	142.1	29.96
17	63,494.08	32	7,171	11.29	1,683.89	139.82	48.59
18	55,982.97	18	5,003.06	8.94	1,634.02	176.07	29.92

Besides, by matching the CNVs of each population with the corresponding CNVRs, we identified 682 CNVRs in the American Duroc pigs and 424 CNVRs in the Canadian Duroc pigs, of which 271 CNVRs were detected in both populations (see Additional file 2: Table S2). CNVs of American Duroc pigs ranged in length from 10.5 kb to 4.1 Mb, averaging 180.0 kb (Fig. 2a). The length change of CNVRs is similar to CNVs, ranging from 11.8Kb to 4.3 Mb (Fig. 2b). In Canadian

Duroc pigs, the length of CNVs ranged from 14.2 kb to 1.9 Mb, with an average of 151.8 kb (Fig. 2c), and the length distribution of CNVRs shown in Fig. 2d ranged in size from 14.2Kb to 2.6 Mb. In summary, most CNVs and CNVRs in both populations are concentrated in the length range of 50–500 kb, and the CNVRs covered ~ 8.54% and 6.54% of the porcine genome (*Sus scrofa* 11.1) in American and Canadian Duroc pigs, respectively. Notably, CNV of gain type is more likely to occur in both populations.

Comparison of CNVRs detected in previous swine studies

We compared the CNVRs identified in this study with the nine previous swine studies (see Additional file 3: Table S3). Results listed in Table 2 showed varying levels of overlapping CNVRs between studies. Due to the differences in breeds, platforms, algorithms, and CNV definitions that have a more significant impact on the results [43], we used a much looser definition of overlap, where two CNVRs were considered to be overlapping as long as they shared at least one bp [44].

Table 2
Comparison of CNVRs identified in this study with other studies (based on the Sscrofa 11.1 genome assembly)

Study	Platform	Software	Breeds (Number[a])	Samples	Number of CNVRs[b] (original CNVRs[c])	Number of overlapped CNVRs in this study
Chen et al. [57]	Porcine SNP60	PennCNV	Duroc, Rongchang, etc. (18)	1,693	243(565)	57
Wang et al. [54]	Porcine SNP60	PennCNV	Duroc, Laiwu, etc. (10)	302	146(348)	38
Wiedmann et al. [59]	Porcine SNP60	PennCNV	α Mixed Breed Swine (1)	1,802	185(502)	38
Wang et al. [58]	1M aCGH	Agilent Genomic Workbench	Duroc, Yorkshire, etc. (9)	12	436(758)	31
Xie et al. [60]	Porcine SNP60	PennCNV	Xiang, Kele (2)	120	75(172)	14
Stafuzza et al. [46]	Porcine SNP80	PennCNV	Duroc (1)	3,520	136(425)	58
Wang et al. [35]	Porcine SNP80	PennCNV	Large White (1)	857	175(312)	81
Keel et al. [44]	Next-generation sequencing	CNVnator & LUMPY	Duroc, Landrace, etc. (3)	240	3,538	177
Zheng et al. [17]	Next-generation sequencing	CNVnator & CNVcaller	Duroc (1)	29	6,700	182

All CNVRs identified in other studies were converted to Sscrofa 11.1 genome assembly using the liftOver tool. [a]Pig breeds used for comparison; [b]Successfully converted CNVRs; [c]Original number of CNVRs.

The most considerable overlap of CNVRs identified between this study and previous studies were found with results reported by the next-generation sequencing platform studies. The percentage of overlapped CNVRs was 21.20% and 21.80% separately [17, 44].

Validation of identified CNVRs by qPCR

To confirm the reliability of the identified CNVRs, we randomly selected nine CNVRs (CNVR 116, 295, 308, 421, 547, 642, 647, 709, and 768) that co-localized with the *ADGRA1*, *PUSL1*, *MAPRE2*, *SGMS2*, *PCID2*, *DSCAM*, *GATD3A*, *ELFN1* and *LIFR* genes, respectively. Finally, seven of these CNVRs (CNVR 116, 295, 308, 421, 642, 647, and 709) were successfully validated (Fig. 3). Details of the primers were listed in Additional file 4: Table S4.

CNVR frequency in two populations

We also calculated the frequencies of the CNVRs in American and Canadian Duroc populations (Fig. 4). For American Duroc pigs, the frequency of CNVR varied from 0.034% (detected in only one individual) to 28.6% (844 individuals out 2945). A similar situation occurs in the Canadian Duroc pigs, CNVR frequencies range from 0.047% (detected in only one individual) to 37.5% (794 individuals out 2116). Moreover, the frequency of CNVRs was mostly concentrated in 0.03% ~ 0.3%, which indicated that low-frequency CNVRs rarely exist in animals and was challenging to measure reliably[45]. Accordingly, an association analysis was performed using CNVRs, with frequencies exceeding 0.5 percent [46].

Phenotypic statistics and CNVR-based association

Before further study on the association analysis of CNVRs, the statistical summary of ADG, AGE, and BFT for two populations are listed in Table 3. All phenotypic data were approximately followed the Normal distribution. In total, the average ± S.D. (standard deviation) of ADG, AGE and BFT for each population were 619.46 ± 31.69, 158.97 ± 8.20, 8.89 ± 0.94 and 611.91 ± 42.08, 161.14 ± 11.14, 9.55 ± 1.77, respectively.

Table 3
The statistics for the phenotypes of growth traits and fatness in two pig populations

Population	Trait	Unit	N[a]	Mean(± SD)[b]	Min[c]	Max[d]	C.V.(%)[e]
American Duroc	ADG	g/day	3,250	619.46 ± 31.69	525.61	716.58	5.12
	AGE	day	3,251	158.97 ± 8.2	134.42	182.7	5.16
	BFT	mm	3,232	8.89 ± 0.94	6.09	12.07	10.57
Canadian Duroc	ADG	g/day	2,569	611.91 ± 42.08	484.52	738.4	6.88
	AGE	day	2,567	161.14 ± 11.14	127.82	195.29	6.91
	BFT	mm	2,550	9.55 ± 1.77	5.1	15.06	18.53

ADG: Average daily gain at 100 kg; AGE: days to 100 kg; BFT: backfat thickness at 100 kg; [a]Number of animals (N); [b]Mean (standard deviation); [c]Minimum (min); [d]Maximum (max); [e]Coefficient of variation (C.V.).

We investigated the relationship between CNVRs and growth, fatness traits in American and Canadian Duroc pigs. 116 CNVRs in American Duroc and 61 CNVRs in Canadian Duroc pigs were selected for association analysis. The Manhattan plots and significant CNVRs ($P < FDR \times n/m$) obtained by separate association analyses for those two populations were shown in Fig. 5 and Fig. 6, Table 4 and Table 5.

Table 4
Significant CNVRs associated with growth traits in two populations

Population	Traits	CNVR ID	Chromosome	Start(bp)	End(bp)	*P*-value	Candidate genes
American Duroc	ADG	CNVR 114	2	150,376,169	150,741,751	7.93E-03	
	ADG	CNVR 118	3	2,457,661	3,480,462	2.42E-03	*SDK1*
	ADG	CNVR 525	11	39,893,652	40,023,260	7.86E-03	
	ADG	CNVR 566	12	24,419,703	24,739,497	8.33E-03	
	ADG; AGE[a]	CNVR 115	2	150,797,343	151,385,319	4.58E-04; 3.23E-04	*PDGFRB*,*PPARGC1B*,*CSF1R*
	ADG; AGE[a]	CNVR 440	9	3,080,614	3,377,860	1.51E-03; 5.70E-04	
	ADG; AGE[a]	CNVR 547	11	77,144,460	78,780,052	1.25E-04; 3.14E-04	*GRTP1*
	ADG; AGE[a]	CNVR 642	13	204,187,794	204,458,963	3.44E-03; 3.04E-03	
	ADG; AGE[a]	CNVR 647	13	206,578,011	208,240,759	1.43E-08; 2.11E-07	*PFKL*
	ADG; AGE[a]	CNVR 819	18	1,697,628	3,331,650	1.34E-03; 5.84E-03	*INSIG1*,*LMBR1*,*DPP6*
Canadian Duroc	AGE	CNVR 162	4	3,149,505	4,034,399	4.60E-03	
	ADG; AGE[a]	CNVR 116	3	162,027	2,026,411	2.49E-03; 2.24E-03	*GPER1*,*PDGFA*,*GNA12*
	ADG; AGE[a]	CNVR 308	6	118,880,845	119,223,012	5.40E-03; 4.33E-03	
	ADG; AGE[a]	CNVR 77	2	22,300,642	23,050,340	2.02E-05; 1.03E-04	

ADG: Average daily gain at 100 kg; AGE: days to 100 kg; [a] CNVRs identified in both traits.

Table 5
Significant CNVRs associated with BFT in American Duroc pigs

CNVR ID	Chromosome	Start(bp)	End(bp)	*P*-value	Candidate genes
CNVR 1	1	437,279	2,292,374	5.73E-04	*TCTE3*,*C6orf120*,*PHF10*,*CCR6*,*THBS2*
CNVR 116[a]	3	162,027	2,026,411	3.14E-04	*GPER1*,*PDGFA*,*GNA12*
CNVR 118[a]	3	2,457,661	3,480,462	4.54E-03	*SDK1*
CNVR 234	5	222,125	1,460,083	4.34E-03	*SCO2*,*HDAC10*,*ALG12*
CNVR 274	6	51,842	1,347,980	1.91E-05	*MC1R*
CNVR 275	6	2,430,990	3,835,699	4.50E-03	*FOXC2*
CNVR 296	6	64,268,261	65,189,989	4.55E-03	*PRDM16*
CNVR 315	6	142,358,005	143,256,333	1.13E-02	
CNVR 445	9	7,165,800	7,567,846	1.52E-03	
CNVR 547[a]	11	77,144,460	78,780,052	7.58E-08	*GRTP1*
CNVR 647[a]	13	206,578,011	208,240,759	1.95E-03	*PFKL*
CNVR 709	14	139,484,309	141,719,266	4.22E-05	
CNVR 745	15	120,041,793	120,478,103	4.16E-03	*SLC11A1*,*VIL1*,*TNS1*
CNVR 785	16	78,550,922	79,365,542	7.55E-03	*LPCAT1*

[a]CNVRs also identified in growth traits.

For growth traits, we obtained a total of ten significant CNVRs for ADG in the American Duroc pigs, where threshold values were set to 9.91E-03. The candidate regions are located on SSC2, 3, 9, 11, 12, 13, and 18. Furthermore, we also found that eight CNVRs, on SSC2, 9, 11, 13, and 18, were exceeded the threshold (9.48E-03) for AGE. These CNVRs are significantly correlated with growth traits that covered 107 protein-coding genes. In the Canadian Duroc pigs, we also identified three, and four CNVRs reached the threshold (8.20E-03; 8.20E-03) for ADG and AGE, respectively, and these CNVRs overlapped 45 protein-coding genes. It should be noted that, due to the high genetic correlation between ADG and AGE [19], six and three CNVRs in American and Canadian Duroc pigs showed significant in both traits separately. However, no shared CNVR was identified in the two experimental populations.

For fatness traits, we identified 14 CNVRs in the BFT trait of American Duroc pigs, where threshold values were set to 1.25E-02. Intriguingly, four CNVRs located on SSC3, 11, and 13 were shared both significant in growth and fatness traits. Nevertheless, we did not find significant CNVRs for the BFT trait in Canadian Duroc pigs.

Based on data of all breeds, we further investigated the function of genes encompassing these significant CNVRs. For ADG and AGE, many common significant CNVRs that are involved in both traits were found to overlap with numerous genes. Consequently, 12 genes were determined as major functional candidates, including *SDK1*, *PDGFRB*, *PPARGC1B*, *CSFIR*, and other candidates. For BTF, we identified 21 candidate genes, such as *TCTE3*, *GPER1*, *PDGFA*, and *LPCAT1*.

Growth and fatness traits-related QTLs overlapped with identified CNVRs

The autosomal QTLs associated with ADG, AGE, and BFT were extracted to explore their overlapping regions on significant CNVRs (see Additional file 5: Table S5). In growth traits, we found that 14 CNVRs overlapped with 11 QTLs for ADG, and the overlapping sequence accounting for ~ 1.26% of these QTLs. Among them, eight QTLs overlapped in American Duroc pigs, and three QTLs overlapped in Canadian Duroc pigs. For AGE, we only found two CNVRs overlapped with two QTLs in the American Duroc population, and the overlapping area accounts for ~ 2.71% of the two QTLs. In terms of fatness, 11 CNVRs were also discovered to overlap with 51 QTLs for BFT in the American Duroc population. These results may help us narrow down the QTL area and anchor range of functional regions associated with complex traits.

Functional analysis of genes involved in the trait-related CNVRs

A total of 521 genes overlapped with 24 significant CNVRs were detected based on the Ensembl annotation of the Sus scrofa 11.1 genome (see Additional file 6: Table S6). Including 342 protein-coding genes, 148 lncRNA genes, as well as some miRNAs, small nucleolar genes (snoRNA), processed pseudogenes, and other genes. To further investigate function genes that affect growth performance and fatness, the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways and Gene Ontology (GO) analysis of protein-coding genes were carried out using KOBAS software (version 3.0).

Gene set enrichment analysis revealed that many terms might be relevant to growth and fatness traits (see Additional file 7: Table S7). In brief, the KEGG analysis uncovers that these genes mainly significantly participated in Rap1 signaling pathway, which was consistent with Zheng et al. [17] and Wang et al. [43] Furthermore, the top terms of GO analysis were primarily enriched in negative regulation of lipid biosynthetic and metabolic process. Given that, we further identified the functions of these genes involved in the critical pathways and biological processes from the GeneCards database and relevant literature. Hence, we highlighted five genes of interest that overlapped with significant CNVRs as well as enriched in Gene set enrichment analysis. Including Platelet-Derived Growth Factor Receptor Beta (*PDGFRB*), Platelet Derived Growth Factor Subunit A (*PDGFA*), G Protein-Coupled Estrogen Receptor 1 (*GPER1*), Insulin Induced Gene 1 (*INSIG1*) and Lysophosphatidylcholine Acyltransferase 1 (*LPCAT1*).

Discussion

Over the past decade, genome-wide association studies (GWAS) have made remarkable contributions to the discovery of common SNPs influencing complex traits [47]. However, most variants explain a small proportion of the heritability, which is called “missing heritability” [48]. To this point, CNVs as an essential source of genetic diversity may provide a new way to explain the genetic variability that GWAS cannot pick up [49].

In this study, we successfully detected 11,100 and 5,042 CNVs in Duroc pigs of American and Canadian and inferred CNVs using rigorous criteria to reduce the risk of false-positive rate. All CNVs were merged to generate 835 CNVRs in two populations, accounting for ~ 10.00% of the pig autosomal genome (*Sus scrofa* 11.1). The results showed that the length and frequency of duplications were much higher than that of deletions in the large fragment (> 10 kb) CNVs (12,449 gain vs. 3,693 loss). Previous CNVs studies have found similar cases. For example, in the CNV study conducted by Long et al. [50] using Porcine SNP60 BeadChip, there were approximately 70.6% gains and 29.4% losses, Zheng et al. [17] also reported that the frequency of duplications was higher than that of deletions in Duroc and Meishan pigs through Next-generation sequencing. This phenomenon suggests that although CNVs can cause duplications or deletions at the same locus in different populations [51], the genome has a greater tolerant for duplication than deletion [52], and is more likely to occur in large-scale CNVs (> 10 kb) [5, 53].

In addition, to evaluate the accuracy of the PennCNV software in determining CNV, we performed qPCR validation for nine randomly selected CNVRs and successfully confirmed seven of nine CNVRs (~ 77.8%). This percentage is similar to Wang et al. [54] (75%), Dong et al. [55] (70%), and Wang et al. [43] (80%). Meanwhile, we also observed that two CNVRs were inconsistent with our expectations. Multiple factors may have contributed to the result of discordance. For example, the

sparse probes on the SNP chip may cause the identified CNVRs to be larger than the actual length. As a consequence, the primers may have been designed outside the exact boundaries of CNVRs [55]. Additionally, these results also indicate that there is a high proportion of singleton CNVs exist in the population [56].

We also compared our results with previous studies on CNVRs, which showed a low overlap rate [17, 35, 44, 46, 54, 57–60]. In brief, a total of 404 CNVRs entirely or partially overlapped with the previously reported CNVRs. A considerable overlap rate was observed in the results of Zheng et al. [17], while that of Xie et al. [60] was the lowest. These discrepant observations may be due to the difference in the breed and population size between our study and others. In this study, the large-scale samples were used for CNV detection, which leads us to identify more novel CNVRs than other studies. It also suggests that a vast amount of CNVs existing in the pig genome has not been discovered [61]. Intriguingly, even within the same breed, different genetic backgrounds may have significant effects on reproducibility. As can be found in our result, only 271 of 835 CNVRs were detected in both populations. Also, previous studies were almost based on the Sscrofa10.2 genome version, while the comparative work in our study was on Sscrofa11.1. Thus, based on the vast differences between the two versions [62], many CNVRs in the Sscrofa10.2 could not be converted to Scrofa11.1 successfully (Table 2). Besides, differences in SNPs density after quality control, as well as different CNV detection platforms, algorithms, and criteria for CNV determination could also explain this outcome [43].

Although CNVs were widespread in pigs and have been reported to be associated with economically relevant traits, the full relevance of CNVs to the genetic architecture of growth rate and fatness across all stages is yet to be elucidated. To further investigate the relationship between CNVs and complex traits (ADG, AGE, and BFT), we performed CNVRs-based association analysis in two populations. For American Duroc pigs, we identified a total of ten CNVRs on ADG and AGE, including six CNVRs that were significant in both traits. A similar pattern occurs in Canadian Duroc pigs. For instance, we detected three CNVRs affecting both traits. The computational formulas of ADG and AGE in this study show an inverse relationship, and both themselves also have relatively high genetic associations [19]. That may explain why most CNVRs were significant in both traits.

However, the results of the association analysis between American and Canadian Duroc pigs differed substantially. Whether ADG or AGE, we found no shared CNVR in the two experimental populations. Moreover, we only detected 14 BTF-related CNVRs in American pigs, but no CNVR was identified in the Canadian population. This finding highlights the complex genetic architecture of growth traits and fatness. Although the Duroc is considered as one breed, substantial genetic differences exist between subpopulations [63], for example, the American and Canadian Duroc pigs in this study. The results were consistent with those of Zhuang et al. [39]. During the differences in natural and human selective pressures, it is presumed that the genetic drift and the exchange of genetic material lead to less consistent in CNVR between the two populations [64]. Therefore, the genetic differentiation that exists between two populations may have a substantial impact on the localization of genetic variant regions [39]. More notably, four CNVRs were associated with growth and fatness traits (CNVR116, CNVR118, CNVR547, and CNVR647). The results suggested that these CNVRs may play a pleiotropic role in regulating pig growth and fat deposition [18, 20].

The GO enrichment and KEGG analysis uncovered that genes in CNVRs are mainly involved in lipid biosynthetic, lipid metabolic process, and O-glycan biosynthesis. Based on our association analysis, gene set enrichment analysis, and the function of relevant candidates, consequently, five genes relevant to growth traits and fatness were highlighted. *INSIG1* gene act as regulators of lipid metabolism in cholesterol and fatty acid synthesis [65, 66]. Yang et al. [67] demonstrated that the pig fed the high-fat diet will increase the *INSIG1* gene expression in backfat tissue. Liu et al. [68] also discovered that polymorphisms in this gene were associated with growth and carcass traits through PCR-RFLP and DNA sequencing. White adipose tissue is recognized as an energy-storing organ, which is closely related to fat deposition and body weight [69]. Human adipose tissue differentiation into beige or white adipocytes depends on *PDGFRβ* expression [70]. Olson et al. [71] identified that increased *PDGFRB* signaling inhibits differentiation of white adipose. The protein encoded by this gene is a Tyrosine-protein kinase that acts as cell-surface receptor for heterodimers formed by *PDGFA* and *PDGFB*. While in

recent years, Gonzalez et al. [72] found that *PDGFA* also plays a vital role in the proliferation and maintenance of adipocyte progenitors in dermal adipose tissues through PI3K-Akt pathways. The *GPER1* gene, known as G protein-coupled estrogen receptor 1, is involved in metabolism and immunity [73]. Sharma et al. [74] reported that weight gain in male *GPER* knockout (KO) mice was associated with visceral and subcutaneous fat. However, *GPER* KO mice showed no differences in food intake or exercise activity compared with wild-type littermates. This observation demonstrated that *GPER* might regulate metabolic parameters associated with obesity. *LPCAT1* plays a role in many disease processes. As an example, *LPCAT1* was down-regulated in the brain tissue of diabetic mice and up-regulated in antidiabetic therapy [75]. Chen et al. [76] also revealed that DNA methylation levels of the *LPCAT1* gene were related to adiposity.

In recent years, the study for the influence of CNV on complex traits has gradually been thrust into the limelight [17, 35]. As far as we know, the present study is the most massive sample case of the genome-wide CNV detection for Duroc pigs. However, due to the sparse makers of the SNP chip, we may overestimate the frequency of large-scale CNV detected in our study. Accordingly, high-density SNP chip or whole-genome sequencing technology should be applied in further CNV detection analysis.

Conclusions

In this study, we performed the genome-wide CNV detection and CNVR-based association between growth traits and fatness in the large-scale population of American and Canadian Duroc pigs. A total of 835 CNVRs were detected in two populations, which account for ~ 10.00% of the pig autosomal genome. Moreover, 24 CNVRs were identified to be associated with growth traits and fatness. However, we found no shared CNVR in the two experimental populations. These findings indicated that the genetic differentiation exists between two populations may have a substantial impact on the localization of genetic variant regions. Also, we identified major candidates that may be related to growth traits and fatness, such as *PDGFRB*, *INSIG1*, *GPER1*, *PDGFA*, and *LPCAT1*. Our results provide valuable insights into the genetic mechanism of growth and fatness traits in pigs.

List Of Abbreviations

CNV: copy number variation; CNVR: copy number variation region; CN: copy number; ADG: average daily gain: AGE: days to 100 kg; BFT: backfat thickness; SSC: *Sus scrofa* chromosome; GWAS: genome-wide association study; SNP: single nucleotide polymorphism; INDEL: insertion and deletion; QTL: quantitative trait locus; qPCR: quantitative PCR;

Declarations

Ethics approval and consent to participate

The animals and experimental methods used in this study are following the guidelines of the Ministry of Agriculture of China and Use Committee of South China Agricultural University (SCAU). The ethics committee of SCAU (Guangzhou, China) approved all of the animal experiments. The informed consent was obtained from Wens Foodstuff Group Co., Ltd. (Guangdong, China) to data collection. There was no use of human participants, data or tissues.

Consent for publication

Not applicable

Availability of data and material

The SNP genotyping data containing variant information for the American (n=3,770) and Canadian (n=2,857) Duroc pigs are not publicly available because the genotyped animals belong to commercial breeding companies, but they can be

obtained from the corresponding author under reasonable requirements.

Competing interests

The authors declare that they have no competing interests.

Funding

This research was supported by the Guangdong YangFan Innovative and Entrepreneurial Research Team Program (2016YT03H062), the Key-Area Research and Development Program of Guangdong Province (2018B020203002), the Natural Science Foundation of Guangdong Province (2018B030315007), and the Pearl River S and T Nova Program of Guangzhou (201906010011). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Authors' contributions

JY and ZW conceived and designed the experiment. YQ, RD, ZZ, JW, MY, SZ, YY, QG, ZX, SH, and GC collected the samples and recorded the phenotypes. ZZ, JW, SZ, YY, and QG extracted the DNA for genotyping. JW and YY conducted the qPCR test. YQ, RD, and ZZ analyzed the data. YQ, RD, and JY wrote the manuscript. ZW contributed to the materials. All authors reviewed and approved the manuscript.

Acknowledgments

The authors would like to thank all staff at the pig core breeding farms of Wens Foodstuff Group Co., Ltd. (Guangdong, China) for the help of sample collection.

References

1. Feuk L, Carson AR, Scherer SW. Structural variation in the human genome. Nat Rev Genet. 2006;7(2):85-97.
2. 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):D986-92.
3. Zarrei M, MacDonald JR, Merico D, Scherer SW. A copy number variation map of the human genome. Nat Rev Genet. 2015;16(3):172-83.
4. 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.
5. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, et al. Global variation in copy number in the human genome. Nature. 2006;444(7118):444-54.
6. Stankiewicz P, Lupski JR. Structural variation in the human genome and its role in disease. Annu Rev Med. 2010;61:437-55.
7. Cooper GM, Nickerson DA, Eichler EE. Mutational and selective effects on copy-number variants in the human genome. Nat Genet. 2007;39(7 Suppl):S22-9.
8. Saitou M, Gokcumen O. An Evolutionary Perspective on the Impact of Genomic Copy Number Variation on Human Health. J Mol Evol. 2020;88(1):104-19.
9. Gonzalez E, Kulkarni H, Bolivar H, Mangano A, Sanchez R, Catano G, et al. The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. Science. 2005;307(5714):1434-40.
10. Marshall CR, Scherer SW. Detection and characterization of copy number variation in autism spectrum disorder. Methods Mol Biol. 2012;838:115-35.

11. Kushima I, Aleksic B, Nakatochi M, Shimamura T, Shiino T, Yoshimi A, et al. High-resolution copy number variation analysis of schizophrenia in Japan. Mol Psychiatry. 2017;22(3):430-40.
12. Aitman TJ, Dong R, Vyse TJ, Norsworthy PJ, Johnson MD, Smith J, et al. Copy number polymorphism in Fcgr3 predisposes to glomerulonephritis in rats and humans. Nature. 2006;439(7078):851-5.
13. Aerts E, Beckers S, Zegers D, Van Hoorenbeeck K, Massa G, Verrijken A, et al. CNV analysis and mutation screening indicate an important role for the NPY4R gene in human obesity. Obesity (Silver Spring). 2016;24(4):970-6.
14. Rubin CJ, Megens HJ, Martinez Barrio A, Maqbool K, Sayyab S, Schwochow D, et al. Strong signatures of selection in the domestic pig genome. Proc Natl Acad Sci U S A. 2012;109(48):19529-36.
15. Giuffra E, Tornsten A, Marklund S, Bongcam-Rudloff E, Chardon P, Kijas JM, et al. A large duplication associated with dominant white color in pigs originated by homologous recombination between LINE elements flanking KIT. Mamm Genome. 2002;13(10):569-77.
16. Ran XQ, Pan H, Huang SH, Liu C, Niu X, Li S, et al. Copy number variations of MTHFSD gene across pig breeds and its association with litter size traits in Chinese indigenous Xiang pig. J Anim Physiol Anim Nutr (Berl). 2018.
17. Zheng X, Zhao P, Yang K, Ning C, Wang H, Zhou L, et al. CNV analysis of Meishan pig by next-generation sequencing and effects of AHR gene CNV on pig reproductive traits. J Anim Sci Biotechnol. 2020;11:42.
18. Revilla M, Puig-Oliveras A, Castello A, Crespo-Piazuelo D, Paludo E, Fernandez AI, et al. 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.
19. Tang Z, Xu J, Yin L, Yin D, Zhu M, Yu M, et al. Genome-Wide Association Study Reveals Candidate Genes for Growth Relevant Traits in Pigs. Front Genet. 2019;10:302.
20. Qiao R, Gao J, Zhang Z, Li L, Xie X, Fan Y, et al. Genome-wide association analyses reveal significant loci and strong candidate genes for growth and fatness traits in two pig populations. Genet Sel Evol. 2015;47:17.
21. Martinez-Montes AM, Fernandez A, Munoz M, Noguera JL, Folch JM, Fernandez AI. Using genome wide association studies to identify common QTL regions in three different genetic backgrounds based on Iberian pig breed. PLoS One. 2018;13(3):e0190184.
22. Liu G, Jennen DG, Tholen E, Juengst H, Kleinwachter T, Holker M, et al. A genome scan reveals QTL for growth, fatness, leanness and meat quality in a Duroc-Pietrain resource population. Anim Genet. 2007;38(3):241-52.
23. Xu L, Yang L, Wang L, Zhu B, Chen Y, Gao H, et al. Probe-based association analysis identifies several deletions associated with average daily gain in beef cattle. BMC Genomics. 2019;20(1):31.
24. Ding R, Yang M, Wang X, Quan J, Zhuang Z, Zhou S, et al. Genetic Architecture of Feeding Behavior and Feed Efficiency in a Duroc Pig Population. Front Genet. 2018;9:220.
25. Suzuki K, Kadowaki H, Shibata T, Uchida H, Nishida A. Selection for daily gain, loin-eye area, backfat thickness and intramuscular fat based on desired gains over seven generations of Duroc pigs. Livestock Production Science. 2005;97(2-3):193-202.
26. Wang Y, Ding X, Tan Z, Ning C, Xing K, Yang T, et al. Genome-Wide Association Study of Piglet Uniformity and Farrowing Interval. Front Genet. 2017;8:194.
27. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015;4:7.
28. Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SF, et al. 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):1665-74.
29. Diskin SJ, Li M, Hou C, Yang S, Glessner J, Hakonarson H, et al. Adjustment of genomic waves in signal intensities from whole-genome SNP genotyping platforms. Nucleic Acids Res. 2008;36(19):e126.

30. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26(6):841-2.
31. Quinlan AR. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. Curr Protoc Bioinformatics. 2014;47:11 2 1-34.
32. Kim JH, Hu HJ, Yim SH, Bae JS, Kim SY, Chung YJ. CNVRuler: a copy number variation-based case-control association analysis tool. Bioinformatics. 2012;28(13):1790-2.
33. Rychlik W. OLIGO 7 primer analysis software. Methods Mol Biol. 2007;402:35-60.
34. Ballester M, Castello A, Ibanez E, Sanchez A, Folch JM. Real-time quantitative PCR-based system for determining transgene copy number in transgenic animals. Biotechniques. 2004;37(4):610-3.
35. 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.
36. Lin CH, Lin YC, Wu JY, Pan WH, Chen YT, Fann CS. A genome-wide survey of copy number variations in Han Chinese residing in Taiwan. Genomics. 2009;94(4):241-6.
37. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. Nat Genet. 2012;44(7):821-4.
38. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society. 1995;57(1):289-300.
39. Zhuang Z, Ding R, Peng L, Wu J, Ye Y, Zhou S, et al. Genome-wide association analyses identify known and novel loci for teat number in Duroc pigs using single-locus and multi-locus models. BMC Genomics. 2020;21(1):344.
40. Hu Z-L, Park CA, Reecy JM. Building a livestock genetic and genomic information knowledgebase through integrative developments of Animal QTLdb and CorrDB. Nucleic Acids Research. 2018;47(D1):D701-D10.
41. Hu ZL, Dracheva S, Jang W, Maglott D, Bastiaansen J, Rothschild MF, et al. A QTL resource and comparison tool for pigs: PigQTLDB. Mamm Genome. 2005;16(10):792-800.
42. Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, et al. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. Nucleic Acids Res. 2011;39(Web Server issue):W316-22.
43. 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. 2019.
44. 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.
45. Pearson TA, Manolio TA. How to interpret a genome-wide association study. JAMA. 2008;299(11):1335-44.
46. Stafuzza NB, Silva RMO, Fragomeni BO, Masuda Y, Huang Y, Gray K, et al. A genome-wide single nucleotide polymorphism and copy number variation analysis for number of piglets born alive. BMC Genomics. 2019;20(1):321.
47. Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res. 2019;47(D1):D1005-D12.
48. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature. 2009;461(7265):747-53.
49. Maher B. Personal genomes: The case of the missing heritability. Nature. 2008;456(7218):18-21.
50. Long Y, Su Y, Ai H, Zhang Z, Yang B, Ruan G, et al. A genome-wide association study of copy number variations with umbilical hernia in swine. Anim Genet. 2016;47(3):298-305.
51. Conrad DF, Hurles ME. The population genetics of structural variation. Nat Genet. 2007;39(7 Suppl):S30-6.
52. Brewer C, Holloway S, Zawalnyski P, Schinzel A, FitzPatrick D. A chromosomal duplication map of malformations: regions of suspected haplo- and triplolethality--and tolerance of segmental aneuploidy--in humans. Am J Hum Genet.

1999;64(6):1702-8.
53. Locke DP, Sharp AJ, McCarroll SA, McGrath SD, Newman TL, Cheng Z, et al. Linkage disequilibrium and heritability of copy-number polymorphisms within duplicated regions of the human genome. Am J Hum Genet. 2006;79(2):275-90.
54. Wang Y, Tang Z, Sun Y, Wang H, Wang C, Yu S, 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.
55. Dong K, Pu Y, Yao N, Shu G, Liu X, He X, et al. Copy number variation detection using SNP genotyping arrays in three Chinese pig breeds. Anim Genet. 2015;46(2):101-9.
56. Upadhyay M, da Silva VH, Megens HJ, Visker M, Ajmone-Marsan P, Balteanu VA, et al. Distribution and Functionality of Copy Number Variation across European Cattle Populations. Front Genet. 2017;8:108.
57. Chen C, Qiao R, Wei R, Guo Y, Ai H, Ma J, et al. 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.
58. 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). 2015;5(6):1253-61.
59. 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.
60. Xie J, Li R, Li S, Ran X, Wang J, Jiang J, et al. Identification of Copy Number Variations in Xiang and Kele Pigs. PLoS One. 2016;11(2):e0148565.
61. Jiang L, Jiang J, Yang J, Liu X, Wang J, Wang H, et al. Genome-wide detection of copy number variations using high-density SNP genotyping platforms in Holsteins. BMC Genomics. 2013;14:131.
62. Warr A, Robert C, Hume D, Archibald AL, Deeb N, Watson M. Identification of Low-Confidence Regions in the Pig Reference Genome (Sscrofa10.2). Front Genet. 2015;6:338.
63. Gorssen W, Meyermans R, Buys N, Janssens S. SNP genotypes reveal breed substructure, selection signatures and highly inbred regions in Pietrain pigs. Anim Genet. 2020;51(1):32-42.
64. Lynch M, Ackerman MS, Gout JF, Long H, Sung W, Thomas WK, et al. Genetic drift, selection and the evolution of the mutation rate. Nat Rev Genet. 2016;17(11):704-14.
65. Yang T, Espenshade PJ, Wright ME, Yabe D, Gong Y, Aebersold R, et al. Crucial step in cholesterol homeostasis: sterols promote binding of SCAP to INSIG-1, a membrane protein that facilitates retention of SREBPs in ER. Cell. 2002;110(4):489-500.
66. Radhakrishnan A, Ikeda Y, Kwon HJ, Brown MS, Goldstein JL. Sterol-regulated transport of SREBPs from endoplasmic reticulum to Golgi: oxysterols block transport by binding to Insig. Proc Natl Acad Sci U S A. 2007;104(16):6511-8.
67. Yang XF, Qiu YQ, Wang L, Gao KG, Jiang ZY. A high-fat diet increases body fat mass and up-regulates expression of genes related to adipogenesis and inflammation in a genetically lean pig. J Zhejiang Univ Sci B. 2018;19(11):884-94.
68. Liu Y, Liu XL, He H, Gu YL. Four SNPs of insulin-induced gene 1 associated with growth and carcass traits in Qinchuan cattle in China. Genet Mol Res. 2012;11(2):1209-16.
69. Sun K, Kusminski CM, Scherer PE. Adipose tissue remodeling and obesity. J Clin Invest. 2011;121(6):2094-101.
70. Gao Z, Daquinag AC, Su F, Snyder B, Kolonin MG. PDGFRalpha/PDGFRbeta signaling balance modulates progenitor cell differentiation into white and beige adipocytes. Development (Cambridge, England). 2018;145(1).
71. Olson LE, Soriano P. PDGFRbeta signaling regulates mural cell plasticity and inhibits fat development. Developmental cell. 2011;20(6):815-26.
72. Rivera-Gonzalez GC, Shook BA, Andrae J, Holtrup B, Bollag K, Betsholtz C, et al. Skin Adipocyte Stem Cell Self-Renewal Is Regulated by a PDGFA/AKT-Signaling Axis. Cell stem cell. 2016;19(6):738-51.

73. Olde B, Leeb-Lundberg LM. GPR30/GPER1: searching for a role in estrogen physiology. Trends Endocrinol Metab. 2009;20(8):409-16.
74. Sharma G, Hu C, Brigman JL, Zhu G, Hathaway HJ, Prossnitz ER. GPER deficiency in male mice results in insulin resistance, dyslipidemia, and a proinflammatory state. Endocrinology. 2013;154(11):4136-45.
75. Stanca E, Serviddio G, Bellanti F, Vendemiale G, Siculella L, Giudetti AM. Down-regulation of LPCAT expression increases platelet-activating factor level in cirrhotic rat liver: potential antiinflammatory effect of silybin. Biochim Biophys Acta. 2013;1832(12):2019-26.
76. Chen L, Dong Y, Wang X, Hao G, Huang Y, Gutin B, et al. Epigenome-Wide Association Study of Dietary Fiber Intake in African American Adolescents. Mol Nutr Food Res. 2018;62(12):e1800155.

Figures

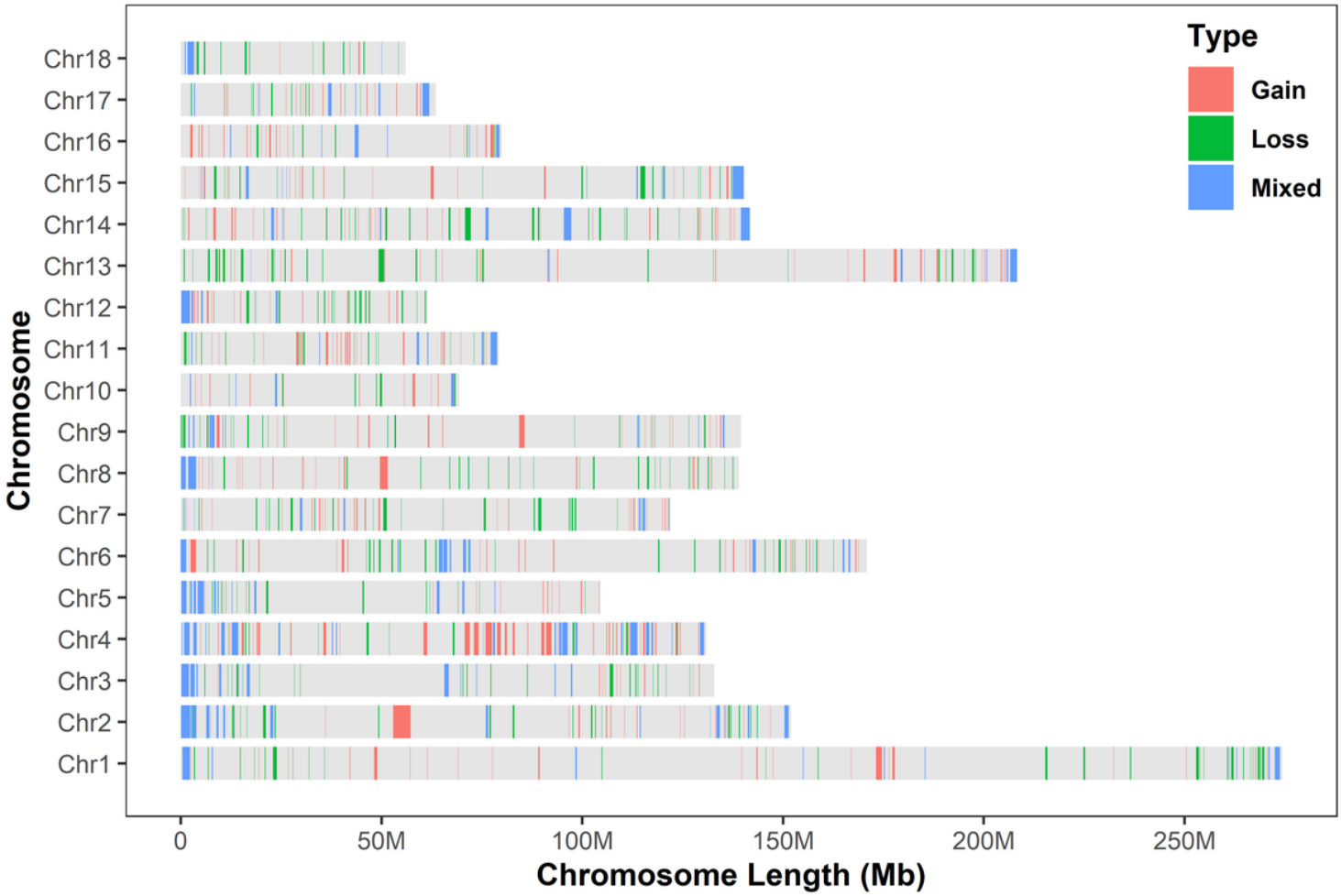

Figure 1

The overall CNVR maps for American and Canadian Duroc pigs in the 18 autosomes. Three types of CNVR are identified, including gain (red), Loss (green), and Mixed (blue). Y-axis values are autosomes, and X-axis values are chromosome position in Mb.

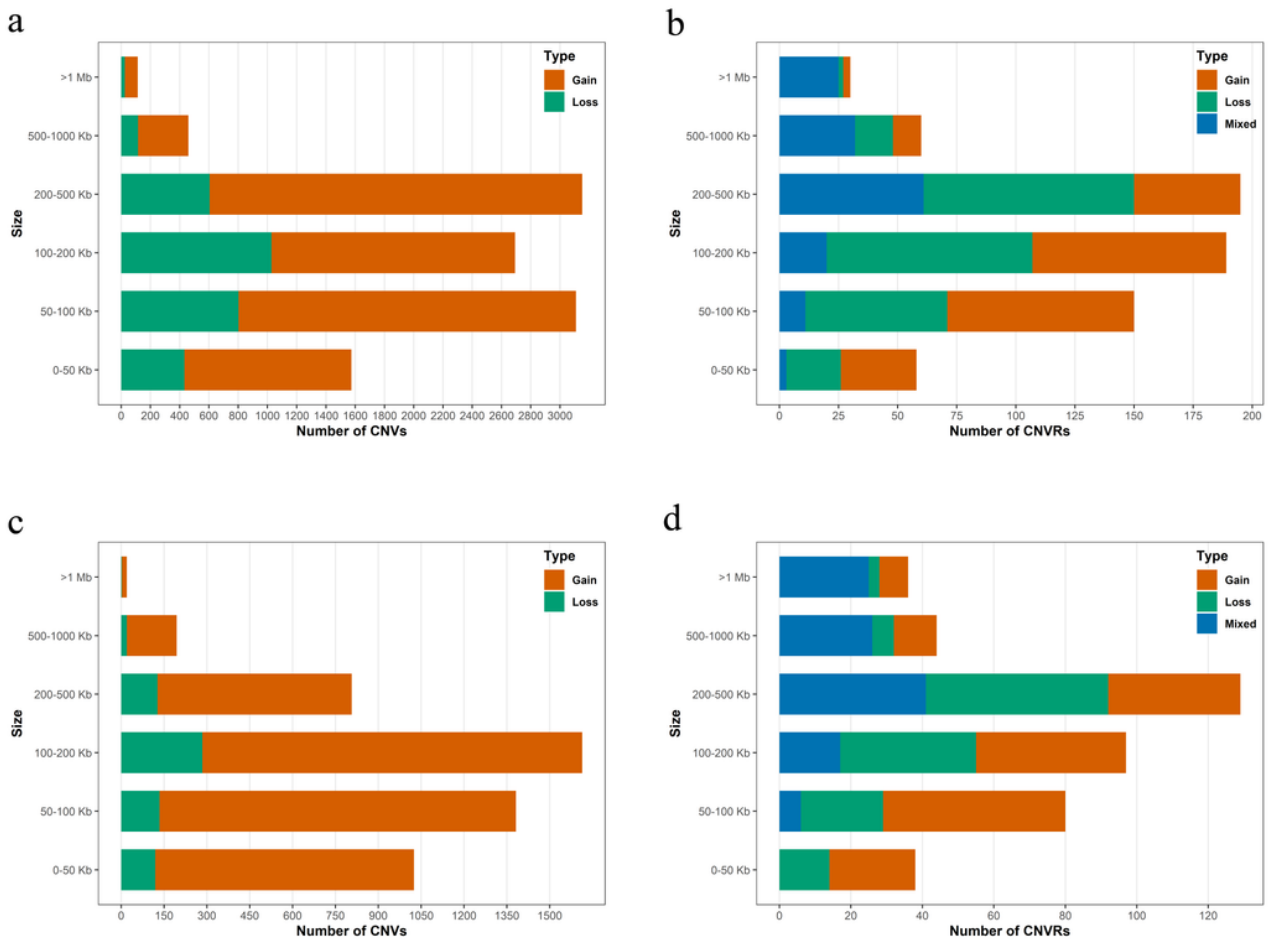

Figure 2

CNV and CNVR distribution of American and Canadian Duroc pigs according to the size interval. The plots of a and b show the CNV and CNVR distribution in American Duroc pigs, respectively. The plots of c and d show the CNV and CNVR distribution in Canadian Duroc pigs, respectively.

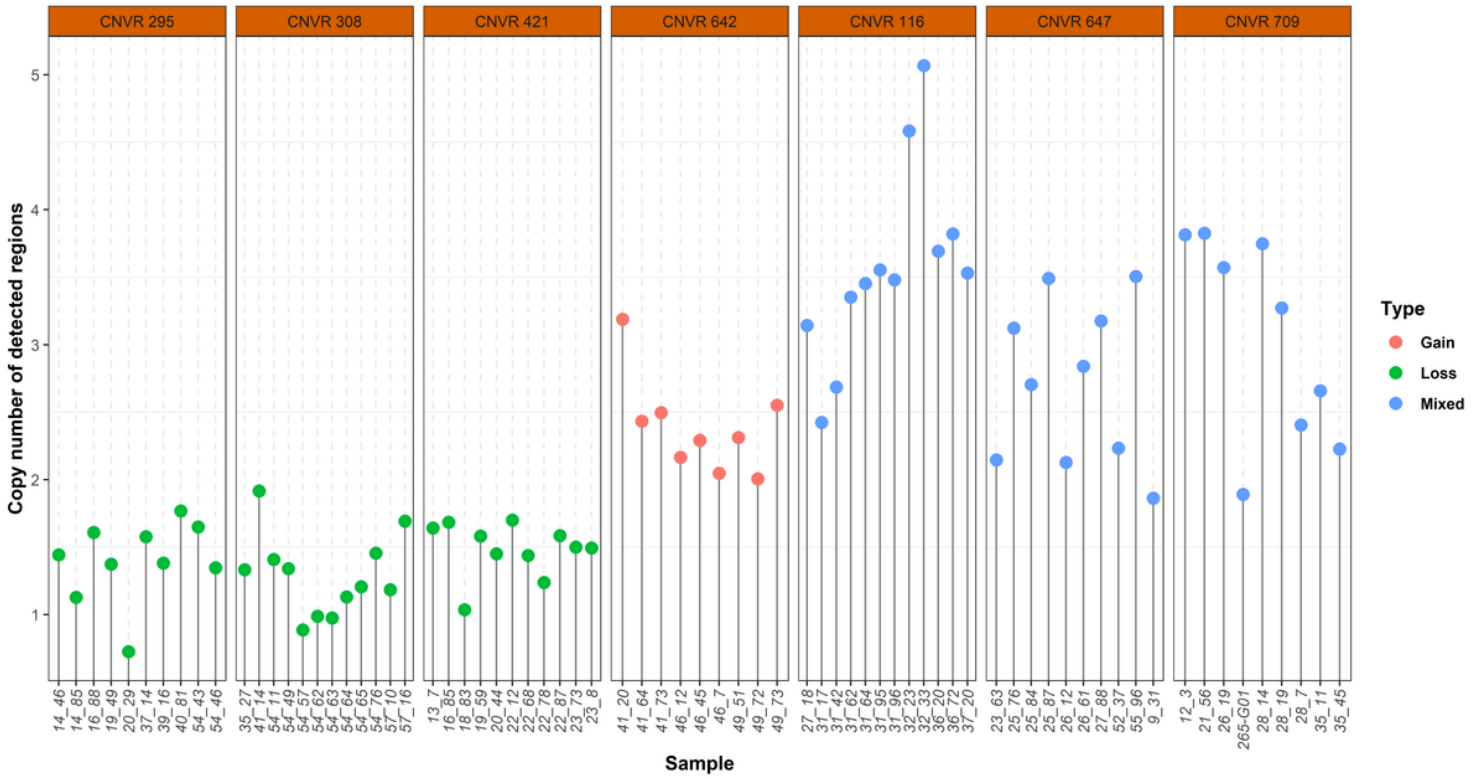

Figure 3

The results of qPCR validation in selected CNVRs. The x-axis represents the tested sample ID. The y-axis represents different copy number, the values around 2 were considered normal. The value of 3 or more means gain and a value of 1 or less represents the status of loss.

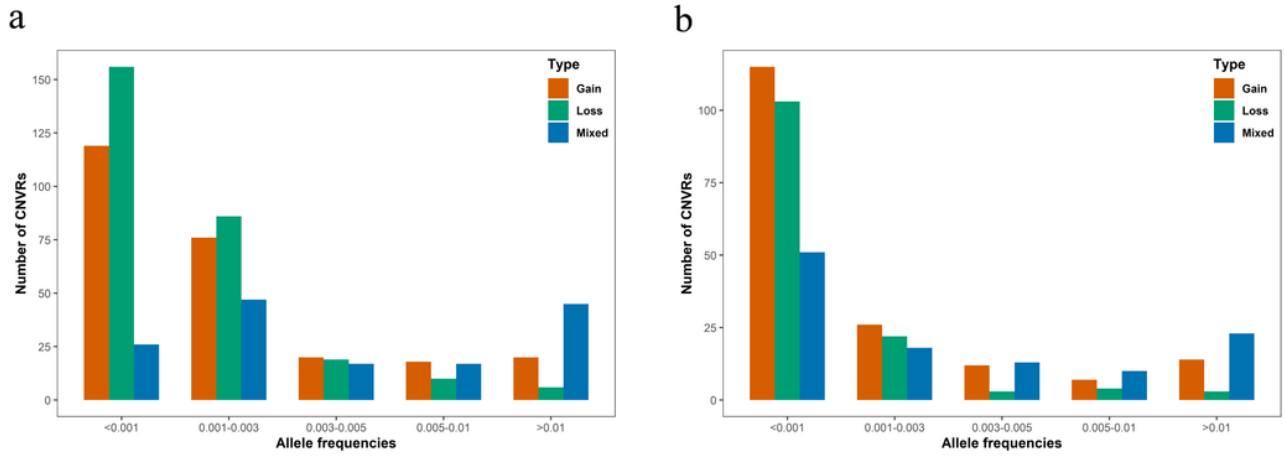

Figure 4

The allele frequencies of CNVRs in the American (a) and Canadian Duroc (b) pigs.

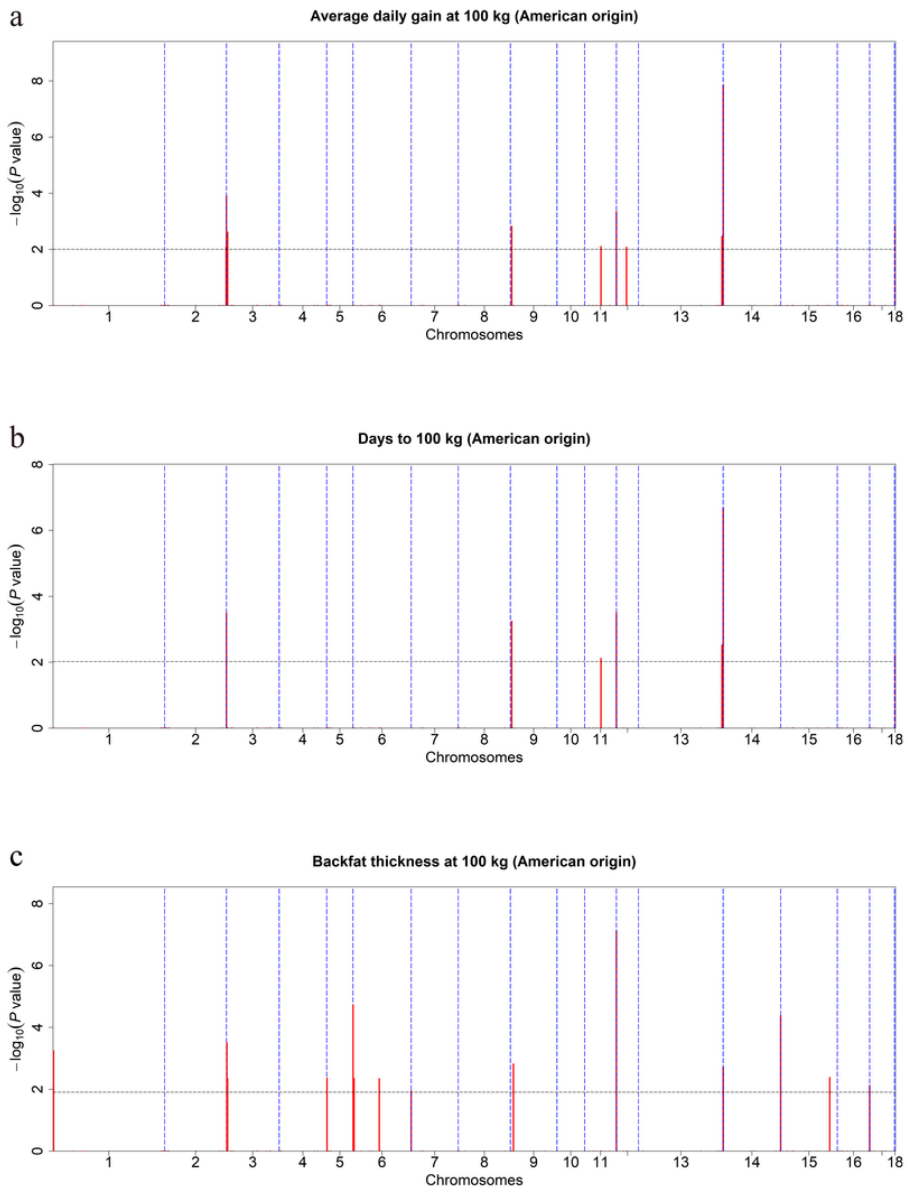

Figure 5

Manhattan plots of CNVR-based association studies in the American Duroc pig population. Manhattan plots consisted of average daily gain at 100 kg (a), days to 100 kg (b), and backfat thickness at 100 kg (c). The x-axis represents the chromosomes, and the y-axis represents the -log10(P-value). The solid lines indicate the FDR corrected thresholds. The FDR corrected thresholds were P < 9.91E-03 (a), P < 9.48E-03 (b) and P < 1.25E-02 (c), respectively.

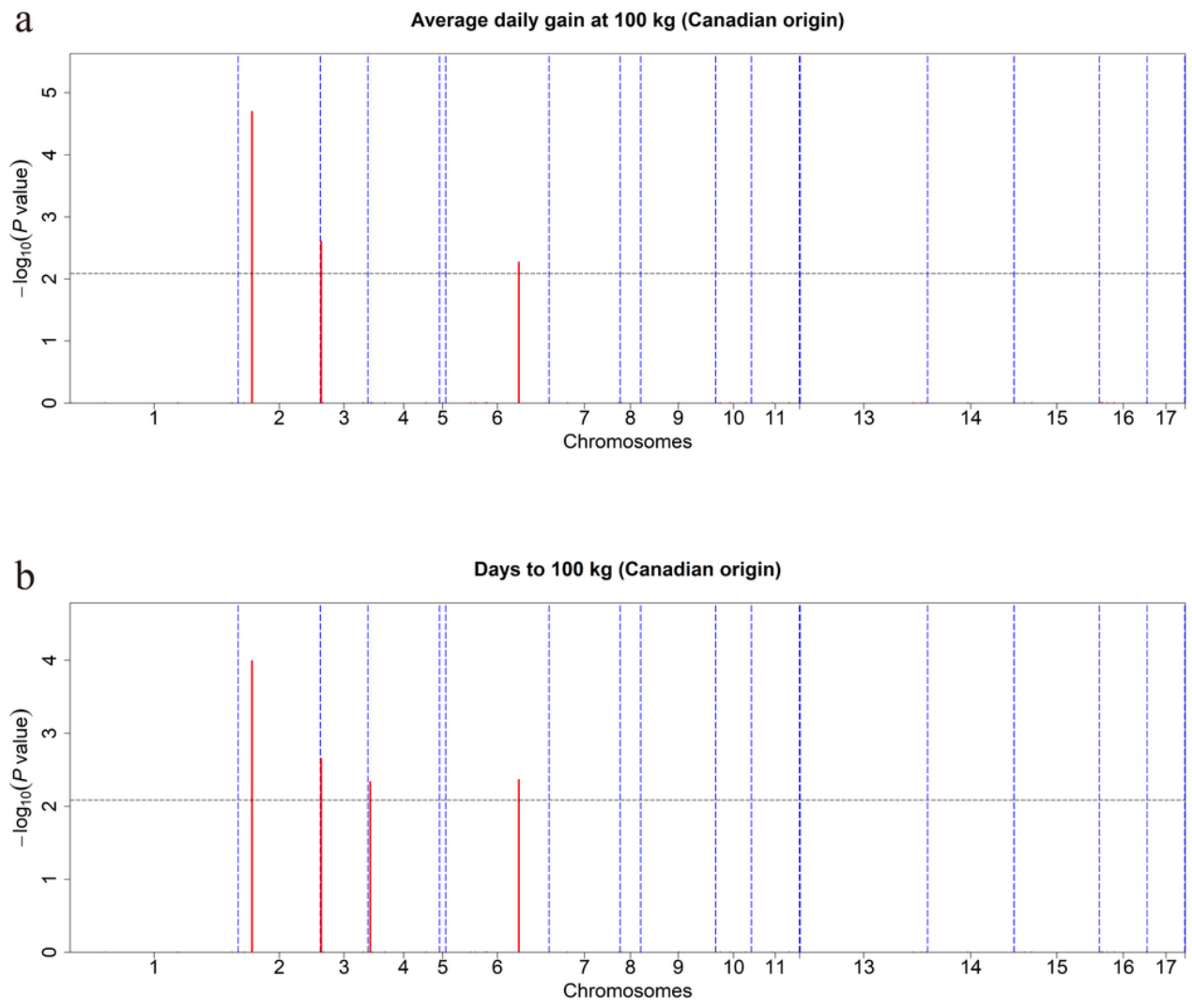

Figure 6

Manhattan plots of CNVR-based association studies in the Canadian Duroc pig population. Manhattan plots consisted of average daily gain at 100 kg (a) and days to 100 kg (b). The x-axis represents the chromosomes, and the y-axis represents the -log10(P-value). The solid lines indicate the FDR corrected thresholds. The FDR corrected thresholds were P < 8.20e-03 (a) and P < 8.20e-03 (b), respectively.

Supplementary Files

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

- Additionalfile1TableS1.xlsx
- Additionalfile2TableS2.xlsx
- Additionalfile3TableS3.xlsx

- Additionalfile4TableS4.xlsx
- Additionalfile5TableS5.xlsx
- Additionalfile6TableS6.xlsx
- Additionalfile7TableS7.xlsx