



A Survey of Copy Number Variation in the Porcine Genome Detected From Whole-Genome Sequence

Brittney N. Keel*, Dan J. Nonneman, Amanda K. Lindholm-Perry, William T. Oliver and Gary A. Rohrer

USDA, ARS, U.S. Meat Animal Research Center, Clay Center, NE, United States

OPEN ACCESS

Edited by:

Denis Milan,
Institut National de la Recherche
Agronomique (INRA), France

Reviewed by:

Martien Groenen,
Wageningen University and Research,
Netherlands
Junwu Ma,
Jiangxi Agricultural University, China

*Correspondence:

Brittney N. Keel
brittney.keel@ars.usda.gov

Specialty section:

This article was submitted to
Livestock Genomics,
a section of the journal
Frontiers in Genetics

Received: 19 February 2019

Accepted: 12 July 2019

Published: 16 August 2019

Citation:

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

Copy number variations (CNVs) are gains and losses of large regions of genomic sequence between individuals of a species. Although CNVs have been associated with various phenotypic traits in humans and other species, the extent to which CNVs impact phenotypic variation remains unclear. In swine, as well as many other species, relatively little is understood about the frequency of CNV in the genome, sizes, locations, and other chromosomal properties. In this work, we identified and characterized CNV by utilizing whole-genome sequence from 240 members of an intensely phenotyped experimental swine herd at the U.S. Meat Animal Research Center (USMARC). These animals included all 24 of the purebred founding boars (12 Duroc and 12 Landrace), 48 of the founding Yorkshire-Landrace composite sows, 109 composite animals from generations 4 through 9, 29 composite animals from generation 15, and 30 purebred industry boars (15 Landrace and 15 Yorkshire) used as sires in generations 10 through 15. Using a combination of split reads, paired-end mapping, and read depth approaches, we identified a total of 3,538 copy number variable regions (CNVRs), including 1,820 novel CNVRs not reported in previous studies. The CNVRs covered 0.94% of the porcine genome and overlapped 1,401 genes. Gene ontology analysis identified that CNV-overlapped genes were enriched for functions related to organism development. Additionally, CNVRs overlapped with many known quantitative trait loci (QTL). In particular, analysis of QTL previously identified in the USMARC herd showed that CNVRs were most overlapped with reproductive traits, such as age of puberty and ovulation rate, and CNVRs were significantly enriched for reproductive QTL.

Keywords: copy number variation, swine, whole-genome sequence, read depth, olfactory receptor

INTRODUCTION

One of the important challenges in post-genomic biology is relating observed phenotypic variation to the underlying genotypic variation. Genome-wide association studies (GWAS) have made thousands of connections between single-nucleotide polymorphisms (SNPs) and phenotypes, implicating regions of the genome that may play a causal role in a variety of complex traits. Despite their success in identifying associated variants, association studies account for only a small percentage of the total heritability (Maher, 2008). Hence, determining other types of variation that may make a substantial contribution to variation in complex traits is a meaningful goal.

Copy number variations (CNVs) are gains and losses of large regions of genomic sequence between individuals of a species, ranging from kilobases to megabases in length (Feuk et al., 2006). It is hypothesized that CNVs represent a significant source of genetic variation, as they have been shown to cover approximately 7% of the mouse genome (Locke et al., 2015), 12% of the human genome (Redon et al., 2006), and 7% of the cattle genome (Keel et al., 2016a). Significant overlap between protein-coding genes and CNV has been reported in a number of species, including human (Bailey et al., 2009), mouse (Locke et al., 2015), cattle (Keel et al., 2016b), and pig (Paudel et al., 2013). Conrad et al. (2010) found that 40% of validated CNV overlapped with at least one gene. In addition, CNVs appear to influence gene expression levels (Stranger et al., 2007; Henrichsen et al., 2009).

In humans and rodents, CNVs have been well studied and linked to various phenotypic traits and diseases (Cook and Scherer, 2008; Almal and Padh, 2012; Girirajan et al., 2013). Initial CNV studies have been performed in a number of domesticated animals: dog (Nicholas et al., 2011; Alvarez and Akey, 2012; Berglund et al., 2012), cattle (Fadista et al., 2010; Liu et al., 2010; Hou et al., 2011; Stothard et al., 2011; Zhan et al., 2011; Bickhart et al., 2012; Hou et al., 2012a; Hou et al., 2012b; Jiang et al., 2012; Choi et al., 2013; Wu et al., 2015; Keel et al., 2016a), sheep (Fontanesi et al., 2011; Liu et al., 2013), chicken (Crooijmans et al., 2013; Yi et al., 2014), and goat (Fontanesi et al., 2010).

Swine CNVs have been reported using a variety of array-based platforms, including comparative genomic hybridization arrays (Fadista et al., 2008; Li et al., 2012; Wang et al., 2014; Wang J. et al., 2015), the Illumina PorcineSNP60 BeadChip (Ramayo-Caldas et al., 2010; Chen et al., 2012; Wang et al., 2012; Wang L. et al., 2013; Schiavo et al., 2014; Wiedmann et al., 2015; Xie et al., 2016; Zhou et al., 2016; Hay et al., 2017), and the Illumina Infinium II Multisample SNP assay (Wang J. et al., 2013; Long et al., 2016). These approaches are known to suffer some drawbacks, including limited coverage of the genome due to low probe density, low resolution, and hybridization noise (Zhao et al., 2013). Ongoing developments and cost decreases in next-generation sequencing (NGS) technology have led to an increased popularity of sequence-based CNV detection. To date, a limited number of studies have utilized NGS data to identify CNV in the porcine genome.

The number and size ranges of CNV detected in previous swine studies utilizing NGS vary dramatically. These discrepancies may be artifact of differences in many aspects of the study, including sequence coverage, sample size, breed, and CNV detection algorithm. In swine, as well as many other species, relatively little is known about the properties of CNV, including their frequency in the genome, sizes, locations, and chromosomal properties. Of all the topics related to CNV, knowledge of their functional impact is the most limited. Despite the wide range of number and size of CNV reported between previous swine NGS studies, the results from functional enrichment analysis of CNV are quite consistent. Gene ontology (GO) terms related to sensory perception (Paudel et al., 2013; Jiang et al., 2014; Paudel et al., 2015), response to stimuli (Paudel et al., 2013; Jiang et al., 2014),

immunity (Jiang et al., 2014; Paudel et al., 2015), and olfactory receptor (OR) activity (Paudel et al., 2015; Revilla et al., 2017) were the most significant in these studies. The same GO terms have been identified in CNV studies in humans and cattle. ORs, which are G-protein-coupled receptors involved in signal transduction, play a role in all the GO terms listed above. The results from previous studies suggest that CNVs may play a role in olfactory ability and sensitivity, which may be related to economically relevant traits in swine including feeding behavior (Connor et al., 2018) and reproduction (Baum and Cherry, 2015).

The CNVs reported in the aforementioned studies represent several diverse pig breeds and wild boars from different regions of the world. Very few animals in these studies (only 37 of 353) represent commercial swine germplasm, which, through domestication, has been shaped by selection for docility and lean meat production. Additionally, previous CNV studies in swine have been conducted using the Sscrofa 9.2 and 10.2 genome builds. The purpose of this study is to identify and characterize CNV regions detected from whole-genome sequence of 240 members of an experimental swine herd at the U.S. Meat Animal Research Center (USMARC), a resource representative of commercial swine germplasm, utilizing the newly released, high-quality Sscrofa 11.1 genome assembly.

MATERIALS AND METHODS

The DNA samples sequenced for this study were extracted from semen, blood, and tail tissue archived under standard operating procedures for the U.S. Meat Animal Research Center tissue repository. The research did not involve experimentation on animals requiring IACUC approval.

Sequencing and Data Acquisition

CNVs were detected from whole-genome sequence of 240 members of an experimental swine herd. This composite population, developed at USMARC, began in 2001 by mating mixed Landrace-Yorkshire sows with 24 purebred founding boars—12 Landrace and 12 Duroc. To produce the second generation, Landrace-sired animals were mated to Duroc-sired animals. Subsequent generations were produced by selecting 1 male and 10 females produced by each founding boar and then randomly mating them, avoiding full-sib and half-sib pairings (Lindholm-Perry et al., 2009). Industry sires were then introduced in generation 10 and used in subsequent generations. This study utilizes whole-genome sequence from all 24 founding boars, 48 of the founding sows, and 109 animals from generations 4 through 9, 29 animals from generation 15, and 30 purebred industry boars (15 Landrace and 15 Yorkshire) used as sires in generations 10 through 15.

DNA extraction and library preparation have been previously described in Keel et al. (2017) and Keel et al. (2018) for the 72 founding animals and the remaining 168 animals, respectively. Libraries were paired-end sequenced (150 bp read length) on an Illumina NextSeq500 (Illumina, San Diego, CA, USA) at USMARC. Bases of the paired-end reads for all sequenced genomes were identified with the Illumina BaseCaller, and

FASTQ files were produced for downstream analysis of the sequence data. Sequence data are available for download from the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) BioProjects PRJNA343658, PRJNA414091, and PRJNA482384.

Sequence Data Processing

The Trimmomatic software (Version 0.35; Bolger et al., 2014) was used to trim Illumina adaptor sequences and low-quality bases from the reads. The quality cutoff was a PHRED33 score of >15. Reads containing any portion with an average PHRED33 score <15 spanning at least 4 bp were removed. The remaining reads were mapped to the Sscrofa 11.1 genome assembly using Burrows-Wheeler Alignment (BWA, Version 0.7.12; Li and Durbin, 2009) with the default parameters.

CNV Detection and Defining CNVRs

A combination of the CNVnator (Version 0.3.2; Abyzov et al., 2011) and LUMPY (Version 0.4.13; Layer et al., 2014) software was used to identify putative CNV in the genome sequence of the 240 pigs. LUMPY is a probabilistic CNV discovery framework that integrates multiple detection signals, including split reads and paired-end mapping, while the CNVnator is a read depth method that uses a mean-shift-based approach to call CNV based on the depth of sequencing.

CNVs were first called for each sample using the CNVnator. The program was run using a window size (bin size) of 1 kb, and all other parameters were set to the default. Next, CNVs were detected using LUMPY with default parameters. CNV breakpoints from the CNVnator output were passed as input into LUMPY using the `-bedpe` option.

In an attempt to reduce the number of false positives, CNVs were also called using the `cn.MOPS` algorithm (Version 1.24.0; Klambauer et al., 2012). `cn.MOPS` is a multiple sample read depth method that applies a Bayesian approach to decompose read variations across multiple samples into integer copy numbers and noise by its mixture components and Poisson distributions, respectively. `cn.MOPS` avoids read count biases along the chromosomes by modeling the depth of coverage across all samples at each genomic position. The `cn.MOPS` program was run using a window length of 1 kb, mean normalization mode, and the default values for all other parameters. Autosomal and sex chromosomes were processed differently due to differences in expected ploidy of the genome. Autosomal CNVs, which are expected to be diploid, were identified using all 240 samples. CNVs on the sex chromosomes were identified by processing the 167 males and 73 females separately, as SSCX is expected to be diploid in female samples and SSCX and SSCY are expected to be haploid in the male samples.

CNVs identified by LUMPY that were at least 10% overlapped by a CNV identified by `cn.MOPS`, meaning that the ratio of the number of bp overlapped between the LUMPY CNV and at least one `cn.MOPS` CNV to the length of the LUMPY CNV was greater than 0.10, were retained for downstream analysis. Next, CNVs were used to construct a set of copy number variable regions (CNVRs). A CNVR was constructed by merging CNVs across

samples that exhibited at least 50% pairwise reciprocal overlap in their genomic coordinates. For example, suppose we have two CNVs, CNV1 beginning at position a and ending at position b and CNV2 running from c to d with $a < c < b < d$. If the reciprocal overlap between the two CNVs is at least 50%, then they are merged into a CNVR that runs from a to d on the genome.

Validation of CNVR Using Data From Sequenced Parent–Offspring Trios

For the transmission rate (paternal and maternal), in each parent–child pair, CNVRs in the parent also called in the child were counted and then divided by the total number of CNVR calls in the parent. For the inheritance rate, CNVR calls in the child also present in at least one parent were counted and then divided by the total number of CNVRs in the child.

Gene Content and GO

Genes from the NCBI *Sus scrofa* annotation (Release 106) overlapping by at least 1 bp with CNVRs were identified. Functions of protein-coding CNV-overlapped genes were determined using the PANTHER classification system (Version 14.0, Mi et al., 2013).

Enrichment analysis of gene function was performed using PANTHER's implementation of the binomial test of overrepresentation. Significance of GO terms was assessed using the default *Sus scrofa* GO annotation as the reference set for the enrichment analysis, and data were considered statistically significant at a Benjamini-Hochberg-corrected P value < 0.05.

Enrichment of Quantitative Trait Loci

Enrichment analysis of quantitative trait loci (QTL) overlapped with CNVR was performed using Fisher's exact test. Data were considered statistically significant at a Benjamini-Hochberg-corrected P value < 0.05.

RESULTS AND DISCUSSION

Sequencing and Read Mapping

Genomic DNA from 240 pigs, from a composite population at USMARC, was sequenced on Illumina HiSeq and NextSeq platforms, generating approximately 72 billion paired-end reads (Table S1). Sequence reads covered each pig's genome at a mean of 13.62-fold (\times) coverage. Individual coverage per animal ranged from 0.97 \times to 31.13 \times ; 24 animals were covered at less than 3 \times , and 44 were covered at more than 20 \times .

When generating our sequence data, we targeted a minimum of 3 \times coverage for each of the founding sows and 10 \times coverage for the remaining 168 animals. However, there was considerable variation around the 3.66 \times and 18.41 \times mean coverage for the founding sows and other animals, respectively. Some of this variation can be attributed to technical aspects of NGS technology, such as the stochasticity of sequencing, DNA quality, and library preparation. The combined sequence from all 240 animals covered 99.99% of the reference genome.

CNVR Discovery and Statistics

CNVs in the genome of the 240 pigs were identified by taking the overlap of two methods: (1) a combination of CNVnator and LUMPY and (2) cn.MOPS. Most of the previous NGS CNV studies in swine have utilized read depth approaches to identify variants (Paudel et al., 2013; Jiang et al., 2014; Paudel et al., 2015; Wang H. et al., 2015; Wang et al., 2016; Revilla et al., 2017). Although read depth can be a powerful tool to identify CNV, often the boundaries are not well determined because of the sliding window approach. The exact boundaries of CNV events can be important for determining their functional effect (e.g., affecting coding sequence). Other CNV detection strategies, such as paired-end mapping or split reads, can be used to fine map CNV and determine more precise boundaries of the variants. The CNVnator–LUMPY combination approach used in this work calls CNV in individual samples utilizing paired-end mapping, split reads, and read depth. Although this method should give more accurate CNV breakpoints than read depth signal alone, single sample CNV callers are known to suffer from decreased detection power and high false-positive rates. A total of 2,079,579 were identified using CNVnator–LUMPY. Utilization of data from multiple samples has been shown to improve CNV detection (Klambauer et al., 2012; Duan et al., 2014). Therefore, as a further error correction step, CNVs were also detected using a multiple sample read depth caller, cn.MOPS (695,741 CNV identified). A total of 39,315 CNVs, overlapping between the two methods, were retained for downstream analysis (Table S2). CNVs were merged across each genome and then across samples into CNVRs, and CNVRs less than 200 bp in length were filtered out. This resulted in a final set consisting of 3,538 CNVRs (Table S3), including 1,820 novel CNVRs that were not reported in previous studies.

Note that approximately 19% (45 of 240) of the animals in this study had low to moderate sequence depth (<5× coverage). The highest sensitivity and resolution in CNV detection are attained through high coverage sequencing (>10×; Alkan et al., 2011). However, until sequencing costs drop dramatically, it is not feasible, in most cases, to generate high coverage genomic sequence on large numbers of animals. We consider low-coverage sequencing data here, because methods for analyzing SNP and CNV in low-coverage data will continue to be relevant in the future in terms of a study's discovery power, where a fixed number of reads should rather be used for sequencing more samples with lower coverage than for sequencing fewer samples with higher coverage (Le and Durbin, 2011). Due to the cost-effectiveness of sequencing at lower coverage, recent studies have explored strategies for using low-coverage sequence to detect common CNV that could explain a significant amount of phenotypic variation (Keel et al., 2016a; Zhou et al., 2018). Both CNVnator and cn.MOPS have been shown to have moderate to high accuracy in detecting CNV from low-coverage sequence in diploid genomes (Keel et al., 2016a; Malekpour et al., 2018), particularly in data sets consisting of samples with mixed levels of coverage. Therefore, the use of these methods, coupled with LUMPY, should provide reasonably accurate results for CNV calling in our 240 animals.

Sizes of the CNVRs ranged from 0.203 to 398.9 kb, with an average of 6.8 kb and a median of 2.9 kb. The CNVR occupied a total of 22.9 unique Mb or 0.94% of the *Sus scrofa* genome. The CNVR coverage of the genome is lower than the results of previous reports in swine (4.0%; Jiang et al., 2014) and other species, including mouse (6.87%; Locke et al., 2015), human (12%; Redon et al., 2006), and cattle (6.7%; Keel et al., 2016a), which may be due to our stringent criteria (e.g., requiring detection with two different approaches). Among the CNVRs, 144 showed copy number gain (duplication), 3372 showed copy number loss (deletion), and 22 showed a mix of copy number loss and gain from different individuals. Clearly, there was a large discrepancy in the numbers of duplication and deletion CNVR. Overall, read-depth methods are more sensitive to deletion CNV calls than duplication calls, especially in mid- to low-coverage sequence data, as it is easier to identify a “missing” segment of the genome than an amplified one with limited sequence reads. In fact, 3.1% (105 of 3372) of deletion calls were identified in only animals with <10× coverage. As low-coverage WGS continues to become more widely utilized, it will be necessary to focus on adapting CNV calling tools to this type of data.

Distribution of CNVR

The distribution of CNVRs along each of the chromosomes is shown in Figure 1. Variants were not uniformly distributed on the chromosomes. The number of CNVRs was strongly correlated with the size of the chromosome (Pearson correlation coefficient $r = 0.77$). SSC1 and SSC13 exhibited the largest numbers of CNVRs (1231 and 231, respectively), while SSCY, SSC18, and SSC12 had the smallest numbers (2, 49, and 52 CNVRs, respectively). On average, 0.79% of each chromosome was covered by CNVRs (Table 1).

The number of CNVRs per animal ranged from 0 to 348, with a mean of 157.8 (Table S4). CNVs spanned up to 0.13% of the genome of each animal, with a mean and median of 0.057% and 0.062%, respectively. This variation across individuals can be partially explained by differences in genomic sequencing coverage. Smaller numbers of CNVRs were identified in samples with low sequencing depth, and the number of identified CNVRs tended to increase as genomic coverage increased (Pearson correlation coefficient $r = 0.84$).

The number of individuals exhibiting each CNVR ranged from 1 to 175. Many CNVRs (~2649) were present in a small percentage (< 5%) of the animals. Three CNVRs (CNVR 2103, 1676, and 2104 in Table S3) were present in more than 60% of the population. The distribution of deletion, duplication, and mixed CNVRs across breeds is shown in Figure 2. The purebred Landrace and Yorkshire boars and the composite animals had more CNVR of all three types than the purebred Duroc boars. This is likely because the *Sscrofa* 11.1 reference genome assembly was obtained from a Duroc animal.

A total of 1620 CNVRs were found to be breed-specific in origin (Table S5). Most (64%) breed-specific CNVRs were present in only one animal. Breed-specific CNVRs that were present in the largest numbers of animals were found in the composite breed. This is likely due to the larger number of

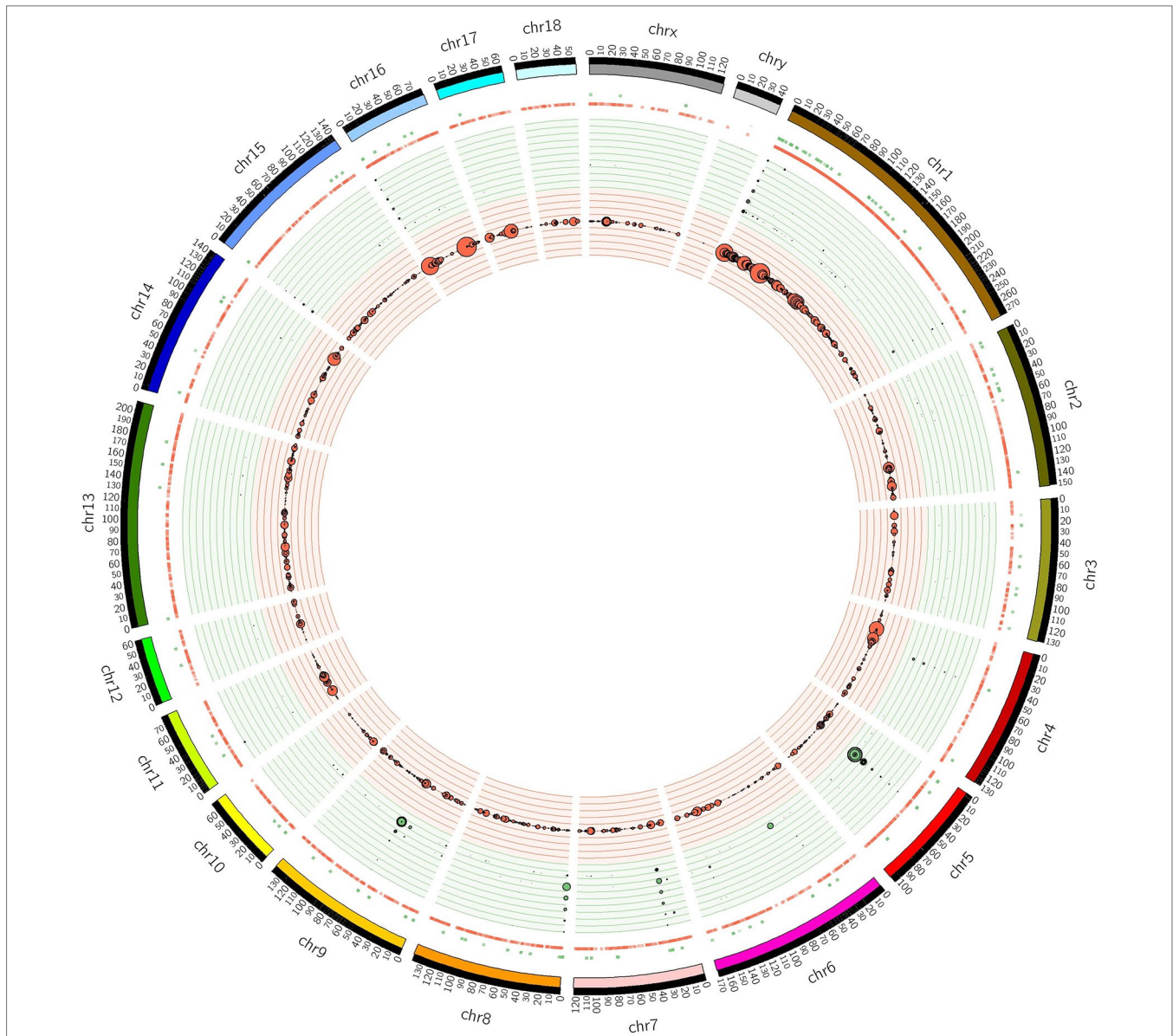


FIGURE 1 | Positions of CNVRs identified from the 240 sequenced swine genomes in Circos format (Krzywinski et al., 2009). The outer ideogram runs clockwise from chromosome 1 to chromosome Y with labels in Mb of physical distance. The copy number data are represented in the inner tracks. The two innermost tracks show scatter plots of the CNVR, where the red track shows copy number loss and the green track shows copy number gain. Concentric circles within these tracks indicate y-axis values in the scatter plot. The 10 concentric circles in the red track mark values $0 \leq y < 2$, with 0 being the innermost track, while the 11 concentric circles in the green track mark values $2 \leq y \leq 8$. The size of the dot in the scatter plot is proportional to the number of samples containing the CNVR. The other track shows a heat map that indicates the parts of the genome that contain copy number gain and loss. This plot simply collapses the scatter plot values onto a single radial position.

composite animals in the data set. Increased numbers of animals in the three pure breeds will be necessary to conduct a complete breed-of-origin analysis. This will be a focus of future work.

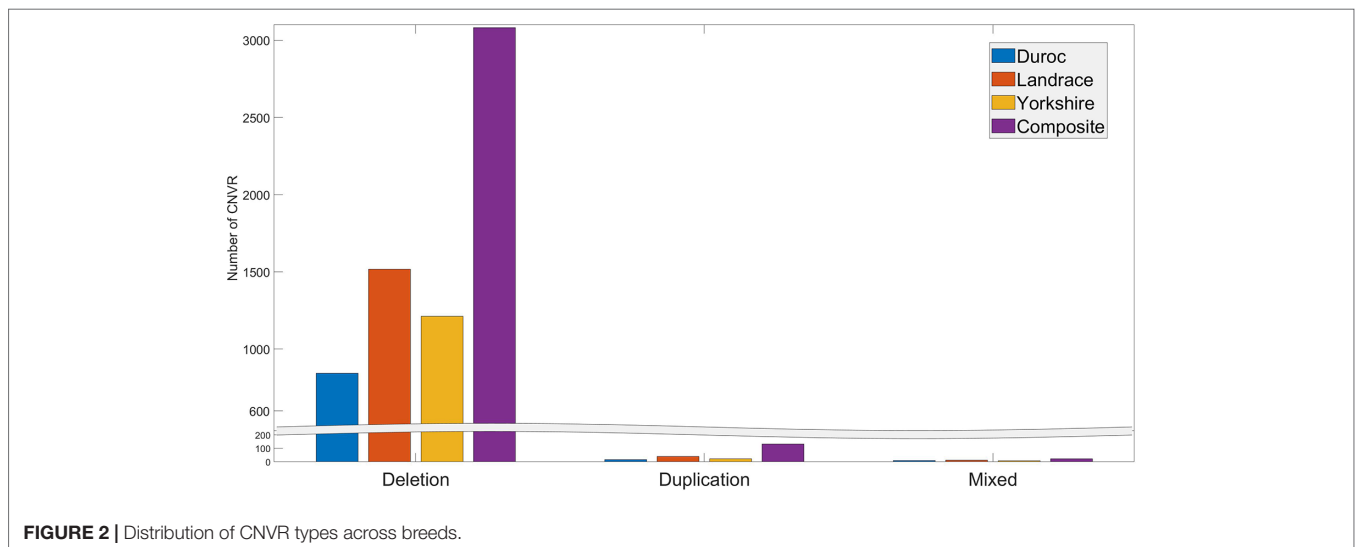
CNV Concordance in Parent–Offspring Trios

Pedigree data from 12 sequenced parent–offspring trios were used as a substitute for molecular validation, which we have

chosen to forego since this work was intended to be a discovery. In a follow-up study, we are planning to look for CNVRs that may associate with phenotypes in our population and validation using PCR methods will be performed for those CNVRs. If a CNV is transmitted from parent to offspring, then it can be considered validated. Although this type of validation is not 100% accurate, it is satisfactory to allow us to estimate error rates. In an ideal data set, paternal and maternal transmission rates would be 50%, and inheritance rates would be 100%. Deviations from this ideal

TABLE 1 | CNVR distribution across the genome.

Chromosome	Chromosome length (bp)	No. of CNVR on chromosome	Unique bp covered by CNVR	Ratio covered by CNVR
SSC1	274330532	1231	7969003	2.90%
SSC2	151935994	192	1318835	0.87%
SSC3	132848913	101	570813	0.43%
SSC4	130910915	137	523004	0.40%
SSC5	104526007	146	1132811	1.08%
SSC6	170843587	152	772899	0.45%
SSC7	121844099	141	638302	0.52%
SSC8	138966237	203	1599178	1.15%
SSC9	139512083	167	1088850	0.78%
SSC10	69359453	82	407652	0.59%
SSC11	79169978	103	1338230	1.69%
SSC12	61602749	52	337802	0.55%
SSC13	208334590	231	1636688	0.79%
SSC14	141755446	126	1099698	0.78%
SSC15	140412725	141	734861	0.52%
SSC16	79944280	127	449040	0.56%
SSC17	63494081	62	277525	0.44%
SSC18	55982971	49	433245	0.77%
SSCX	125939595	93	582165	0.46%
SSCY	43547828	2	17943	0.04%

**FIGURE 2** | Distribution of CNVR types across breeds.

could be explained by multiple factors. Both false-positive and false-negative CNV calls will cause a decrease in transmission and inheritance rates. Another possible factor is *de novo* mutations in offspring, which will not affect transmission rates, but will affect inheritance rates. Additionally, there is the possibility of somatic mutation in one or more of the parents, essentially a *de novo* mutation in parents as they age. Somatic mutations would affect transmission rates but not inheritance rates. All of these factors could potentially affect the data simultaneously. Therefore, they cannot be individually estimated. However, assuming that *de novo* and somatic mutations are rare compared to CNVR calling errors, we can use transmission and inheritance rates to estimate error rates.

Table 2 shows the paternal and maternal transmission rates and the inheritance rate for each of the 12 sets of trios. The

average transmission rates were 37.7% and 41.4% for maternal and paternal, respectively. These rates are much closer to the ideal 50% transmission rate than what was reported in a similar study in humans (27% for maternal and 28% for paternal; Zheng et al., 2012). The average inheritance rate was 52%, which falls between inheritance rates reported in previous studies, 42% in Zheng et al. (2012) and 74.8% in Wang et al. (2007). Therefore, under the assumption that the *de novo* and somatic mutations are rare, we approximate the error rate in CNVR calls to be 48% (100% minus the inheritance rate). This error rate is comparable to previously published results for several different CNV-calling algorithms for whole-genome sequence data (26%–77%; Legault et al., 2015). This consistency suggests that high error rates may be due to algorithmic issues rather than the input data. Clearly, further development of bioinformatics protocols and tools for

TABLE 2 | Transmission and inheritance rates in parent–offspring trios.

Trio number	Paternal transmission rate	Maternal transmission rate	Inheritance rate
1	0.277	0.270	0.549
2	0.414	0.352	0.453
3	0.351	0.519	0.524
4	0.370	0.369	0.356
5	0.396	0.714	0.417
6	0.373	0.448	0.576
7	0.393	0.382	0.570
8	0.428	0.403	0.482
9	0.428	0.367	0.603
10	0.401	0.390	0.441
11	0.360	0.408	0.668
12	0.335	0.343	0.594
Mean	0.377	0.414	0.520

producing high-confidence, consistent CNV calls is necessary to improve the quality of CNV discovery studies.

Comparison of CNVRs with Previous Studies

Comparison of our results with CNVRs identified in several previous swine studies showed varying levels of overlapping CNVRs between studies (Table 3). Here, we used a much less stringent definition of overlap than that used in identifying

overlapping CNV, where two CNVRs were considered overlapped as long as they shared at least one base.

Generally speaking, percentages of overlap in CNV events identified between this work and previous studies were low (average of 4.33% overlap). This result is very similar to what has been observed in cattle CNV studies, where typically <40% overlap exists between studies (Keel et al. 2016b). These discrepancies are likely driven by many technical aspects of the experiments, including vastly different sample sizes, differences in breeds and the number of breeds represented, detection platform (array-based vs. NGS), and CNV detection algorithms. Many of the CNV discovery studies in swine have involved pure and half Chinese breeds. Therefore, it is likely that many of the CNVRs identified in those studies do not segregate in our population. Our population represents commercial swine germplasm and, because of domestication and selection for lean meat production and reproductive efficiency, has diverged from germplasm studied in other experiments.

It should be noted that two of the three studies with highest overlap percentages, Wang et al. (2013) and Wiedmann et al. (2015), were those that had high representations of Yorkshire, Landrace, and Duroc animals. In fact, the study of Wiedmann et al. (2015) was performed on animals from the same population used in this study. The discrepancy in CNV identified in their study and ours is likely due to differences in platform (SNP beadchip vs. whole-genome sequence), detection algorithm, and genome build (Sscrofa 10.2 vs. Sscrofa 11.1).

TABLE 3 | Comparison of CNVRs identified in this study to results from other studies (based on the Sscrofa 11.1 genome assembly).

Platform	Findings from other studies				CNVR overlap with this study	
	Study ^a	Breeds	Samples	No. of CNVRs (No. before mapping)	No. of overlapped CNVRs from this study	Ratio of overlapped CNVRs from this study
CGH-based study	Fadista et al., 2008**	1	12	31 (37)	1	0.00%
	Li et al., 2012*	8	12	241 (259)	47	1.33%
	Wang et al., 2014**	9	12	48 (52)	8	0.22%
	Wang J. et al., 2015**	9	12	602 (689)	129	3.65%
SNP-based study (PorcineSNP60)	Ramayo-Caldas et al., 2010*	2	55	47 (49)	104	2.94%
	Chen et al., 2012**	18	1693	537 (565)	262	7.41%
	Wang et al., 2012*	3	474	357 (382)	130	3.67%
	Wang L. et al., 2013**	2	585	234 (249)	670	18.94%
	Schiavo et al., 2014**	1	305	166 (170)	100	2.83%
	Wiedmann et al., 2015**	3	1802	480 (502)	579	16.37%
	Long et al., 2016**	3	905	3746 (6193)	182	5.14%
	Xie et al., 2016**	2	120	166 (172)	84	2.37%
	Hay et al., 2017**	1	660	256 (271)	189	5.34%
	Wang J. et al., 2013**	10	14	57 (63)	20	0.57%
SNP-based study (Infimum II)	Paudel et al., 2013**	13	16	2880 (3118)	70	1.98%
	Jiang et al., 2014**	10	13	2820 (3131)	238	6.73%
	Paudel et al., 2015**	5	16	1238 (1408)	33	0.93%
	Wang H. et al., 2015**	13	49	2359 (3131)	53	1.50%
	Wang et al., 2016**	6	252	433 (455)	13	0.37%
Next-Generation Sequencing	Revilla et al., 2017**	2	7	508 (540)	153	4.32%
	This Study	3	240	3538		

^aOriginal data set was mapped to Sscrofa 9.2 or Sscrofa 10.2 assembly, denoted by * and **, respectively. CNVRs were converted to Sscrofa 11.1 coordinates using the NCBI Genome Remapping tool. Successfully mapped CNVRs are shown in the CNVR column with the original number of published CNVRs shown in parentheses.

Function of CNV-Overlapped Genes

A total of 1401 genes from the NCBI annotation of the Sscrofa 11.1 genome were identified to be overlapping with our detected CNVRs (Table S3), including 911 protein-coding genes, 58 pseudogenes, 273 non-coding RNA, and 160 miscellaneous RNA. CNV-overlapped genes were overlapped with 2314 CNVRs. Using PANTHER's functional annotation tool to inspect GO slim terms mapping to protein-coding CNV-overlapped genes, we identified that many of these genes were involved in binding (34.7%), catalytic activity (35.7%), metabolic process (23.1%), biological regulation (20.3%), cellular process (11.4%), localization (9.3%), and molecular transducer activity (9.2%).

Enrichment analysis was performed, using the *Sus scrofa* GO database to identify GO terms that were significantly enriched in our gene set. GO enrichment analysis showed that biological process terms related to regulation of ion transport, cell adhesion, signaling, nervous system development, neurogenesis, and locomotion, as well as molecular function terms related to glutamate receptor activity, protein binding, enzyme binding, ATP binding, and neurotransmitter receptor activity, were significantly overrepresented in the genes overlapped by CNVR (Benjamini-Hochberg-corrected P value <0.05 ; Table S6).

Approximately 3.6% of the CNV-overlapped genes belonged to the OR gene family, one of the largest gene families in the porcine genome (Groenen et al., 2012; Nguyen et al., 2012). ORs are G-protein-coupled receptors involved in signal transduction and have been found to be copy number variable in many mammalian species, including human (Young et al., 2008), rat (Guryev et al., 2008), mouse (Pezer et al., 2015), swine (Chen et al., 2012; Wang et al., 2012; Paudel et al., 2013; Wang J. et al., 2013; Paudel et al., 2015), and cattle (Liu et al., 2010; Keel et al., 2016b; Xu et al., 2016). Young et al. (2008) showed that OR genes displayed varying copy numbers among 50 people, and that this variation may play a role in olfactory ability and sensitivity. It is also thought that ORs may play a chemosensory role as they are expressed on sperm and thought to direct them to the egg *via* chemotaxis (Spehr et al., 2006). Paudel et al. (2015) identified that OR genes were overrepresented among CNVRs across several members of the *Sus* genus. These genes may have been important components of swine evolution, as scent would have been critical for foraging for food, avoiding predators, and finding a mate.

Overlap and Enrichment of Known QTL in CNVRs

To reveal the potential relationships between CNVRs and QTL, we analyzed the overlap between our CNVRs and known swine QTL and performed QTL enrichment analyses. Swine QTL from the Sscrofa 11.1 genome build were downloaded from the Animal QTL database (Release 34; <http://www.animalgenome.org/cgi-bin/QTLdb/SS/index>), which includes 26,076 known QTL for 647 different traits. QTL overlapping with CNVRs were identified (Table S7A), traits were ranked according to the number of QTL/CNVR overlaps (Table S7B), and QTL enrichment analysis was performed (Table S7B). The 10 highest ranked traits included drip loss (519 overlaps), average daily gain (235 overlaps), average backfat thickness (195 overlaps), loin

muscle area (179 overlaps), backfat at last rib (153 overlaps), teat number (127 overlaps), carcass length (95 overlaps), ham weight (81 overlaps), backfat at tenth rib (75 overlaps), and lean meat percentage (73 overlaps).

QTL enrichment analysis, using QTL from the Animal QTL database overlapping with CNVR ($n = 525$ traits), identified that QTL for 132 traits were significantly enriched. The most significantly enriched QTL was drip loss ($P = 4.09E-99$). Several meat quality traits, including average back fat thickness, loin muscle area, ham weight, carcass weight, and dressing percentage, were also found to be among the most significantly enriched.

Approximately 840 QTL have been previously reported from GWAS utilizing animals from the same experimental herd used in this study (Table 4). QTL/CNVR overlaps were identified (Table S8A), traits were ranked according to the number of overlaps (Table S8B), and QTL enrichment analysis was performed (Table S8B). The highest ranked traits included vertebra number (28 overlaps), as well as several reproductive traits including age of puberty (41 overlaps), ovulation rate (18 overlaps), % stillborn ignoring the last piglet (18 overlaps), and last birth interval (17 overlaps). It should be noted that, in this work, CNVRs were not tested for statistical association with QTL, but rather the overlapping genomic positions of the latter were used as one indicator of the potential function of the CNVRs.

Of the 20 GWAS traits that had QTL overlapping with CNVR, 7 of them were found to be significantly enriched. These included vertebra number ($P = 4.35E-07$), percent stillborn ignoring the last piglet ($P = 4.64E-07$), last birth interval ($P = 2.79E-06$), number stillborn in the last birth position ($P = 1.58E-05$), litter average birth interval minus the last birth ($P = 8.13E-05$), kyphosis ($P = 1.49E-05$), and number stillborn ignoring the last piglet ($P = 1.49E-04$).

These results are similar to those from a study conducted by Revay et al. (2015), where age of puberty and teat number were found to be the most abundant reproductive QTL overlapped by swine CNVRs. This coupled with the overrepresentation of GO terms such as cell motility, nervous system development, and organ development in CNVR-overlapped genes suggests that CNVR may play a role in shaping various reproductive traits in swine.

CONCLUSION

CNV continues to gain considerable interest as a source of genetic variation that may play a role in phenotypic diversity. Swine CNV research has made significant progress in the last 5 years. Genome-wide surveys of CNV have been conducted using a variety of platforms and algorithms. Studies that utilize NGS data have been limited in swine. Moreover, much of the NGS-based studies have focused on diverse pig breeds and wild boars from different regions of the world rather than commercial breeds. To capture CNV present in commercial swine germplasm, we utilized whole-genome sequence from 240 animals. Our study is one of the largest sequence-based swine CNV studies to date.

We identified 1401 genes overlapping with CNVRs. GO enrichment analysis showed that our set of CNV-overlapped genes was enriched with genes involved in organism development, and QTL analysis showed that CNVRs overlapped with many QTL

TABLE 4 | QTL identified in USMARC swine population from previously published GWAS.

Trait	Number of reported QTL (number overlapping with CNVR)	Publication
Age of puberty	222 (41)	Nonneman et al., 2016
Litter average birth interval (minus last birth)	25 (15)	Schneider et al., 2015
Last birth interval	25 (17)	Schneider et al., 2015
Number stillborn (ignoring last piglet)	25 (13)	Schneider et al., 2015
Number stillborn in last birth position	25 (16)	Schneider et al., 2015
Percent stillborn (ignoring last piglet)	25 (18)	Schneider et al., 2015
Total number born	11 (2)	Schneider et al., 2012
Number born alive	14 (4)	Schneider et al., 2012
Number born dead	1 (0)	Schneider et al., 2012
Litter birth weight	33 (5)	Schneider et al., 2012
Average piglet birth weight	65 (9)	Schneider et al., 2012
Temperature humidity feeding behavior comparison normal-alert temperatures	17 (6)	Cross et al., 2018
Temperature humidity feeding behavior comparison normal-danger temperatures	13 (3)	Cross et al., 2018
Temperature humidity feeding behavior comparison normal-emergency temperatures	13 (6)	Cross et al., 2018
Temperature humidity feeding behavior comparison alert-danger temperatures	13 (4)	Cross et al., 2018
Temperature humidity feeding behavior comparison alert-emergency temperatures	6 (3)	Cross et al., 2018
Temperature humidity feeding behavior comparison danger-emergency temperatures	4 (2)	Cross et al., 2018
Immunocrit	36 (0)	Rohrer et al., 2014
Shear force	3 (0)	Nonneman et al., 2013
Intramuscular fat	31 (0)	Nonneman et al., 2013
Minolta color score L*	3 (0)	Nonneman et al., 2013
Minolta color score b*	2 (0)	Nonneman et al., 2013
Cookloss	11 (0)	Nonneman et al., 2013
pH	10 (0)	Nonneman et al., 2013
Purge	8 (0)	Nonneman et al., 2013
Ovulation rate	96 (18)	Schneider et al., 2014
Teat number	36 (0)	Rohrer and Nonneman, 2017
Vertebra number	49 (28)	Rohrer et al., 2015
Kyphosis	16 (11)	Lindholm-Perry et al., 2010

for reproductive traits. These results are consistent with findings in other swine CNV studies, which suggests that CNV may play a role in shaping reproductive traits. Understanding the exact role that CNV plays in reshaping gene structure, modulating gene expression, and ultimately contributing to phenotypic variation are open questions. The focus of our future work will be to develop strategies for CNV imputation, identify CNVs that associate with phenotypes in our population, and validate those CNVs using methods such as digital droplet PCR, with the long-term goal of discovering the extent to which CNVs affect economic traits of interest and developing strategies for incorporating them into genomic selection systems.

DATA AVAILABILITY

The raw data supporting the conclusions of this manuscript will be made available by the authors, without undue reservation, to any qualified researcher.

DISCLAIMER

Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture.

The U.S. Department of Agriculture (USDA) prohibits discrimination in all its programs and activities on the basis of race, color, national origin, age, disability, and where applicable, sex, marital status, familial status, parental status, religion, sexual orientation, genetic information, political beliefs, reprisal, or because all or part of an individual's income is derived from any public assistance program. (Not all prohibited bases apply to all programs.) Persons with disabilities who require alternative means for communication of program information (Braille, large print, audiotape, etc.) should contact USDA's TARGET Center at (202) 720-2600 (voice and TDD). To file a complaint of discrimination, write to USDA, Director, Office of Civil Rights, 1400 Independence Avenue, S.W., Washington, D.C. 20250-9410, or call (800) 795-3272 (voice) or (202) 720-6382 (TDD). USDA is an equal opportunity provider and employer.

AUTHOR CONTRIBUTIONS

BK conceived the study, and BK, DN, and GR participated in its design and coordination. DN, GR, WO, and AL-P were involved in the acquisition of data, and BK performed all data analysis. BK drafted the manuscript, and DN, GR, WO, and AL-P contributed to the writing and editing. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

The authors would like to thank Kris Simmerman for sample collection, DNA extraction, and library preparation; Sue Hauver for library preparation; the USMARC Core Laboratory and the Iowa State DNA Core facility for performing the sequencing; and Rebecca Anderson for her assistance with the project during her summer internship.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.00737/full#supplementary-material>

TABLE S1 | Sequencing statistics.

TABLE S2 | CNV overlapping between Lumpy and cn.MOPS analyses.

TABLE S3 | CNVR and their overlapping genes.

TABLE S4 | Distribution of CNVR in individual animals.

TABLE S5 | Breed specific CNVR.

TABLE S6 | Results from GO enrichment analysis.

TABLE S7 | Results from QTL enrichment analysis using QTL from the AnimalQTL Database.

TABLE S8 | Results from QTL enrichment analysis using previously identified QTL in the USMARC swine population.

REFERENCES

- Abyzov, A., Urban, A. E., Snyder, M., and Gernstein, M. (2011). CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Res.* 21, 974–984. doi: 10.1101/gr.114876.110
- Alkan, C., Coe, B. P., and Eichler, E. E. (2011). Genome structural variation discovery and genotyping. *Nat. Rev. Genet.* 12, 363–376. doi: 10.1038/nrg2958
- Almal, S. H., and Padh, H. (2012). Implications of gene copy-number variation in health and diseases. *J. Hum. Genet.* 57 (1), 6–13. doi: 10.1038/jhg.2011.108
- Alvarez, C. E., and Akey, J. M. (2012). Copy number variation in the domestic dog. *Mamm. Genome.* 23 (1), 144–163. doi: 10.1007/s00335-011-9369-8
- Bailey, J. A., Kidd, J. M., and Eichler, E. E. (2009). Human copy number polymorphic genes. *Cytogenet. Genome Res.* 123 (1–4), 234–243. doi: 10.1159/000184713
- Baum, M. J., and Cherry, J. A. (2015). Processing by the main olfactory system of chemosignals that facilitate mammalian reproduction. *Hormones Behav.* 68, 53–64. doi: 10.1016/j.yhbeh.2014.06.003
- Berglund, J., Nevalainen, E. M., Molin, A. M., Perloski, M., The LUPA Consortium, André, C., et al. (2012). Novel origins of copy number variation in the dog genome. *Genome Biol.* 13 (8), R73. doi: 10.1186/gb-2012-13-8-r73
- Bickhart, D. M., Hou, Y., Schroeder, S. G., Alkan, C., Cardone, M. F., Matukumalli, L. K., et al. (2012). Copy number variation of individual cattle genomes using next-generation sequencing. *Genome Res.* 22 (4), 778–790. doi: 10.1101/gr.133967.111
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30 (15), 2114–2120. doi: 10.1093/bioinformatics/btu170
- Chen, C., Qiao, R., Wei, R., Guo, Y., Ai, H., Ma, J., et al. (2012). 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* 13 (1), 733. doi: 10.1186/1471-2164-13-733
- Choi, J.-W., Lee, K.-T., Liao, X., Stothard, P., An, H.-S., Ahn, S., et al. (2013). Genome-wide copy number variation in Hanwoo, Black Angus, and Holstein cattle. *Mamm. Genome.* 24 (3), 151–163. doi: 10.1007/s00335-013-9449-z
- Conrad, D. F., Pinto, D., Redon, R., Feuk, L., Gokcumen, O., Zhang, Y., et al. (2010). Origins and functional impact of copy number variation in the human genome. *Nature* 464 (7289), 704–712. doi: 10.1038/nature08516
- Connor, E. E., Zhou, Y., and Liu, G. E. (2018). The essence of appetite: does olfactory receptor variation play a role? *J. Anim. Sci.* 96 (4), 1551–1558. doi: 10.1093/jas/sky068
- Cook, E. H., Jr., and Scherer, S. W. (2008). Copy-number variations associated with neuropsychiatric conditions. *Nature* 455 (7215), 919–923. doi: 10.1038/nature07458
- Crooijmans, R. P., Fife, M. S., Fitzgerald, T. W., Strickland, S., Cheng, H. H., Kaiser, P., et al. (2013). Large scale variation in DNA copy number in chicken breeds. *BMC Genomics* 14 (1), 398. doi: 10.1186/1471-2164-14-398
- Cross, A. J., Keel, B. N., Brown-Brandl, T. M., Cassady, J. P., and Rohrer, G. A. (2018). Genome-wide association of changes in swine feeding behavior due to heat stress. *Genet. Sel. Evol.* 50, 11. doi: 10.1186/s12711-018-0382-1
- Duan, J., Deng, H.-W., and Wang, Y.-P. (2014). Common copy number variation detection from multiple sequenced samples. *IEEE Transact. Biomed. Engr.* 61 (3), 928–937. doi: 10.1109/TBME.2013.2292588
- Fadista, J., Nygaard, M., Holm, L.-E., Thomsen, B., and Bendixen, C. (2008). A snapshot of CNVs in the pig genome. *PLoS One* 3 (12), e3916. doi: 10.1371/journal.pone.0003916
- Fadista, J., Thomsen, B., Holm, L.-E., and Bendixen, C. (2010). Copy number variation in the bovine genome. *BMC Genomics* 11 (1), 284. doi: 10.1186/1471-2164-11-284
- Feuk, L., Carson, A. R., and Scherer, S. W. (2006). Structural variation in the human genome. *Nat. Genet. Rev.* 7, 85–97. doi: 10.1038/nrg1767
- Fontanesi, L., Martelli, P. L., Beretti, F., Riggio, V., Dall'Olio, S., Colombo, M., et al. (2010). An initial comparative map of copy number variations in the goat (*Capra hircus*) genome. *BMC Genomics* 11 (1), 639. doi: 10.1186/1471-2164-11-639
- Fontanesi, L., Beretti, F., Martelli, P. L., Colombo, M., Dall'Olio, S., Occidente, M., et al. (2011). A first comparative map of copy number variations in the sheep genome. *Genomics* 97 (3), 158–165. doi: 10.1016/j.ygeno.2010.11.005
- Girirajan, S., Dennis, M. Y., Baker, C., Malig, M., Coe, B. P., Campbell, C. D., et al. (2013). Refinement and discovery of new hotspots of copy-number variation associated with autism spectrum disorder. *Am. J. Hum. Genet.* 92 (2), 221–237. doi: 10.1016/j.ajhg.2012.12.016
- Groenen, M. A. M., Archibald, A. L., Uenishi, H., Tuggle, S. K., Takeuchi, Y., Rothschild, M. F., et al. (2012). Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* 491 (7424), 393–398. doi: 10.1038/nature11622
- Guryev, V., Saar, K., Adamovic, T., Verheul, M., van Heesch, S. A. A., Cook, S., et al. (2008). Distribution and functional impact of DNA copy number variation in the rat. *Nat. Genet.* 40, 538–545. doi: 10.1038/ng.141
- Hay, E. H., Choi, I., Xu, L., Zhou, Y., Rowland, R. R., Lunney, J. K., et al. (2017). CNV analysis of host responses to porcine reproductive and respiratory syndrome virus infection. *J. Genomics* 5, 58–63. doi: 10.7150/jgen.20358
- Henrichsen, C. N., Vinckenbosch, N., Zöllner, S., Chaignat, E., Pradervand, S., Schütz, F., et al. (2009). Segmental copy number variation shapes tissue transcriptomes. *Nat. Genet.* 41 (4), 424–429. doi: 10.1038/ng.345
- Hou, Y., Liu, G. E., Bickhart, D. M., Cardone, M. F., Wang, K., Kim, E. S., et al. (2011). Genomic characteristics of cattle copy number variations. *BMC Genomics* 12 (1), 127. doi: 10.1186/1471-2164-12-127
- Hou, Y., Bickhart, D. M., Hvinden, M. L., Li, C., Song, J., Boichard, D. A., et al. (2012a). Fine mapping of copy number variations on two cattle genome assemblies using high density SNP array. *BMC Genomics* 13 (1), 376. doi: 10.1186/1471-2164-13-376
- Hou, Y., Liu, G. E., Bickhart, D. M., Matukumalli, L. K., Li, C., Song, J., et al. (2012b). Genomic regions showing copy number variations associate with resistance or susceptibility to gastrointestinal nematodes in Angus cattle. *Funct. Integr. Genomics* 12 (1), 81–92. doi: 10.1007/s10142-011-0252-1
- Jiang, L., Jiang, J., Wang, J., Ding, X., Liu, J., and Zhang, Q. (2012). Genome-wide identification of copy number variations in Chinese Holstein. *PLoS One* 7 (11), e48732. doi: 10.1371/journal.pone.0048732
- Jiang, J., Wang, J., Wang, H., Zhang, Y., Kang, H., Feng, X., et al. (2014). Global copy number analyses by next generation sequencing provide insight into pig genome variation. *BMC Genomics* 15, 593. doi: 10.1186/1471-2164-15-593
- Keel, B. N., Keele, J. W., and Snelling, W. M. (2016a). Genome-wide copy number variation in the bovine genome detected using low coverage sequence of popular beef breeds. *Anim. Genet.* 48 (2), 141–150. doi: 10.1111/age.12519
- Keel, B. N., Lindholm-Perry, A. K., and Snelling, W. M. (2016b). Evolutionary and functional features of copy number variation in the cattle genome. *Front Genet.* 7, 207. doi: 10.3389/fgene.2016.00207
- Keel, B. N., Nonneman, D. J., and Rohrer, G. A. (2017). A survey of single nucleotide polymorphisms identified from whole-genome sequencing and their functional effect in the porcine genome. *Anim. Genet.* 48 (4), 404–411. doi: 10.1111/age.12557
- Keel, B. N., Nonneman, D. J., Lindholm-Perry, A. K., Oliver, W. T., and Rohrer, G. A. (2018). Porcine single nucleotide polymorphisms and their functional effect: an update. *BMC Res. Notes.* 11, 860. doi: 10.1186/s13104-018-3973-6
- Klambauer, G., Schwarzbauer, K., Mayr, A., Clevert, D.-A., Mitterecker, A., Bodenhofer, U., et al. (2012). cn.MOPS: mixture of poisson for discovering

- copy number variations in next-generation sequencing data with a low false discovery rate. *Nucleic Acids Res.* 40 (9), e69. doi: 10.1093/nar/gks003
- Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., et al. (2009). Circos: an information aesthetic for comparative genomics. *Genome Res.* 19 (9), 1639–1645. doi: 10.1101/gr.092759.109
- Layer, R. M., Chiang, C., Quinlan, A. R., and Hall, I. M. (2014). LUMPY: a probabilistic framework for structural variant discovery. *Genome Biol.* 15, R84. doi: 10.1186/gb-2014-15-6-r84
- Le, S. Q., and Durbin, R. (2011). SNP detection and genotyping from low-coverage sequencing data of multiple diploid samples. *Genome Res.* 21, 952–960. doi: 10.1101/gr.113084.110
- Legault, M. A., Girard, S., Perreault, L. P. L., Rouleau, G. A., and Dubé, M. P. (2015). Comparison of sequencing based CNV discovery methods using monozygotic twin quartets. *PLoS One* 10 (3), e0122287. doi: 10.1371/journal.pone.0122287
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324
- Li, Y., Mei, S., Zhang, X., Peng, X., Liu, G., Tao, H., et al. (2012). Identification of genome-wide copy number variations among diverse pig breeds by array CGH. *BMC Genomics* 13 (1), 725. doi: 10.1186/1471-2164-13-725
- Lindholm-Perry, A. K., Rohrer, G. A., Holl, J. W., Shackelford, S. D., Wheeler, T. J., Koochmarie, M., et al. (2009). Relationships among calpastatin single nucleotide polymorphisms, calpastatin expression and tenderness in pork longissimus. *Anim. Genet.* 40, 713–721. doi: 10.1111/j.1365-2052.2009.01903.x
- Lindholm-Perry, A. K., Rohrer, G. A., Kuehn, L. A., Keele, J. W., Holl, J. W., Shackelford, S. D., et al. (2010). Genomic regions associated with kyphosis in swine. *BMC Genetics* 11, 112. doi: 10.1186/1471-2156-11-112
- Liu, G. E., Hou, Y., Zhu, B., Cardone, M. F., Jiang, L., Cellamare, A., et al. (2010). Analysis of copy number variations among diverse cattle breeds. *Genome Res.* 20 (5), 693–703. doi: 10.1101/gr.105403.110
- Liu, J., Zhang, L., Xu, L., Ren, H., Lu, J., Zhang, X., et al. (2013). Analysis of copy number variations in the sheep genome using 50K SNP BeadChip array. *BMC Genomics* 14 (1), 229. doi: 10.1186/1471-2164-14-229
- Locke, M. E. O., Milojevic, M., Eitutus, S. T., Patel, N., Wishart, A. E., Daley, M., et al. (2015). Genomic copy number variation in *Mus musculus*. *BMC Genomics* 16, 497. doi: 10.1186/s12864-015-1713-z
- Long, Y., Su, Y., Ai, H., Zhang, Z., Yang, B., Ruan, G., et al. (2016). A genome-wide association study of copy number variations with umbilical hernia in swine. *Anim. Genet.* 47 (3), 298–305. doi: 10.1111/age.12402
- Maher, B. (2008). The case of the missing heritability. *Nature* 456 (7218), 18–21. doi: 10.3389/fgene.2013.00225
- Malekpour, S. A., Pezeshk, H., and Sadeghi, M. (2018). MSeq-CNV: Accurate detection of copy number variation from sequencing of multiple samples. *Nat. Sci. Rep.* 8, 4009. doi: 10.1038/s41598-018-22323-8
- Mi, H., Muruganujan, A., Casagrande, J. T., and Thomas, P. D. (2013). Large-scale gene function analysis with the PANTHER classification system. *Nat. Protoc.* 8 (8), 1551–1566. doi: 10.1038/nprot.2013.092
- Nicholas, T. J., Baker, C., Eichler, E. E., and Akey, J. M. (2011). A high-resolution integrated map of copy number polymorphisms within and between breeds of the modern domesticated dog. *BMC Genomics* 12 (1), 414. doi: 10.1186/1471-2164-12-414
- Nguyen, D. T., Lee, K., Choi, H., Choi, M. K., Le, M. T., Song, N., et al. (2012). The complete swine olfactory subgenome: expansion of the olfactory gene repertoire in the pig genome. *BMC Genomics* 13, 584. doi: 10.1186/1471-2164-13-584
- Nonneman, D. J., Schneider, J. F., Lents, C. A., Wiedmann, R. T., Vallet, J. L., and Rohrer, G. A. (2016). Genome-wide association and identification of candidate genes for age of puberty in swine. *BMC Genetics* 17, 50. doi: 10.1186/s12863-016-0352-y
- Nonneman, D. J., Wiedmann, R. T., Snelling, W. M., and Rohrer, G. A. (2013). Genome-wide association of meat quality traits and tenderness in swine. *J. Anim. Sci.* 91, 4043–4050. doi: 10.2527/jas.2013-6255
- Paudel, Y., Madsen, O., Megens, H.-J., Frantz, L. A., Bosse, M., Bastiaansen, J. W., et al. (2013). Evolutionary dynamics of copy number variation in pig genomes in the context of adaptation and domestication. *BMC Genomics* 14 (1), 449. doi: 10.1186/1471-2164-14-449
- Paudel, Y., Madsen, O., Megens, H.-J., Frantz, L. A. F., Bosse, M., Crooijmans, R. P. M. A., et al. (2015). Copy number variation in the speciation of pigs: A possible prominent role for olfactory receptors. *BMC Genomics* 16 (1), 330. doi: 10.1186/s12864-015-1449-9
- Pezer, Ž., Harr, B., Teschke, M., Babikar, H., and Tautz, D. (2015). Divergence patterns of genic copy number variation in natural populations of the house mouse (*Mus musculus domesticus*) reveal three conserved genes with major population-specific expansions. *Genome Res.* 25 (8), 1114–1124. doi: 10.1101/gr.187187.114
- Ramayo-Caldas, Y., Castelló, A., Pena, R. N., Alves, E., Mercadé, A., Souza, C. A., et al. (2010). Copy number variation in the porcine genome inferred from a 60 k SNP BeadChip. *BMC Genomics* 11 (1), 593. doi: 10.1186/1471-2164-11-593
- Redon, R., Ishikawa, S., Fitch, K. R., Feuk, L., Perry, G. H., Andrews, T. D., et al. (2006). Global variation in copy number in the human genome. *Nature* 444, 444–454. doi: 10.1038/nature05329
- Revay, T., Quach, A. T., Maignel, L., Sullivan, B., and King, W. A. (2015). Copy number variations in high and low fertility breeding boars. *BMC Genomics* 16 (1), 280. doi: 10.1186/s12864-015-1473-9
- Revilla, M., Puig-Oliveras, A., Castelló, A., Crespo-Piazuelo, D., Paludo, E., Fernández, A. I., et al. (2017). A global analysis of CNVs in swine using whole genome sequence data and association analysis with fatty acid composition and growth traits. *PLoS One* 12 (5), e0177014. doi: 10.1371/journal.pone.0177014
- Rohrer, G. A., and Nonneman, D. J. (2017). Genetic analysis of teat number in pigs reveals some developmental pathways independent of vertebra number and several loci which only affect a specific side. *Genet. Sel. Evol.* 49, 4. doi: 10.1186/s12711-016-0282-1
- Rohrer, G. A., Nonneman, D. J., Wiedmann, R. T., and Schneider, J. F. (2015). A study of vertebra number in pigs confirms the association of vertin and reveals additional QTL. *BMC Genetics* 16, 129. doi: 10.1186/s12863-015-0286-9
- Rohrer, G. A., Rempel, L. A., Miles, J. R., Keele, J. W., Wiedmann, R. T., and Vallet, J. L. (2014). Identifying genetic loci controlling neonatal passive transfer of immunity using a hybrid genotyping strategy. *Anim. Genet.* 45, 340–349. doi: 10.1111/age.12131
- Schiavo, G., Dolezal, M. A., Scotti, E., Bertolini, F., Calò, D. G., Galimberti, G., et al. (2014). Copy number variants in Italian Large White pigs detected using high-density single nucleotide polymorphisms and their association with back fat thickness. *Anim. Genet.* 45 (5), 745–749. doi: 10.1111/age.12180
- Schneider, J. F., Miles, J. R., Brown-Brandl, T. M., Nienaber, J. A., Rohrer, G. A., and Vallet, J. L. (2015). Genomewide association analysis for average birth interval and stillbirth in swine. *J. Anim. Sci.* 93, 529–540. doi: 10.2527/jas.2014-7899
- Schneider, J. F., Nonneman, D. J., Wiedmann, R. T., Vallet, J. L., and Rohrer, G. A. (2014). Genomewide association and identification of candidate genes for ovulation rate in swine. *J. Anim. Sci.* 92, 3792–3803. doi: 10.2527/jas.2014-7788
- Schneider, J. F., Rempel, L. A., and Rohrer, G. A. (2012). Genome-wide association study of swine farrowing traits. Part II: Bayesian analysis of marker data. *J. Anim. Sci.* 90, 3360–3367. doi: 10.2527/jas.2011-4759
- Spehr, M., Schwane, K., Riffell, J. A., Zimmer, R. K., and Hatt, H. (2006). Odorant receptors and olfactory-like signaling mechanisms in mammalian sperm. *Mol. Cell. Endocrinol.* 250 (1–2), 128–136. doi: 10.1016/j.mce.2005.12.035
- Stothard, P., Choi, J.-W., Basu, U., Sumner-Thomson, J. M., Meng, Y., Liao, X., et al. (2011). Whole genome resequencing of Black Angus and Holstein cattle for SNP and CNV discovery. *BMC Genomics* 12 (1), 559. doi: 10.1186/1471-2164-12-559
- Stranger, B. E., Forrest, M. S., Dunning, M., Ingle, C. E., Beazley, C., Thorne, N., et al. (2007). Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* 315 (5813), 848–853. doi: 10.1126/science.1136678
- Wang, H., Wang, C., Yang, K., Liu, J., Wang, Y., Xu, X., et al. (2015). Genome wide distributions and functional characterization of copy number variations between Chinese and western pigs. *PLoS One* 10 (7), e0131522. doi: 10.1371/journal.pone.0131522
- Wang, J., Jiang, J., Fu, W., Ding, X., Liu, J.-F., Zhang, Q., et al. (2012). A genome-wide detection of copy number variations using SNP genotyping arrays in swine. *BMC Genomics* 13 (1), 273. doi: 10.1186/1471-2164-13-273
- Wang, J., Wang, H., Jiang, J., Kang, H., Feng, X., Zhang, Q., et al. (2013). Identification of genome-wide copy number variations among diverse pig breeds using SNP genotyping arrays. *PLoS One* 8 (7), e68683. doi: 10.1371/journal.pone.0068683
- Wang, J., Jiang, J., Wang, H., Kang, H., Zhang, Q., and Liu, J. F. (2014). Enhancing genome-wide copy number variation identification by high density array CHH

- using diverse resources of pig breeds. *PLoS One* 9 (1), e87571. doi: 10.1371/journal.pone.0087571
- Wang, J., Jiang, J., Wang, H., Kang, H., Zhang, Q., and Liu, J. F. (2015). Improved detection and characterization of copy number variations among diverse pig breeds by array CGH. *G3: Genes[Genomes]Genetics* 5 (6), 1253–1261. doi: 10.1534/g3.115.018473
- Wang, K., Li, M., Hadley, D., Liu, R., Glessner, J., Grant, S. F., et al. (2007). PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res.* 17, 1665–1674. doi: 10.1101/gr.6861907
- Wang, L., Liu, X., Zhang, L., Yan, H., Luo, W., et al. (2013). Genome-wide copy number variations inferred from SNP genotyping arrays using a large white and Minzhu intercross population. *PLoS One* 8 (10), e74879. doi: 10.1371/journal.pone.0074879
- Wang, Z., Chen, Q., Liao, R., Zhang, L., Zhang, X., Liu, X., et al. (2016). Genome-wide genetic variation discovery in Chinese Taihu pig breeds using next generation sequencing. *Anim. Genet.* 48 (1), 38–47. doi: 10.1111/age.12465
- Wiedmann, R. T., Nonneman, D. J., and Rohrer, G. A. (2015). Genome-wide copy number variations using SNP genotyping in a mixed breed swine population. *PLoS One* 10 (7), e0133529. doi: 10.1371/journal.pone.0133529
- Wu, Y., Fan, H., Jing, S., Xia, J., Chen, Y., Zhang, L., et al. (2015). A genome-wide scan for copy number variations using high-density single nucleotide polymorphism array in Simmental cattle. *Anim. Genet.* 46 (3), 289–298. doi: 10.1111/age.12288
- Xie, J., Li, R., Li, S., Ran, X., Wang, J., Jiang, J., et al. (2016). Identification of copy number variations in Xiang and Kele pigs. *PLoS One* 11 (2), e0148565. doi: 10.1371/journal.pone.0148565
- Xu, L., Hou, H., Bickhart, D. M., Zhou, Y., Hay, E. H., Song, J., et al. (2016). Population-genetic properties of differentiated copy number variations in cattle. *Sci. Rep.* 6, 23161. doi: 10.1038/srep23161
- Yi, G., Qu, L., Liu, J., Yan, Y., Xu, G., and Yang, N. (2014). Genome-wide patterns of copy number variation in the diversified chicken genomes using next-generation sequencing. *BMC Genomics* 15 (1), 962. doi: 10.1186/1471-2164-15-962
- Young, J. M., Endicott, R. M., Parghi, S. S., Walker, M., Kidd, J. M., and Trask, B. J. (2008). Extensive copy-number variation of the human olfactory receptor gene family. *Am. J. Hum. Genet.* 83 (2), 228–242. doi: 10.1016/j.ajhg.2008.07.005
- Zhan, B., Fadista, J., Thomsen, B., Hedegaard, J., Panitz, F., and Bendixen, C. (2011). Global assessment of genomic variation in cattle by genome resequencing and high-throughput genotyping. *BMC Genomics* 12 (1), 557. doi: 10.1186/1471-2164-12-557
- Zhao, M., Wang, Q., Wang, Q., Jia, P., and Zhao, Z. (2013). Computational tools for copy number variation (CNV) detection using next-generation sequencing data: features and perspectives. *BMC Bioinformatics* 14, S1. doi: 10.1186/1471-2105-14-S11-S1
- Zhou, L. S., Li, J., Yang, J., Liu, C. L., Xie, X. H., He, Y. N., et al. (2016). Genome-wide mapping of copy number variations in commercial hybrid pigs using high-density SNP genotyping array. *Genetika* 52 (1), 97–105. doi: 10.7868/S0016675815120140
- Zheng, X., Shaffer, J. R., McHugh, C. P., Laurie, C. C., Feenstra, B., Melbye, M., et al. (2012). 1.0 using family as a standard to evaluate copy number variation calling strategies for genetic association studies. *Genet. Epidemiol* 36 (3), 253–262. doi: 10.1002/gepi.21618
- Zhou, B., Ho, S. S., Zhang, X., Pattni, R., Haraksingh, R. R., and Urban, A. E. (2018). Whole-genome sequencing analysis of CNV using low-coverage and paired-end strategies is efficient and outperforms array-based CNV analysis. *J. Med. Genet.* 11, 735. doi: 10.1136/jmedgenet-2018-105272

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

Copyright © 2019 Keel, Nonneman, Lindholm-Perry, Oliver and Rohrer. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.