

RESEARCH ARTICLE

Open Access



A genome-wide single nucleotide polymorphism and copy number variation analysis for number of piglets born alive

Nedenia Bonvino Stafuzza^{1,2*} , Rafael Medeiros de Oliveira Silva^{2,3}, Breno de Oliveira Fragomeni⁴, Yutaka Masuda², Yijian Huang⁵, Kent Gray⁵ and Daniela A. Lino Lourenco²

Abstract

Background: In this study we integrated the CNV (copy number variation) and WssGWAS (weighted single-step approach for genome-wide association) analyses to increase the knowledge about number of piglets born alive, an economically important reproductive trait with significant impact on production efficiency of pigs.

Results: A total of 3892 samples were genotyped with the Porcine SNP80 BeadChip. After quality control, a total of 57,962 high-quality SNPs from 3520 Duroc pigs were retained. The PennCNV algorithm identified 46,118 CNVs, which were aggregated by overlapping in 425 CNV regions (CNVRs) ranging from 2.5 Kb to 9718.4 Kb and covering 197 Mb (~ 7.01%) of the pig autosomal genome. The WssGWAS identified 16 genomic regions explaining more than 1% of the additive genetic variance for number of piglets born alive. The overlap between CNVR and WssGWAS analyses identified common regions on SSC2 (4.2–5.2 Mb), SSC3 (3.9–4.9 Mb), SSC12 (56.6–57.6 Mb), and SSC17 (17.3–18.3 Mb). Those regions are known for harboring important causative variants for pig reproductive traits based on their crucial functions in fertilization, development of gametes and embryos. Functional analysis by the Panther software identified 13 gene ontology biological processes significantly represented in this study such as reproduction, developmental process, cellular component organization or biogenesis, and immune system process, which plays relevant roles in swine reproductive traits.

Conclusion: Our research helps to improve the understanding of the genetic architecture of number of piglets born alive, given that the combination of GWAS and CNV analyses allows for a more efficient identification of the genomic regions and biological processes associated with this trait in Duroc pigs. Pig breeding programs could potentially benefit from a more accurate discovery of important genomic regions.

Keywords: CNV, Complex trait, GWAS, SNP panel, *Sus scrofa domesticus*

Background

Reproductive performance in pig production systems is usually quantified by several economically important traits. The number of piglets born alive is an important trait in pig breeding programs due to its significant impact on the production efficiency; however, this is a difficult trait to improve because of low prediction accuracy and

heritability [1–3]. Although the heritability of reproductive traits is moderate to low, genetic improvements can be obtained using genomic tools to explore the chromosomal regions and genes that explain the variation in reproductive traits. In addition to that, genomic acts as an extra source of information increasing prediction accuracy even for lowly heritable traits [4].

Until now, a total of 27,465 quantitative trait loci (QTL) have been mapped in the porcine genome for 663 different traits (Pig QTLdb Release 35). From about 2058 QTL described for reproduction traits, a total of 1004 QTL were described for litter traits, of which 163 QTL are related to number of piglets born alive [5].

* Correspondence: nedeniab@gmail.com

¹Department of Exact Science, School of Agricultural and Veterinarian Sciences (FCAV), Sao Paulo State University (UNESP), Jaboticabal, SP 14884-900, Brazil

²Department of Animal and Dairy Science, University of Georgia, Athens, GA, USA

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



Studies have reported the association of several genes and genomic regions with number of piglets born alive in several pig breeds [1–3, 6–10]. Genome-wide association studies (GWAS) became very powerful tools to investigate genetic architecture of economically important traits in livestock, allowing detection of genomic regions and genes controlling polygenic traits related to reproductive performance in pigs [2, 3, 8].

The combination of GWAS with other genomic approaches can provide a better understanding about genes and pathways involved in complex traits [11]. In several livestock species, it has been demonstrated useful to integrate GWAS and copy number variation (CNV) analysis to advance the knowledge of economically important complex traits [12, 13].

Structural variants such as CNV represent an important source of genomic variation in mammalian genomes. They can be defined as segments of DNA (conventionally > 1 Kb) that differ in number of copies compared to a reference genome [14]. Compared with SNPs, CNVs cover wider chromosomal regions and may potentially be responsible for changes in gene structure, modifications in gene regulation, changes in gene dosage, and exposing recessive alleles, resulting in large phenotypic effects. A well-characterized phenotypic variation affected by CNV in pig is the white coat phenotype generated by the duplication of the *KIT* gene [15].

Although several methodologies have been applied to detect CNVs in the pig genome, SNP arrays have better properties, such as lower cost compared to next generation sequencing (NGS) methodology and can be used for both GWAS and CNV detections. Employing SNP for CNV detection, previous studies have identified thousands of porcine CNVs [12, 16–28]. Because most of these studies comprised relatively small numbers of animals, the impacts of CNVs on phenotypes are still relatively poorly understood.

In animal breeding and genetics, the commercial populations are usually large and comprise phenotypes, pedigree, and genotypes for a fraction of pedigreed animals. Single-step genomic best linear unbiased prediction (ssGBLUP) was developed to predict breeding values when this type of data is available [29]. This is the method of choice for such populations because traditional GWAS methods cannot be implemented directly due to the necessity of combining results with pedigree structure to create pseudo-observations [30, 31]. When the structure of the genotyped dataset is complex, problems such as double counting of contributions from pedigree and phenotypes, and preselection bias [32] reduce accuracy. Lately, ssGBLUP was extended to GWAS [33, 34]. The GWAS under the ssGBLUP framework is called ssGWAS and allows the combination of phenotypes, pedigree and genotypes in one single analysis with

no need of calculating pseudo-observations [35, 36]. In ssGBLUP, the main assumption is that all SNP explain the same proportion of additive genetic variance; however, this is not biologically true, especially when the traits are affected by large QTL. To account for the fact SNP may explain different proportion of variance, weighted ssGBLUP can be used. The WssGBLUP (weighted single-step approach for genome-wide association) method weighs SNP according to their effects in an iterative way [33]. The WssGWAS is fast, accurate and simple to implement for genome-wide association studies [33].

The aim of this study was to perform a WssGWAS to effectively identify genomic regions and biological processes related to number of piglets born alive in the Duroc breed and perform a CNV analyses to detect potential regions affecting the phenotype through changes in gene dosage. The elucidation of genes and molecular mechanisms controlling this trait should result in a better understanding of the genetic regulation of reproductive performance.

Results

CNV detection

After applying stringent filtering criteria, 653 samples did not pass the filtering and were discarded. The 2867 remaining samples were explored to search for CNV.

A total of 46,118 CNVs (4892 gains and 41,226 losses) were detected by PennCNV, of which 8152 (645 gain and 7507 loss) were non-redundant CNVs among the total of 2865 samples. CNVs were not identified in two samples. All CNVs detected by PennCNV were used to infer CNVR by aggregating overlapping CNVs. Thus, a total of 425 CNVRs were obtained (Additional file 1).

The size of the CNVRs averaged 463,621 bp and ranged from 2552 bp to 9718,410 bp (Fig. 1). Among these regions, 342 corresponded to copy losses, 19 to copy gains and 64 to both (the same fragment showed losses or gains in different animals). It corresponds to a loss:gain ratio of 4.89.

The CNVRs inferred in our study covered 197,038,894 bp (7.01%) of the autosomal genome sequence, and their frequencies ranged from 0.5 to 53.61% in this Duroc population (Table 1).

Although CNVRs were identified in all autosomes, the number and proportion of chromosomes covered by CNVRs varied considerably (Fig. 2, Table 1). Chromosome 1 showed the largest number of CNVRs (49 CNVRs), which covered only 3.94% of its sequence; the lowest coverage observed for a single chromosome. Although the SSC11 showed the highest coverage of a chromosome sequence (16.03%), this chromosome showed a small number of CNVR (22), but with bigger sizes. The SSC17 presented the smallest number of CNVR (10), covering only 5.05% of its sequence.

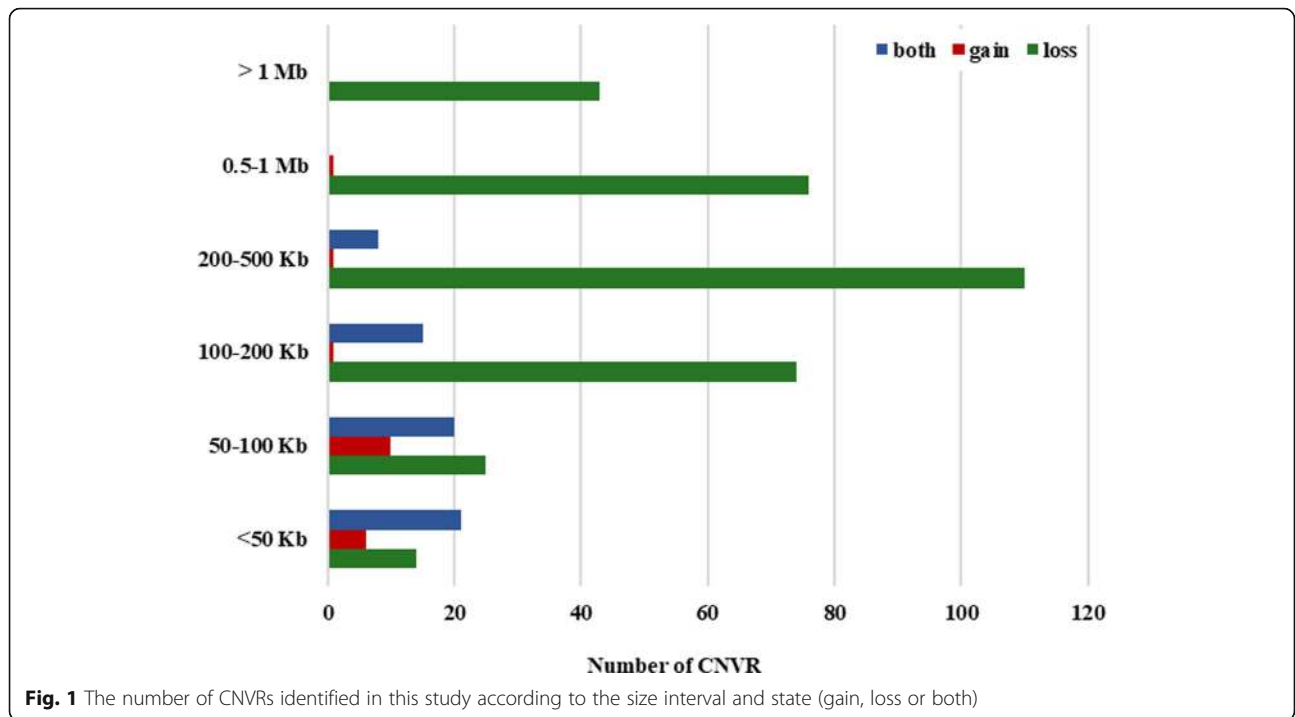


Table 1 Chromosome distribution of all 425 CNVRs detected in the porcine genome

| Chr | Chr length (bp) | CNVR number | Length of CNVR (bp) | % |
|-------|-----------------|-------------|---------------------|-------|
| SSC1 | 315,321,322 | 49 | 12,420,297 | 3.94 |
| SSC2 | 162,569,375 | 33 | 17,317,379 | 10.65 |
| SSC3 | 144,787,322 | 30 | 12,626,965 | 8.72 |
| SSC4 | 143,465,943 | 15 | 10,565,509 | 7.36 |
| SSC5 | 111,506,441 | 17 | 9,232,945 | 8.28 |
| SSC6 | 157,765,593 | 22 | 12,493,722 | 7.92 |
| SSC7 | 134,764,511 | 26 | 10,659,114 | 7.91 |
| SSC8 | 148,491,826 | 36 | 10,789,394 | 7.26 |
| SSC9 | 153,670,197 | 19 | 12,793,931 | 8.32 |
| SSC10 | 79,102,373 | 16 | 8,515,041 | 10.76 |
| SSC11 | 87,690,581 | 22 | 14,058,242 | 16.03 |
| SSC12 | 63,588,571 | 25 | 8,687,209 | 13.66 |
| SSC13 | 218,635,234 | 24 | 11,779,507 | 5.38 |
| SSC14 | 153,851,969 | 27 | 9,624,904 | 6.25 |
| SSC15 | 157,681,621 | 26 | 23,006,893 | 14.59 |
| SSC16 | 86,898,991 | 12 | 5,426,107 | 6.24 |
| SSC17 | 69,701,581 | 10 | 3,519,539 | 5.05 |
| SSC18 | 61,220,071 | 16 | 3,522,196 | 5.75 |
| Total | 2,808,509,378 | 425 | 197,038,894 | 7.01 |

Weighted single-step genome-wide association study

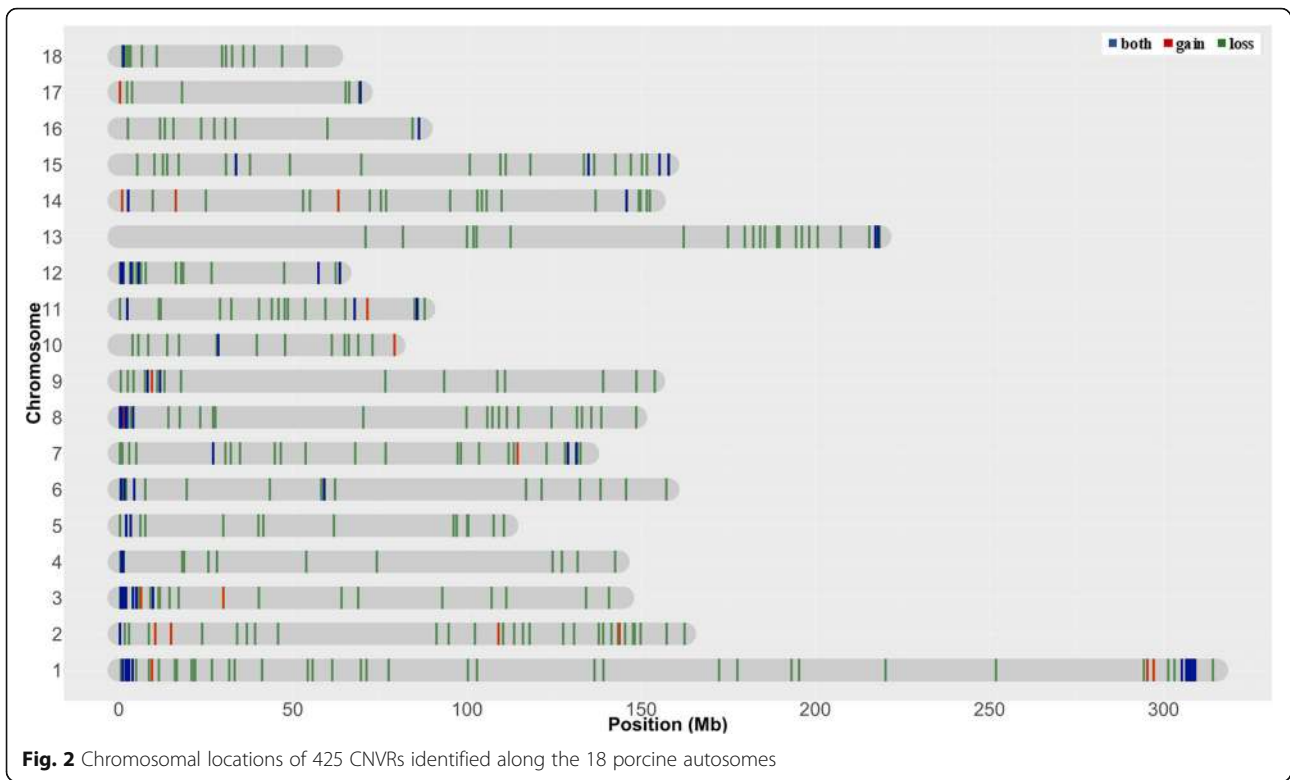
The estimated heritability for number of piglets born alive was 0.11 ± 0.008 . The additive genetic direct and permanent environmental variances were 0.78 and 0.57, respectively, whereas the residual variance was 5.68.

A total of 16 windows were detected each explaining more than 1% of the additive genetic variance for number of piglets born alive on chromosomes 2, 3, 4, 11, 12, 13, 14, 15, 16, and 17 (Fig. 3, Table 2). These significant windows explained a total of 22.54% of additive genetic variance for number of piglets born alive in the Duroc breed.

We identified common regions between GWAS and CNVR analyses on SSC2 (4.2–5.2 Mb), SSC3 (3.9–4.9 Mb), SSC12 (56.6–57.6 Mb), and SSC17 (17.3–18.3 Mb), comprising a total of 56 protein-coding genes (Table 2). These regions are very interesting because their gene content could affect the number of pigs born alive through changes in gene dosage. The windows on SSC2, SSC3 and SSC17 are in deletion CNVR, whereas the window on SSC12 is in deletion and duplication CNVR.

Gene enrichment analysis

Functional annotation revealed genes involved in 13 gene ontology (GO) biological processes such as reproduction (GO:0000003), developmental process (GO:0032502), cellular process (GO:0009987), cellular component organization or biogenesis (GO:0071840), biological regulation (GO:0065007), immune system process (GO:0002376), and rhythmic process (GO:0048511), which may play relevant roles in reproductive traits (Fig. 4, Additional file 2).



Discussion

The number of animals used in this study is greater than previous in studies on pigs [12, 22, 23, 25, 27, 28, 37, 38], which could decrease the amount of false-positive CNVs.

Several studies also detected deletions more frequently than duplications in pigs [17, 18, 22, 23, 39] which could be partially explained by biological factors because non-allelic homologous recombination tends to create more deletions than duplications [40], and partially by technical bias

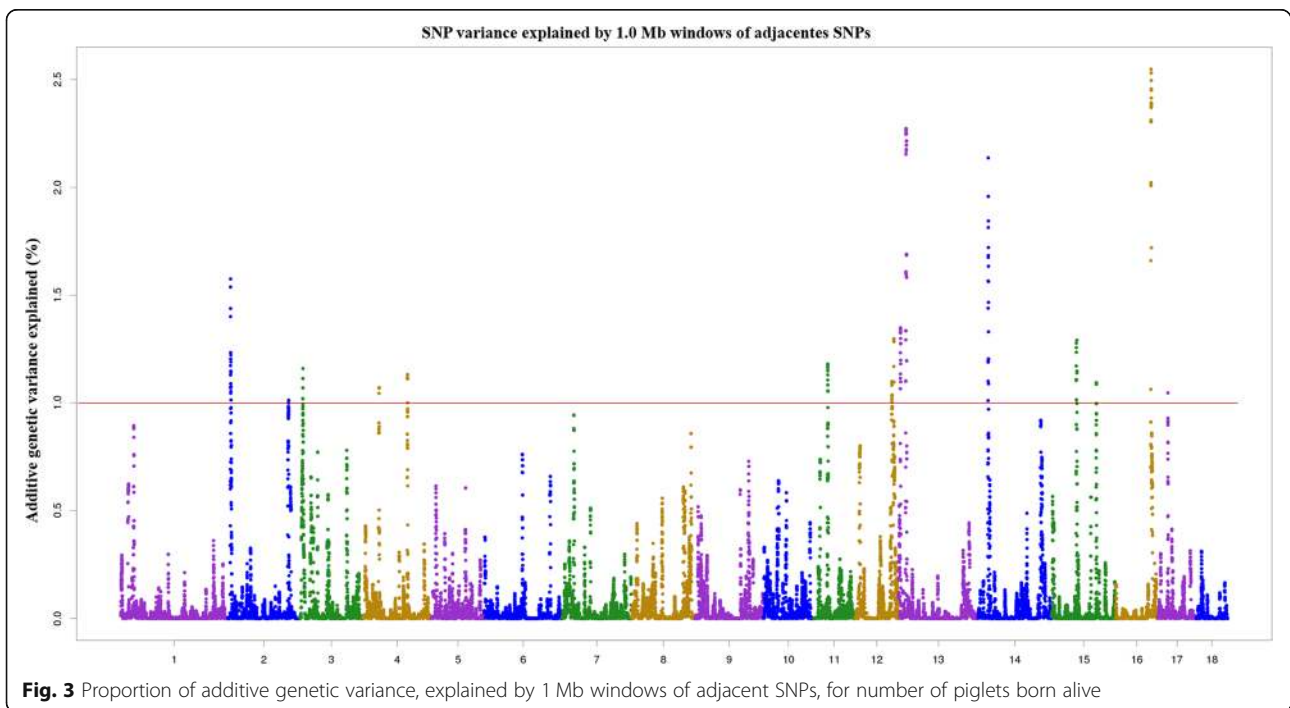


Table 2 Identification of protein-coding genes based on the additive genetic variance (Var) explained by 1 Mb windows of adjacent SNPs

| Chr | Region (bp) | Genes | Var |
|--------------------|-------------------------|---|------|
| SSC2 ^a | 4,239,252-5,222,613 | <i>KDM2A, RHOD, C2H1orf86, SYT12, PC, RCE1, C2H1orf80, CCDC87, CCS, CTSF, ACTN3, ZDHHC24, BBS1, DPP3, PELI3, MRPL11, NPAS4, SLC29A2, B4GAT1, BRMS1, RIN1, CD248, TMEM151A, YIF1A, CNIH2, RAB1B, KLC2, PACS1, SF3B2</i> | 1.57 |
| SSC2 | 145,012,848-146,011,865 | <i>SPOCK1, KLHL3, HNRNPA0, LOC100517462, MYOT, PKD2L2, LOC100520183, FAM13B, WNT8A, NME5, BRD8, KIF20A, CDC23, GFRA3, CDC25C, SLBP2</i> | 1.01 |
| SSC3 ^a | 3,981,003-4,980,011 | <i>SDK1, FOXK1, AP5Z1, RADIL, PAPOLB, MMD2, WIPI2, SLC29A4, TNRC18, FBXL18, ACTB, FSCN1, RNF216</i> | 1.16 |
| SSC4 | 20,774,634-21,762,434 | <i>COLEC10, TNFRSF11B, SAMD12, EXT1</i> | 1.07 |
| SSC4 | 104,642,633-105,640,667 | <i>SLC39A1, CRTC2, DENND4B, GATAD2B, SLC27A3, INTS3, NPR1, ILF2, SNAPIN, CHTOP, S100A5, S100A4, S100A3, S100A2, S100A16, S100A14, S100A13, PPGRP-S, S100A9, S100A12, S100A8, LOC102161828, S100A7, PGLYRP3, LOR, PRR9, LELP1, SPRR2E, LOC100737840, SPRR1A, SPRP, IVL, SMCP, KPRP</i> | 1.13 |
| SSC11 | 25,254,733-26,232,874 | <i>TNFSF11, AKAP11, DGKH, VWA8, RGCC, NAA16</i> | 1.18 |
| SSC12 | 54,066,400-55,065,792 | <i>GP1BA, MINK1, PLD2, ZMYND15, CXCL16, MED11, ARRB2, PSMB6, GLTPD2, VMO1, TM4SF5, PELP1, ALOX15, SLC16A11, SLC16A13, BCL6B, C12H17orf49, RNASEK, ALOX12, CLEC10A, ASGR2, SLC2A4, CLDN7, ELP5, GABARAP, PHF23, DVL2, DLG4, YBX2, NEURL4, ACAP1, KCTD11, TMEM95, TNK1, PLSCR3, POLR2A, SLC35G6, ZBTB4, CHRN1, FGF11, TMEM102, SPEM2, DNAH2</i> | 1.10 |
| SSC12 ^a | 56,619,955-57,608,007 | <i>CCDC42, MFSD6L, PIK3R6, PIK3R5, NTN1, STX8, CFAP52, USP43, DHRS7C, GSG1L2, GLP2R, RCVRN, GAS7</i> | 1.30 |
| SSC13 | 3,171,064-4,162,557 | <i>BTD, ANKRD28, GALNT15, DPH3, OXNAD1, RFTN1, DAZL</i> | 1.35 |
| SSC13 | 10,788,202-11,761,268 | <i>UBE2E2</i> | 2.27 |
| SSC14 | 13,059,455-14,032,461 | <i>SCARA5, NUGGC, ELP3, PNOG, ZNF395, FBXO16, FZD3, EXTL3, INTS9, HMBOX1, KIF13B</i> | 2.14 |
| SSC15 | 55,188,321-56,161,595 | <i>FGFR1, LETM2, NSD3, PLPP5, DDHD2, BAG4, ADRB3, STAR, ASH2L, EIF4EBP1, LSM1, BRF2, ADGRA2, PLPBP, ERLIN2, ZNF703, KCNU1</i> | 1.28 |
| SSC15 | 57,551,770-58,532,159 | – | 1.29 |
| SSC15 | 131,996,123-132,969,736 | <i>TNS1, RUFY4, CXCR2, CXCR1</i> | 1.09 |
| SSC16 | 77,028,962-78,013,471 | <i>GLRA1, G3BP1, ATOX1, LOC100514500, SPARC, FAT2, SLC36A1, SLC36A2, SLC36A3, GM2A, CCDC69, ANXA6</i> | 2.55 |
| SSC17 ^a | 17,325,007-18,300,615 | <i>BMP2</i> | 1.04 |

^aregion also identified in a CNVR

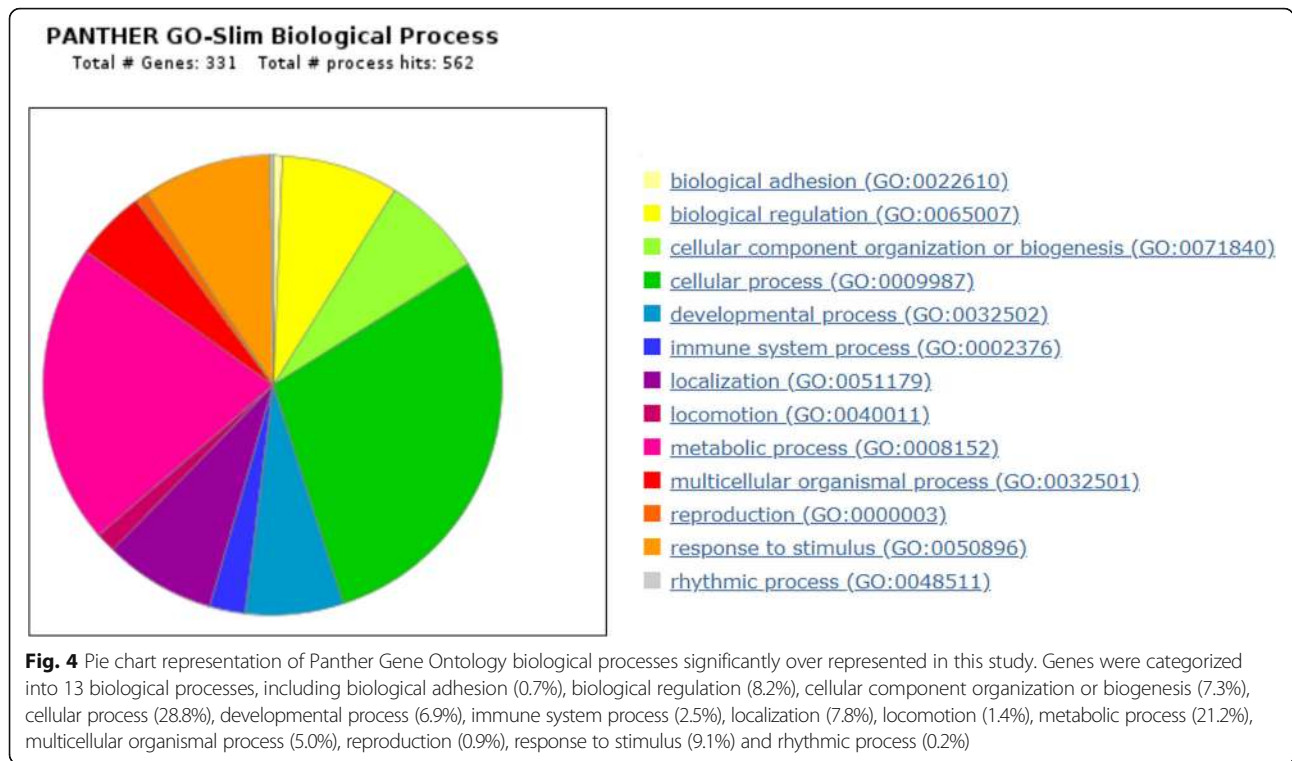
because deletions tend to be under stronger purifying selection, i.e., deletions are more deleterious than duplications [41].

Although several studies have reported CNVs in the porcine genome, CNVs have been described to have breed-specific characteristics [22, 23, 25, 27, 28, 37, 39, 42]. By using the Illumina Porcine SNP60 BeadChip to detect CNVR, a total of 170 CNVRs (7 gains, 161 losses and 2 both) were detected in 293 Large White pigs [39], 249 CNVRs (70 gains, 43 losses and 136 both) in 585 Large White X Minzhu pigs [42], 65 CNVRs (21 gains, 32 losses and 12 both) in 223 Iberian pigs [37], and 348 CNVRs (88 gains, 243 losses and 17 both) in 302 animals from ten Chinese pig breeds [22]. By using arrayCGH technology, 37 CNVRs (18 gains and 19 losses) were identified in 12 Duroc boars [43]. Besides breed, it is important to highlight that the number of samples analyzed, as well as the differences in calling technology (NGS, SNP genotyping or arrayCGH), resolution, genome coverage and/or quality control to filter CNVs used in the aforementioned

studies directly influence the results obtained, as described previously [44–46].

Some genes identified in common regions between GWAS and CNVR analyses were highlighted according to their function. The *KDM2A*, *ACTN3*, and *RHOD* genes are mapped on SSC2 (4.2–5.2 Mb) region. The *KDM2A* gene, also known as *FBXL11* and *JHDM1A*, plays an important role in gene silencing, cell cycle, and cell growth through histone demethylation modification. This gene was identified as differentially expressed in porcine embryonic skeletal muscle, being therefore involved in skeletal muscle development and growth [47]. The *ACTN3* gene also encodes a protein which exhibits an important function in muscle metabolism [48]. The *RHOD* gene is a regulator of reorganization of the actin cytoskeleton and consequently, regulates several cellular processes such as vesicle trafficking, chemotaxis, cell migration and proliferation [49].

The *ACTB* gene is located on SSC3 window (3.9–4.9 Mb). This gene encodes β -actin, a member of the actin family of proteins that are related to cell motility, structure,



and integrity, fundamental processes for embryonic development. β -actin protein is required for meiosis in mouse oocytes [50] and for early embryonic development because knocking out this gene resulted in embryonic lethality [51].

According to function, we identified some important genes mapped on SSC12 (56.6–57.6 Mb), such as *CCDC42*, *PIK3R5*, and *NTN1*. The *CCDC42* gene plays a crucial role in sperm development and male fertility in mouse [52], while the *PIK3R5* gene has been implicated in several cell functions such as proliferation, survival, differentiation, growth, motility, and intracellular trafficking [53, 54]. The *NTN1* gene codifies the netrin-1 which is an essential protein of embryonic development with important functions in cell migration, axon guidance, angiogenesis and morphogenesis [55]. Netrin-1 also is a crucial factor for regulating angiogenesis in placenta [56] and osteoclast differentiation [57]. Basini et al. [58] identified the *NTN1* gene as a potential modulator of swine follicular function.

The *BMP2* gene is the only gene mapped on SSC17 window (17.3–18.3 Mb). BMP proteins exhibit wide spectrum of activities in several tissues (cartilage, bone, blood vessels, liver, lung, kidney, heart and neurons) and perform multiple roles in regulation of growth, differentiation, and apoptosis, playing important functions during embryonic development and tissue morphogenesis [59]. The *BMP2* gene plays a critical role in mesenchymal cells influencing adipogenesis, myogenesis, chondrogenesis, and osteogenesis [60–63]. Homozygous knockout mice for *BMP2* gene exhibits embryonic lethality with defects in extra-embryonic and

embryonic tissues [64], whereas heterozygous knockout mice have defects in cartilage, bone and heart development [63]. As the *BMP2* is mapped to a window that explains part of the additive genetic variance for number of piglets born alive, which is located on a deletion CNVR, the *BMP2* becomes an important gene to influence the number of piglets born alive, especially because subnormal levels of *BMP2* negatively impacted embryonic development [63] and in the adult, it is required for uterine decidual response during embryo implantation [65].

Additionally, we highlighted the SSC12 (54.0–55.0 Mb) and SSC14 (13.0–14.0 Mb) chromosome regions identified only by WssGWAS due its gene content. The *ZMYND15*, *YBX2* and *TMEM95* genes are mapped on SSC12 (54.0–55.0 Mb) and are related with reproductive traits. The *ZMYND15* gene is primordial for sperm production and male fertility [66]. The *YBX2* gene codifies a protein required for mammal development and fertility because knockout mice for this gene presented disruption of spermatogenesis and oogenesis [67–69]. The *TMEM95* gene codifies a protein located at the surface of spermatozoa and deficiency of *TMEM95* severely compromises male reproductive performance, resulting in subfertility in cattle [70]. The *ELP3* gene, mapped on SSC14 window (13.0–14.0 Mb), plays important roles in embryonic stem cell maintenance and early development in mouse [71].

The *SCIMP*, *TNPI*, *MUS81*, *KCNU1*, and *ILF2* genes were identified as related with reproduction (GO:0000003) process. Among these genes, we highlighted the *TNPI* gene

which codifies a transition nuclear protein that acts in spermiogenesis. To detect the role of *TPN1* in vivo, a study produced knockout mice and identified that mutations had no effect on female fertility, although had reduced litter sizes [72]. The *KCNUI* gene, also known as *SLO3*, is expressed only in mammalian testis and plays an important role in male fertility [73].

Members of S100 protein family (*S100A3*, *S100A4*, *S100A5*, *S100A9*, *S100A12*, *S100A14*, *S100A16*) were identified acting in cellular process (GO:0009987) and metabolic process (GO:0008152). These genes act in the regulation of cellular processes such as differentiation, proliferation, motility, chemotaxis, apoptosis, and inflammatory response [74]. The *S100A14* and *S100A16* were identified as candidate genes for human embryo adhesion and implantation [75].

Deleted in Azoospermia (DAZ) gene family encodes proteins with essential roles in male and female gametogenesis. The protein encoded by *DAZL* gene, related with metabolic process (GO:0008152), was detected in fetal germ cells and developing oocytes [76]. Mutations in *DAZL* gene have been associated to infertility in man and women [77, 78].

The *FGF11*, *FGRF1* and *WNT8A* genes are related to developmental process (GO:0032502), response to stimulus (GO:0050896), and cellular process (GO:0009987). The *FGF11* and *FGRF1* gene encodes a member of the fibroblast growth factor (FGF) family and a member of the fibroblast growth factor receptor (FGFR) family, respectively. The FGF family members and their receptors influence mitogenesis and differentiation with essential roles in biological processes, including preimplantation of embryos, embryonic development, and organogenesis [79]. The *WNT8A* gene, expressed during early embryogenesis, belongs to WNT gene family that has been related to developmental processes, including differentiation, proliferation, apoptosis, regulation of cell fate and patterning during embryonic development [80].

The *TNKL1*, *PNOG* and *AKAP11* genes are related with biological regulation (GO:0065007) and cellular process (GO:0009987). The *TNKL1* gene encodes a tyrosine protein kinase family member highly expressed in fetal tissues signaling pathways widely utilized during fetal development [81]. Tyrosine protein kinases are crucial regulators of intracellular signal transduction pathways, cell growth, differentiation, survival, and migration [82]. The *AKAP11* gene is highly expressed during spermatogenesis and in mature sperm with assumed role in spermatogenesis and sperm functions, which may play functions in cell cycle control of germ and somatic cells [83, 84]. The *PNOG* gene encodes the precursor for biologically active peptides, such as nociceptin, which have been related to several physiological roles in the central nervous system. Additionally, a study suggested

that nociceptin plays a key function in meiosis during spermatogenesis [85].

Conclusion

In this study we integrated WssGWAS and CNV analyses to improve the investigation of genetic factors determining number born alive in Duroc pigs. Our study was the first to provide a map of 425 CNV regions in the pig genome, which is a substantial source of information for further studies that aim to explore the association between reproductive traits and CNV regions. The overlapping regions between WssGWAS and CNVR analyses harbor important causative variants related to pig reproductive traits based on their critical roles in fertilization, development of gametes and embryos, which may be valuable for additional validation and consideration in future selection programs aiming to improve number of piglets born alive and other reproductive traits.

Methods

Phenotype and pedigree information

The phenotypic information was collected by Smithfield Premium Genetics from five farms of Duroc pigs. Number of piglets born alive was recorded from sows born between 2008 and 2017. A total of 39,427 records from 13,845 sows spanning 1 to 12 parities were used, and number of piglets born alive in each parity ranged from 1 and 19. Animals were grouped into nine contemporary groups which were formed by concatenating farm, month and year of farrowing. Pedigree information was available for 772,779 animals.

SNP genotyping and quality control

A total of 3892 DNA samples were genotyped using the GeneSeek® Genomic Profiler Porcine HD (<https://support.illumina.com/downloads/geneseek-ggp-porcine-hd-product-files.html>) which contained 68,528 SNPs across 18 autosomes and two sex chromosomes. Aiming to eliminate poor-quality DNA samples and decrease false-positive CNVs, only the samples with a call rate greater than 98% and call frequency greater than 90% were retained. The SNPs mapped to the sex chromosomes and those not mapped to any of the chromosomes were discarded. A total of 3520 samples genotyped for 57,962 SNPs remained after quality control for CNV and WssGWAS analyses.

CNV detection and statistical analysis

Individual-based CNV detection was conducted using PennCNV software [86] based on a hidden Markov model, which is widely used for detecting CNV based on SNP array data due its relatively low false-positive rate [87].

Multiple sources of information were used simultaneously for obtaining accurate CNV detections such as distance between SNPs, log R ratio (LRR), population frequency of the B allele (PBF), and B allele frequency

(BAF). The LRR and BAF measures were automatically computed by GenomeStudio software v2.0 (Illumina, Inc., USA) from the signal intensity files of the SNP data. The PBF file was calculated from the signal files using the compile_pbf.pl routine present in the PennCNV software [86]. In addition, we performed a wave adjustment procedure for genomic waves due to guanine-cytosine content effect applying the gcmol option in the PennCNV software to eliminated false positive CNVs detected from the differentiating signal intensities generated by genomic waves [88, 89]. The porcine gcmol was generated by calculating the guanine-cytosine content in genomic regions surrounding each SNP (500 kb each side).

A quality control of signal intensity was performed to reduce false-positive CNVs originated from poor-quality DNA and to increase the confidence in CNV identification, which included a BAF drift < 0.01, standard deviation of LRR < 0.30 and/or GC wave factor < 0.05 (after genomic waves were corrected by guanine-cytosine content) to generate raw CNV calls. The CNVs identified in only one sample and with less than three consecutive SNPs were discarded.

In order to eliminate inconsistent calling of CNV boundaries, CNV regions (CNVR) were inferred through concatenation of filtered individual CNVs identified in more than one animal. Regions with less than 0.5% allele frequency and with very low density of overlapping (recurrence < 0.1) were discarded for a more precise definition of CNVR.

Weighted single-step genome-wide association study (WssGWAS)

Variance components for number of piglets born alive were estimated by AIREMLF90 software [90], which uses the Average Information Restricted Maximum Likelihood method to estimates variance components as well as solutions for fixed and random effects. The single-trait model included contemporary group as fixed effect, random animal genetic effect (containing inbreeding), permanent environmental effect, and the residual effect. In matrix notation, the model is described as:

$$y = Xb + Wa + Kpe + e$$

where **y** is the vector of phenotypic records; **b** is the vector of fixed effect of contemporary groups; **a** is the vector of direct additive genetic effects, **pe** is the vector of permanent environmental effects; and **X**, **W**, and **K** are the incidence matrices for the effects contained in **b**, **a**, and **pe**, respectively. Narrow sense heritability was estimated as $h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_{pe}^2 + \sigma_e^2}$, where σ_i^2 is the variance of the *i*-th effect.

The same animal model as described previously was used to estimate the genomic breeding values using the ssGBLUP (single-step genomic BLUP) approach [91], which combines genomic and pedigree relationships into

a realized relationship matrix (**H**). Therefore, the difference between the regular BLUP and ssGBLUP is that the inverse of the pedigree relationship matrix (**A**⁻¹) is replaced by **H**⁻¹, which is represented as follows:

$$H^{-1} = A^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & G^{-1} - A_{22}^{-1} \end{bmatrix}$$

where **G**⁻¹ is the inverse of the genomic relationship matrix and **A**₂₂⁻¹ is the inverse pedigree relationship matrix for genotyped animals. The **G** matrix was constructed as in [30]:

$$G = \frac{ZDZ'}{2 \sum p_i(1-p_i)},$$

where **Z** is a matrix of genotypes centered by twice the current allele frequencies of each SNP (*p*); *i* is the *i*th locus; **D** is a diagonal matrix of weights (variances) for SNP, which is an identity matrix for the regular ssGBLUP. To avoid singularity problems, **G** was blended with 5% of **A**₂₂.

After genomic breeding values (GEBV) were estimated by ssGBLUP, they were back solved to obtain SNP effects as described by [33]:

$$\hat{u} = \lambda DZ'G^{-1}\hat{a}_g$$

where \hat{a}_g is GEBV for genotyped animals; λ is the ratio of SNP to additive genetic variances ($\frac{\sigma_u^2}{\sigma_a^2} = \frac{1}{\sum_{i=1}^M 2p_i(1-p_i)}$).

The weight for each SNP was calculated based on SNP effects as follows [33]:

$$d_i = \hat{u}_i^2 2p_i(1-p_i)$$

where d_i is the weight for the *i*-th SNP.

The WssGWAS is an iterative procedure that involves the following steps [33]: (i) set **D** = **I**; (ii) construct **G** matrix as described by [30]; (iii) estimate GEBVs for all animals using ssGBLUP; (iv) estimate the SNP effect; (v) estimate weight for each SNP individually; (vi) normalize **D** to maintain the additive genetic variance constant; (vii) iterate from step ii.

The analyses were performed using BLUPF90 software [92] and the results were obtained as the percentage of additive genetic variance explained by 1 Mb sliding SNP-windows. The percentage of the additive genetic variance explained by *i*th window was calculated as in [34]:

$$\frac{var(a_i)}{\sigma_a^2} * 100 = \frac{var\left(\sum_{j=1}^n z_j \hat{u}_j\right)}{\sigma_a^2} * 100$$

where a_i is genetic value of the *i*-th region consisting of 1 Mb window length physical size, σ_a^2 is the total genetic variance, z_j is vector of genotype of the *j*-th SNP for all

animals, \hat{u}_j is SNP effect of the j -th SNP within the i -th region, n is the number of SNP in a window of 1 Mb.

Gene annotation and enrichment analysis

The genomic regions exhibiting more than 1% of the additive genetic variance were prospected for possible QTL related to number of piglets born alive. The gene content of the windows was identified using the Ensembl Biomart tool [93]. The search for biologically relevant functions was performed with Panther v.13.1 database [94] selecting a 500 Kb window around a significant region (upstream and downstream) and using Sscrofa10.2 assembly as genome reference. P -values generated by Panther were Bonferroni-corrected for the number of conducted comparisons.

Additional files

Additional file 1: Description of 425 CNVRs detected in the porcine genome. (DOCX 75 kb)

Additional file 2: Gene ontology biological processes revealed by Panther analysis. (DOCX 40 kb)

Abbreviations

BAF: Ballele frequency; BLUP: Best linear unbiased prediction; CNV: Copy number variation; CNVR: Copy number variation region; GO: Gene ontology; GWAS: Genome-wide association study; LRR: Log R ratio; NGS: Next generation sequencing; PBF: Population frequency of the B allele; QTL: Quantitative trait loci; SNP: Single nucleotide polymorphism; SSC: *Sus scrofa* chromosome; ssGBLUP: Single-step genomic best linear unbiased prediction; ssGWAS: Single-step genomic association study; WssGWAS: Weighted single-step approach for genome-wide association study

Acknowledgements

The authors gratefully acknowledge Smithfield Premium Genetics for making this study possible by providing the data and support.

Funding

No funding was obtained for this study.

Availability of data and materials

The data that support the findings of this study are available from Smithfield Premium Genetics (Rose Hill, NC) but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of Smithfield Premium Genetics. A request to Smithfield Premium Genetics for accessing data may be sent to Kent Gray, Director of Genetic Research and Development (kgray@smithfield.com).

Authors' contributions

Conceptualization: YH, KG, DALL. Experimental design: NBS, DALL. Data curation: NBS, RMOS, BOF, YM. Statistical analysis: NBS, RMOS, BOF, YM, DALL. Results interpretation: NBS, DALL. Manuscript preparation: NBS. Manuscript review and editing: RMOS, BOF, YH, DALL. All authors contributed, read and approved the final manuscript.

Ethics approval and consent to participate

Animal Care and Use Committee approval was not obtained for this study because the data were from an existing database provided by Smithfield Premium Genetics (Rose Hill, NC). We did not collect any new samples for this study.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

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

Author details

¹Department of Exact Science, School of Agricultural and Veterinarian Sciences (FCAV), Sao Paulo State University (UNESP), Jaboticabal, SP 14884-900, Brazil. ²Department of Animal and Dairy Science, University of Georgia, Athens, GA, USA. ³National Center for Cool and Cold Water Aquaculture (NCCCW), Agricultural Research Service, United States Department of Agriculture, Kearneysville, WV, USA. ⁴Department of Animal Science, University of Connecticut, Storrs-Mansfield, CT, USA. ⁵Smithfield Premium Genetics Group, Rose Hill, NC, USA.

Received: 12 July 2018 Accepted: 11 April 2019

Published online: 27 April 2019

References

- Bergfelder-Drüing S, Grosse-Brinkhaus C, Lind B, Erbe M, Schellander K, Simianer H, Tholen E. A genome-wide association study in large white and landrace pig populations for number piglets born alive. *PLoS One*. 2015; 10(3):e0117468. <https://doi.org/10.1371/journal.pone.0117468>.
- Wu P, Yang Q, Wang K, Zhou J, Ma J, Tang Q, Jin L, Xiao W, Jiang A, Jiang Y, Zhu L, Li X, Tang G. Single step genome-wide association studies based on genotyping by sequence data reveals novel loci for the litter traits of domestic pigs. *Genomics*. 2018;110(3):171–9. <https://doi.org/10.1016/j.ygeno.2017.09.009>.
- Wang Y, Ding X, Tan Z, Xing K, Yang T, Wang Y, Sun D, Wang C. Genome-wide association study for reproductive traits in a large white pig population. *Anim Genet*. 2018;49(2):127–31. <https://doi.org/10.1111/age.12638>.
- García-Ruiz A, Cole JB, VanRaden PM, Wiggans GR, Ruiz-López FJ, Van Tassel CP. Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection. *Proc Natl Acad Sci*. 2016;113(28):E3995–4004. <https://doi.org/10.1073/pnas.1519061113>.
- Pig Quantitative Trait Locus Database (Pig QTLdb), Release 35. <https://www.animalgenome.org/cgi-bin/QTLdb/SS/index>. Accessed 21 Jun 2018.
- Onteru SK, Fan B, Du ZQ, Garrick DJ, Stalder KJ, Rothschild MF. A whole-genome association study for pig reproductive traits. *Anim Genet*. 2011; 43(1):18–26. <https://doi.org/10.1111/j.1365-2052.2011.02213.x>.
- Hernandez SC, Finlayson HA, Ashworth CJ, Haley CS, Archibald AL. A genome-wide linkage analysis for reproductive traits in F2 large white × Meishan cross gilts. *Anim Genet*. 2014;45(2):191–7. <https://doi.org/10.1111/age.12123>.
- Guo X, Su G, Christensen OF, Janss L, Lund MS. Genome-wide association analyses using a Bayesian approach for litter size and piglet mortality in Danish landrace and Yorkshire pigs. *BMC Genomics*. 2016;17:468. <https://doi.org/10.1186/s12864-016-2806-z>.
- Laliotis GP, Marantidis A, Avdi M. Association of BF, RBP4, and ESR2 genotypes with litter size in an autochthonous pig population. *Anim Biotechnol*. 2017; 28(2):138–43. <https://doi.org/10.1080/10495398.2016.1242490>.
- An SM, Hwang JH, Kwon S, Yu GE, Park DH, Kang DG, Kim TW, Park HC, Ha J, Kim CW. Effect of single nucleotide polymorphisms in IGFBP2 and IGFBP3 genes on litter size traits in berkshire pigs. *Anim Biotechnol*. 2018;29(4):301–8. <https://doi.org/10.1080/10495398.2017.1395345>.
- Ritchie MD, Holzinger ER, Li R, Pendergrass SA, Kim D. Methods of integrating data to uncover genotype-phenotype interactions. *Nat Rev Genet*. 2015;16(2):85–97. <https://doi.org/10.1038/nrg3868>.
- Wang J, Wang H, Jiang J, Kang H, Feng X, Zhang Q, Liu JF. Identification of genome-wide copy number variations among diverse pig breeds using SNP genotyping arrays. *PLoS One*. 2013;8(7):e68683. <https://doi.org/10.1371/journal.pone.0068683>.
- Santana MHA, Junior GAO, Cesar AS, Freua MC, Gomes RC, Silva SL, Leme PR, Fukumasu H, Carvalho ME, Ventura RV, Coutinho LL, Kadarmideen HN, Ferraz JB. Copy number variations and genome-wide associations reveal putative genes and metabolic pathways involved with the feed conversion ratio in beef cattle. *J Appl Genet*. 2016;57(4):495–504. <https://doi.org/10.1007/s13353-016-0344-7>.
- Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W, Cho EK, Dallaire S, Freeman JL, González JR, Gratacòs M, Huang J, Kalaitzopoulos D, Komura D, JR MD, Marshall CR,

- Mei R, Montgomery L, Nishimura K, Okamura K, Shen F, Somerville MJ, Tchinda J, Valsesia A, Woodwark C, Yang F, Zhang J, Zerjal T, Zhang J, Armengol L, Conrad DF, Estivill X, Tyler-Smith C, Carter NP, Aburatani H, Lee C, Jones KW, Scherer SW, Hurles ME. Global variation in copy number in the human genome. *Nature*. 2006;444(7118):444–54.
15. Giuffra E, Törnsten A, Marklund S, Bongcam-Rudloff E, Chardon P, Kijas JM, Anderson SI, Archibald AL, Andersson L. A large duplication associated with dominant white color in pigs originated by homologous recombination between LINE elements flanking KIT. *Mamm Genome*. 2002;13(10):569–77.
 16. Ramayo-Caldas Y, Castelló A, Pena RN, Alves E, Mercadé A, Souza CA, Fernández AI, Perez-Enciso M, Folch JM. Copy number variation in the porcine genome inferred from a 60 k SNP BeadChip. *BMC Genomics*. 2010;11:593. <https://doi.org/10.1186/1471-2164-11-593>.
 17. Chen C, Qiao R, Wei R, Guo Y, Ai H, Ma J, Ren J, Huang L. A comprehensive survey of copy number variation in 18 diverse pig populations and identification of candidate copy number variable genes associated with complex traits. *BMC Genomics*. 2012;13:733. <https://doi.org/10.1186/1471-2164-13-733>.
 18. Li Y, Mei S, Zhang X, Peng X, Liu G, Tao H, Wu H, Jiang S, Xiong Y, Li F. Identification of genome-wide copy number variations among diverse pig breeds by array CGH. *BMC Genomics*. 2012;13:725. <https://doi.org/10.1186/1471-2164-13-725>.
 19. Wang J, Jiang J, Fu W, Jiang L, Ding X, Liu JF, Zhang Q. A genome-wide detection of copy number variations using SNP genotyping arrays in swine. *BMC Genomics*. 2012;13:273. <https://doi.org/10.1186/1471-2164-13-273>.
 20. Paudel Y, Madsen O, Megens HJ, Frantz LA, Bosse M, Bastiaansen JW, Crooijmans RP, Groenen MA. Evolutionary dynamics of copy number variation in pig genomes in the context of adaptation and domestication. *BMC Genomics*. 2013;14:449. <https://doi.org/10.1186/1471-2164-14-449>.
 21. Jiang J, Wang J, Wang H, Zhang Y, Kang H, Feng X, Wang J, Yin Z, Bao W, Zhang Q, Liu JF. Global copy number analyses by next generation sequencing provide insight into pig genome variation. *BMC Genomics*. 2014;15:593. <https://doi.org/10.1186/1471-2164-15-593>.
 22. Wang Y, Tang Z, Sun Y, Wang H, Wang C, Yu S, Liu J, Zhang Y, Fan B, Li K, Liu B. Analysis of genome-wide copy number variations in Chinese indigenous and western pig breeds by 60 K SNP genotyping arrays. *PLoS One*. 2014;9(9):e106780. <https://doi.org/10.1371/journal.pone.0106780>.
 23. Wang J, Jiang J, Wang H, Kang H, Zhang Q, Liu JF. Enhancing genome-wide copy number variation identification by high density array CGH using diverse resources of pig breeds. *PLoS One*. 2014;9(1):e87571. <https://doi.org/10.1371/journal.pone.0087571>.
 24. Wang H, Wang C, Yang K, Liu J, Zhang Y, Wang Y, Xu X, Michal JJ, Jiang Z, Liu B. Genome wide distributions and functional characterization of copy number variations between Chinese and Western pigs. *PLoS One*. 2015;10(7):e0131522. <https://doi.org/10.1371/journal.pone.0131522>.
 25. Dong K, Pu Y, Yao N, Shu G, Liu X, He X, Zhao Q, Guan W, Ma Y. Copy number variation detection using SNP genotyping arrays in three Chinese pig breeds. *Anim Genet*. 2015;46(2):101–9. <https://doi.org/10.1111/age.12247>.
 26. Paudel Y, Madsen O, Megens HJ, Frantz LA, Bosse M, Crooijmans RP, Groenen MA. Copy number variation in the speciation of pigs: a possible prominent role for olfactory receptors. *BMC Genomics*. 2015;16:330. <https://doi.org/10.1186/s12864-015-1449-9>.
 27. Long Y, Su Y, Ai H, Zhang Z, Yang B, Ruan G, Xiao S, Liao X, Ren J, Huang L, Ding N. A genome-wide association study of copy number variations with umbilical hernia in swine. *Anim Genet*. 2016;47(3):298–305. <https://doi.org/10.1111/age.12402>.
 28. Xie J, Li R, Li S, Ran X, Wang J, Jiang J, Zhao P. Identification of copy number variations in Xiang and Kele pigs. *PLoS One*. 2016;11(2):e0148565. <https://doi.org/10.1371/journal.pone.0148565>.
 29. Misztal I, Legarra A, Aguilar I. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. *J Dairy Sci*. 2009;92:4648–55. <https://doi.org/10.3168/jds.2009-2064>.
 30. VanRaden PM. Efficient methods to compute genomic predictions. *J Dairy Sci*. 2008;91(11):4414–23. <https://doi.org/10.3168/jds.2007-0980>.
 31. Garrick DJ, Taylor JF, Fernando RL. Deregressing estimated breeding values and weighting information for genomic regression analyses. *Genet Sel Evol*. 2009;41:55. <https://doi.org/10.1186/1297-9686-41-55>.
 32. Legarra A, Christensen OF, Aguilar I, Misztal I. Single step, a general approach for genomic selection. *Livest Sci*. 2014;166:54–65. <https://doi.org/10.1016/j.livsci.2014.04.029>.
 33. Wang H, Misztal I, Aguilar I, Legarra A, Muir WM. Genome-wide association mapping including phenotypes from relatives without genotypes. *Genet Res*. 2012;94(2):73–83. <https://doi.org/10.1017/S0016672312000274>.
 34. Wang H, Misztal I, Aguilar I, Legarra A, Fernando RL, Vitezica Z, Okimoto R, Wing T, Hawken R, Muir WM. Genome-wide association mapping including phenotypes from relatives without genotypes in a single-step (ssGWAS) for 6-week body weight in broiler chickens. *Front Genet*. 2014;5:134. <https://doi.org/10.3389/fgene.2014.00134>.
 35. Chen CY, Misztal I, Aguilar I, Legarra A, Muir WM. Effect of different genomic relationship matrices on accuracy and scale. *J Anim Sci*. 2011;99:2673–9. <https://doi.org/10.2527/jas.2010-3555>.
 36. Forni S, Aguilar I, Misztal I. Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. *Genet Sel Evol*. 2011;43:1–7. <https://doi.org/10.1186/1297-9686-43-1>.
 37. Fernández AI, Barragán C, Fernández A, Rodríguez MC, Villanueva B. Copy number variants in a highly inbred Iberian porcine strain. *Anim Genet*. 2014;45(3):357–66. <https://doi.org/10.1111/age.12137>.
 38. Zhou LS, Li J, Yang J, Liu CL, Xie XH, He YN, Liu XX, Xin WS, Zhang WC, Ren J, Ma JW, Huang LS. Genome-wide mapping of copy number variations in commercial hybrid pigs using a high-density SNP genotyping array. *Genetika*. 2016;52(1):97–105. <https://doi.org/10.1134/S1022795415120145>.
 39. Schiavo G, Dolezal MA, Scotti E, Bertolini F, Calò DG, Galimberti G, Russo V, Fontanesi L. Copy number variants in Italian large white pigs detected using high-density single nucleotide polymorphisms and their association with back fat thickness. *Anim Genet*. 2014;45(5):745–9. <https://doi.org/10.1111/age.12180>.
 40. Fadista J, Thomsen B, Holm LE, Bendixen C. Copy number variation in the bovine genome. *BMC Genomics*. 2010;11:284. <https://doi.org/10.1186/1471-2164-11-284>.
 41. Schrider DR, Hahn MW. Gene copy-number polymorphism in nature. *Proc Biol Sci*. 2010;277:3213–21. <https://doi.org/10.1098/rspb.2010.1180>.
 42. Wang L, Liu X, Zhang L, Yan H, Luo W, Liang J, Cheng D, Chen S, Ma X, Song X, Zhao K, Wang L. Genome-wide copy number variations inferred from SNP genotyping arrays using a large white and Minzhu intercross population. *PLoS One*. 2013;8(10):e74879. <https://doi.org/10.1371/journal.pone.0074879>.
 43. Fadista J, Nygaard M, Holm LE, Thomsen B, Bendixen C. A snapshot of CNVs in the pig genome. *PLoS One*. 2008;3(12):e3916. <https://doi.org/10.1371/journal.pone.0003916>.
 44. Winchester L, Yau C, Ragoussis J. Comparing CNV detection methods for SNP arrays. *Brief Funct Genomic Proteomic*. 2009;8(5):353–66. <https://doi.org/10.1093/bfpg/elp017>.
 45. Clop A, Vidal O, Amills M. Copy number variation in the genomes of domestic animals. *Anim Genet*. 2012;43(5):503–17. <https://doi.org/10.1111/j.1365-2052.2012.02317.x>.
 46. Liu GE, Bickhart DM. Copy number variation in the cattle genome. *Funct Integr Genomics*. 2012;12(4):609–24. <https://doi.org/10.1007/s10142-012-0289-9>.
 47. Peng YB, Fan B, Han XL, Xu XW, Rothschild MF, Yerle M, Liu B. Molecular characterization of the porcine JHD1M1A gene associated with average daily gain: evaluation its role in skeletal muscle development and growth. *Mol Biol Rep*. 2011;38(7):4697–704. <https://doi.org/10.1007/s11033-010-0604-2>.
 48. Berman Y, North KN. A gene for speed: the emerging role of alpha-actinin-3 in muscle metabolism. *Physiology*. 2010;25(4):250–9. <https://doi.org/10.1152/physiol.00008.2010>.
 49. Blom M, Reis K, Heldin J, Kreuger J, Aspenström P. The atypical rho GTPase RhoD is a regulator of actin cytoskeleton dynamics and directed cell migration. *Exp Cell Res*. 2017;352(2):255–64. <https://doi.org/10.1016/j.yexcr.2017.02.013>.
 50. Brockmann C, Huarte J, Dugina V, Challet L, Rey E, Conne B, Swetloff A, Nef S, Chaponnier C, Vassalli JD. Beta- and gamma-cytoplasmic actins are required for meiosis in mouse oocytes. *Biol Reprod*. 2011;85(5):1025–39. <https://doi.org/10.1095/biolreprod.111.091736>.
 51. Bunnell TM, Burbach BJ, Shimizu Y, Ervasti JM. β -Actin specifically controls cell growth, migration, and the G-actin pool. *Mol Biol Cell*. 2011;22(21):4047–58. <https://doi.org/10.1091/mbc.E11-06-0582>.
 52. Pasek RC, Malarkey E, Berbari NF, Sharma N, Kesterson RA, Tres LL, Kierszenbaum AL, Yoder BK. Coiled-coil domain containing 42 (Ccdc42) is necessary for proper sperm development and male fertility in the mouse. *Dev Biol*. 2016;412(2):208–18. <https://doi.org/10.1016/j.ydbio.2016.01.042>.
 53. Vanhaesebroeck B, Guillermet-Guibert J, Graupera M, Bilanges B. The emerging mechanisms of isoform-specific PI3K signalling. *Nat Rev Mol Cell Biol*. 2010;11(5):329–41. <https://doi.org/10.1038/nrm2882>.

54. Klempner SJ, Myers AP, Cantley LC. What a tangled web we weave: emerging resistance mechanisms to inhibition of the phosphoinositide 3-kinase pathway. *Cancer Discov.* 2013;3(12):1345–54. <https://doi.org/10.1158/2159-8290.CD-13-0063>.
55. Bradford D, Cole SJ, Cooper HM. Netrin-1: diversity in development. *Int J Biochem Cell Biol.* 2009;41(3):487–93. <https://doi.org/10.1016/j.biocel.2008.03.014>.
56. Wang Q, Zhu J, Zou L, Yang Y. Role of axonal guidance factor netrin-1 in human placental vascular growth. *J Huazhong Univ Sci Technolog Med Sci.* 2011;31(2):246–50. <https://doi.org/10.1007/s11596-011-0261-2>.
57. Mediero A, Ramkhalawon B, Perez-Aso M, Moore KJ, Cronstein BN. Netrin-1 is a critical autocrine/paracrine factor for osteoclast differentiation. *J Bone Miner Res.* 2015;30(5):837–54. <https://doi.org/10.1002/jbmr.2421>.
58. Basini G, Cortimiglia C, Baioni L, Bussolati S, Grolli S, Ramoni R, Grasselli F. The axonal guidance factor netrin-1 as a potential modulator of swine follicular function. *Mol Cell Endocrinol.* 2011;331(1):41–8. <https://doi.org/10.1016/j.mce.2010.08.001>.
59. Miyazono K, Kamiya Y, Morikawa M. Bone morphogenetic protein receptors and signal transduction. *J Biochem.* 2010;147(1):35–51. <https://doi.org/10.1093/jb/mvp148>.
60. Peng H, Usas A, Olshanski A, Ho AM, Gearhart B, Cooper GM, Huard J. VEGF improves, whereas sFlt1 inhibits, BMP2-induced bone formation and bone healing through modulation of angiogenesis. *J Bone Miner Res.* 2005;20(11):2017–27. <https://doi.org/10.1359/JBMR.050708>.
61. Bandyopadhyay A, Tsuji K, Cox K, Harfe BD, Rosen V, Tabin CJ. Genetic analysis of the roles of BMP2, BMP4, and BMP7 in limb patterning and skeletogenesis. *PLoS Genet.* 2006;2(12):e216. <https://doi.org/10.1371/journal.pgen.0020216>.
62. Bae JS, Gutierrez S, Narla R, Pratap J, Devados R, van Wijnen AJ, Stein JL, Stein GS, Lian JB, Javed A. Reconstitution of Runx2/Cbfa1-null cells identifies a requirement for BMP2 signaling through a Runx2 functional domain during osteoblast differentiation. *J Cell Biochem.* 2007;100(2):434–49. <https://doi.org/10.1002/jcb.21039>.
63. Shah TA, Zhu Y, Shaikh NN, Harris MA, Harris SE, Rogers MB. Characterization of new bone morphogenetic protein (Bmp)-2 regulatory alleles. *Genesis.* 2017;55(7). <https://doi.org/10.1002/dvg.23035>.
64. Zhang H, Bradley A. Mice deficient for BMP2 are nonviable and have defects in amnion/chorion and cardiac development. *Development.* 1996;122(10):2977–86.
65. Lee KY, Jeong JW, Wang J, Ma L, Martin JF, Tsai SY, Lydon JP, DeMayo FJ. Bmp2 is critical for the murine uterine decidual response. *Mol Cell Biol.* 2007;27(15):5468–78.
66. Yan W, Si Y, Slaymaker S, Li J, Zheng H, Young DL, Aslanian A, Saunders L, Verdin E, Charo IF. Zmynd15 encodes a histone deacetylase-dependent transcriptional repressor essential for spermiogenesis and male fertility. *J Biol Chem.* 2010;285(41):31418–26. <https://doi.org/10.1074/jbc.M110.116418>.
67. Yang J, Medvedev S, Yu J, Tang LC, Agno JE, Matzuk MM, Schultz RM, Hecht NB. Absence of the DNA–RNA-binding protein MSY2 results in male and female infertility. *Proc Natl Acad Sci.* 2005;102(16):5755–60.
68. Medvedev S, Pan H, Schultz RM. Absence of MSY2 in mouse oocytes perturbs oocyte growth and maturation, RNA stability, and the transcriptome. *Biol Reprod.* 2011;85(3):575–83. <https://doi.org/10.1095/biolreprod.111.091710>.
69. Kleene KC. Position-dependent interactions of Y-box protein 2 (YBX2) with mRNA enable mRNA storage in round spermatids by repressing mRNA translation and blocking translation-dependent mRNA decay. *Mol Reprod Dev.* 2016;83(3):190–207. <https://doi.org/10.1002/mrd.22616>.
70. Pausch H, Kölle S, Wurms C, Schwarzenbacher H, Emmerling R, Jansen S, Trottmann M, Fuerst C, Götz KU, Fries R. A nonsense mutation in TMEM95 encoding a nondescript transmembrane protein causes idiopathic male subfertility in cattle. *PLoS Genet.* 2014;10(1):e1004044. <https://doi.org/10.1371/journal.pgen.1004044>.
71. Yoo H, Son D, Jang YJ, Hong K. Indispensable role for mouse ELP3 in embryonic stem cell maintenance and early development. *Biochem Biophys Res Commun.* 2016;478(2):631–6. <https://doi.org/10.1016/j.bbrc.2016.07.120>.
72. Meistrich ML, Mohapatra B, Shirley CR, Zhao M. Roles of transition nuclear proteins in spermiogenesis. *Chromosoma.* 2003;111(8):483–8.
73. Santi CM, Martínez-López P, de la Vega-Beltrán JL, Butler A, Alisio A, Darszon A, Salkoff L. The SLO3 sperm-specific potassium channel plays a vital role in male fertility. *FEBS Lett.* 2010;584(5):1041–6. <https://doi.org/10.1016/j.febslet.2010.02.005>.
74. Donato R, Cannon BR, Sorci G, Riuzzi F, Hsu K, Weber DJ, Geczy CL. Functions of S100 proteins. *Curr Mol Med.* 2013;13(1):24–57.
75. Munch EM, Sparks AE, Gonzalez Bosquet J, Christenson LK, Devor EJ, Van Voorhis BJ. Differentially expressed genes in preimplantation human embryos: potential candidate genes for blastocyst formation and implantation. *J Assist Reprod Genet.* 2016;23(8):1017–25. <https://doi.org/10.1007/s10815-016-0745-x>.
76. He J, Stewart K, Kinnell HL, Anderson RA, Childs AJ. A developmental stage-specific switch from DAZL to BOLL occurs during fetal oogenesis in humans, but not mice. *PLoS One.* 2013;8(9):e73996. <https://doi.org/10.1371/journal.pone.0073996>.
77. Chen P, Wang X, Xu C, Xiao H, Zhang WH, Wang XH, Zhang XH. Association of polymorphisms of A260G and A386G in DAZL gene with male infertility: a meta-analysis and systemic review. *Asian J Androl.* 2016;18(1):96–101. <https://doi.org/10.4103/1008-682X.153542>.
78. Rosario R, Adams IR, Anderson RA. Is there a role for DAZL in human female fertility? *Mol Hum Reprod.* 2016;22(6):377–83. <https://doi.org/10.1093/molehr/gaw024>.
79. Ornitz DM, Itoh N. The fibroblast growth factor signaling pathway. *Wiley Interdiscip Rev Dev Biol.* 2015;4(3):215–66. <https://doi.org/10.1002/wdev.176>.
80. Tepekoy F, Akkoyunlu G, Demir R. The role of Wnt signaling members in the uterus and embryo during pre-implantation and implantation. *J Assist Reprod Genet.* 2015;32(3):337–46. <https://doi.org/10.1007/s10815-014-0409-7>.
81. Tanaka TS, Jaradat SA, Lim MK, Kargul GJ, Wang X, Grahovac MJ, Pantano S, Sano Y, Piao Y, Nagaraja R, Doi H, Wood WH 3rd, Becker KG, Ko MS. Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray. *Proc Natl Acad Sci.* 2000;97(16):9127–32.
82. Heldin, C-H. Protein tyrosine kinase receptor signaling overview. In: *Handbook of Cell Signaling*, 2nd edition, Bradshaw RA and Dennis EA, eds. Academic Press, San Diego, chapter: 59, pps. 419–426, 2010. doi: <https://doi.org/10.1016/B978-0-12-374145-5.00059-0>.
83. Logue JS, Whiting JL, Tunquist B, Sacks DB, Langeberg LK, Wordeman L, Scott JD. AKAP220 protein organizes signaling elements that impact cell migration. *J Biol Chem.* 2011;286(45):39269–81. <https://doi.org/10.1074/jbc.M111.277756>.
84. Reinton N, Collas P, Haugen TB, Skålhegg BS, Hansson V, Jahnesen T, Taskén K. Localization of a novel human A-kinase-anchoring protein, hAKAP220, during spermatogenesis. *Dev Biol.* 2000;223(1):194–204.
85. Eto K. Nociceptin and meiosis during spermatogenesis in postnatal testes. *Vitam Horm.* 2015;97:167–86. <https://doi.org/10.1016/bs.vh.2014.10.003>.
86. Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SF, Hakonarson H, Bucan M. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res.* 2007;17(11):1665–74.
87. Dellinger AE, Saw SM, Goh LK, Seielstad M, Young TL, Li YJ. Comparative analyses of seven algorithms for copy number variant identification from single nucleotide polymorphism arrays. *Nucleic Acids Res.* 2010;38(9):e105. <https://doi.org/10.1093/nar/gkq040>.
88. Hou Y, Liu GE, Bickhart DM, Cardone MF, Wang K, Kim ES, Matukumalli LK, Ventura M, Song J, VanRaden PM. Genomic characteristics of cattle copy number variations. *BMC Genomics.* 2011;12:127. <https://doi.org/10.1186/1471-2164-12-127>.
89. Liu J, Zhang L, Xu L, Ren H, Lu J, Zhang X, Zhang S, Zhou X, Wei C, Zhao F, Du L. Analysis of copy number variations in the sheep genome using 50K SNP BeadChip array. *BMC Genomics.* 2013;14:229. <https://doi.org/10.1186/1471-2164-14-229>.
90. Misztal I, Tsuruta S, Strabel T, Auvray B, Druet T, Lee DH, Ducrocq V, Elsen JM, Minvielle F. BLUPF90 and related programs (BGF90). In: proceedings of the 7th world congress on genetics applied to livestock production. Montpellier (France), 19–23 august 2002, communication no. 28-07.; 2002.
91. Aguilar I, Misztal I, Johnson DL, Legarra A, Tsuruta S, Lawlor TJ. Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J Dairy Sci.* 2010;93(2):743–52. <https://doi.org/10.3168/jds.2009-2730>.
92. Misztal I, Tsuruta S, Lourenco D, Masuda Y, Aguilar I, Legarra A, Vitezica Z. Manual for BLUPF90 family of programs. Athens: University of Georgia, 2016. 142p. Available: http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all5.pdf.
93. Haidar S, Ballester B, Smedley D, Zhang J, Rice P, Kasprzyk A. BioMart Central Portal—unified access to biological data. *Nucleic Acids Res.* 2009;37:W23–7. <https://doi.org/10.1093/nar/gkp265>.
94. Mi H, Huang X, Muruganujan A, Tang H, Mills C, Kang D, Thomas PD. PANTHER version 11: expanded annotation data from gene ontology and Reactome pathways, and data analysis tool enhancements. *Nucleic Acids Res.* 2017;45(D1):D183–9. <https://doi.org/10.1093/nar/gkw1138>.