

RESEARCH ARTICLE

# Detection of genomic structural variations in Guizhou indigenous pigs and the comparison with other breeds

Chang Liu<sup>1</sup>, Xueqin Ran<sup>1\*</sup>, Jiafu Wang<sup>1,2\*</sup>, Sheng Li<sup>1</sup>, Jianfeng Liu<sup>3</sup>

**1** Institute of Agro-Bioengineering/College of Animal Science, Guizhou University, Guiyang, China, **2** Tongren University, Tongren, China, **3** College of Animal Science and Technology, China Agricultural University, Beijing, China

\* [jfwang@gzu.edu.cn](mailto:jfwang@gzu.edu.cn) (JW); [xqran@gzu.edu.cn](mailto:xqran@gzu.edu.cn) (XR)



## Abstract

Genomic structural variation (SV) is noticed for the contribution to genetic diversity and phenotypic changes. Guizhou indigenous pig (GZP) has been raised for hundreds of years with many special characteristics. The present paper aimed to uncover the influence of SV on gene polymorphism and the genetic mechanisms of phenotypic traits for GZP. Eighteen GZPs were chosen for resequencing by Illumina sequencing platform. The confident SVs of GZP were called out by both programs of pindel and softSV simultaneously and compared with the SVs deduced from the genomic data of European pig (EUP) and the native pig outside of Guizhou, China (NPOG). A total of 39,166 SVs were detected and covered 27.37 Mb of pig genome. All of 76 SVs were confirmed in GZP pig population by PCR method. The SVs numbers in NPOG and GZP were about 1.8 to 1.9 times higher than that in EUP. And a SV hotspot was found out from the 20 Mb of chromosome X of GZP, which harbored 29 genes and focused on histone modification. More than half of SVs was positioned in the intergenic regions and about one third of SVs in the introns of genes. And we found that SVs tended to locate in genes produced multi-transcripts, in which a positive correlation was found out between the numbers of SV and the gene transcripts. It illustrated that the primary mode of SVs might function on the regulation of gene expression or the transcripts splicing process. A total of 1,628 protein-coding genes were disturbed by 1,956 SVs specific in GZP, in which 93 GZP-specific SV-related genes would lose their functions due to the SV interference and gathered in reproduction ability. Interestingly, the 1,628 protein-coding genes were mainly enriched in estrogen receptor binding, steroid hormone receptor binding, retinoic acid receptor binding, oxytocin signaling pathway, mTOR signaling pathway, axon guidance and cholinergic synapse pathways. It suggested that SV might be a reason for the strong adaptability and low fecundity of GZP, and 51 candidate genes would be useful for the configuration phenotype in Xiang pig breed.

## OPEN ACCESS

**Citation:** Liu C, Ran X, Wang J, Li S, Liu J (2018) Detection of genomic structural variations in Guizhou indigenous pigs and the comparison with other breeds. PLoS ONE 13(3): e0194282. <https://doi.org/10.1371/journal.pone.0194282>

**Editor:** Mingzhi Liao, Northwest A&F University, CHINA

**Received:** October 30, 2017

**Accepted:** February 28, 2018

**Published:** March 20, 2018

**Copyright:** © 2018 Liu et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** This work was supported by the National High Technology Research and Development Program of China (863 Program) [2013AA102503] (<http://program.most.gov.cn/>), National Natural Science Foundation of China (31672390, 31401091) (<http://www.nsf.gov.cn/>), the Guizhou Province "Hundred" Innovative Talents Project [2016-4012] (<http://kjt.gzst.gov.cn/>), and the Guizhou Agriculture Research program (QKHZC

[2017]2585, QKHZC[2017]2587) (<http://kjt.gzst.gov.cn>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

## Introduction

Indigenous pig breed shows great phenotypic varieties for hair and color pattern, morphology, reproduction, growth and adaptability [1]. There present seven native pig breeds in Guizhou province, China. Some of them have borne natural and artificial selection for hundreds of years, including Xiang, Kele, Qianbei black, Guanling, Luobo pig breeds, and so on. They share preponderant features including better disease resistance, strong adaptability and favorite meat quality. But many reports show that Guizhou pigs give much lower litter sizes compared with European pig breeds [2]. For example, the average litter size is 9–11 piglets in Large White breed while only 6.6–6.9 piglets in Xiang pig and 6–8 in Kele pig [3].

It has been found that the insertion or deletion in the pivotal regions of gene did change the gene structure and expression and have a link to phenotype trait in pig. A 12-bp insertion/deletion (indel) polymorphism in exon 1 of the secreted folate binding protein (sFBP) gene is confirmed to be associated with the uterine capacity, the number of corpora lutea and the litter size in gilt [4]. An insertion in 51 bp is also found out from exon in the Testis expressed 14 (TEX14) gene and causes infertility of boar [5]. It is thought that the insertion or deletion in exon regions might affect the folding and stability of the mRNA or the translation efficiency of these genes for fecundity regulation. Additionally, a 304 bp insertion in promoter region increases the expression of the mitochondrial NAD<sup>+</sup>-dependent isocitrate dehydrogenase  $\beta$  subunit (*IDH3 $\beta$* ) gene and is correlated with a higher backfat thickness of pig [6]. Two transcripts are resulted from the insertion or deletion of 574-bp spanned exon 5 and part of 3'-UTR of dopamine D2 receptor gene [7]. Additional STAT binding site is created by a 23-bp insertion in the promoter of Toll-like receptor 5 (TLR5), which is used to recognize flagellin in the flagella of gram-positive and gram-negative bacteria [8]. Both of dopamine D2 receptor and TLR5 genes are important in pathogen susceptibility or resistance patterns in animal. PRE-1, one of SINE element specific in pig genome, is detected in the intron of vertnin gene of pig with increased numbers of vertebrae [9]. If the PRE-1 presented in the 3'-untranslated regions of the porcine prolactin receptor short form, the protein expression would be downregulated [10]. However, it is far from clear in trait regulation mechanism just focused on insertion and deletion of several genes in pig.

Structural variation (SV) in genome includes insertion, duplication, deletion, translocation and inversion with length of more than 50 bp [11–13]. It is estimated that SVs have been manifested accounting for 83.6 percent of total genetic variation [14]. Studies in human showed that SVs have been associated with schizophrenia [15], cancer [16], and complex genetic disorders [17]. Next generation sequencing (NGS) technology provides a chance for SV detection to elucidate genetic complexity and variations contributing to diverse traits on the whole genome level. Plenty of SVs were identified based on the NGS data in sheep and candidate genes functionally related to energy metabolism and body size [18]. A SV hotspot spanning 35 Mb regions on the X chromosome is identified specifically in Chinese pig by NGS [19]. Similarly, a comprehensive survey of small- and intermediate- SVs constructed a single-nucleotide resolution map in Tibetan, and contribution to pig diversity and phenotypic changes [20]. Many genetic variations might still remain under cover in other native pig breeds.

Thus, we performed whole-genome resequencing to identify SVs of five Guizhou pig breeds, including Xiang, Kele, Qianbei, Guanling, Luobo pig breeds. In addition, we download the resequencing data of other pigs from China and European and analyzed simultaneously. It was looking forward to finding the specific changes in GZP genome and the relationship with their economic traits.

## Materials and methods

### Animal ethics

All animal procedures were approved by the Institutional Animal Care and Use Committee of Guizhou University and were conducted in accordance with the National Research Council Guide for the Care and Use of Laboratory Animals.

### Animal collection

Eighteen unrelated Guizhou pigs (GZPs) were utilized for resequencing. Xiang pig (XP,  $n = 6$ ) was from Congjiang county, Kele pig (KL,  $n = 3$ ) from Bijie county, Qianbei black pig (QB,  $n = 3$ ) from Zunyi county, Guanling pig (GL,  $n = 3$ ) from Anshun county, Luobo pig (LB,  $n = 3$ ) from Tongren county. Blood samples were collected from the precaval vein according to standard procedures. The information for age and farm coordinate of these five breeds was shown in [S1 Table](#).

### DNA extraction and sequencing

Genomic DNA was extracted from blood using SQ Blood DNA Kit (OMEGA, USA) and the qualified DNA was used for genome resequencing. Two paired-end libraries were constructed for each sample and the libraries were sequenced on Illumina HiSeq2500 instrument (Illumina, USA). Reference genome sequence of pig (Sscrofa 11.1) was downloaded from Ensembl ([ftp://ftp.ensembl.org/pub/release-90/fasta/sus\\_scrofa/dna/](ftp://ftp.ensembl.org/pub/release-90/fasta/sus_scrofa/dna/)). The raw sequencing reads was filtered by NGS QC Toolkit with default parameters [21]. Clean reads were mapped to the pig reference genome sequence using the Burrows-Wheeler Alignment software with default parameters [22]. SAMtools was used to convert the files in SAM format to BAM format [23]. Then, duplicate marking were removed using Picard package, and base quality recalibration was performed using the Genome Analysis ToolKit (GATK) program [24].

### Identification of SVs

Bioinformatics detection of genomic variation was performed on the eighteen BAM files by Pindel [25] and SoftSV softwares [26]. The default parameters were used for both programs. Since Pindel is not applicable for translocations and translocations inverted, we only chose SV types of deletions (DEL), insertions (INS), inversions (INV) and tandem duplications (DUP). Two standards were used to filter the raw data. Firstly, the short read appeared at least three paired-ends. Secondly, two softwares, pindel and softSV, were applying for SV calling and breakpoint prediction. Two SVs overlapped more than 25 bp were merged into one SV if both of SVs were belong to the same variation type at the same chromosome [27]. We only retained SV detected by both Pindel and SoftSV programs. Furthermore, to eliminate the gender effect on SV detection, data from the chromosome Y were excluded.

### SV validation

To evaluate the reliability of the data, 76 randomly selected SVs were validated using a PCR assay and direct sequencing method. SVs primer pairs for PCR were designed based on 500-bp upstream or downstream of the insertion/deletion breakpoints of the SVs based on the reference genome sequence using the primer3 algorithm (<http://frodo.wi.mit.edu/primer3/>) ([S2 Table](#)). The genomic DNAs were taking as /templates for validation by PCR detection. A total volume of 20  $\mu$ L was used for PCR taking 1  $\mu$ L of genomic DNA (80–120 ng/ $\mu$ L) as templates, 10  $\mu$ L of 2 $\times$ PCR Master Mix (Tiangen, Beijing), 0.4  $\mu$ L of 0.1  $\mu$ g/ $\mu$ L primers, and 8.2  $\mu$ L of ultrapure water. The PCR program was set at 94°C for 5 min; 30 cycles of 94°C for 30 s, 60°C

for 30 s and 72°C for 35 s; and a final extension of 10 min at 72°C. The PCR products were purified and then subjected to Sanger sequencing. The genotyping of six candidate genes was also performed for 284 pigs, including XP (n = 48), KL (n = 34), QB (n = 30), LB (n = 32), GL (n = 24), Large White pig (LW, n = 48), Duroc (DU, n = 24), Rongchang pig (RC, n = 44) using the same PCR approach.

### SV calling specific to Guizhou pig

To screen the specific genomic structures of Guizhou indigenous pig, we downloaded publicly available NGS data of 36 pigs from the NCBI database (<https://www.ncbi.nlm.nih.gov/sra/?term=pig>) (S3 Table), including eighteen native pigs outside of Guizhou (NPOG) from six breeds (Min, Rongchang, Neijiang, Tongcheng, Jinhua, Tibetan pig) in China, eighteen European pigs (EUP) from five breeds (Landrace, Large White, Hampshire, Duroc, Berkshire pig). The confident SVs of NPOG and EUP were detected using the same method and standard for GZP. Then, we applied SVmerge approach to merge SVs among different individual [28]. Only SVs detected by two or more individuals were selected into final call set. To compare the population structure, all of SVs, detected from GZP, NOPG and EUP groups, were collected for Principal Components Analysis (PCA) using the OmicShare tools online (<http://www.omicshare.com/tools>).

### Annotation of SV regions

The SVs were evaluated their function using the Ensembl Variant Effect Predictor tool (<http://www.ensembl.org/info/docs/tools/vep/index.html>). Variant annotation classified by the VEP tool as high (disruptive impact in the protein) or modifier (non-disruptive variant) severity consequences to be used in genetic analysis of phenotypic differences observed.

To test on high conserved genes overlapped with SV, we collected a set of 458 core eukaryotic genes (CEG) that exist in a wide range of species [29]. We downloaded the RefSeq peptide ID (<http://korflab.ucdavis.edu/Datasets>) and used the BioMart data management system (<http://www.ensembl.org/biomart/martview/>) to convert the RefSeq peptide ID to homologous pig Ensembl gene IDs. Then the convert results can be overlapped with our detected genes with SV. We downloaded the total pig reference genes from BioMart, a Chi squared test was performed to test whether the conserved genes regions contained less SV than the other genes.

Then, we performed Gene Ontology (GO) functional annotation and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis of the genes (i.e. breed-specific or non-breed-specific with SV in GZP, and overlapping or non-overlapping with SV) using KOBAS 3.0 tool (<http://kobas.cbi.pku.edu.cn/>) [30,31], and P value of <0.10 determined by Fisher's exact test was set as the criteria for significance.

## Results

### Summary of sequencing and mapping

A total of 578.03 Gb raw sequences were generated by the Illumina HiSeq2500 platform. After removing the adaptor contamination and low-quality reads, we collected 545.4 Gb (94.35%) of clean sequences. On average, a pig individual obtained approximately 30.3 Gb clean reads ranged from 26.77 Gb (LB3) to 42.68 Gb (XP6). Total of 545.4 Gb clean reads were mapped to the pig reference genome assembly (Sscrofa11.1, ~2.445 Gb) using BWA software [22]. The mapped reads occupied 96.42% of the reference with an average 11.93× of sequencing depth (Table 1). In addition, we downloaded the genome data of 36 individuals originated across the world using Illumina next generation sequencing technology from NCBI database, including 18 native pigs outside of Guizhou province in China (NPOG), 18 European pigs (EUP) from

**Table 1. Summary of sequencing and mapping statistics.**

Sample	Raw Base(G)	Clean Base(G)	Map Reads(G)	Map ratio(%)	Depth(X)	Q20(%)	GC (%)
KL1	29.51	28.64	27.31	95.36	11.17	97.66	42.77
KL2	31.20	30.32	29.21	96.34	11.95	97.75	42.62
KL3	34.76	33.59	32.51	96.78	13.30	97.4	42.44
QB1	29.94	28.85	27.91	96.74	11.42	97.39	42.06
QB2	30.72	29.69	28.75	96.83	11.76	97.38	42.30
QB3	28.90	27.81	26.72	96.08	10.93	97.33	42.19
GL1	29.20	28.01	27.15	96.93	11.10	96.56	42.82
GL2	29.42	28.49	27.59	96.84	11.28	96.95	42.95
GL3	28.38	27.33	26.51	97.00	10.84	96.73	42.99
LB1	28.86	27.39	26.47	96.64	10.83	96.76	42.96
LB2	28.10	26.97	26.15	96.96	10.70	96.83	43.05
LB3	28.23	26.77	26.02	97.20	10.64	96.61	42.72
XP1	30.11	28.31	27.42	96.86	11.21	96.30	42.52
XP2	33.52	31.69	30.63	96.66	12.53	96.50	42.26
XP3	48.46	42.25	39.82	94.25	16.29	92.90	42.42
XP4	29.22	27.41	26.56	96.90	10.87	96.50	42.64
XP5	30.88	29.20	28.22	96.64	11.54	96.19	42.94
XP6	48.62	42.68	40.23	94.26	16.45	92.75	41.73

<https://doi.org/10.1371/journal.pone.0194282.t001>

five breeds. Sequencing of 36 pigs generated a total of 670.18 Gb clean reads, and the mapped read depth for 36 individuals ranged from 4.05 to 11.90 × (S3 Table).

### Identification of SVs

Two algorithms of pindel and softSV reported different numbers of SV for eighteen GZPs, with 400,779 SVs by Pindel and 453,682 SVs by softSV (S4 Table). Based on the guideline of overlap at least 25 bp at the same variation type and chromosomal coordinate [27], two or more SVs were merged, and only those SVs detected by both of pindel and softSV were retained. The confident SVs were dropped to 190,411 for GZP. Taking the same analysis strategy for GZP, the confident SVs were identified to be 56,562 SVs for EUP and 163,876 SVs for NPOG. The SV numbers of individual ranged from 970 in LA3 to 19,109 in XP3.

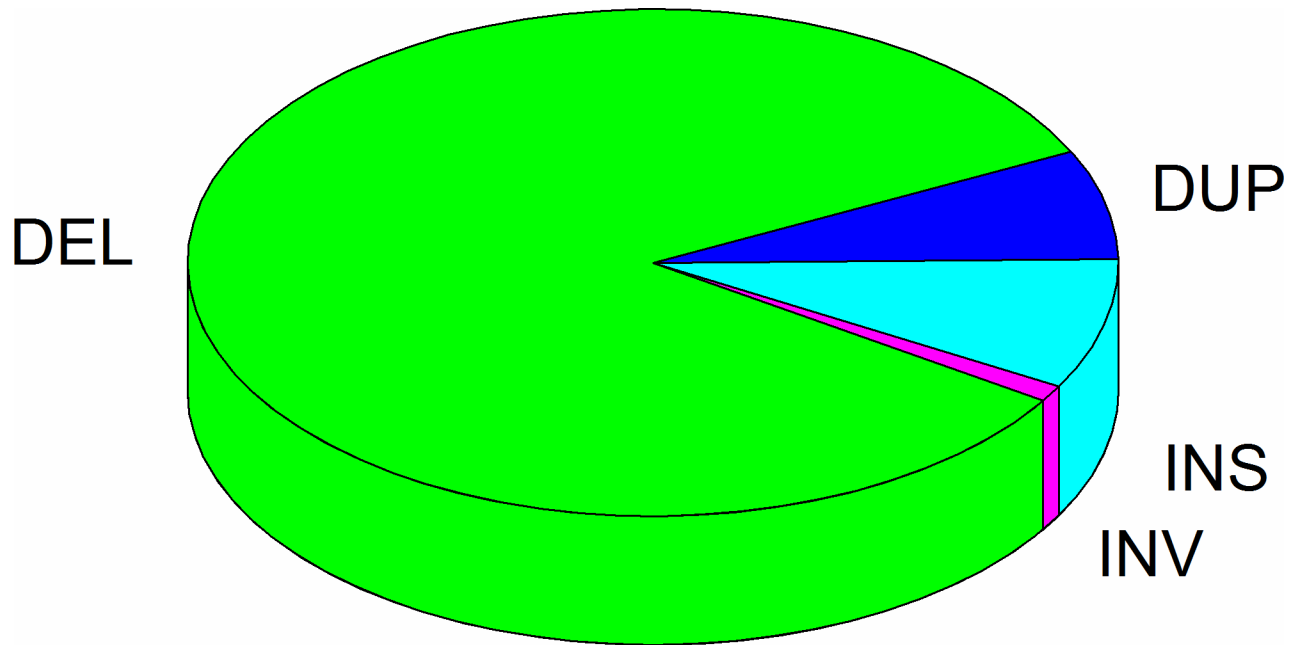
To get non-redundant SV, we applied SVmerge approach to merge SVs among different individual [28], in which only SVs detected by two or more individuals were selected into final call set. In total, 39,166 non-redundant SVs, named GZsv00001-GZsv39166 (S5 Table), were obtained from the 54 pig datasets, which consisted of 32,750 deletions (DEL, 83.62%), 3,268 insertions (INS, 8.34%), 2747 tandem duplications (DUP, 7.02%), 401 inversions (INV, 1.02%) (Fig 1), the most type of identified SVs was deletion.

### SV validation

To verify the efficiency of our approach and the authenticity of the identified SVs, 76 SVs were randomly selected for validation using PCR method. The deletion or insertion genotypes of 76 SVs were confirmed (S1 Fig) and further proofed by Sanger sequencing.

### Genomic distribution of SVs

The length distribution of 39,166 SVs (Table 2) revealed that 90% DELs (29326/32750) are 50 to 1,000 bp and affected 8397.430 Kb of genome sequence, but only 3424 DELs which were



**Fig 1. Pie chart of all types of SVs.** Each slice of the pie represented one type of SV.

<https://doi.org/10.1371/journal.pone.0194282.g001>

larger than 1000 bp covered 12253.185 Kb of genome sequence. DELs identified in this study covered a total length near to 20.65 Mb, and the largest deletion was 16,273 bp in length. INVs and DUPs covered a total length up to 6057.886 Kb. The largest inversion and tandem duplication were 16,391 and 16,220 bp in length, respectively. The majority of inversions (44.88%) were ranged between 1 Kb and 10 Kb. However, INS identified in this study only covered a total length of 243.903 Kb, with a period from 50 to 132 bp. A total of 39,166 SVs covered 27.37 Mb pig genome.

For the chromosomal distribution of SVs, the most number of SVs presented at chr1 in a ratio of 9.37%, followed by chr13 (7.54%), chr15 (6.81%), chr6 (6.28%), chr9 (6.11%), chr2 (5.84%), chr8 (5.56%), chr16 (5.51%), chr4 (5.35%), chr14 (5.25%), chr3 (5.24%), chr7 (5.219%), chr5 (5.10%), chrX (4.78%), chr11 (3.85%), chr10 (3.51%), chr12 (3.01%), chr17 (2.99%) and chr18 (2.68%) (Fig 2). Further, the chr1 contained the highest density of SVs than the others. The density of variants within each chromosome was proportional to the chromosome length, except for SVs at chr14, chrX, chr15, and chr16. Both of chr15 and chr16 were found out to give a high number of INS variants (Fig 3).

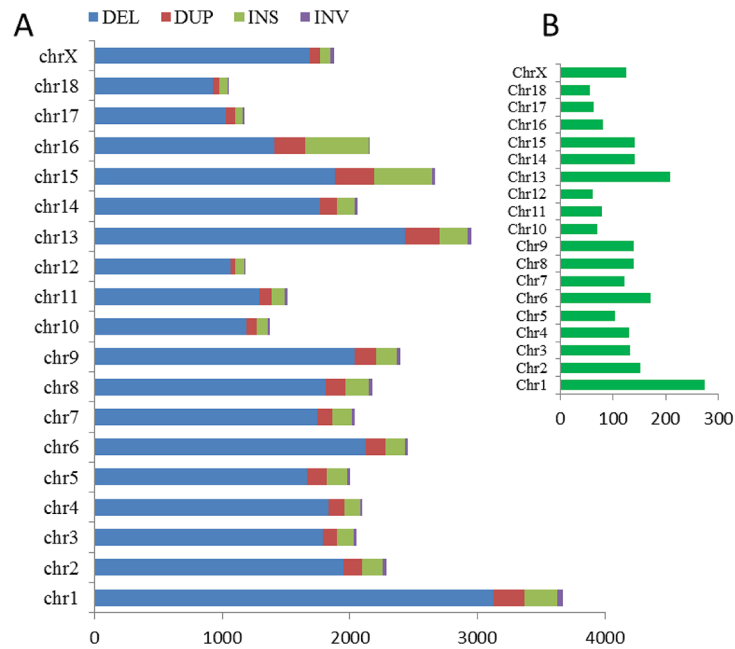
### Distribution of SV in pig population

Overall, 2,741 SVs were found only in single pig breeds, while 183 SVs presented in all sixteen pig breeds. The SV numbers were close to each other with 34,159 SVs in GZP breeds and 31,752 SVs in NPOG breeds. But it was much less in EUP breeds with 17,639 SVs (Table 3). It

**Table 2. Distribution of SVs length.** N: number of SV. L: SV length (bp).

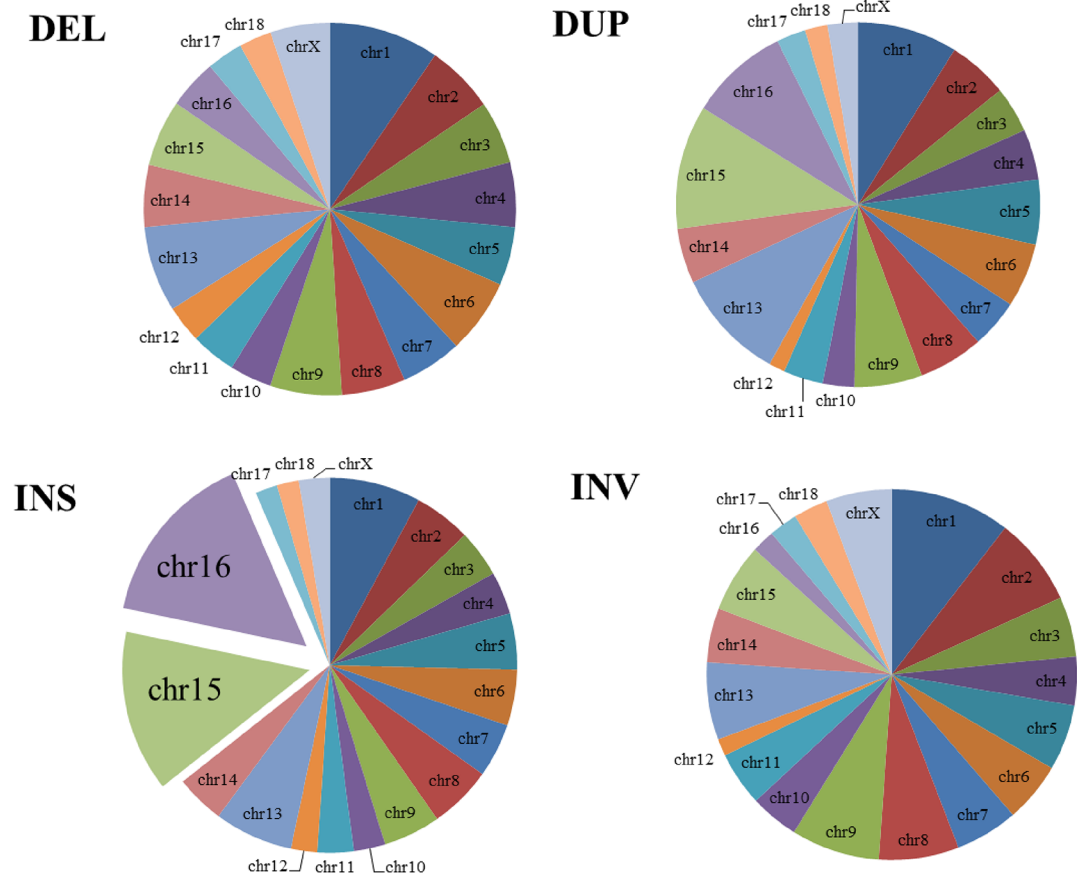
Region	DEL_N	DEL_L	DUP_N	DUP_L	INS_N	INS_L	INV_N	INV_L
50–1000 bp	29326	8397430	2080	454370	3268	243903	120	58038
1–10 kb	3293	10531040	551	1912944	0	0	180	754322
10–100 kb	131	1722145	116	2367314	0	0	101	1338534

<https://doi.org/10.1371/journal.pone.0194282.t002>



**Fig 2. SV distribution on chromosomes.** (A) Chromosomal distribution of SVs. (B) The chromosomal length of pig reference genome based on Sscrofa11.1.

<https://doi.org/10.1371/journal.pone.0194282.g002>



**Fig 3. The percentage distribution of each SV type on chromosomes.**

<https://doi.org/10.1371/journal.pone.0194282.g003>

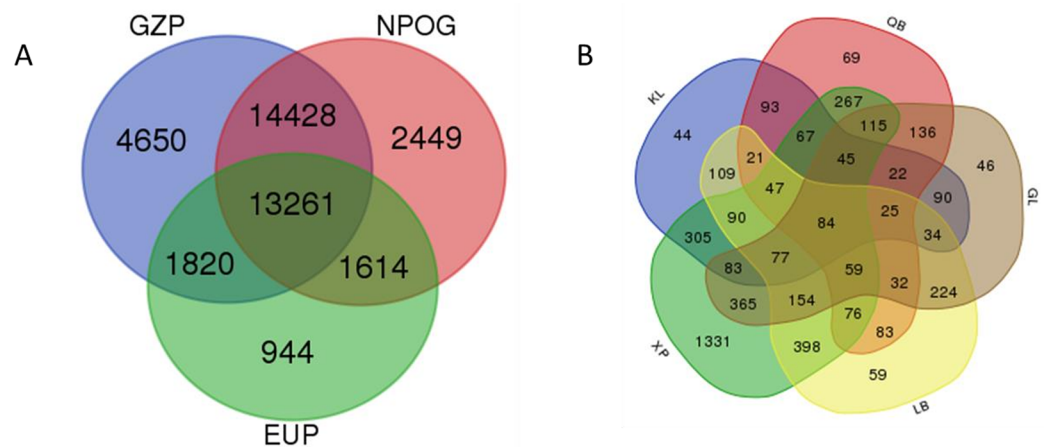
**Table 3. Number of SV detected in different pig breeds.**

Group	Breed	Total	DEL	DUP	INS	INV
GZP	KL	15796	14044	448	1224	80
	QB	15752	14124	434	1101	93
	GL	17260	15415	471	1291	83
	XP	28040	24352	1252	2177	259
	LB	17028	15320	479	1128	101
NPOG	NJ	13844	12658	618	517	51
	JH	12712	11870	465	335	42
	RC	15167	14088	566	455	58
	TI	17773	16340	768	600	65
	TC	17305	15875	711	645	74
	MP	19598	17733	901	819	145
EUP	DU	10998	9739	659	545	55
	LA	1983	1783	114	78	8
	LW	10822	9784	520	482	36
	HP	1559	1393	102	59	5
	BK	7644	6951	369	297	27

<https://doi.org/10.1371/journal.pone.0194282.t003>

was noticed that the highest number of SVs was in the XP and the lowest one in the HP pig breed. In other words, the genome structure of HP is the nearest to the reference of Sscrofa 11.1, and the XP genome is the most diverse one in this study.

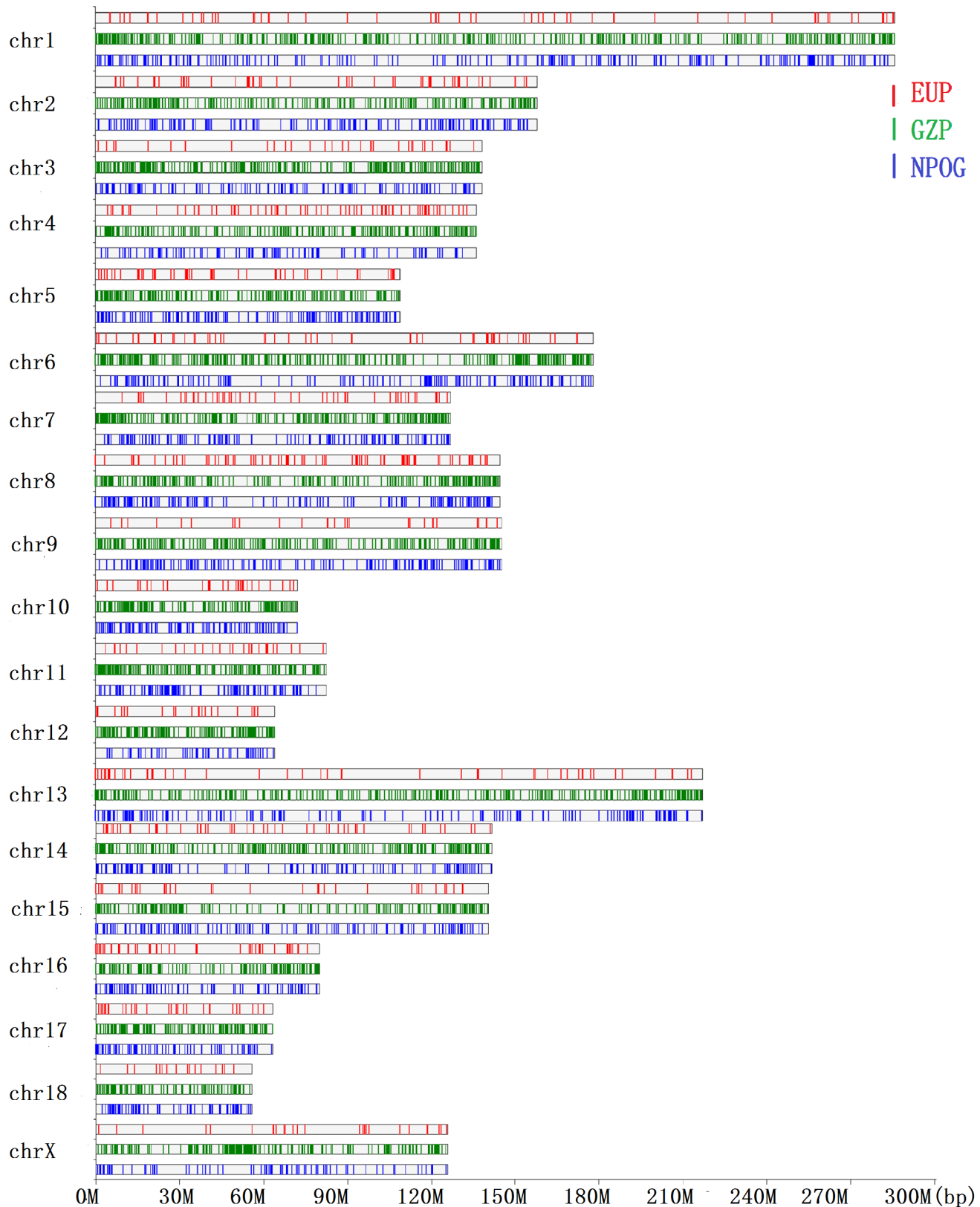
Compared SV distribution among three groups (Fig 4A), a total of 13,261 SVs were shared by three groups while the SVs with no overlap with any other breed (breed-specific) represented 4650 SVs in GZP, 2449 SVs in NPOG, and 944 SVs in EUP. SVs specific in GZP were the most abundant, while SV specific in EUP were the fewest. Of 4650 SVs specific in GZP breed, 84 SVs were common in all five breeds while the specific SV were much different in each GZP breed: 44 SVs in KL, 69 SVs in QB, 1331 SVs in XP, 46 SVs in GL and 59 SVs in LB breed (Fig 4B). The number of breed-specific SVs identified in the XP was the highest, and this finding showed that the XP breed contained plenty of variation. The distribution of these breed-specific SVs was presented in Fig 5.



**Fig 4. Venn diagram showing the overlap of identified SVs. (A)** Venn diagram showing the overlap of identified SVs in the GZP, NPOG, and EUP breed. **(B)** Venn diagram showing the overlap of identified SVs in five GZP.

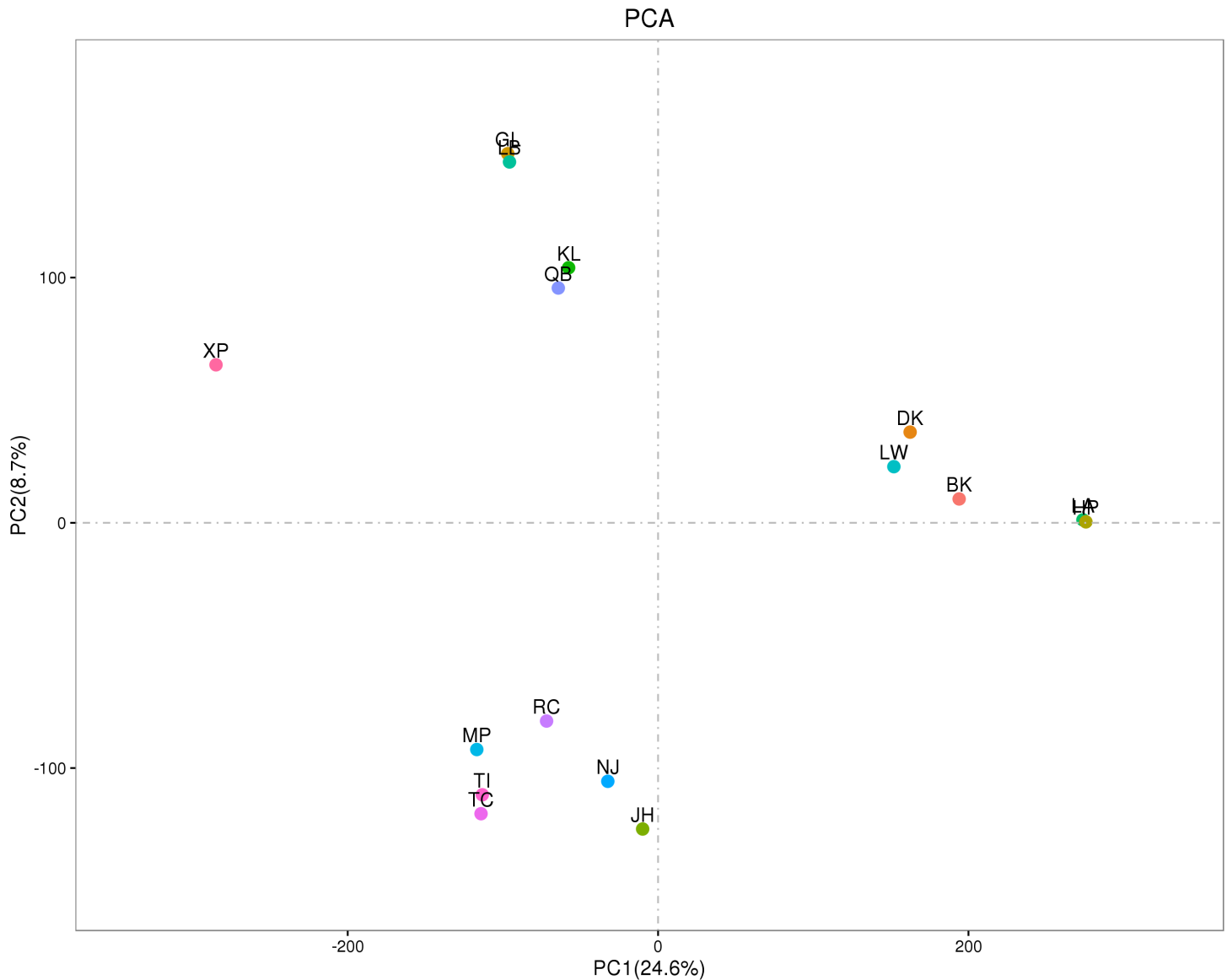
<https://doi.org/10.1371/journal.pone.0194282.g004>





**Fig 5. The chromosome distribution of the breed-specific in GZP, NPOG, and EUP pig breed.** The vertical bars indicated different pig breed with the red, green and blue color for EUP, GZP and NPOG, respectively.

<https://doi.org/10.1371/journal.pone.0194282.g005>



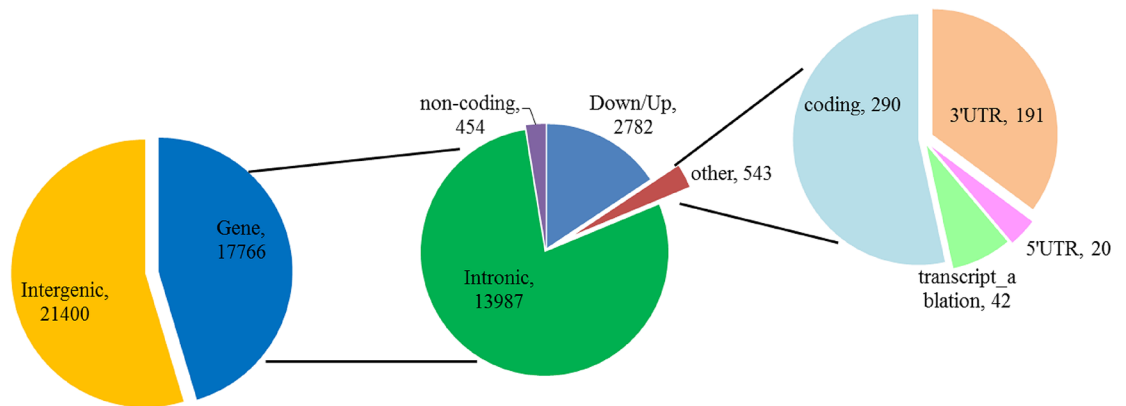
**Fig 6. Principal component analyses for all of 16 pig breeds.**

<https://doi.org/10.1371/journal.pone.0194282.g006>

This SV pattern was clear based on by principle component analysis (PCA) analysis (Fig 6). The PC1 geographically distinguished 11 pig breeds in China from five pig breeds in Europe, whereas the PC2 captured the biological differentiation between five GZP in Guizhou and six NOPG outside of Guizhou originated from China. These findings revealed genetically distinct clusters that related to geographic locations. We also found that XP breed was detected as an outlier and had a phylogenetic distance from the other four pig breed from Guizhou province, China.

### Annotation of SVs

Functional annotation of the identified 39,166 SVs was performed by Variant Effect Predictor program at the Ensembl website (S5 Table). In total, majority of SVs dispersed in intergenic regions (21400/39166, 54.64%) or intronic regions (13987/39166, 35.71%). And a small number of variants were annotated in the gene of exons or untranslated flank regions (Fig 7 and S6

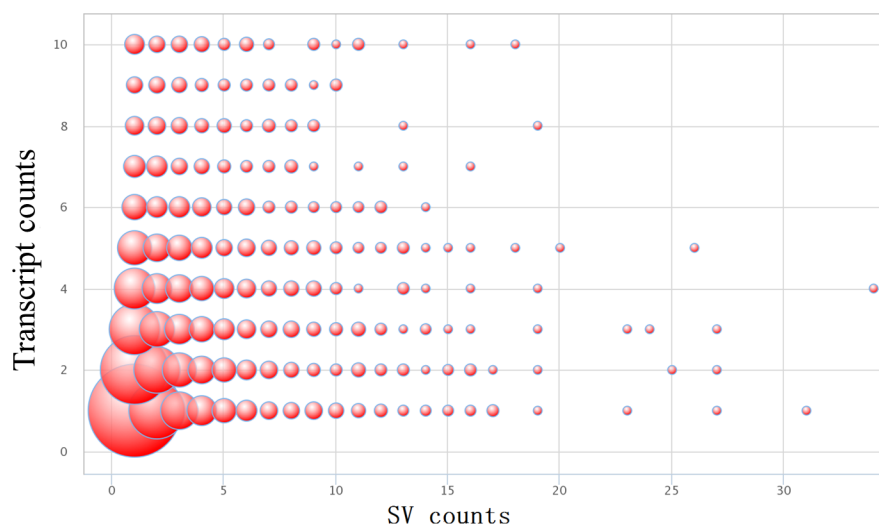


**Fig 7. Summary of functional classification of SVs identified.**

<https://doi.org/10.1371/journal.pone.0194282.g007>

Table). Of the 17,766 SVs in the genic regions, the DEL was the predominant type with 15,372 SVs. Types DUP and INS ranged in the middle with 1070 and 1136 SVs, and type INV were the fewest with 188 SVs. A total of 7,881 unique Ensembl genes were overlapped or nearby to those SVs (S5 Table). Notably, most of genes contained one or two SVs (5801/7881, 73.61%). There were also a small number of genes possessed many SVs, in which both of *DIAPH2* (diaphanous related formin 2) and *CCSER1* (coiled-coil serine rich protein 1) contained 30 and 34 SVs, respectively. *DIAPH2* may play a role in the development and normal function of the ovary [32]. Most of these genes contained many SVs belonged to large multigene families, including the olfactory receptor (OR), KRT, zinc finger protein, TMEM and HOXB families.

To the conserved eukaryotes genes, 437 unique pig genes were selected after conversion for the 458 core eukaryotic genes. About 112 core eukaryotic genes could be found out from the 7881 genes contained SV in present study. There involved 212 SVs in the 112 conserved genes. The other genes (7769) contained 17554 SVs. And conserved genes contained less SV numbers than the other genes ( $\chi^2 = 4.88$ ,  $P = 0.027$ ). Another interested finding was that SV tended to present in genes with many transcripts (Fig 8). For example, *TIAM1* gene contained 18 SVs



**Fig 8. The pattern of transcripts counts changed with the SV numbers.** The largest bubble showed the SV with most transcripts counts.

<https://doi.org/10.1371/journal.pone.0194282.g008>

which produced ten transcripts and FARS2 gene contained 16 SVs which possessed ten transcripts. Based on STATA analysis, it showed that the SV numbers were positive correlated with the transcript numbers (Spearman = 0.231,  $P < 0.05$ ).

For GZP, the effect situation on gene of 34,159 SV was much similar. There were 54.04 percent (18459/34159) of SVs located in intergenic region and 15,700 SVs mapped in 7,356 Ensembl genes, including 7023 protein-coding genes, 16 pseudogenes. The remaining genes coded for 5S rRNA, snoRNA, snRNA, microRNA, lincRNA, *et al.* Of the 4,650 GZP-specific SVs, 56.10% SVs were mapped to intergenic region (2609/4650), the remaining 2041 SVs covered 1,705 Ensembl genes. For the 1,628 protein-coding genes, the ratios were 79.42% in introns, 20.15% in the regulatory regions and 0.43% in exons. Further, we identified 93 loss of function (LoF) variation from 93 protein-coding genes including 54 exon\_variations of 53 genes and 39 UTR\_variations of 39 genes in GZP pig breed. Of the LoF genes, we identified some interested one that might impact economic traits of GZP breed, in which five genes are involved in fertility: *CDH5*, *FTL*, *KLF3*, *BOLL*, and *ZNF608* [33–37], and the gene *MAN2B2* (Mannosidase alpha class 2B member 2) is associated with the litter size of pig [38]. We also detected *PLCL2* gene related to immune response [39] and *LBR* (Lamin B receptor) involved in the cholesterol synthesis [40].

Furthermore, we found a SV hotspot region on the GZP chrX from 39 Mb to 59 Mb, harboring 104 SVs. In this region, 29 genes were annotated and some of them have been confirmed to associate with histone modification (S7 Table). The *TRO* (Trophinin) gene, with one frameshift variant in exon 12 due to SV (GZsv37822), encodes a membrane protein and involved in blastocyst implantation and associated with ovarian cancer [41]. *MTMR8* (Myotubularin-related protein 8) gene was damaged by GZsv37892 for two exonic variants, which is essential for the endothelial cell differentiation and vasculature development [42]. Three affected genes, *HUWE1*, *PHF8* and *KLF8*, are related to breast or ovary cancers [43–45]. Androgen receptor (*AR*) gene is critical for the ovarian development [46]. *ZC3H12B* gene is negative regulator in macrophage activation and may involve in host immunity and inflammatory diseases [47].

### Functional enrichment analysis for variation genes

The 1,628 genes affected by GZP-specific SVs were further used for GO and KEGG enrichment analysis ( $P < 0.05$ , S8 Table). Six GO terms related with reproduction biology process were detected only from GZP but was not disturbed in both of NPOG and EUP groups. Seven genes were enriched in the six GO terms, including estrogen receptor binding (GO: 0030331), intracellular estrogen receptor signaling pathway (GO: 0030520), steroid hormone receptor binding (GO: 0035258), maternal process involved in female pregnancy (GO: 0060135), regulation of intracellular estrogen receptor signaling pathway (GO: 0033146), and retinoic acid receptor binding (GO: 0042974). Another impressed enrichment terms were associated with immunization, such as, inflammatory response to antigenic stimulus (GO: 0002437), interleukin-4 production (GO: 0032633), regulation of interleukin-1 production (GO: 0032652), T-helper 1 type immune response (GO: 0042088), macrophage activation (GO: 0042116), positive regulation of leukocyte cell-cell adhesion (GO: 1903039), positive regulation of lymphocyte activation (GO: 0051251), interleukin-1 production (GO: 0032612), positive regulation of T cell activation (GO: 0050870). Fourteen GO terms associated with adaptability, including cell projection part (GO: 0044463), axon (GO: 0030424), catecholamine transport (GO: 0051937), ATPase activity, coupled to transmembrane movement of ions, rotational mechanism (GO: 0044769), axo-dendritic transport (GO: 0008088), regulation of catecholamine secretion (GO: 0050433), cell projection cytoplasm (GO: 0032838), catecholamine secretion

(GO: 0050432), dopamine transport (GO: 0015872), regulation of amine transport(GO: 0051952), dendrite (GO: 0030425), amine transport (GO: 0015837), cell projection (GO: 0042995), and positive regulation of response to biotic stimulus (GO: 0002833). Interestingly, the genes affected by SV in GZP enriched in the KEGG pathway mainly comprised metabolism and biosynthesis, reproduction, immune and adaptability, involved in oxytocin signaling pathway (ssc04921), mTOR signaling pathway (ssc04150), axon guidance (ssc04360), cholinergic synapse (ssc04725), fructose and mannose metabolism (ssc00051), glycerophospholipid metabolism (ssc00564), mucin type O-Glycan biosynthesis (ssc00512), and Glycosaminoglycan biosynthesis-heparan sulfate / heparin (ssc00534).

The seven genes involved in reproduction biology process were *ZNF366*, *LEF1*, *CNOT1*, *MED1*, *CTSB*, *HAVCR2*, and *VDR* (Table 4). Fourteen genes affected by SV in GZP enriched in the KEGG pathway of oxytocin signaling pathway, including *MEF2C*, *EEF2K*, *NFATC3*, *ROCK1*,

**Table 4. Genes affected by GZP-specific SVs enriched in GO terms and KEGG pathway related with fertility of GZP pig.**

NO.	Chr	SV Start	SV End	SV Length	SV Type	Gene	Symbol	SV Location	GO/kegg ID
GZsv12513	6	20311318	20311590	272	DEL	ENSSSCG00000002799	CNOT1	Intron	GO:0030331 GO:0033146 GO:0035258 GO:0030520 GO:0042974
GZsv11540	5	78225835	78226139	304	DEL	ENSSSCG000000020864	VDR	Intron	GO:0060135 GO:0042974
GZsv18234	8	113860056	113860306	250	DEL	ENSSSCG00000009148	LEF1	Intron	GO:0030331;GO:0035258
GZsv24487	12	22838005	22838322	317	DEL	ENSSSCG000000017505	MED1	Intron	GO:0030331 GO:0033146 GO:0060135 GO:0035258 GO:0030520 GO:0042974
GZsv28563	14	15027500	15027652	152	DUP	ENSSSCG000000023666	CTSB	Intron	GO:0060135
GZsv34554	16	48808965	48809055	90	DEL	ENSSSCG000000016976	ZNF366	Intron	GO:0030331 GO:0033146 GO:0035258 GO:0030520
GZsv34801	16	66140226	66140303	77	DEL	ENSSSCG000000028875	HAVCR2	Intron	GO:0060135
GZsv04997	2	96288258	96288329	71	DEL	ENSSSCG000000014149	MEF2C	Intron	ssc04921
GZsv06434	3	23971756	23972484	730	INV	ENSSSCG000000007839	EEF2K	Intron	ssc04921
GZsv12631	6	28764371	28764445	74	DEL	ENSSSCG000000029578	NFATC3	Intron	ssc04921
GZsv13502	6	106214603	106214934	331	DEL	ENSSSCG000000021893	ROCK1	Intron	ssc04921
GZsv14469	6	165382096	165383623	1527	DEL	ENSSSCG000000003909	Novel gene*	Intron	ssc04921
GZsv16830	8	11130507	11130784	277	DEL	ENSSSCG000000008742	CD38	3'UTR	ssc04921
GZsv20391	9	98239262	98239451	189	DEL	ENSSSCG000000015402	CACNA2D1	Intron	ssc04921
GZsv24985	12	49955326	49955592	266	DEL	ENSSSCG000000017873	CAMKK1	Upstream	ssc04921
GZsv27142	13	137089754	137089859	105	INS	ENSSSCG000000027952	ADCY5	Intron	ssc04921
GZsv27144	13	137107559	137108002	443	DEL	ENSSSCG000000027952	ADCY5	Intron	ssc04921
GZsv34990	16	74352283	74352805	522	DEL	ENSSSCG000000017101	ADCY2	Intron	ssc04921
GZsv35481	17	17421770	17421871	101	INS	ENSSSCG000000007056	PLCB1	Intron	ssc04921
GZsv35498	17	18213874	18214792	918	DEL	ENSSSCG000000007058	PLCB4	Intron	ssc04921
GZsv36345	18	5484398	5484686	288	DEL	ENSSSCG000000016432	PRKAG2	Intron	ssc04921
GZsv36357	18	6232010	6232110	100	DEL	ENSSSCG000000016450	NOS3	Upstream	ssc04921

\*: Uncharacterized protein.

<https://doi.org/10.1371/journal.pone.0194282.t004>



**Fig 9. Results of Genetic population analysis.** (A) Allele D frequency in pigs. (B) Allele I frequency in pigs. (C) The distribution frequencies of the six genes in 284 pigs.

<https://doi.org/10.1371/journal.pone.0194282.g009>

*CD38*, *CACNA2D1*, *CAMKK1*, *ADCY5*, *ADCY2*, *PLCB1*, *PLCB4*, *PRKAG2*, *NOS3* and a novel gene (Table 4). For all of 21 genes except for *CD38*, their harboured SVs located in the intron region or nearby to the gene (Table 4). These SVs might not change the coded peptides even though *CD38* gene hold a DEL SV at the last exon, which located downstream of the stop codon.

Besides, compared between genes with SV and without SV in GZP, we found nine GO terms related with ion transport process ( $P < 0.10$ , S9 Table), including ammonium transport (GO:0015696, GO:0072488, GO:0008519), lactate transport (GO:0015129, GO:0015727, GO:0035873, GO:0035879), proton-transporting two-sector ATPase complex (GO:0033178), cargo receptor activity (GO:0038024). It suggested that ammonium and lactate transports might be affected by those SVs in GZP genome by genes such as *SLC16A4*, *SLC16A5*, *SLC16A7*, *SLC16A12*, *SLC5A12*, *SLC12A2*, *SLC22A2*, *SLC22A3*, *SLC44A1*, *SLC44A4*, *RHAG* and *RHCG*. The proton-transporting ATPase included *ATP5A1*, *ATP6V1A*, *ATP6V1B2*, *ATP6V1C1*, *ATP6V1C2*, *ATP6V1E1* and *ATP6V1H*.

### Identify the candidate genes associated with the body configuration in Xiang pig

Candidate genes were identified according to the following criteria: (1) genes were specific mutation only in Xiang pig (XP); (2) genes were enriched in pathways related to development and metabolism; (3) genes were associated with growth traits reported by previous studies (<https://www.ncbi.nlm.nih.gov/gene/?term=growth+Sus+scrofa>). The application of these criteria led to the identification of 51 candidate genes associated with body configuration in XP pig breeds, including some well-known genes such as *APOD*, *INO80*, *IGF2BP3*, *GSK3B*, *AKT3*, *MEF2C*, and so on (S10 Table).

To estimate the candidate genes pattern in pig population, the six candidate genes verified by PCR were genotyped using enlarged samples of 284 pigs. The genotype and allele frequencies were calculated based on the gel electrophoresis patterns. The genotype DD frequency was higher in XP than the LW, DU, KL, LB, QB, GL and RC breeds (S11 Table and Fig 9,  $P < 0.05$ ). It indicated that the six genes could be involved in the regulation on body configuration in Xiang pig. In addition, the gene homozygosity ( $H_o$ ), heterozygosity ( $H_e$ ), polymorphism information content ( $PIC$ ), and effective allele numbers ( $N_e$ ) was calculated across genotype number (<http://www.msrcall.com/Gdical.aspx>). The observed  $H_e$  mean was  $0.605 \pm 0.096$  and  $PIC$  mean was  $0.312 \pm 0.058$  in XP breed, and showed that the six markers might be informative.

## Discussion

In the present study, we performed genome resequencing for eighteen Guizhou pigs in China by NGS technology. The raw reads of GZP were filtered and combined with resequencing data from other 36 pigs of Chinese native pig outside of Guizhou and European pigs download from public database. We identified 39,166 SVs from 56 pig data, which might affected 7881 genes representing 30.45% (7881/25880) of the total genes based on the reference annotation in Sscrofa11.1. And more than 13300 SVs detected in our data were overlapped with previous SV data from thirteen Chinese and European breeds [19]. About 76 SVs were confirmed in a large of Xiang pig by PCR method.

Compared among three groups, SV numbers in GZP and NPOG were close to each other while they were much more than that in EUP. It was further confirmed by PCA results in Fig 6. Previous reports also find that Chinese indigenous pig breeds possess more plenty of genetic diversity than European breeds [19,20]. Correspondingly, the highest number of breed-specific SVs was attributed to GZP with 4,650 SV, and then NPOG with 2,449 SV, EUP with 944 SV. Inside of the five GZP, it was Xiang pig contained the highest number of breed-specific SVs (1,331) while the specific number of other four GZPs was less than a hundred. It illustrated that the genome carried many variation in Xiang pig population. It is also notable that Chinese wild boars have many variation compared with Europe pigs [19].

Notably, we identified a GZP-specific SV hotspot from 39 Mb to 59 Mb of chrX. Within the region of 20 Mb, there presented 104 SVs and contained 29 genes. Some of them have been tested for a link with reproduction including *TRO*, *MTMR8*, *HUWE1*, *PHF8*, *KLF8*, *AR* and *ZC3H12B* genes [41–47]. In addition, another work suggests a selective sweep signals region presented in chrX from 40–80 Mb of Chinese pig populations, the region involves a high *Fst* estimate between Chinese indigenous and European pig breeds [48]. It may supported that a specific SV hotspot of 39 Mb to 59 Mb on the chrX can be considered as a hotspot region to genetic diverse between Guizhou pig breed and the others.

To better understand the functions of these SV variants in GZP, we performed VEP online analysis from Ensembl. It was similar to the previous reports that most of SVs present in the intergenic region but the protein-coding region had a major influence on gene function [49]. And SV tended to distribute in genes produced multi-transcripts and showed a significant positive correlation between the SV number and the transcripts number. For example, *FARS2* gene harbored 16 SVs, encodes for mitochondrial phenylalanyl-tRNA synthetase, and could generate 10 transcripts in pig. In human, a patient suffers with global developmental delay, dysarthria and tremor caused by a deletion at chromosome 6p25.1 includes all of exon 6 and parts of introns 5 and 6 of *FARS2* [50]. It has been confirmed that intron 44 retention of the von Willebrand factor (*VWF*) gene resulting from a silent mutation in the *VWF* gene that structurally influences the splice site [51]. In our study, the *VWF* gene overlapped with six SVs

in GZP (GZsv11270, GZsv11271, GZsv11272, GZsv11273, GZsv11274, and GZsv11275). We applied RegRNA 2.0 to predict the effects of SV on the changes of intron including splicing donor/acceptor sites, exon splicing enhancer (ESE), exon splicing silencer (ESS), intron splicing enhancer (ISE), and intron splicing silencer (ISS) [52]. We founded that GZsv11272, GZsv11273 and GZsv11275 contained ESE, ESS and ISE. GZsv11273 also contained a splicing acceptor site. It is reported that the splicing donor/acceptor sites may change the 3'-end of an intron or the 5'-end splice site of the intron, and lead to the production of different isoforms of transcript [53]. Transcript isoforms resulting from alternative splicing (AS) events can be viewed as having "internal-paralogs" in the same gene [54]. These "internal-paralogs" may have different functions especially in gene evolution. Taken together, SVs in the gene region might be a reason for the alternative splicing event and resulted in multi-transcripts. Beside, we found only a small parts of genes (112) contained SV from 437 CEG, which is highly conserved across eukaryotes species [29]. And these conserved genes contained less SV than the other non-conserved genes.

In addition, those genes affected by GZP-specific SVs were used for gene family annotation and function enrichment analysis by Kobas 3.0. We found seven genes (*ZNF366*, *LEF1*, *CNOT1*, *MED1*, *CTSB*, *HAVCR2*, and *VDR*) affected by SVs involved in GO terms of reproduction processes (Table 4). *ZNF366*, which encodes an evolutionarily conserved zinc finger protein, interacts with the estrogen receptor- $\alpha$  DNA binding domain (*ER $\alpha$*  DBD), represses *ER $\alpha$*  activity and regulating the expression of genes in response to ER [55]. *CNOT1* contains several LXMs and interact directly in a ligand-binding domain (LBD) fashion with *ER $\alpha$* , and represses the LBD transcriptional activation function of *ER $\alpha$*  [56]. ER mediates the function of estrogen in reproductive systems of the female and the male [57]. *LEF1* (lymphoid enhancer-binding factor 1) has been shown to be regulated in the embryo [58] and uterus [59] during pregnancy primarily. *MED1* (Mediator complex subunit 1) can promote nuclear hormone receptor-mediated transcription in a ligand-dependent manner [60], and regulate meiotic progression during spermatogenesis in mice [61]. *CTSB* (Cathepsins B) may modify proteins for fluid-phase transport across porcine uterine, placental, and neonatal gut epithelia [62]. *VDR* (Vitamin D3 receptor) expressed throughout central and peripheral organs of reproduction [63]. Many papers show that GZP give much lower litter sizes compared to European pig breeds with 6.6–6.9 piglets in Xiang pig, 6–8 in Kele pig, and 9–11 in Large White pig [3]. It suggested that these genes affected by SV might change the reproductive performance of GZP breed.

Interestingly, the genes affected by SV in GZP enriched in several mainly pathway related to reproduction, immune and adaptability, containing oxytocin signaling pathway, mTOR signaling pathway, axon guidance and cholinergic synapse. This oxytocin signaling pathway has a role in uterine contractions during parturition and milk release during lactation [64]. The mTOR signaling pathway regulated in porcine reproductive and respiratory syndrome virus (PRRSV) infected porcine alveolar macrophages at different activation statuses [65]. The mTOR signaling pathway is also related to change synaptic plasticity in stress and depression [66], and synaptic plasticity is basic for the adaptability of the mammalian brain [67]. Axon guidance is a key stage for formation of neuronal network, and it is guided to its proper target by sensing extracellular cues in the local environment [68]. Cholinergic synapse involved in the afferent neuronal regulation of gonadotropin-releasing hormone neurons (GnRH) in rat, and GnRH is the common pathway in the hypothalamic regulation of reproduction [69]. The cholinergic system also seems likely to positively promote proliferation, differentiation, integration and affect cortical development and adult neurogenesis [70]. Therefore, these pathways might regulate reproduction, immune and adaptability, and contribution to phenotype different in GZP pig breed.



Compared the GO biological process between genes with SVs and without SVs in GZPs, there were nine ion transports processes might be affected by SVs significantly. Five involved genes were associated with the lactate transport, including *SLC16A4*, *SLC16A12*, *SLC5A12*, *SLC16A5*, and *SLC16A7*. For all of the five genes involved DEL variation located in the intron region. However, we retrieved 252 GO terms related with ion transports (AmiGO 2, <http://amigo.geneontology.org/amigo/landing>). It seemed that the effects of SV on the ion transports might be compensated by other members of SLC family.

The Xiang pig is well known Chinese miniature breed for its small body size. To identify candidate genes associated with special phenotype in Xiang pig breeds, the 51 candidate genes were collected from the integration of genes enriched in the KEGG pathway and the known genes deposited in the Gene database of NCBI. Population of 284 pigs from eight pig breeds was detected for the genotype frequencies of six genes. The deletion type of SV site, GZsv04997 in *MEF2C* gene, was detected only from XP breed. *MEF2C* (myocyte enhancer factor 2 C) is expressed in skeletal muscle and control of overall body size in mice [71]. We found that a 300 bp deletion in *APOD* gene (GZsv27094) was mainly present in XP breed, and *APOD* (Apolipoprotein D) is associated to high bone turnover, low bone mass and influences bone metabolism [72]. The remained 4 genes of genotype DD frequency were higher in XP than the LW, DU, KL, LB, QB, GL and RC breeds. *IGF2BP3* (IGF2-binding protein 3) involved in transcriptional regulation of IGF2 [73] and related to bone formation [74]. *AKT3* gene polymorphisms associated with myofiber characteristics in chickens [75]. *GSK3B* (glycogen synthase kinase 3 beta) plays a vital role in the muscle growth and differentiation [76–77]. *AKT3* and *GSK3B* gene involved in both of thyroid hormone signaling pathway and insulin signaling pathway. It has been well-documented that the thyroid hormone contributed to the growth velocity in children with idiopathic short stature [78]. *INO80* is required for promotion of mesenchymal stem cells (MSC) osteogenic differentiation [79]. These genes take part in the growth of muscle and bone and could be taken as a marker to phenotype characteristic in Xiang pig, but the specific effects of these genes during development and metabolism were needed further to be clear.

## Conclusion

The whole genome resequencing for five GZP breeds and comparison with data from NPOG and EUP breeds lead to identification of 39,166 SVs. This study had three highlights. Firstly, SV tended to located in the genes with multi-transcripts and the number of SV was positive correlated with that of gene transcripts. It suggested that SV might be a reason for the splice variant of pig gene. Secondly, we applied the SVs to access the population structures of these pig breeds, and got the pattern that the similarity of SV in five GZP much closer to each other than the other two groups. Thirdly, we identified 4,650 GZP-specific SVs overlapped with 1,628 protein-coding genes, in which a few SVs reshaped the coding frame of genes and about 93 genes lose function due to SV variations. Moreover, a SV hotspot was detected in 20 Mb of chrX in GZP and harbored 29 protein coding genes. The functional enrichment analysis suggested that these genes affected by GZP-specific SVs gathered in reproduction, nervous system and immune functions. Further, we identified 51 candidate genes associated with body configuration in Xiang pigs. These results provided worthwhile genomic region related to economically traits in pig and suggested that specific SVs might be a reason for the strong adaptability and low fecundity of GZP.

## Supporting information

**S1 Fig. Representative gel images for the confirmation of SVs in GZP pig.**  
(PDF)

**S1 Table. Information of the pig breeds used for resequencing.**

(XLSX)

**S2 Table. Primers used for validation by PCR method.**

(XLSX)

**S3 Table. The information of 36 pig resequencing data downloaded from the NCBI.**

(XLSX)

**S4 Table. Counts of raw SVs in pig genomes detected by pindel and softsv programs.**

(XLSX)

**S5 Table. Confident SVs identified from sixteen pig breeds.**

(XLSX)

**S6 Table. Annotation of all SVs and the location of SVs in pig genome.**

(XLSX)

**S7 Table. The genes in the SV hotspot on chrX of GZP.**

(XLSX)

**S8 Table. KEGG Pathway analysis and Gene ontology (GO) annotation in breed-specific SVs.**

(XLSX)

**S9 Table. KEGG and GO annotations between gene with SV and gene without SV in GZP.**

(XLSX)

**S10 Table. Candidate genes associated with the body configuration in Xiang pig.**

(XLSX)

**S11 Table. The frequency of the six genes distributed in the pig populations.**

(XLSX)

## Acknowledgments

This work is funded by the National High Technology Research and Development Program of China (863 Program) [2013AA102503], the National Natural Science Foundation of China (31672390, 31401091), the Guizhou Province "Hundred" Innovative Talents Project [2016–4012], and the Guizhou Agriculture Research program (QKHZC[2017]2585, QKHZC[2017]2587).

## Author Contributions

**Conceptualization:** Xueqin Ran, Jiafu Wang.

**Data curation:** Chang Liu, Xueqin Ran.

**Formal analysis:** Chang Liu, Xueqin Ran.

**Funding acquisition:** Jiafu Wang.

**Investigation:** Chang Liu, Sheng Li.

**Methodology:** Xueqin Ran, Jiafu Wang, Jianfeng Liu.

**Project administration:** Xueqin Ran, Jiafu Wang, Jianfeng Liu.

**Resources:** Chang Liu, Xueqin Ran.

**Software:** Chang Liu, Xueqin Ran.

**Supervision:** Xueqin Ran.

**Validation:** Xueqin Ran, Sheng Li.

**Writing – original draft:** Chang Liu.

**Writing – review & editing:** Xueqin Ran, Jiafu Wang.

## References

1. Zhang C, Plastow G. Genomic Diversity in Pig (*Sus scrofa*) and its Comparison with Human and other Livestock. *Curr Genomics*. 2011; 12: 138–146. <https://doi.org/10.2174/138920211795564386> PMID: 21966252.
2. China National Commission of Animal Genetic Resources. *Animal Genetic Resources in China* Agriculture Press, Beijing. 2011.
3. Liu JJ, Ran XQ, Li S, Feng Y, Wang JF. Polymorphism in the first intron of follicle stimulating hormone beta gene in three Chinese pig breeds and two European pig breeds. *Anim Reprod Sci*. 2009; 111(2–4):369–75. <https://doi.org/10.1016/j.anireprosci.2008.03.004> PMID: 18436398
4. Vallet JL, Freking BA, Leymaster KA, Christenson RK. Allelic variation in the secreted folate binding protein gene is associated with uterine capacity in swine. *J Anim Sci*. 2005; 83(8):1860–7. <https://doi.org/10.2527/2005.8381860x> PMID: 16024705
5. Sironen A, Uimari P, Venhoranta H, Andersson M, Vilkki J. An exonic insertion within Tex14 gene causes spermatogenic arrest in pigs. *BMC Genomics*. 2011; 12:591. <https://doi.org/10.1186/1471-2164-12-591> PMID: 22136159
6. Ren Z, Liu W, Zheng R, Zuo B, Xu D, Lei M, et al. A 304 bp insertion/deletion mutation in promoter region induces the increase of porcine IDH3 $\beta$  gene expression. *Mol Biol Rep*. 2012; 39(2):1419–26. <https://doi.org/10.1007/s11033-011-0876-1> PMID: 21617947
7. Xu HP, He XM, Fang MX, Hu YS, Jia XZ, Nie QH, et al. Molecular cloning, expression and variation analyses of the dopamine D2 receptor gene in pig breeds in China. *Genet Mol Res*. 2011; 10(4):3371–84. <https://doi.org/10.4238/2011.December.5.6> PMID: 22180074
8. Domínguez MA, Landi V, Martínez A, Garrido JJ. Identification and functional characterization of novel genetic variations in porcine TLR5 promoter. *DNA Cell Biol*. 2014; 33(7):469–76. <https://doi.org/10.1089/dna.2013.2318> PMID: 24684626
9. Mikawa S, Sato S, Nii M, Morozumi T, Yoshioka G, Imaeda N, et al. Identification of a second gene associated with variation in vertebral number in domestic pigs. *BMC Genet*. 2011; 12:5. <https://doi.org/10.1186/1471-2156-12-5> PMID: 21232157
10. Trott JF, Freking BA, Hovey RC. Variation in the coding and 3' untranslated regions of the porcine prolactin receptor short form modifies protein expression and function. *Anim Genet*. 2014; 45(1):74–86. <https://doi.org/10.1111/age.12100> PMID: 24350599
11. Feuk L, Carson AR, Scherer SW. Structural variation in the human genome. *Nat Rev Genet*. 2006; 7(2):85–97. <https://doi.org/10.1038/nrg1767> PMID: 16418744
12. Alkan C, Coe BP, Eichler EE. Genome structural variation discovery and genotyping. *Nat Rev Genet*. 2011; 12(5):363–76. <https://doi.org/10.1038/nrg2958> PMID: 21358748
13. Mills RE, Walter K, Stewart C, Handsaker RE, Chen K, Alkan C, et al. Mapping copy number variation by population-scale genome sequencing. *Nature*. 2010; 470(7332):59–65. <https://doi.org/10.1038/nature09708> PMID: 21293372
14. Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, Thorne N, et al. Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science*. 2007; 315, 848–853. <https://doi.org/10.1126/science.1136678> PMID: 17289997
15. Marshall CR, Howrigan DP, Merico D, Thiruvahindrapuram B, Wu W, Greer DS, et al. Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. *Nat Genet*. 2017; 49(1):27–35. <https://doi.org/10.1038/ng.3725> PMID: 27869829
16. Liu W, Sun J, Li G, Zhu Y, Zhang S, Kim ST, et al. Association of a germ-line copy number variation at 2p24.3 and risk for aggressive prostate cancer. *Cancer Res*. 2009; 69(6):2176–9. <https://doi.org/10.1158/0008-5472.CAN-08-3151> PMID: 19258504

17. Lupski JR. Genomic disorders ten years on. *Genome Med.* 2009; 1(4):42. <https://doi.org/10.1186/gm42> PMID: 19439022
18. Yang J, Li WR, Lv FH, He SG, Tian SL, Peng WF, et al. Whole-Genome Sequencing of Native Sheep Provides Insights into Rapid Adaptations to Extreme Environments. *Mol Biol Evol.* 2016; 33(10):2576–92. <https://doi.org/10.1093/molbev/msw129> PMID: 27401233
19. Zhao P, Li J, Kang H, Wang H, Fan Z, Yin Z, et al. Structural Variant Detection by Large-scale Sequencing Reveals New Evolutionary Evidence on Breed Divergence between Chinese and European Pigs. *Sci Rep.* 2016; 6:18501. <https://doi.org/10.1038/srep18501> PMID: 26729041
20. Chen L, Jin L, Li M, Tian S, Che T, Tang Q, et al. Snapshot of structural variations in the Tibetan wild boar genome at single-nucleotide resolution. *J Genet Genomics.* 2014; 41(12):653–7. <https://doi.org/10.1016/j.jgg.2014.10.001> PMID: 25527106
21. Patel RK, Jain M. NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PLoS One.* 2012; 7(2):e30619. <https://doi.org/10.1371/journal.pone.0030619> PMID: 22312429
22. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009; 25(14):1754–60. <https://doi.org/10.1093/bioinformatics/btp324> PMID: 19451168
23. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009; 25(16):2078–9. <https://doi.org/10.1093/bioinformatics/btp352> PMID: 19505943
24. McKernan KJ, Peckham HE, Costa GL, McLaughlin SF, Fu Y, Tsung EF, et al. Sequence and structural variation in a human genome uncovered by short-read, massively parallel ligation sequencing using two-base encoding. *Genome Res.* 2009; 19(9):1527–41. <https://doi.org/10.1101/gr.091868.109> PMID: 19546169
25. Ye K, Schulz MH, Long Q, Apweiler R, Ning Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics.* 2009; 25(21):2865–71. <https://doi.org/10.1093/bioinformatics/btp394> PMID: 19561018
26. Bartenhagen C, Dugas M. Robust and exact structural variation detection with paired-end and soft-clipped alignments: SoftSV compared with eight algorithms. *Brief Bioinform.* 2016; 17(1):51–62. <https://doi.org/10.1093/bib/bbv028> PMID: 25998133
27. Chen L, Chamberlain AJ, Reich CM, Daetwyler HD, Hayes BJ. Detection and validation of structural variations in bovine whole-genome sequence data. *Genet Sel Evol.* 2017; 49(1):13. <https://doi.org/10.1186/s12711-017-0286-5> PMID: 28122487
28. Wong K, Keane TM, Stalker J, Adams DJ. Enhanced structural variant and breakpoint detection using SVMerge by integration of multiple detection methods and local assembly. *Genome Biol.* 2010; 11(12):R128. <https://doi.org/10.1186/gb-2010-11-12-r128> PMID: 21194472
29. Parra G, Bradnam K, Korf I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics.* 2007; 23: 1061–7. <https://doi.org/10.1093/bioinformatics/btm071> PMID: 17332020
30. Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, et al. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res.* 2011; 39:W316–22. <https://doi.org/10.1093/nar/gkr483> PMID: 21715386
31. Wu J, Mao X, Cai T, Luo J, Wei L. KOBAS server: a web-based platform for automated annotation and pathway identification. *Nucleic Acids Res.* 2006; 34:W720–4. <https://doi.org/10.1093/nar/gkl167> PMID: 16845106
32. Mandon-Pépin B, Oustry-Vaiman A, Vigier B, Piumi F, Crihiu E, Cotinot C. Expression profiles and chromosomal localization of genes controlling meiosis and follicular development in the sheep ovary. *Biol Reprod.* 2003; 68(3):985–95. PMID: 12604652
33. Fry SA, Robertson CE, Swann R, Dwek MV. Cadherin-5: a biomarker for metastatic breast cancer with optimum efficacy in oestrogen receptor-positive breast cancers with vascular invasion. *Br J Cancer.* 2016; 114(9):1019–26. <https://doi.org/10.1038/bjc.2016.66> PMID: 27010749
34. Su Q, Lei T, Zhang M. Association of ferritin with prostate cancer. *J BUON.* 2017; 22(3):766–770. PMID: 28730787
35. Ma DD, Wang DH, Yang WX. Kinesins in spermatogenesis. *Biol Reprod.* 2017; 96(2):267–276. <https://doi.org/10.1095/biolreprod.116.144113> PMID: 28203733
36. VanGompel MJ, Xu EY. A novel requirement in mammalian spermatid differentiation for the DAZ-family protein Boule. *Hum Mol Genet.* 2010; 19(12):2360–9. <https://doi.org/10.1093/hmg/ddq109> PMID: 20335278
37. Báez-Vega PM, Echevarría Vargas IM, Valiyeva F, Encarnación-Rosado J, Roman A, Flores J, et al. Targeting miR-21-3p inhibits proliferation and invasion of ovarian cancer cells. *Oncotarget.* 2016; 7(24):36321–36337. <https://doi.org/10.18632/oncotarget.9216> PMID: 27166999

38. Dall'Olio S, Fontanesi L, Tognazzi L, Russo V. Genetic structure of candidate genes for litter size in Italian Large White pigs. *Vet Res Commun.* 2010; 34 Suppl 1:S203–6. <https://doi.org/10.1007/s11259-010-9380-7> PMID: 20473567
39. Takenaka K, Fukami K, Otsuki M, Nakamura Y, Kataoka Y, Wada M, et al. Role of phospholipase C-L2, a novel phospholipase C-like protein that lacks lipase activity, in B-cell receptor signaling. *Mol Cell Biol.* 2003; 23(20):7329–38. <https://doi.org/10.1128/MCB.23.20.7329-7338.2003> PMID: 14517301
40. Tsai PL, Zhao C, Turner E, Schlieker C. The Lamin B receptor is essential for cholesterol synthesis and perturbed by disease-causing mutations. *Elife.* 2016; 5. pii: e16011. <https://doi.org/10.7554/eLife.16011> PMID: 27336722
41. Baba T, Mori S, Matsumura N, Kariya M, Murphy SK, Kondoh E, et al. Trophinin is a potent prognostic marker of ovarian cancer involved in platinum sensitivity. *Biochem Biophys Res Commun.* 2007; 360(2):363–9. <https://doi.org/10.1016/j.bbrc.2007.06.070> PMID: 17597582
42. Mei J, Liu S, Li Z, Gui JF. Mtmr8 is essential for vasculature development in zebrafish embryos. *BMC Dev Biol.* 2010; 10:96. <https://doi.org/10.1186/1471-213X-10-96> PMID: 20815916
43. Yang D, Sun B, Zhang X, Cheng D, Yu X, Yan L, et al. Huwe1 Sustains Normal Ovarian Epithelial Cell Transformation and Tumor Growth through the Histone H1.3-H19 Cascade. *Cancer Res.* 2017; 277(18):4773–4784. <https://doi.org/10.1158/0008-5472.CAN-16-2597> PMID: 28687618
44. Shao P, Liu Q, Maina PK, Cui J, Bair TB, Li T, et al. Histone demethylase PHF8 promotes epithelial to mesenchymal transition and breast tumorigenesis. *Nucleic Acids Res.* 2017; 45(4):1687–1702. <https://doi.org/10.1093/nar/gkw1093> PMID: 27899639
45. Mukherjee D, Lu H, Yu L, He C, Lahiri SK, Li T, Zhao J. Krüppel-like factor 8 activates the transcription of C-X-C cytokine receptor type 4 to promote breast cancer cell invasion, transendothelial migration and metastasis. *Oncotarget.* 2016; 7(17):23552–68. <https://doi.org/10.18632/oncotarget.8083> PMID: 26993780
46. Tanaka R, Izumi H, Kuroiwa A. Androgens and androgen receptor signaling contribute to ovarian development in the chicken embryo. *Mol Cell Endocrinol.* 2017; 443:114–120. <https://doi.org/10.1016/j.mce.2017.01.008> PMID: 28087386
47. Liang J, Wang J, Azfer A, Song W, Tromp G, Kolattukudy PE, et al. A novel CCCH-zinc finger protein family regulates proinflammatory activation of macrophages. *J Biol Chem.* 2008; 283(10):6337–46. <https://doi.org/10.1074/jbc.M707861200> PMID: 18178554
48. Yang S, Li X, Li K, Fan B, Tang Z. A genome-wide scan for signatures of selection in Chinese indigenous and commercial pig breeds. *BMC Genet.* 2014; 15:7. <https://doi.org/10.1186/1471-2156-15-7> PMID: 24422716
49. Ng PC, Levy S, Huang J, Stockwell TB, Walenz BP, Li K, et al. Genetic variation in an individual human exome. *PLoS Genet.* 2008; 4(8):e1000160. <https://doi.org/10.1371/journal.pgen.1000160> PMID: 18704161
50. Vernon HJ, McClellan R, Batista DA, Naidu S. Mutations in FARS2 and non-fatal mitochondrial dysfunction in two siblings. *Am J Med Genet A.* 2015; 167A(5):1147–51. <https://doi.org/10.1002/ajmg.a.36993> PMID: 25851414
51. Yadegari H, Biswas A, Akhter MS, Driesen J, Ivaskевичius V, Marquardt N, et al. Intron retention resulting from a silent mutation in the VWF gene that structurally influences the 5' splice site. *Blood.* 2016; 128(17):2144–2152. <https://doi.org/10.1182/blood-2016-02-699686> PMID: 27543438
52. Chang TH, Huang HY, Hsu JB, Weng SL, Horng JT, Huang HD. An enhanced computational platform for investigating the roles of regulatory RNA and for identifying functional RNA motifs. *BMC Bioinformatics.* 2013; 14 Suppl 2:S4. <https://doi.org/10.1186/1471-2105-14-S2-S4> PMID: 23369107
53. Sugnet CW, Kent WJ, Ares M Jr, Haussler D. Transcriptome and genome conservation of alternative splicing events in humans and mice. *Pac Symp Biocomput.* 2004:66–77. PMID: 14992493
54. Kopelman NM, Lancet D, Yanai I. Alternative splicing and gene duplication are inversely correlated evolutionary mechanisms. *Nat Genet.* 2005; 37(6):588–9. <https://doi.org/10.1038/ng1575> PMID: 15895079
55. Lopez-Garcia J, Periyasamy M, Thomas RS, Christian M, Leao M, Jat P, et al. ZNF366 is an estrogen receptor corepressor that acts through CtBP and histone deacetylases. *Nucleic Acids Res.* 2006; 34(21):6126–36. <https://doi.org/10.1093/nar/gkl875> PMID: 17085477
56. Winkler GS, Mulder KW, Bardwell VJ, Kalkhoven E, Timmers HT. Human Ccr4-Not complex is a ligand-dependent repressor of nuclear receptor-mediated transcription. *EMBO J.* 2006; 25(13):3089–99. <https://doi.org/10.1038/sj.emboj.7601194> PMID: 16778766
57. Hewitt SC, Harrell JC, Korach KS. Lessons in estrogen biology from knockout and transgenic animals. *Annu Rev Physiol.* 2005; 67:285–308. <https://doi.org/10.1146/annurev.physiol.67.040403.115914> PMID: 15709960

58. Oosterwegel M, van de Wetering M, Timmerman J, Kruisbeek A, Destree O, Meijlink F, et al. Differential expression of the HMG box factors TCF-1 and LEF-1 during murine embryogenesis. *Development*. 1993; 118(2):439–48. PMID: [8223271](#).
59. Hayashi K, Burghardt RC, Bazer FW, Spencer TE. WNTs in the ovine uterus: potential regulation of periimplantation ovine conceptus development. *Endocrinology*. 2007; 148(7):3496–506. <https://doi.org/10.1210/en.2007-0283> PMID: [17431004](#)
60. Chen W, Roeder RG. Mediator-dependent nuclear receptor function. *Semin Cell Dev Biol*. 2011; 22(7):749–58. <https://doi.org/10.1016/j.semcdb.2011.07.026> PMID: [21854863](#)
61. Huszar JM, Jia Y, Reddy JK, Payne CJ. Med1 regulates meiotic progression during spermatogenesis in mice. *Reproduction*. 2015; 149(6):597–604. <https://doi.org/10.1530/REP-14-0483> PMID: [25778538](#)
62. Song G, Bailey DW, Dunlap KA, Burghardt RC, Spencer TE, Bazer FW, et al. Cathepsin B, cathepsin L, and cystatin C in the porcine uterus and placenta: potential roles in endometrial/placental remodeling and in fluid-phase transport of proteins secreted by uterine epithelia across placental areolae. *Biol Reprod*. 2010; 82(5):854–64. <https://doi.org/10.1095/biolreprod.109.080929> PMID: [20107207](#)
63. Nandi A, Sinha N, Ong E, Sonmez H, Poretsky L. Is there a role for vitamin D in human reproduction? *Horm Mol Biol Clin Investig*. 2016; 25(1):15–28. <https://doi.org/10.1515/hmbci-2015-0051> PMID: [26943610](#)
64. Kim SH, Bennett PR, Terzidou V. Advances in the role of oxytocin receptors in human parturition. *Mol Cell Endocrinol*. 2017; 449:56–63. <https://doi.org/10.1016/j.mce.2017.01.034> PMID: [28119132](#)
65. Liu Q, Miller LC, Blecha F, Sang Y. Reduction of infection by inhibiting mTOR pathway is associated with reversed repression of type I interferon by porcine reproductive and respiratory syndrome virus. *J Gen Virol*. 2017; 98(6):1316–1328. <https://doi.org/10.1099/jgv.0.000802> PMID: [28613152](#)
66. Marsden WN. Synaptic plasticity in depression: molecular, cellular and functional correlates. *Prog Neuropsychopharmacol Biol Psychiatry*. 2013; 43:168–84. <https://doi.org/10.1016/j.pnpbp.2012.12.012> PMID: [23268191](#)
67. McCann CM, Tapia JC, Kim H, Coggan JS, Lichtman JW. Rapid and modifiable neurotransmitter receptor dynamics at a neuronal synapse in vivo. *Nat Neurosci*. 2008; 11(7):807–15. <https://doi.org/10.1038/nn.2145> PMID: [18568021](#)
68. Negishi M, Oinuma I, Katoh H. Plexins: axon guidance and signal transduction. *Cell Mol Life Sci*. 2005; 62(12):1363–71. <https://doi.org/10.1007/s00018-005-5018-2> PMID: [15818466](#)
69. Turi GF, Liposits Z, Hrabovszky E. Cholinergic afferents to gonadotropin-releasing hormone neurons of the rat. *Neurochem Int*. 2008; 52(4–5):723–8. <https://doi.org/10.1016/j.neuint.2007.09.001> PMID: [17933433](#)
70. Bruel-Jungerman E, Lucassen PJ, Francis F. Cholinergic influences on cortical development and adult neurogenesis. *Behav Brain Res*. 2011; 221(2):379–88. <https://doi.org/10.1016/j.bbr.2011.01.021> PMID: [21272598](#)
71. Anderson CM, Hu J, Barnes RM, Heidt AB, Cornelissen I, Black BL. Myocyte enhancer factor 2C function in skeletal muscle is required for normal growth and glucose metabolism in mice. *Skelet Muscle*. 2015; 5:7. <https://doi.org/10.1186/s13395-015-0031-0> PMID: [25789156](#)
72. Martineau C, Najyb O, Signor C, Rassart É, Moreau R. Apolipoprotein D deficiency is associated to high bone turnover, low bone mass and impaired osteoblastic function in aged female mice. *Metabolism*. 2016; 65(9):1247–58. <https://doi.org/10.1016/j.metabol.2016.05.007> PMID: [27506732](#)
73. Jones JL, Clemmons DR. Insulin-like growth factors and their binding proteins: biological actions. *Endocr Rev*. 1995; 16(1):3–34. <https://doi.org/10.1210/edrv-16-1-3> PMID: [7758431](#)
74. Alam I, Sun Q, Liu L, Koller DL, Liu Y, Edenberg HJ, et al. Genomic expression analysis of rat chromosome 4 for skeletal traits at femoral neck. *Physiol Genomics*. 2008; 35(2):191–6. <https://doi.org/10.1152/physiolgenomics.90237.2008> PMID: [18728226](#)
75. Chen S, An J, Lian L, Qu L, Zheng J, Xu G, et al. Polymorphisms in AKT3, FIGF, PRKAG3, and TGF- $\beta$  genes are associated with myofiber characteristics in chickens. *Poult Sci*. 2013; 92(2):325–30. <https://doi.org/10.3382/ps.2012-02766> PMID: [23300296](#)
76. van der Velden JL, Langen RC, Kelders MC, Wouters EF, Janssen-Heininger YM, Schols AM. Inhibition of glycogen synthase kinase-3 $\beta$  activity is sufficient to stimulate myogenic differentiation. *Am J Physiol Cell Physiol*. 2006; 290(2):C453–62. <https://doi.org/10.1152/ajpcell.00068.2005> PMID: [16162663](#)
77. van der Velden JL, Langen RC, Kelders MC, Willems J, Wouters EF, Janssen-Heininger YM, et al. Myogenic differentiation during regrowth of atrophied skeletal muscle is associated with inactivation of GSK-3 $\beta$ . *Am J Physiol Cell Physiol*. 2007; 292(5):C1636–44. <https://doi.org/10.1152/ajpcell.00504.2006> PMID: [17166938](#)

78. Ocaranza P, Lammoglia JJ, Iñiguez G, Román R, Cassorla F. Effects of thyroid hormone on the GH signal transduction pathway. *Growth Horm IGF Res.* 2014; 24(1):42–6. <https://doi.org/10.1016/j.ghir.2014.01.001> PMID: [24439614](https://pubmed.ncbi.nlm.nih.gov/24439614/)
79. Zhou C, Zou J, Zou S, Li X. INO80 is Required for Osteogenic Differentiation of Human Mesenchymal Stem Cells. *Sci Rep.* 2016; 6:35924. <https://doi.org/10.1038/srep35924> PMID: [27804957](https://pubmed.ncbi.nlm.nih.gov/27804957/)