


RESEARCH

Open Access



# Associations of genome-wide structural variations with phenotypic differences in cross-bred Eurasian pigs

Wencheng Zong<sup>1</sup>, Jinbu Wang<sup>1</sup>, Runze Zhao<sup>1,2</sup>, Naiqi Niu<sup>1</sup>, Yanfang Su<sup>1</sup>, Ziping Hu<sup>1,3</sup>, Xin Liu<sup>1</sup>, Xinhua Hou<sup>1</sup>, Ligang Wang<sup>1</sup>, Lixian Wang<sup>1\*</sup> and Longchao Zhang<sup>1\*</sup> 

## Abstract

**Background** During approximately 10,000 years of domestication and selection, a large number of structural variations (SVs) have emerged in the genome of pig breeds, profoundly influencing their phenotypes and the ability to adapt to the local environment. SVs ( $\geq 50$  bp) are widely distributed in the genome, mainly in the form of insertion (INS), mobile element insertion (MEI), deletion (DEL), duplication (DUP), inversion (INV), and translocation (TRA). While studies have investigated the SVs in pig genomes, genome-wide association studies (GWAS)-based on SVs have been rarely conducted.

**Results** Here, we obtained a high-quality SV map containing 123,151 SVs from 15 Large White and 15 Min pigs through integrating the power of several SV tools, with 53.95% of the SVs being reported for the first time. These high-quality SVs were used to recover the population genetic structure, confirming the accuracy of genotyping. Potential functional SV loci were then identified based on positional effects and breed stratification. Finally, GWAS were performed for 36 traits by genotyping the screened potential causal loci in the F2 population according to their corresponding genomic positions. We identified a large number of loci involved in 8 carcass traits and 6 skeletal traits on chromosome 7, with *FKBP5* containing the most significant SV locus for almost all traits. In addition, we found several significant loci in intramuscular fat, abdominal circumference, heart weight, and liver weight, etc.

**Conclusions** We constructed a high-quality SV map using high-coverage sequencing data and then analyzed them by performing GWAS for 25 carcass traits, 7 skeletal traits, and 4 meat quality traits to determine that SVs may affect body size between European and Chinese pig breeds.

**Keywords** Body size, GWAS, Pig, Skeleton, Structural variations

## Background

Genome rearrangements generate an abundance of structural variations (SVs) that, despite occurring mainly in non-coding regions, can determine the binding of transcriptional regulatory elements, mRNA splicing and processing, genome folding and higher order structures, and translational alterations due to their size and location [1, 2]. In general, SVs can be divided into two types based on changes in the DNA content of the genome: 1) unbalanced copy number variants (CNVs), including deletions (DELs), duplications (DUPS), insertions (INSs),

\*Correspondence:

Lixian Wang  
iaswlx@263.net  
Longchao Zhang  
zhlichias@163.com

<sup>1</sup> State Key Laboratory of Animal Biotech Breeding, Institute of Animal Sciences, Chinese Academy of Agricultural Sciences, Beijing 100193, China

<sup>2</sup> College of Animal Science, Shanxi Agricultural University, Jinzhong 030801, China

<sup>3</sup> College of Animal Science and Technology, Qingdao Agricultural University, Qingdao 266109, China



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

and mobile element insertions (MEIs); and 2) balanced rearrangements, including inversions (INVs) and translocations (TRAs) [3]. In livestock research, SVs have been shown to be associated with adaptability and production traits [4, 5]. Compared to SNPs, SVs contribute to a higher proportion of complex phenotypes [6]. As a consensus, SVs are defined as a significant mutational force shaping genome evolution and function [7].

During the long process of domestication and selection, which began about 10,000 years ago in Europe (Near East) and Asia (China), pig breeds with independent biological traits and breed-specific genomic variants have emerged [8]. The Large White pig is a common Western commercial breed with a long carcass, fast growth rate, high lean meat ratio, and high feed utilization efficiency [9]. In contrast, Min pigs distributed in Northeastern China perform relatively poorly for these traits, but show better tolerance to harsh conditions, roughage, and have high intramuscular fat [10]. The construction of reference populations has become an important method for determining the associations between genomic variants and phenotypic differences in agricultural research, and genome-wide association studies (GWAS) have provided a wealth of new information in the last decade [11, 12]. Here, an F2 population was constructed based on a cross between 4 Large White and 15 Min pigs, which were selected for traits according to Mendel's law of free association and provided an opportunity to subsequently study phenotypic differences.

In recent years, several SV studies on pig genome have been reported [13–16]. Nevertheless, GWAS based on SVs are rarely reported, which limits our understanding of the potential function and exploitation of SVs as genetic markers. Previous SV studies were performed at a low sequencing depth, which may reduce the sensitivity and accuracy of SV identifications. As a consensus, most SV studies rely on multiple software to increase the number and accuracy of SV identifications, but this approach is usually computationally resource-intensive and time-consuming in population-scale studies. Moreover, most SV-calling software does not recognize insertion variants or subsequently does not genotype them accurately, which has led most SV studies to ignore the contribution of insertion variants to animal phenotypes.

In this context, a high-quality SV map containing 4 SV types was constructed using resequencing data with an average depth of more than 35×. Subsequently, potential causal loci that were screened between breeds were rapidly genotyped in a large population of offspring according to our own designed strategy. Finally, we performed GWAS for 36 traits to determine the effect of SVs on phenotype. To our knowledge, this may be the most comprehensive trait association study of the pig genome using

SV markers to date. In conclusion, our study provides a new strategy for SV research at the population scale that has been demonstrated to be reliable and efficient. Meanwhile, this study will provide a new theoretical basis for using SVs as molecular markers or developing marker-assisted selection and deepen the understanding of the potential function of SVs in the pig genome.

## Methods

### Animal collection

The pigs used in the experiment were all from the Large White×Min pig resource population, raised at the Changping pig farm of the Institute of Animal Sciences, Chinese Academy of Agricultural Sciences. The Large White×Min pig resource population included 19 F0 individuals and 513 F2 individuals, among which the F0 individuals included 4 Large White and 15 Min pigs. The 15 Min pigs were collected from the Jilin Academy of Agricultural Sciences (JAAS) and the rural areas of Northeastern China, and the 4 Large White pigs were from the United Kingdom. All F2 individuals were raised to market age ( $240 \pm 7$  d) and slaughtered for commercial purposes. Paired-end sequencing was performed using Illumina Hi-seq 2500, with a sequencing depth of  $>30\times$  for F0 individuals and  $5-7\times$  for F2 individuals. The sequencing data of the F0 and F2 individuals used in this study have been submitted to the Genome Sequence Archive (GSA) with the accession number CRA002451. We downloaded data for additional 11 Large White pigs from the NCBI Sequence Read Archive (SRA, <https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJEB39374>), and information for all individuals is shown in Table S1, Additional file 1.

### Phenotype determination

The F2 population constructed with Large White and Min pigs collected abundant quantitative traits. For the convenience of this study, we divided all the traits into 25 carcass traits, 7 skeletal traits, and 4 meat quality traits. Among them, carcass traits include carcass length, body length, body height, cannon circumference, scapular width, chest width, chest depth, abdominal circumference, waist width, hip width, hip length, hip circumference, bone rate, total weight of front bone, total weight of middle bone, total weight of hind bone, total weight of front lean meat, total weight of middle lean meat, total weight of hind lean meat, total weight of front fat, total weight of middle fat, total weight of hind fat, heart weight, liver weight, and lung weight, skeletal traits include scapula length, humerus length, forearm bone length, hip bone length, femur length, calf bone length, and vertebral number, and meat quality traits include marbling, intramuscular fat, tenderness, and moisture

percentage. All phenotypic characteristics were defined according to the Animal Genetic Resources in China (pig) [10], Wikipedia (<https://en.wikipedia.org>), published breed genetic resource studies, and the official websites of pig breeds.

#### Data processing

The raw reads were trimmed with Trimmomatic [17], and high-quality trimmed reads were aligned against the pig reference genome (Sscrofa11.1) with Bwa [18]. Samtools [19] was employed to convert sam files to bam format and subsequently sort and index the bam files. The PCR duplicates were marked with Picard [20]. Samples that had more than one associated bam file were merged with samtools. The sequencing depth statistics of all samples were calculated using mosdepth [21].

#### SV calling, filtering, and validation

The primary research approach was to employ several software to detect SVs, essential to maximize the obtained SV loci. Thus, five SV software were selected for SV discovery: Delly v0.9.1, Smoove v0.2.8, Manta v1.6.0, Breakdancer v1.4.5, and MELT v2.2.2. Among them, Delly, Smoove, Manta, and Breakdancer were used to identify DELs, DUPs, and INVs, whereas MELT was used to identify MEIs. Previous studies have relied on the overlap of results from several different SV software, although this strategy does not reliably improve detection and may even aggravate false discoveries [22]. In this study, we merged the results of several software analyses instead of including only overlaps, to maximize the number of SV loci obtained. Among these software, Smoove, Delly, Breakdancer, and Manta called SVs according to default parameters. For MELT, Repeatmasker [23] was employed to annotate the pig reference genome for MEI using the RepeatMasker libraries (2018-10-26) of Repbase database (<https://www.girinst.org/>), and the three most widely distributed types of ERV, LINE, and SINE were selected for MEI calls, and the reference sequences are displayed in Table S2. Survivor [24] software was used to merge the SV datasets of each software, which were defined and merged according to a distance of 1,000 bp between breakpoints, taking into account the strands and types of SVs. Subsequently, all individuals were merged to generate the SV map. Genotyping of DELs, DUPs, and INVs was performed using SVtyper [25], whereas that of MEIs was performed using MELT.

To control the quality of all SV loci, we also set a strict filter for each SV locus, keeping only loci with  $QUAL > 200$  for DELs, DUPs, and INVs and only loci with "PASS" for MEIs, to ensure the quality of each SV locus. For INVs, variants within 1,000 bp could not be merged

because of the reverse position of breakpoints output by different software. We merged the loci with more than 75% overlap between each breakpoint using our own written script. Meanwhile, these loci were divided into two groups,  $< 100$  kb and  $> 100$  kb, and merged separately in order to avoid large variants covering small variants. For SVs of 1–10 Mb, we used Samplot [26] for visualization and randomly selected three loci of each SV type for PCR validation to assess accuracy. The primer design schematics and primer sequence information for all SV types are shown in Fig. S1 (Additional file 2) and Table S3 (Additional file 1) respectively. To test the accuracy of SV genotyping, we randomly selected one SV locus per chromosome and verified them by PCR. For DEL, DUP, and MEL, we determined the primers based on around 500 bp of the breakpoints on both sides of the SV locus, whereas for INV, one primer is set about 500 bp upstream or downstream of the SV breakpoint and the other primer is set inside the INV. PCR was then performed using DNA from 19 F0 individuals, and the primer sequences are listed in Table S4, Additional file 1.

#### Population genetics, SV functional, and $F_{ST}$ analysis

The sample geographic distribution map was produced using the ggplot2 [27] and ggspatial packages in R. Principal component analysis (PCA) was performed using the GCTA [28] software. Population structure was evaluated using Admixture [29] and three possible populations ( $K=2-4$ ) were calculated. Next, the ggplot2 package was employed to plot the PCA and population structure results. The neighbor-joining trees were constructed using Phylip [30] and visualized by MEGA11 [31]. SV distribution locations were determined based on gene location annotations from the Ensemble database, and SV effects were estimated by Snpeff [32] based on the locations of SV breakpoints. In order to identify the breed-stratified SVs, the VCFtools [33] was used to calculate the fixation index ( $F_{ST}$ ) values of all SVs with the Weir and Cockerham method comparing 15 Min vs. 15 Large White pigs. All SVs within the differentiated region with  $F_{ST}$  values in the top 5% were selected, and the corresponding genes overlapping with SVs were considered as candidates for breed stratification based on the gene information annotated by Ensemble. The candidate genes were annotated using Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) functional enrichment analysis by DAVID online webloci (<https://david.ncicrf.gov/>). The wordcloud webloci (<https://www.jasondavies.com/wordcloud/>) was used to create word clouds for the top 10 significantly enriched terms ( $P < 0.05$ ) to reveal underlying molecular mechanisms.

### Population-specific SV screening, genotyping, and GWAS

In order to obtain potential causal loci affecting specific traits, a screening strategy for differentially genotyped loci was used. Six relatively purebred Min pigs (M1, M3–M7) and 15 Large White pigs were selected for this screening strategy. Two conditions were set: first, one genotype was selected for one breed and the remaining two genotypes were set for the other breed; second, the frequency of genotypes was more than 80% for Min pigs ( $\geq 5$  individuals) and more than 50% for Large White pigs ( $\geq 8$  individuals) (Fig. S2, Additional file 2). An SV locus was considered a candidate locus when it matched the above two conditions. The screened SV loci were genotyped using the bam file of the F2 individuals, among which, DELs, DUPs, and INVs were genotyped using SVtyper, whereas MEIs were genotyped using paragraph [34]. Plink [35] was performed to filter all genotyped SV loci with the following specific parameters: a sample call rate of  $>90\%$ , an SV call rate of  $>90\%$ , and a minor allele frequency of  $>5\%$ . EMMAX [36] was employed to perform GWAS using a mixed linear model for all filtered SV loci. The sex and slaughter batch were used as fixed effects, and the PCA was used as a covariate. The significance cutoff was defined as the Bonferroni test threshold, which was set as  $0.05/(\text{total number of SVs})$ . All GWAS results were visualized using the rMVP [37] package.

## Result

### The landscape-wide SV discovery in the genomes of the Large White and Min pig

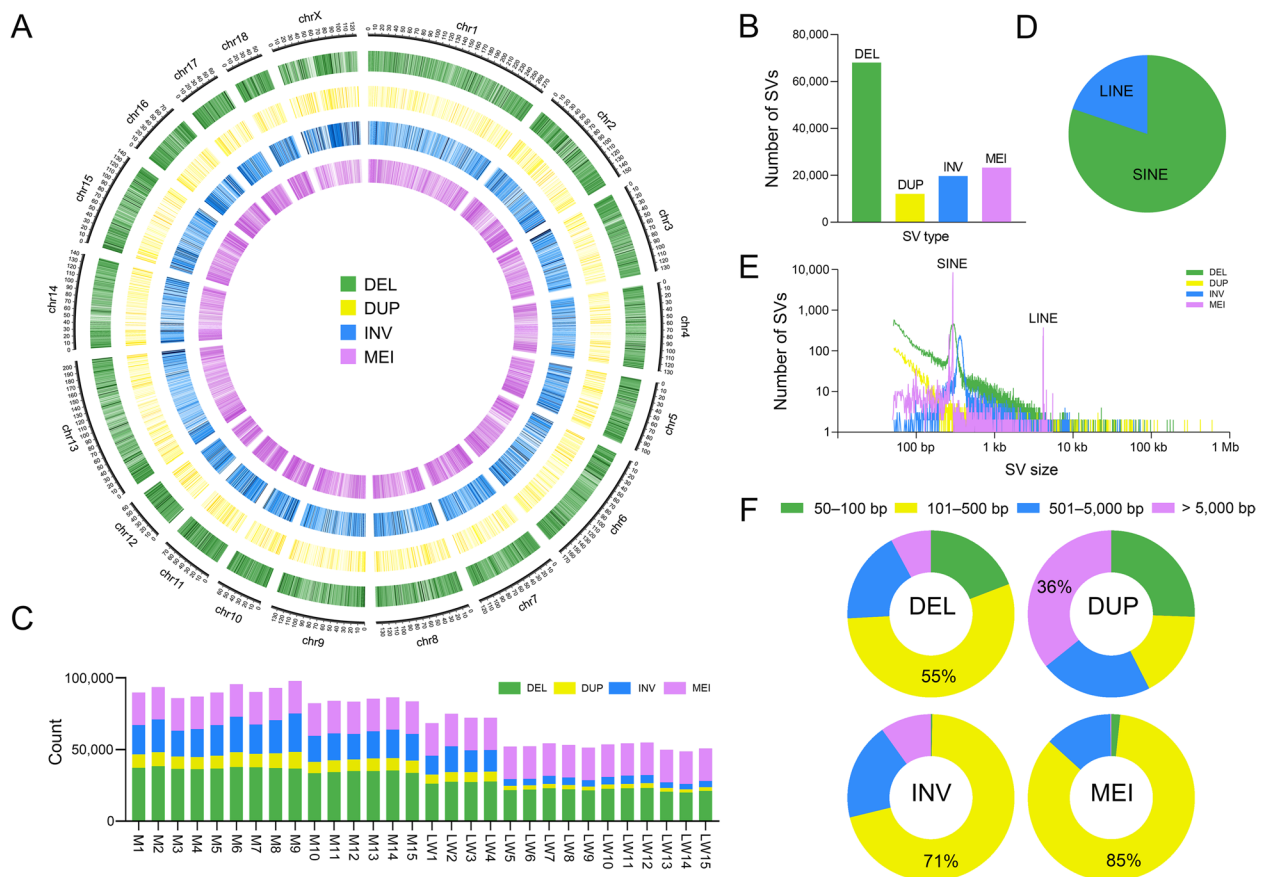
We developed an SV study for the Large White  $\times$  Min pig resource population, which included 19 F0 individuals and 513 F2 individuals. Among them, the F0 individuals involved 4 Large White and 15 Min pigs. In order to obtain a more comprehensive SV landscape, we collected additional 11 Large White pigs (PRJEB39374). Then, whole-genome resequencing data were obtained for the 15 Large White and 15 Min pigs, with a total of approximately 1,396 G data obtained after quality control and an average sequencing depth of more than  $35\times$ . The sequencing data were aligned to the Sscrofa11.1 using BWA-mem, and SVs were called by Delly [38], Manta [39], Smoove (<https://github.com/brentp/smoove>), Breakdancer [40], and MELT [41]. All variants from the results of each SV tool were merged individually and then further merged at the population level based on SV breakpoint locations, types, and orientations, thus generating an initial SV dataset. After quality control, manual merging, and genotyping to produce the final SV landscape, a detailed SV analysis pipeline was created, which is presented in Fig. S3, Additional file 2. To exclude false positives generated from NGS data due to length limitations, most previous studies limited the

SV length to 50 bp–10 Mb [13, 14, 42], but recent study used a range of 50 bp–1 Mb [43]. In this study, we visualized all loci from 1–10 Mb using Samplot and randomly selected three loci from DELs, DUPs, and INVs for PCR validation. We found that all SVs in the range of 1–10 Mb did not show the expected bands, and the genomic coverage map could not determine whether variants were present (Fig. S4–S6, Additional file 2). This result showed that SVs in the range of 1–10 Mb were not reliable, so loci  $>1$  Mb were excluded from this analysis. In addition, to assess the quality of the SVs identified, one locus per chromosome was selected randomly for PCR analysis using all F0 individuals. The results showed that almost all loci showed bands of the expected size with an accuracy of 94.46% (Fig. S7, Additional file 2).

A high-quality landscape with 123,151 SV loci, including 68,121 DELs, 12,045 DUPs, 19,727 INVs, and 23,258 MEIs (Fig. 1A and B) was obtained from the analysis. We compared the identified SV loci with the Ensemble public SV database (version 22-08-26) and filtered them with a 75% overlap rate, and a total of 66,435 new SV loci were found, which greatly enriched the public SV database (Fig. S8, Additional file 2). This result suggests that the sensitivity of each software differs for different genomic regions and that combining the results of multiple software for identifying SVs and using high-coverage sequencing data would significantly increase the number of new SV loci discovered. Among the 19 F0 individuals, more SVs were identified in Min pigs than in Large White pigs (Fig. 1C and Table S5, Additional file 1). Compared to European pigs, the genomes of Chinese local pig breeds confers higher genetic diversity [44]. In addition, fewer SVs were identified in the remaining 11 Large White pigs (LW5–LW15) than in the 19 F0 individuals, which may be related to the sequencing depth (Fig. 1C). SINEs and LINEs contributed to all MEI types (Fig. 1D), with SINEs accounting for more than 80% of the major insertion types. Notably, SINEs have previously been reported to contribute a large number of polymorphisms in the pig genome [45]. We further investigated the size distribution of the identified SVs with length between 50 bp and 1 Mb. Most SVs were small ( $<500$  bp), with a large number of SINEs and LINEs identified as the variant size increased (Fig. 1E and F). The DELs, INVs, and MEIs were mainly within the range of 100–500 bp, while the DUPs were mainly large SVs of  $>5,000$  bp (Fig. 1F and Table S6, Additional file 1).

### Population structure inference

To further confirm data quality, we used the discovered SVs to infer the population genetic structure of 15 Large White and 15 Min pigs (Fig. 2A–D). PCA was performed uniformly for all SV genotypes. The results



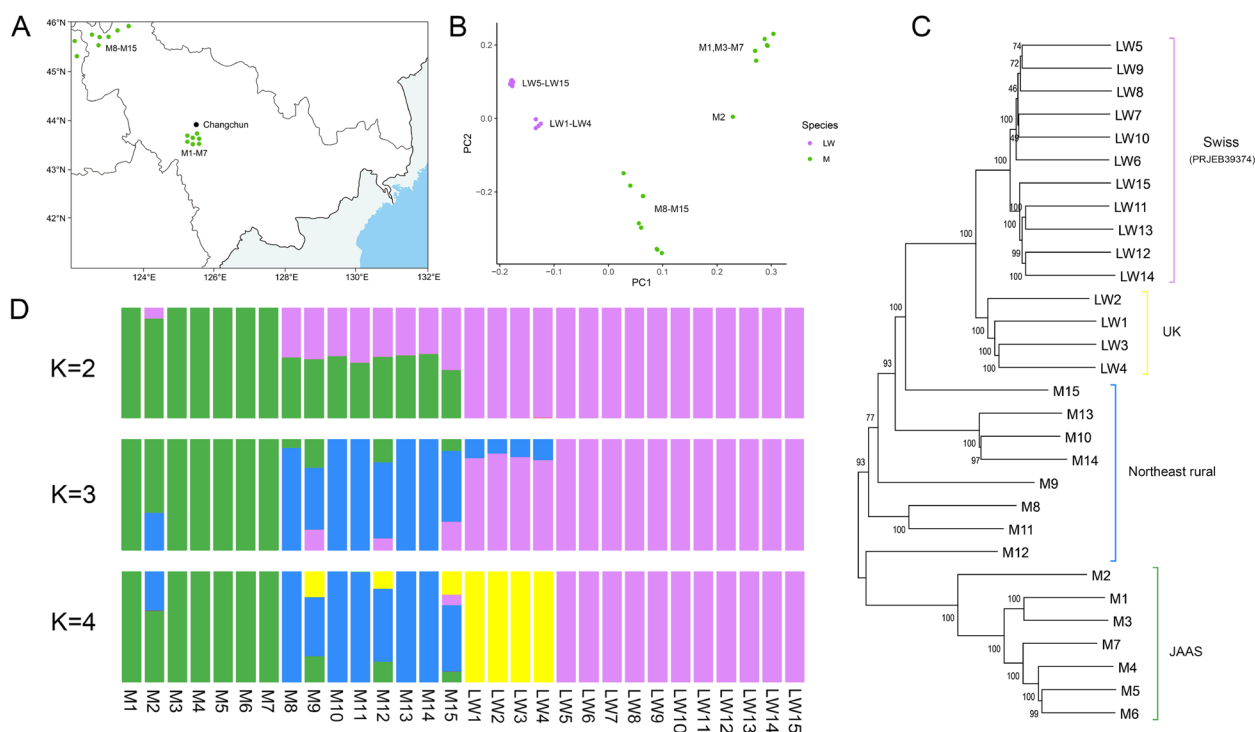
**Fig. 1** The SV landscape of Large White and Min pigs. **A** Distribution of the discovered SVs in the pig genome. The circle diagram shows the distribution of SVs in the chromosomes, where concentric circles show the following from outside to inside: DELs, DUPs, INVs, and MEIs. **B** Total number of SVs identified per type. Statistics on the number of loci per SV type in the final generated SV dataset. **C** Number of each SV type for 15 Large White and 15 Min pigs. Stacked bar graph showing the number of SVs initially called for each sample, containing 18 autosomes and X chromosome. **D** Percentage of transposon types in MEIs. The pie chart shows the percentage statistics of ERV, LINE, and SINE transposons. **E** SV size distribution per SV type with x-axis and y-axis shown in log<sub>10</sub> scale. **F** Distribution of length range per SV type. Four length ranges are labeled above the doughnut chart

confirmed the separation of the Min pigs into distinct groups from the Large White pigs (Fig. 2B). Depending on the sampling location (Fig. 2A), the Min pigs from the JAAS and those from other areas in Northeast China also imply different pedigrees. In addition, 4 Large White pigs of F<sub>0</sub> generation (UK) and 11 Large White pigs from Swiss (PRJEB39374) are also clearly separated. The above results, which are approximately the same as the results of the 50 k chip analysis, reconfirm the accuracy of SV genotyping (Fig. S9, Additional file 2). We also constructed a phylogenetic tree, which can be divided into four clusters (Fig. 2C): the first and second clusters are Swiss and UK Large White pigs, respectively, and the fourth cluster is JAAS Min pigs. The third cluster does not form a single cluster, implying that these individuals have a complex pedigree. Therefore, we further executed a population structure analysis (Fig. 2D) and found that the Min pigs

(M8–M15) sampled in rural areas of Northeastern China may have undergone crossbreeding, showing a clear exotic bloodline infiltration. This may be due to the introduction of commercial pig breed pedigrees by local people to improve economic efficiency.

#### Functional relevance of SVs

To explore the potential functions of SVs, we investigated their locations in the genome, including gene downstream, exon, intergenic, intron, gene upstream, and untranslated regions (UTR3 and UTR5). All four SV types are located mainly at intergenic and intron positions, with DELs, DUPs, INVs, and MEIs accounting for 96.75%, 95.15%, 97.40%, and 96.18%, respectively (Fig. 3A). The remaining SVs were located in the coding region, the untranslated region, and within 1 kb upstream and downstream of the gene. Approximately 42.40% of SVs overlapped, with one or more Ensemble genes. The



**Fig. 2** Population genetic analysis using SV markers. **A** The geographical distribution of Min pigs used in this study. **B** PCA derived from SVs. Purple represents Large White pigs and green represents Min pigs. **C** Phylogenetic tree constructed for Large White and Min pigs based on whole-genome SV data. Green represents Min pigs from JAAS, blue represents Min pigs from rural areas of Northeast China, yellow represents Large White pigs from the UK, and purple represents Large White pigs from Swiss. **D** Genome-wide admixture analyses inferred from SVs ( $K = 2, 3$ , and  $4$ ). Each individual is a vertical rectangle with different colors implying different genetic populations

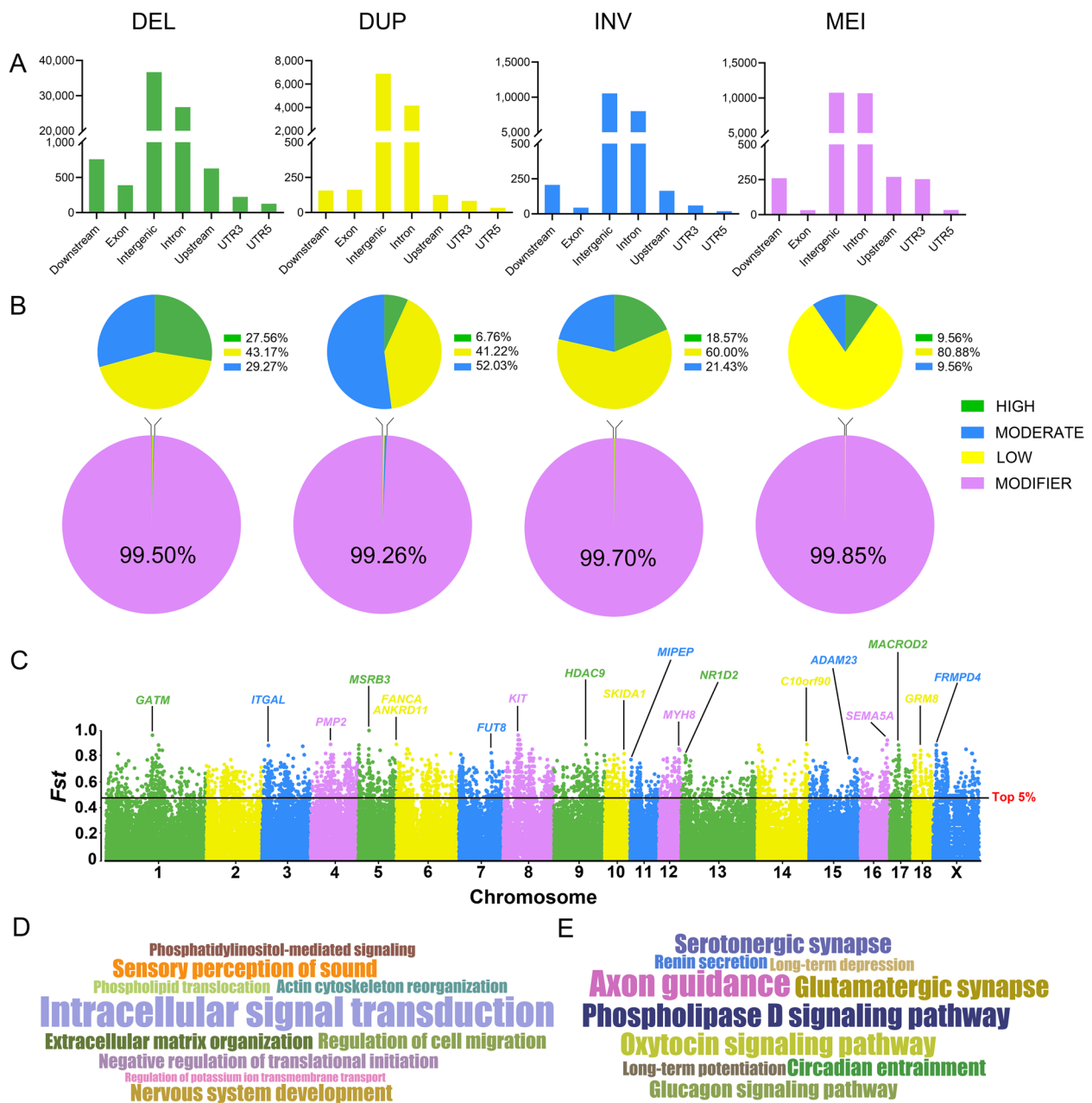
different types of SVs did not show a statistically specific preference for positional distribution, implying that the distribution of SVs was independent of the SV type.

We further predicted the effects for four SV types according to their distribution in the genome. Most of the SV effects were defined as "MODIFIER", implying they generally had no effect on genes (Fig. 3B). The remaining SV effects were defined as "HIGH", "MODERATE", and "LOW". Among them, the proportion of DELs, DUPs, INVs, and MEIs with "HIGH" effects was 27.56%, 6.76%, 18.57%, and 9.56%, respectively (Fig. 3B and Table S7, Additional file 1). The SVs with "HIGH" effects were annotated, which revealed that they involve several disease-related pathways, including Coronavirus disease-COVID-19 (*ssc05171*), Parkinson's disease (*ssc05012*), Non-alcoholic fatty liver disease (*ssc04932*), and Alzheimer disease (*ssc05010*) (Fig. S10, Additional file 2).

#### Breed-stratified SVs

In order to discover candidate adaptive SVs, we calculated the  $F_{ST}$  between 15 Large White and 15 Min pigs. The top 5% were identified as potential breed-stratified SV loci with a total of 3,797 DELs, 271 DUPs, 231 INVs, and 525 MEIs (Fig. 3C and Table S8, Additional

file 1). We annotated the SV loci with the highest  $F_{ST}$  values on each chromosome to identify potentially functional genes that may be affected. Among them, the SV locus with the highest  $F_{ST}$  was *MSRB3* on chromosome 5, which was previously reported to have a key role in pig ear size [46]. *MYH8* has been reported to be associated with muscle development and meat quality traits [47], and *NR1D2* is responsible for adipogenesis and lipid accumulation in the myocardium [48, 49]. The *KIT* locus is a key gene in determining coat color in different pig breeds [50]. *GATM* and *SEMA5A* are involved in placental development and embryonic development, respectively [51, 52]. *HDAC9* and *GRM8* are associated with eye muscle area and the relative area of type I fibers, respectively [53, 54]. *ITGAL* is immune-related and involved in leukocyte recruitment processes [55]. *FANCA* is associated with cell meiosis and germ cell development, and its mutation leads to reduced fertility and follicular reduction [56]. *ADAM23*, *ANKRD11*, and *MACROD2* function in the nervous system and are associated with several neurological disorders [57–61]. *FUT8* disruption leads to growth retardation, early postnatal developmental death, and emphysema-like changes in the lungs



**Fig. 3** Position effect estimation and  $F_{ST}$  screening. **A** Distribution of the location per SV type in the genome. The x-axis is the genomic positions and the y-axis is the number of SVs. **B** Proportion of effects predicted for each SV type. Predicted effects based on SV distribution locations in the pig genome were, in order, "MODIFIER", "LOW", "MODERATE", and "HIGH". **C** Manhattan plot based on Weir and Cockerham's fixed index ( $F_{ST}$ ) statistics. The nearest gene of the SV locus with the highest  $F_{ST}$  value was marked on each chromosome. **D** GO and **E** KEGG enrichment analysis based on the top 5% of  $F_{ST}$  loci overlapping genes. The font size of GO terms and KEGG pathways correlates with the number of enriched genes

[62]. *MIPEP* expression is up-regulated in response to heat stress [63], and *FRMPD4* is highly expressed in pig breeds with high teat numbers [64]. In addition, *SKIDA1* is related to the survival of human embryonic stem cells [65]. Then, the top 5% of SV loci-associated

genes were analyzed using GO and KEGG. A total of 4,824 common SV regions overlapped with 1,440 functional genes, contributing to the enriched terms and pathways. The top 10 significant GO terms and KEGG pathways were enriched for cellular processes and biological regulation, as well as for pathways associated

with nervous system function and the endocrine system (Fig. 3D and E). We found that most breed-stratified SVs focused on neurological-related pathways, which may emphasize a special role of these pathways in the domestication and selection of Large White and Min pigs.

#### GWAS found that SVs were mainly associated with the body size difference between Large White and Min pigs

The F2 population data processing resulted in a total of approximately 5,110 G of data at a depth of 5–7×. To identify SV loci with phenotypic variation due to differences between breeds, we performed a locus screen for differential genotypes between Large White and Min pigs, and then selected these SV loci for genotyping in the F2 population based on the corresponding genomic positions (Fig. 4A), as described in Methods. Finally, a total of 33,909 loci were screened, of which 97.15% (16,898,906/17,395,317) were successfully genotyped and then GWAS were performed. Bonferroni's multiple testing method was employed for *P*-value correction, which was defined as  $0.05/n$ , where *n* represents the number of SVs for each independent GWAS. A total of 36 traits were involved in the GWAS, including 25 carcass traits, 7 skeletal traits, and 4 meat quality traits.

We found that SVs may have an effect on pig body size, and the GWAS identified overlapping strong association peaks for carcass length, body length, body height, cannon circumference, and bone rate involving 87 significant loci (Fig. 4B–F and Table 1) overlapping with 25 protein-coding genes (Fig. 4G and H, and Table 1), including intron variants, as well as upstream and downstream variants of genes. Among them, a DEL of the intron region located in *FKBP5* is the most significant loci for all five traits (Fig. 4I and J, Fig. S11A, Additional file 2 and Table S9, Additional file 1), and this gene has been reported to be involved in osteoclast differentiation [66, 67]. A SINE insertion upstream of *ILRUN* is also one of the most significant loci (Fig. S11B and Table S9), and this gene has been frequently reported to be associated with human height [68–71] as well as carcass length,

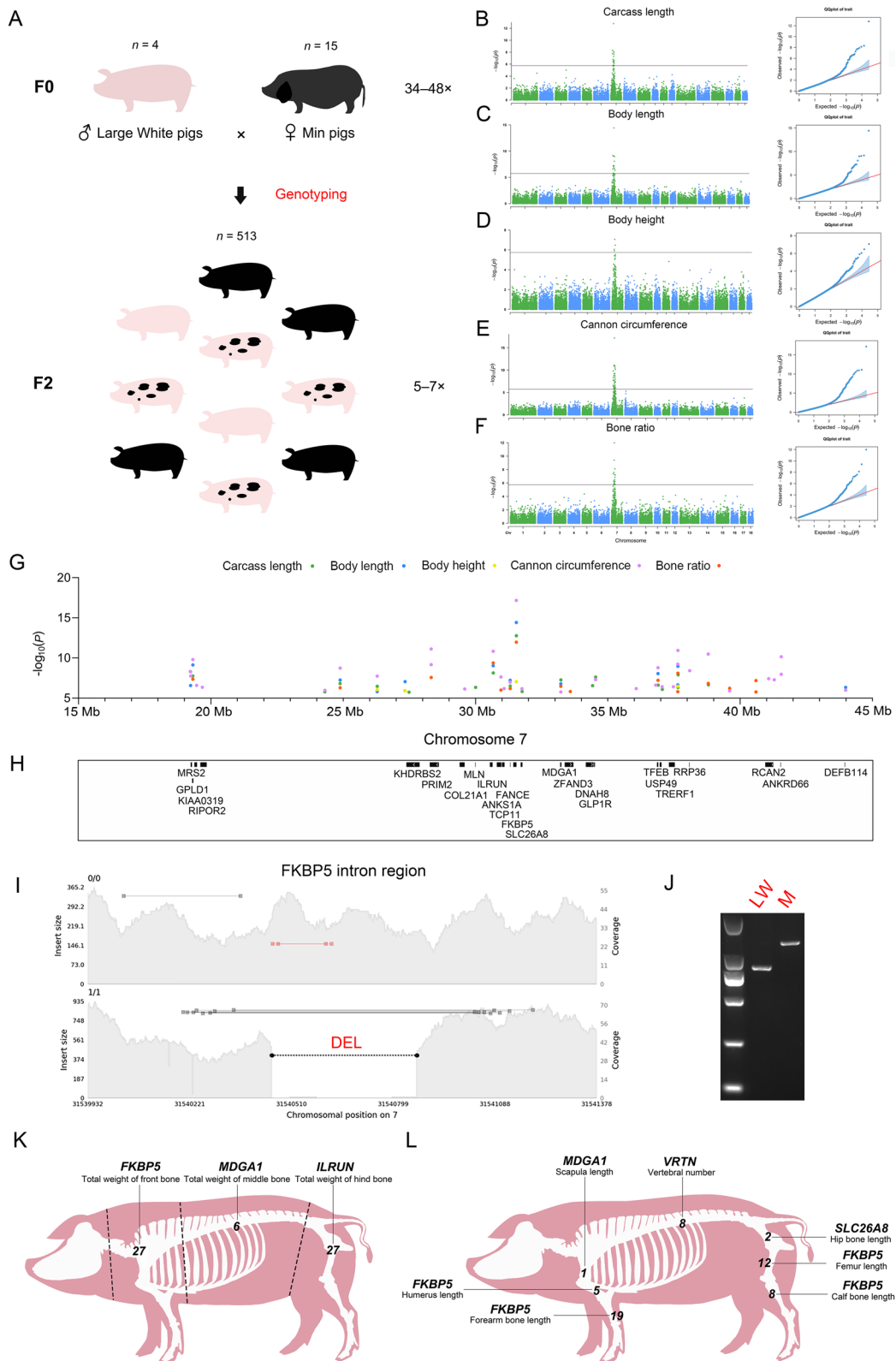
body length, and cannon circumference in pigs [72, 73]. A DEL was identified in the intron region of *TFEB* (Fig. S11C and Table S9), *RCAN2* (Fig. S11D and Table S9), and *ANKS1A* (Fig. S11E and Table S9), which have been reported to be associated with osteoblast differentiation [74], osteoblast function [75], and bone mineral density [76], respectively. In addition, a DEL was found upstream of *MRS2* (Fig. S11F and Table S9) and downstream of *GLPIR* (Fig. S11G and Table S9), respectively. Among them, *MRS2* is associated with  $Mg^{2+}$  expression, and lower  $Mg^{2+}$  levels stimulate osteoclast formation [77]. *GLPIR* plays a key role in bone strength and quality [78]. The GWAS results for the above five traits included multiple genes associated with skeleton, suggesting that SVs may affect skeletal size and thus pig body size. We also performed GWAS on scapular width, chest width, chest depth, abdominal circumference, waist width, hip width, hip length, and hip circumference (Fig. S12A–H, Additional file 2). The results showed no significant SV loci for the traits except two significant loci for abdominal circumference (Fig. S12D), which may imply that these carcass traits are poorly correlated with skeleton.

To verify whether there was position specificity in the effect of SVs on skeleton, we divided the pig carcass into three sections, namely front, middle, and hind, according to the position of 4–5 ribs and the lumbosacral joint after removing the head. We then performed GWAS for total weight of front bone, total weight of middle bone, total weight of hind bone, scapula length, humerus length, forearm bone length, hip bone length, femur length, calf bone length, and vertebral number. The results showed that all three sections of bone weight and six bone length traits showed strong association peaks (Fig. S13A–I, Additional file 2 and Table 1), which almost overlapped with previous GWAS results for carcass length, body length, body height, cannon circumference, and bone rate (Fig. S14, Additional file 2). The 7\_31540442 ( $P = 2.25255E-12$ ), 7\_33224291 ( $P = 6.19796E-09$ ), and 7\_30669698 ( $P = 7.95851E-13$ ) were the most significant SV loci in the three sections, corresponding to the three genes: *FKBP5*, *MDGA1*, and *ILRUN* (Fig. 4K). Moreover, *FKBP5* is the most significant gene for humerus, forearm

(See figure on next page.)

**Fig. 4** SV-based GWAS. **A** Schematic diagram of Large White × Min pig resource population. The numerals on the top of the pig image represent the number of samples, and the numerals on the right represent the range of sequencing depth. **B–F** Manhattan and quantum quantum (QQ) plots of associated SVs for carcass length (**B**), body length (**C**), body height (**D**), cannon circumference (**E**), and bone rate (**F**). **G** and **H** Manhattan plot of five carcass phenotypes of significant loci at 15–45 Mb on chromosome 7 with corresponding protein-coding genes. **I** The result of genome coverage visualization for *FKBP5* intron region (chr7:31,539,932–31,541,378). The left vertical coordinate shows the insert size of the reads, and the right vertical coordinate shows the genome coverage. The black dotted line marks the location of DEL. **J** Electropherogram of the DEL in the *FKBP5* intronic region. Electropherogram showing the results of PCR amplification for the Large White and Min pig. The size of electrophoretic bands was indicated with TaKaRa DL2000. **K** GWAS for bone weight in the front, middle, and hind sections of pig carcass. The dotted line represents the segmentation position of the pig carcass. The genes closest to the most significant loci are labeled above each phenotype, with Arabic numerals representing the number of significant SV loci for each phenotype. **L** GWAS for seven pig skeletal phenotypes. The description of this figure is consistent with **K**





**Fig. 4** (See legend on previous page.)

**Table 1** Summary of significant SV loci in 36 traits

Phenotype	Chr	Position	Gene name	SV location	Type	Length	P-value
Abdominal circumference	1	260,511,205	<i>CDK5RAP2</i>	Intron	DEL	283	4.24E-07
	1	271,010,710	<i>LAMC3</i>	Intron	DEL	284	5.41E-07
Body height	7	31,540,442	<i>FKBP5</i>	Intron	DEL	441	8.75128E-08
	7	37,650,773	LncRNA	Intron	DEL	311	3.3894E-07
	7	26,286,305	-	Intergenic	DEL	291	8.85504E-07
	7	27,338,884	-	Intergenic	DEL	2,446	1.23795E-06
Body length	7	31,540,442	<i>FKBP5</i>	Intron	DEL	441	3.78792E-15
	7	19,322,838	<i>GPLD1</i>	Intron	DEL	101	7.1424E-10
	7	30,669,698	<i>ILRUN</i>	Gene upstream	MEI	294	8.99621E-10
	7	37,650,773	LncRNA	Intron	DEL	311	1.05991E-09
	7	36,897,444	<i>TFEB</i>	Intron	DEL	136	8.61888E-09
	7	24,883,903	<i>ENSSSCG00000061569, ENSSSCG00000001455</i>	Coverage	DUP	19,228	5.58603E-08
	7	31,311,879	<i>FANCE</i>	Gene upstream	DEL	288	5.9659E-08
	7	27,338,884	-	Intergenic	DEL	2,446	8.84893E-08
	7	33,224,291	<i>MDGA1</i>	Intron	DEL	75	1.61736E-07
	7	37,641,459	LncRNA	Intron	DEL	127	2.19766E-07
	7	19,232,872	<i>ENSSSCG00000057632</i>	Gene upstream	MEI	292	2.62207E-07
	7	43,989,593	<i>DEFB114</i>	Intron	DEL	1,677	4.49759E-07
	7	26,286,305	-	Intergenic	DEL	291	1.53065E-06
	Bone ratio	7	31,540,442	<i>FKBP5</i>	Intron	DEL	441
7		30,669,698	<i>ILRUN</i>	Gene upstream	MEI	294	4.02735E-10
7		37,650,773	LncRNA	Intron	DEL	311	7.68871E-09
7		28,324,725	<i>PRIM2</i>	Intron	INV	388	2.68482E-08
7		19,322,838	<i>GPLD1</i>	Intron	DEL	101	4.31208E-08
7		36,897,444	<i>TFEB</i>	Intron	DEL	136	5.94847E-08
7		40,599,238	-	Intergenic	DUP	309	6.52153E-08
7		38,794,225	-	Intergenic	DEL	357	1.46567E-07
7		33,224,291	<i>MDGA1</i>	Intron	DEL	75	3.36078E-07
7		24,883,903	<i>ENSSSCG00000061569, ENSSSCG00000001455</i>	Coverage	DUP	19,228	5.02475E-07
7		39,605,564	-	Intergenic	DEL	694	5.9881E-07
7		31,311,879	<i>FANCE</i>	Gene upstream	DEL	288	6.393E-07
7		30,962,999	<i>ANKS1A</i>	Intron	DEL	92	9.76289E-07
7		27,336,953	-	Intergenic	DUP	5,095	1.18774E-06
7		33,582,502	<i>ZFAND3</i>	Intron	DEL	65	1.49752E-06
7		37,641,459	LncRNA	Intron	DEL	127	1.54425E-06
Calf bone length		7	40,599,295	-	Intergenic	DEL	267
	7	38,794,225	-	Intergenic	DEL	357	1.54116E-10
	7	31,540,442	<i>FKBP5</i>	Intron	DEL	441	2.16357E-08
	7	19,322,838	<i>GPLD1</i>	Intron	DEL	101	2.68006E-08
	7	37,641,459	LncRNA	Intron	DEL	127	3.68605E-08
	7	36,897,444	<i>TFEB</i>	Intron	DEL	136	7.87112E-08
	7	19,232,872	<i>ENSSSCG00000057632</i>	Gene upstream	MEI	292	4.2437E-07
	7	33,093,365	-	Intergenic	DEL	1,547	4.70256E-07
	7	30,669,698	<i>ILRUN</i>	Gene upstream	MEI	294	6.1988E-07

**Table 1** (continued)

Phenotype	Chr	Position	Gene name	SV location	Type	Length	P-value
Cannon circumference	7	31,540,442	<i>FKBP5</i>	Intron	DEL	441	6.71872E-18
	7	28,324,725	<i>PRIM2</i>	Intron	INV	388	7.63587E-12
	7	37,650,773	LncRNA	Intron	DEL	311	1.14929E-11
	7	30,669,698	<i>ILRUN</i>	Gene upstream	MEI	294	1.41436E-11
	7	38,794,225	-	Intergenic	DEL	357	3.23829E-11
	7	41,553,724	<i>ANKRD66</i>	Intron	DEL	673	7.02795E-11
	7	19,322,838	<i>GPLD1</i>	Intron	DEL	101	1.56476E-10
	7	37,641,459	LncRNA	Intron	DEL	127	5.70722E-10
	7	28,324,716	<i>PRIM2</i>	Intron	DEL	130	6.7774E-10
	7	36,897,444	<i>TFEB</i>	Intron	DEL	136	1.62437E-09
	7	24,883,903	<i>ENSSSCG00000061569, ENSSSCG0000001455</i>	Coverage	DUP	19,228	1.80288E-09
	7	38,094,955	<i>RRP36</i>	Gene downstream	DEL	309	3.80358E-09
	7	19,232,872	<i>ENSSSCG00000057632</i>	Gene upstream	MEI	292	5.08074E-09
	7	41,553,727	<i>ANKRD66</i>	Intron	INV	721	1.0657E-08
	7	19,239,684	<i>MRS2</i>	Gene upstream	DEL	278	1.77718E-08
	7	26,286,305	-	Intergenic	DEL	291	1.80886E-08
	7	30,962,999	<i>ANKS1A</i>	Intron	DEL	92	2.25339E-08
	7	41,078,897	<i>RCAN2</i>	Intron	DEL	314	3.8015E-08
	7	34,536,073	<i>GLP1R</i>	Gene downstream	DEL	221	4.84358E-08
	7	41,288,870	-	Intergenic	DEL	327	5.08201E-08
	7	31,311,879	<i>FANCE</i>	Gene upstream	DEL	288	1.01069E-07
	7	36,811,392	-	Intergenic	DEL	53	2.35974E-07
	7	19,463,307	<i>KIAA0319</i>	Intron	DEL	309	2.56879E-07
	7	37,476,936	<i>TRERF1</i>	Intron	DEL	294	3.72867E-07
	7	37,055,115	<i>USP49</i>	Gene upstream	DEL	304	3.76061E-07
	7	19,678,493	<i>RIPOR2</i>	Intron	DEL	308	4.41254E-07
	7	36,074,049	-	Intergenic	DEL	268	6.33459E-07
	7	31,084,916	<i>TCP11</i>	Intron	DEL	289	6.50236E-07
	7	31,757,248	<i>SLC26A8</i>	Intron	DEL	2,450	6.78635E-07
	7	29,593,390	<i>COL21A1</i>	Intron	DEL	316	7.1008E-07
	7	43,989,593	<i>DEFB114</i>	Intron	DEL	1677	1.00074E-06
	7	24,311,068	<i>ENSSSCG00000030901</i>	Coverage	DUP	31,737	1.05293E-06
	7	39,605,564	-	Intergenic	DEL	694	1.22437E-06
7	33,224,291	<i>MDGA1</i>	Intron	DEL	75	1.70098E-06	

**Table 1** (continued)

Phenotype	Chr	Position	Gene name	SV location	Type	Length	P-value
Carcass length	7	31,540,442	<i>FKBP5</i>	Intron	DEL	441	1.65E-13
	7	19,232,872	<i>ENSSSCG00000057632</i>	Gene upstream	MEI	292	4.82488E-09
	7	30,669,698	<i>ILRUN</i>	Gene upstream	MEI	294	7.31789E-09
	7	37,650,773	LncRNA	Intron	DEL	311	1.22581E-08
	7	19,322,838	<i>GPLD1</i>	Intron	DEL	101	1.74936E-08
	7	34,536,073	<i>GLP1R</i>	Gene downstream	DEL	221	2.34715E-08
	7	33,224,291	<i>MDGA1</i>	Intron	DEL	75	5.21348E-08
	7	24,883,903	<i>ENSSSCG00000061569, ENSSSCG00000001455</i>	Coverage	DUP	19,228	1.51579E-07
	7	38,794,225	-	Intergenic	DEL	357	2.21327E-07
	7	36,897,444	<i>TFEB</i>	Intron	DEL	136	2.41208E-07
	7	34,419,953	<i>DNAH8</i>	Intron	DEL	302	2.72373E-07
	7	31,311,879	<i>FANCE</i>	Gene upstream	DEL	288	3.10433E-07
	7	26,286,305	-	Intergenic	DEL	291	3.37836E-07
	7	30,002,117	<i>MLN</i>	Gene downstream	DEL	396	4.64132E-07
	7	37,641,459	LncRNA	Intron	DEL	127	4.78724E-07
	7	37,055,115	<i>USP49</i>	Gene upstream	DEL	304	8.04377E-07
	7	31,757,248	<i>SLC26A8</i>	Intron	DEL	2,450	1.5249E-06
	7	24,311,068	<i>ENSSSCG00000030901</i>	Coverage	DUP	31,737	1.6098E-06
	7	27,489,396	<i>KHDRBS2</i>	Intron	DEL	879	1.79273E-06
	Femur length	7	31,540,442	<i>FKBP5</i>	Intron	DEL	441
7		19,322,838	<i>GPLD1</i>	Intron	DEL	101	4.4582E-09
7		37,641,459	LncRNA	Intron	DEL	127	1.41047E-08
7		30,962,999	<i>ANKS1A</i>	Intron	DEL	92	1.74416E-08
7		38,794,225	-	Intergenic	DEL	357	3.2897E-08
7		36,897,444	<i>TFEB</i>	Intron	DEL	136	7.98636E-08
3		125,943,216	<i>NOL10</i>	Intron*	DEL	252	2.17341E-07
7		38,094,955	<i>RRP36</i>	Gene downstream	DEL	309	6.88085E-07
7		37,055,115	<i>USP49</i>	Gene Upstream	DEL	304	7.03532E-07
7		33,224,291	<i>MDGA1</i>	Intron	DEL	75	7.45747E-07
3		125,953,211	<i>NOL10</i>	Intron*	DEL	2,681	7.80005E-07
7		37,650,773	LncRNA	Intron	DEL	311	8.90967E-07

**Table 1** (continued)

Phenotype	Chr	Position	Gene name	SV location	Type	Length	P-value
Forearm bone length	7	31,540,442	<i>FKBP5</i>	Intron	DEL	441	7.15594E-12
	7	37,641,459	LncRNA	Intron	DEL	127	1.46387E-10
	7	30,962,999	<i>ANKS1A</i>	Intron	DEL	92	3.50128E-10
	7	37,650,773	LncRNA	Intron	DEL	311	1.4709E-09
	7	36,897,444	<i>TFEB</i>	Intron	DEL	136	2.61096E-09
	7	37,488,354	<i>TRERF1</i>	Intron	DEL	300	3.35499E-09
	7	38,794,225	-	Intergenic	DEL	357	3.63713E-09
	7	19,232,872	<i>ENSSSCG00000057632</i>	Gene upstream	MEI	292	4.92423E-09
	7	19,322,838	<i>GPLD1</i>	Intron	DEL	101	2.78437E-08
	7	33,224,291	<i>MDGA1</i>	Intron	DEL	75	1.06476E-07
	7	30,002,117	<i>MLN</i>	Gene downstream	DEL	396	2.4688E-07
	7	19,239,684	<i>MRS2</i>	Gene upstream	DEL	278	3.66744E-07
	7	37,261,717	<i>C6orf132</i>	Gene upstream	DEL	297	4.22894E-07
	7	34,536,073	<i>GLP1R</i>	Gene downstream	DEL	221	6.18251E-07
	7	28,324,716	<i>PRIM2</i>	Intron	DEL	130	1.04137E-06
	7	30,669,698	<i>ILRUN</i>	Gene upstream	MEI	294	1.09207E-06
	7	24,883,903	<i>ENSSSCG00000061569, ENSSSCG00000001455</i>	Coverage	DUP	19,228	1.19239E-06
	7	36,074,049	-	Intergenic	DEL	268	1.19749E-06
	7	37,055,115	<i>USP49</i>	Gene upstream	DEL	304	1.76888E-06
	Heart weight	3	125,953,211	<i>NOL10</i>	Intron*	DEL	2,681
Hip bone length	7	31,757,248	<i>SLC26A8</i>	Intron	DEL	2,450	6.10178E-07
	7	24,311,068	<i>ENSSSCG00000030901</i>	Coverage	DUP	31,737	7.92587E-07
Humerus length	7	31,540,442	<i>FKBP5</i>	Intron	DEL	441	1.35763E-08
	7	34,536,073	<i>GLP1R</i>	Gene downstream	DEL	221	4.81907E-08
	7	19,232,872	<i>ENSSSCG00000057632</i>	Gene upstream	MEI	292	1.04481E-07
	7	37,641,459	LncRNA	Intron	DEL	127	2.51234E-07
Intramuscular fat	1	264,535,617	<i>DENND1A</i>	Intron	DEL	311	5.70534E-07
	12	57,675,134	<i>HS3ST3A1</i>	Intron	DEL	122	8.67028E-07
	12	54,564,984	<i>CFAP52</i>	Intron	MEI	291	9.75823E-07
	12	54,407,453	<i>STX8</i>	Intron	DEL	302	1.19263E-06
Liver weight	7	19,239,684	<i>MRS2</i>	Gene upstream	DEL	278	1.50207E-06
Scapula length	7	33,224,291	<i>MDGA1</i>	Intron	DEL	75	2.38035E-07

**Table 1** (continued)

Phenotype	Chr	Position	Gene name	SV location	Type	Length	P-value
Total weight of front bone	7	31,540,442	<i>FKBP5</i>	Intron	DEL	441	2.25255E-12
	7	30,669,698	<i>ILRUN</i>	Gene upstream	MEI	294	5.21713E-12
	7	33,224,291	<i>MDGA1</i>	Intron	DEL	75	3.8738E-11
	7	37,641,459	LncRNA	Intron	DEL	127	1.47312E-10
	7	36,897,444	<i>TFEB</i>	Intron	DEL	136	5.58282E-10
	7	37,650,773	LncRNA	Intron	DEL	311	1.06535E-09
	7	38,794,225	-	Intergenic	DEL	357	2.91302E-09
	7	24,311,068	<i>ENSSSCG00000030901</i>	Coverage	DUP	31,737	4.48617E-09
	7	37,055,115	<i>USP49</i>	Gene upstream	DEL	304	1.64413E-08
	7	24,883,903	<i>ENSSSCG00000061569, ENSSSCG00000001455</i>	Coverage	DUP	19,228	1.70589E-08
	7	28,324,725	<i>PRIM2</i>	Intron	INV	388	2.21799E-08
	7	19,232,872	<i>ENSSSCG00000057632</i>	Gene upstream	MEI	292	2.60468E-08
	7	34,419,953	<i>DNAH8</i>	Intron	DEL	302	6.0505E-08
	7	36,074,049	-	Intergenic	DEL	268	7.33407E-08
	7	30,962,999	<i>ANKS1A</i>	Intron	DEL	92	1.05105E-07
	7	38,094,955	<i>RRP36</i>	Gene downstream	DEL	309	1.45388E-07
	7	19,322,838	<i>GPLD1</i>	Intron	DEL	101	2.07241E-07
	7	37,261,717	<i>C6orf132</i>	Gene upstream	DEL	297	2.68119E-07
	7	37,488,354	<i>TRERF1</i>	Intron	DEL	300	4.77823E-07
	7	34,536,073	<i>GLP1R</i>	Gene downstream	DEL	221	7.19729E-07
	7	28,324,716	<i>PRIM2</i>	Intron	DEL	130	7.81429E-07
	7	30,002,117	<i>MLN</i>	Gene downstream	DEL	396	7.89703E-07
	7	40,599,295	-	Intergenic	DEL	267	1.01179E-06
	7	43,989,593	<i>DEFB114</i>	Intron	DEL	1,677	1.17865E-06
	7	39,605,564	-	Intergenic	DEL	694	1.31736E-06
	7	27,489,396	<i>KHDRBS2</i>	Intron	DEL	879	1.49335E-06
	7	35,332,237	LncRNA	Gene upstream	DEL	306	1.59814E-06
Total weight of front lean meat	7	41,494,812	<i>PLA2G7</i>	Intron	MEI	294	6.03091E-07

**Table 1** (continued)

Phenotype	Chr	Position	Gene name	SV location	Type	Length	P-value	
Total weight of hind bone	7	30,669,698	<i>ILRUN</i>	Gene upstream	MEI	294	7.95851E-13	
	7	38,794,225	-	Intergenic	DEL	357	3.41618E-11	
	7	36,897,444	<i>TFEB</i>	Intron	DEL	136	3.51948E-11	
	7	31,540,442	<i>FKBP5</i>	Intron	DEL	441	9.62978E-11	
	7	37,650,773	LncRNA	Intron	DEL	311	9.81726E-11	
	7	37,641,459	LncRNA	Intron	DEL	127	2.8606E-10	
	7	33,224,291	<i>MDGA1</i>	Intron	DEL	75	9.30272E-10	
	7	19,232,872	<i>ENSSSCG00000057632</i>	Gene upstream	MEI	292	3.80882E-09	
	7	37,055,115	<i>USP49</i>	Gene upstream	DEL	304	4.22536E-09	
	7	34,419,953	<i>DNAH8</i>	Intron	DEL	302	4.37384E-09	
	7	30,962,999	<i>ANKS1A</i>	Intron	DEL	92	1.12907E-08	
	7	37,488,354	<i>TRERF1</i>	Intron	DEL	300	2.10081E-08	
	7	19,322,838	<i>GPLD1</i>	Intron	DEL	101	2.53408E-08	
	7	31,757,248	<i>SLC26A8</i>	Intron	DEL	2,450	2.65487E-08	
	7	36,074,049	-	Intergenic	DEL	268	6.61502E-08	
	7	34,536,073	<i>GLP1R</i>	Gene downstream	DEL	221	8.30738E-08	
	7	31,311,879	<i>FANCE</i>	Gene upstream	DEL	288	9.31817E-08	
	7	37,261,717	<i>C6orf132</i>	Gene upstream	DEL	297	2.37267E-07	
	7	39,605,564	-	Intergenic	DEL	694	2.55411E-07	
	7	30,002,117	<i>MLN</i>	Gene downstream	DEL	396	4.61386E-07	
	7	24,311,068	<i>ENSSSCG00000030901</i>	Coverage	DUP	31,737	4.69345E-07	
	7	33,093,365	-	Intergenic	DEL	1547	5.24922E-07	
	7	19,463,307	<i>KIAA0319</i>	Intron	DEL	309	6.42425E-07	
	7	38,094,955	<i>RRP36</i>	Gene downstream	DEL	309	7.07833E-07	
	7	36,811,392	-	Intergenic	DEL	53	1.32686E-06	
	7	28,324,725	<i>PRIM2</i>	Intron	INV	388	1.32768E-06	
	7	27,338,884	-	Intergenic	DEL	2,446	1.46419E-06	
	Total weight of middle bone	7	33,224,291	<i>MDGA1</i>	Intron	DEL	75	6.19796E-09
		7	30,669,698	<i>ILRUN</i>	Gene upstream	MEI	294	1.16594E-07
		7	37,650,773	LncRNA	Intron	DEL	311	1.39685E-07
7		34,419,953	<i>DNAH8</i>	Intron	DEL	302	1.59434E-07	
7		31,540,442	<i>FKBP5</i>	Intron	DEL	441	1.89109E-07	
6		158,747,897	<i>GLIS1</i>	Intron	DEL	85	1.49265E-06	
Vertebral number	7	97,615,896	<i>VRTN</i>	Intron	MEI	294	5.41E-11	
	1	266,822,698	-	Intergenic	MEI	294	2.51235E-09	
	1	268,351,956	<i>AK1</i>	Gene downstream	DEL	258	2.35774E-07	
	1	268,687,871	<i>DNM1</i>	Intron	DEL	60	3.94325E-07	
	1	267,960,645	<i>ZNF79</i>	Exon	MEI	294	7.2393E-07	
	7	97,653,457	<i>SYNDIG1L</i>	Gene downstream	DEL	765	8.30871E-07	
	1	261,580,807	<i>ENSSSCG0000005518</i>	Intron	DEL	291	1.05102E-06	
	1	262,031,225	<i>TTLL11</i>	Intron	DEL	955	1.51808E-06	

\* Indicates that the variation is mostly located in this region

bone, femur, and calf bone length (Fig. 4L). Regarding vertebral number (Fig. S13J), the most significant locus was the 291 bp intron variant ( $P = 5.41E-11$ ) of *VRTN*, which is consistent with a previous study using SNP markers [79]. We also identified a SINE insertion locus ( $P = 7.2393E-07$ ) in the exonic region of the *ZNF79* gene, which was previously associated with bone mineral density [80]. Further analysis, we checked whether SVs affected tissues other than bone by performing GWAS of total lean and fat weight in the front, middle, and hind sections. The results showed only one significant locus for total weight of front lean meat, with no significant association peaks for the remaining traits (Fig. S15A–F, Additional file 2 and Table 1). This result indicates that SVs mainly involve bone tissue and are poorly associated with other tissues.

Among the remaining other traits, GWAS were performed for four meat quality traits: marbling, intramuscular fat, tenderness, and moisture percentage (Fig. S16A–G, Additional file 2). The results showed that only three SV loci were significant for intramuscular fat (Fig. S16B and Table 1), which were located in the intron regions of *HS3ST3A1*, *CFAP52*, and *STX8*. Heart, liver, and lung weight were also investigated for their associations with SVs, identifying one significant locus in each of the heart (Fig. S16E) and liver weight (Fig. S16F), which were a 2,681-bp DEL overlapping 3 bp with the *NOL10* exon and a 278-bp DEL upstream of *MRS2*, respectively. Considering that a certain tolerance is needed for the determination of SV breakpoint locations [81], the exonic variants still need further validation.

## Discussion

Here, we performed an SV study based on a resource population constructed from Large White and Min pigs. A typical approach to SV research is to take the results of multiple software intersections to improve the accuracy when identifying variants. This strategy has been reported to not reliably improve performance and in some cases even aggravate false discoveries [22]. Therefore, merging the analysis results of several software instead of including only overlapping regions is expected to maximize the performance of each software, which improves the sensitivity of SV identification to obtain more new SV loci. Using this approach, we developed a high-quality SV map with 53.95% newly discovered SV loci compared to the Ensemble public SV database, which will greatly enrich the public SV database. Then, the genotyping accuracy of the SV loci was validated by PCR to be more than 94%. We suggest that it will be necessary to perform multiple SV software in future studies and retain specific results from each software which will not only allow

for the identification of more new SV loci, but also maintain accuracy.

Generally, when constructing a segregating population, high depth sequencing is allocated to the parental generation and low depth sequencing to the F2 generation or even more distant generations for cost reasons. In contrast, low coverage sequencing for SV identification appears to reduce sensitivity and accuracy. Here, we designed a novel approach for population genetic studies, which identified reliable SV loci in F0 individuals at high sequencing depths and then used these loci to genotype the F2 population according to the corresponding genomic positions for GWAS to identify causal SV loci due to breed differences. Our results show that almost all loci were successfully genotyped, confirming the reliability of this approach. This approach improves the accuracy of SV identification and increases the efficiency of the analysis by avoiding the identified variants in a large population. The analysis detected genes identified in previous studies using SNP markers including *ILRUN*, *TFEB*, *RCAN2*, and *VRTN* [72, 73, 79], which confirmed the accuracy of SV genotyping in the F2 population and the potential of SVs as markers. To our knowledge, this is the first report of this method in livestock studies. In addition, third-generation sequencing has begun to be applied in the study of animal and plant genomes, which has more potential to identify larger structural variants. However, its application is limited due to its high cost, especially within large populations. Genotyping in second-generation sequencing data using SV loci identified by third-generation sequencing data is a potential solution, although the reported recall of genotyping is currently only about 50% [82].

Compared to SNPs, SVs in non-coding regions are more likely to alter gene expression and phenotype through dosage effects, and SVs can also modify expression levels by directly altering gene copy numbers [83–86]. Therefore, using SVs as markers for directly performing GWAS is expected to identify causal loci affecting phenotypes. There have been several previous studies on SVs in the pig genome, but SV-based GWAS in the pig genome have been rarely reported [13–15]. The Large White × Min pig resource population provides an opportunity to deepen the understanding of the potential and biological role of SVs as markers for association studies. Moreover, insertional variants in particular have rarely been included in SV studies due to their complex genotyping process, and the phenotypic impact of insertional variants remains largely unknown. While transposon insertion identification depends on reference sequences, which facilitates subsequent genotyping, and a previous study has confirmed that approximately 80% of the variation in the pig genome overlaps with transposable elements [15], which



provided an opportunity to investigate the contribution of insertional variants to the phenotype in the present study. Therefore, we performed GWAS on four SV types in the pig genome, providing new insights into the contribution of different SV types to the phenotype. To our knowledge, this is perhaps the most comprehensive SV-based GWAS of the pig genome to date.

Bone is a highly complex and active mineralized material. Bone tissue undergoes a continuous cycle of osteoclast bone resorption and osteoblast bone formation [87] and then receives mechanical loads from the musculoskeletal system and interactions with other biological systems (such as the endocrine, nervous, and immune systems) in order to maintain the shape, volume, and density of the bone [88]. During the growth period, there is intense bone formation to increase body size [89]. In this study, we performed GWAS and revealed a large number of candidate genes associated with skeleton in pigs. Among them, *FKBP5* and *MRS2* are involved in the differentiation and formation of osteoclasts, respectively. In contrast, *TFEB* and *RCAN2* are associated with the differentiation and function of osteoblasts. We hypothesize that, during domestication and selection, these key candidate genes may be affected due to surrounding or internal SVs regulation, resulting in differential body size between breeds. Based on previous studies, SV in the intron region can cause alternative splicing of RNA [90, 91] and play a promoter or enhancer role [92, 93], whereas SV located upstream or downstream of a gene may be associated with transcriptional regulation of that gene, especially transposon insertions, which have been reported to play a functional role in carrying cis-regulatory elements [94–97].

## Conclusion

In this study, we constructed a high-quality SV map using high-coverage resequencing data from the Large White and Min pigs. More than half of the SV loci were reported for the first time by merging the results of 5 SV tools, suggesting the need to use multiple software for SV analysis and to retain the specific variants identified by different SV tools. GWAS for 36 traits showed that SVs were mainly associated with skeletal size, which may contribute to the differences in body size between European and Chinese pig breeds.

## Abbreviations

DEL	Deletion
DUP	Duplication
GWAS	Genome-wide association studies
INV	Inversion
MEI	Mobile element insertion
SNP	Single nucleotide polymorphism
SV	Structural variation

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40104-023-00929-x>.

**Additional file 1: Table S1.** Sample information. **Table S2.** TE reference sequence. **Table S3.** Primers for 1–10 Mb SV validation. **Table S4.** Primers for SV genotype accuracy validation. **Table S5.** Number of all SV types identified per individual. **Table S6.** SV distribution for different length ranges. **Table S7.** Information on SV loci annotated as HIGH impact. **Table S8.** Top 5% fixed index ( $F_{5\%}$ ) statistics based on Weir and Cockerham. **Table S9.** Primers for significant loci of skeletal-related traits.

**Additional file 2: Fig. S1.** Schematic design of SV primers for 1–10 Mb. **Fig. S2.** Schematic diagram of differential genotype screening. **Fig. S3.** SV Pipeline. **Fig. S4.** Validation of DEL variants at 1–10 Mb loci. **Fig. S5.** Validation of DUP variants at 1–10 Mb loci. **Fig. S6.** Validation of INV variants at 1–10 Mb loci. **Fig. S7.** The accuracy of SV genotyping was verified by agarose electrophoresis. **Fig. S8.** Venn diagram of SV loci in public databases and this study. **Fig. S9.** PCA plot of the 50k chip in 19 F0 individuals. **Fig. S10.** KEGG pathway. **Fig. S11.** Validation of SVs involving carcass traits using Sanger sequencing. **Fig. S12.** Manhattan plots for scapular width (A), chest width (B), chest depth (C), abdominal circumference (D), waist width (E), hip width (F), hip length (G), and hip circumference (H). **Fig. S13.** Manhattan plots for total weight of front bone (A), total weight of middle bone (B), total weight of hind bone (C), scapula length (D), humerus length (E), forearm bone length (F), hip bone length (G), femur length (H), calf bone length (I), and vertebral number (J). **Fig. S14.** Venn diagrams of three bone weight and six bone length traits overlapping with carcass length, body length, body height, cannon circumference, and bone rate. **Fig. S15.** Manhattan plots for total weight of front lean meat (A), total weight of middle lean meat (B), total weight of hind lean meat (C), total weight of front fat (D), total weight of middle fat (E), and total weight of hind fat (F). **Fig. S16.** Manhattan plots for marbling (B), tenderness (C), moisture percentage (D), heart weight (E), liver weight (F), and lung weight (G).

## Acknowledgements

We thank the researchers at our laboratories for their dedication and hard work.

## Authors' contributions

LXW and LCZ conceived the project and designed the research; WCZ, RZZ, YFS and ZPH performed the experimental validation; WCZ, NQN and JBW performed data processing; WCZ and LCZ performed the data analysis and wrote manuscript; LCZ, LXW, WCZ, LGW, XHH and XL performed edits and revisions to the paper; All authors read and approved the final version of the manuscript.

## Funding

This research was supported by the National Key R&D Program of China (2021YFD1301101), National Swine Industry Technology System (CARS-35), and Agricultural Science and Technology Innovation Program (ASTIP-IAS02).

## Availability of data and material

The sequencing data of the F0 and F2 individuals used in this study have been submitted to the Genome Sequence Archive (GSA) with the accession number CRA002451.

## Declarations

### Ethics approval and consent to participate

All experiments and procedures were carried out following the regulations from the Animal Care and Ethics Committee for animal experiments, Institute of Animal Science, Chinese Academy of Agricultural Sciences (Permit Number: IAS2020-39).

### Consent for publication

Not applicable.

**Competing interests**

The authors have declared no competing interests.

Received: 23 May 2023 Accepted: 3 August 2023

Published online: 07 October 2023

**References**

- Roses AD, Akkari PA, Chiba-Falek O, Lutz MW, Gottschalk WK, Saunders AM, et al. Structural variants can be more informative for disease diagnostics, prognostics and translation than current SNP mapping and exon sequencing. *Expert Opin Drug Metab Toxicol*. 2016;12(2):135–47.
- Chiang C, Scott AJ, Davis JR, Tsang EK, Li X, Kim Y, et al. The impact of structural variation on human gene expression. *Nat Genet*. 2017;49(5):692–9.
- Ho SS, Urban AE, Mills RE. Structural variation in the sequencing era. *Nat Rev Genet*. 2020;21(3):171–89.
- Yan CL, Lin J, Huang YY, Gao QS, Piao ZY, Yuan SL, et al. Population genomics reveals that natural variation in PRDM16 contributes to cold tolerance in domestic cattle. *Zool Res*. 2022;43(2):275–84.
- Yuan Y, Zhang WY, Yang BG, Zhou DK, Xu L, He YM, et al. A 1.1 Mb duplication CNV on chromosome 17 contributes to skeletal muscle development in Boer goats. *Zool Res*. 2023;44:303–14.
- Dermitzakis ET, Stranger EB, Forrest MS, Dunning M, Ingle CE, Beazley C, et al. Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science*. 2007;315(5813):848–53.
- Collins RL, Brand H, Karczewski KJ, Zhao X, Alfoldi J, Francioli LC, et al. A structural variation reference for medical and population genetics. *Nature*. 2020;581(7809):444–51.
- Frantz LAF, Schraiber JG, Madsen O, Megens HJ, Cagan A, Bosse M, et al. Evidence of long-term gene flow and selection during domestication from analyses of Eurasian wild and domestic pig genomes. *Nat Genet*. 2015;47(10):1141–8.
- White BR, Lan YH, McKeith FK, Novakofski J, Wheeler MB, McLaren DG. Growth and body composition of Meishan and Yorkshire barrows and gilts. *J Anim Sci*. 1995;73(3):738–49.
- China National Commission of Animal Genetic Resources. *Animal genetic resources in China: pigs*. Beijing: China Agriculture Press; 2011.
- Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 years of GWAS discovery: Biology, function, and translation. *Am J Hum Genet*. 2017;101(1):5–22.
- Loos RJF. 15 years of genome-wide association studies and no signs of slowing down. *Nat Commun*. 2020;11:5900.
- 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.
- Du H, Zheng X, Zhao Q, Hu Z, Wang H, Zhou L, et al. Analysis of structural variants reveal novel selective regions in the genome of Meishan pigs by whole genome sequencing. *Front Genet*. 2021;12:550676.
- Gong H, Liu W, Wu Z, Zhang M, Sun Y, Ling Z, et al. Evolutionary insights into porcine genomic structural variations based on a novel-constructed dataset from 24 worldwide diverse populations. *Evol Appl*. 2022;15:1264–80.
- Chen JQ, Zhang MP, Tong XK, Li JQ, Zhang Z, Huang F, et al. Scan of the endogenous retrovirus sequences across the swine genome and survey of their copy number variation and sequence diversity among various Chinese and Western pig breeds. *Zool Res*. 2022;43:423–41.
- Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114–20.
- Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754–60.
- 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.
- Broad Institute. Picard toolkit. GitHub repository. 2019. <https://broadinstitute.github.io/picard>. Accessed 4 Mar 2023.
- Pedersen BS, Quinlan AR. Mosdepth: Quick coverage calculation for genomes and exomes. *Bioinformatics*. 2018;34(5):867–8.
- Cameron DL, Di Stefano L, Papenfuss AT. Comprehensive evaluation and characterisation of short read general-purpose structural variant calling software. *Nat Commun*. 2019;10:3240.
- Chen N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinformatics*. 2004;25:4.10.1–14.
- Jeffares DC, Jolly C, Hoti M, Speed D, Shaw L, Rallis C, et al. Transient structural variations have strong effects on quantitative traits and reproductive isolation in fission yeast. *Nat Commun*. 2017;8:14061.
- Chiang C, Layer RM, Faust GG, Lindberg MR, Rose DB, Garrison EP, et al. SpeedSeq: Ultra-fast personal genome analysis and interpretation. *Nat Methods*. 2015;12(10):966–8.
- Belyeu JR, Chowdhury M, Brown J, Pedersen BS, Cormier MJ, Quinlan AR, et al. Samplot: a platform for structural variant visual validation and automated filtering. *Genome Biol*. 2021;22:161.
- Ginestet C. ggplot2: Elegant graphics for data analysis. *J Stat Soft*. 2010;35:1–3.
- Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: A tool for genome-wide complex trait analysis. *Am J Hum Genet*. 2011;88(1):76–82.
- Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*. 2009;19(9):1655–64.
- Baum BR. PHYLP: Phylogeny Inference Package. Version 3.2. Joel Felsenstein. *Q Rev Biol*. 1989;1989:539–41.
- Tamura K, Stecher G, Kumar S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol Biol Evol*. 2021;38(7):3022–7.
- Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms. SnpEff Fly (Austin). 2012;581(7809):444–51.
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics*. 2011;27(15):2156–8.
- Chen S, Krusche P, Dolzhenko E, Sherman RM, Petrovski R, Schlesinger F, et al. Paragraph: A graph-based structural variant genotyper for short-read sequence data. *Genome Biol*. 2019;20(1):291.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559–75.
- Kang HM, Sul JH, Service SK, Zaitlen NA, Kong SY, Freimer NB, et al. Variance component model to account for sample structure in genome-wide association studies. *Nat Genet*. 2010;42(4):348–54.
- Yin L, Zhang H, Tang Z, Xu J, Yin D, Zhang Z, et al. rMVP: A memory-efficient, visualization-enhanced, and parallel-accelerated tool for genome-wide association study. *Genom Proteom Bioinf*. 2021;19(4):619–28.
- Rausch T, Zichner T, Schlattl A, Stütz AM, Benes V, Korbel JO. DELLY: Structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics*. 2012;28(18):333–9.
- Chen X, Schulz-Trieglaff O, Shaw R, Barnes B, Schlesinger F, Källberg M, et al. Manta: Rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics*. 2016;32(8):1220–2.
- Chen K, Wallis JW, McLellan MD, Larson DE, Kalicki JM, Pohl CS, et al. BreakDancer: An algorithm for high-resolution mapping of genomic structural variation. *Nat Methods*. 2009;6(9):677–81.
- Gardner EJ, Lam VK, Harris DN, Chuang NT, Scott EC, Stephen Pittard W, et al. The mobile element locator tool (MELT): Population-scale mobile element discovery and biology. *Genome Res*. 2017;27(11):1916–29.
- Guo J, Cao K, Deng C, Li Y, Zhu G, Fang W, et al. An integrated peach genome structural variation map uncovers genes associated with fruit traits. *Genome Biol*. 2020;21(1):258.
- Lv FH, Cao YH, Liu GJ, Luo LY, Lu R, Liu MJ, et al. Whole-genome resequencing of worldwide wild and domestic sheep elucidates genetic diversity, introgression, and agronomically important loci. *Mol Biol Evol*. 2022;39(2):msab353.
- Ai H, Fang X, Yang B, Huang Z, Chen H, Mao L, et al. Adaptation and possible ancient interspecies introgression in pigs identified by whole-genome sequencing. *Nat Genet*. 2015;47(3):217–25.
- Chen C, D'Alessandro E, Murani E, Zheng Y, Giosa D, Yang N, et al. SINE jumping contributes to large-scale polymorphisms in the pig genomes. *Mob DNA*. 2021;12:17.
- Chen C, Liu C, Xiong X, Fang S, Yang H, Zhang Z, et al. Copy number variation in the MSRB3 gene enlarges porcine ear size through a mechanism involving miR-584-5p. *Genet Sel Evol*. 2018;50:72.
- Wang K, Wu P, Wang S, Ji X, Chen D, Jiang A, et al. Genome-wide DNA methylation analysis in Chinese Chenghua and Yorkshire pigs. *BMC Genomic Data*. 2021;22:21.

48. Huang W, Zhang X, Li A, Xie L, Miao X. Differential regulation of mRNAs and lncRNAs related to lipid metabolism in two pig breeds. *Oncotarget*. 2017;8(50):87539–53.
49. Huang W, Zhang X, Li A, Xie L, Miao X. Genome-wide analysis of mRNAs and lncRNAs of intramuscular fat related to lipid metabolism in two pig breeds. *Cell Physiol Biochem*. 2018;50(6):2406–22.
50. Rubin CJ, Megens HJ, Barrio AM, Maqbool K, Sayyab S, Schwochow D, et al. Strong signatures of selection in the domestic pig genome. *Proc Natl Acad Sci U S A*. 2012;109(48):19529–36.
51. Khan A, Tian L, Zhang C, Yuan K, Xu S. Genetic diversity and natural selection footprints of the glycine amidinotransferase gene in various human populations. *Sci Rep*. 2016;6:18755.
52. Fiore R, Rahim B, Christoffels VM, Moorman AFM, Püschel AW. Inactivation of the *Sema5a* gene results in embryonic lethality and defective remodeling of the cranial vascular system. *Mol Cell Biol*. 2005;40(20):e00409–20.
53. Hou R, Chen L, Liu X, Liu H, Shi G, Hou X, et al. Integrating genome-wide association study with RNA-sequencing reveals *HDAC9* as a candidate gene influencing loin muscle area in Beijing Black pigs. *Biology (Basel)*. 2022;11:1635.
54. Guo T, Gao J, Yang B, Yan G, Xiao S, Zhang Z, et al. A whole genome sequence association study of muscle fiber traits in a White Duroc × Erhualian F2 resource population. *Asian-Australasian J Anim Sci*. 2020;33(5):704–11.
55. Huang J, Yang Y, Tian M, Deng D, Yu M. Spatial transcriptomic and miRNA analyses revealed genes involved in the mesometrial-biased implantation in pigs. *Genes (Basel)*. 2019;10(10):808.
56. Yang X, Zhang X, Jiao J, Zhang F, Pan Y, Wang Q, et al. Rare variants in *FANCA* induce premature ovarian insufficiency. *Hum Genet*. 2019;138(11–12):1227–36.
57. Fukaya Y, Lovero KL, Iwanaga T, Watanabe A, Yokoi N, Tabuchi K, et al. Disruption of *LGII*-linked synaptic complex causes abnormal synaptic transmission and epilepsy. *Proc Natl Acad Sci U S A*. 2010;107(8):3799–804.
58. Gallagher D, Voronova A, Zander MA, Cancino GI, Bramall A, Krause MP, et al. *Ankrd11* is a chromatin regulator involved in autism that is essential for neural development. *Dev Cell*. 2015;32(1):31–42.
59. Anney R, Klei L, Pinto D, Regan R, Conroy J, Magalhaes TR, et al. A genome-wide scan for common alleles affecting risk for autism. *Hum Mol Genet*. 2010;19(20):4072–82.
60. Tsang KM, Croen LA, Torres AR, Kharrazi M, Delorenze GN, Windham GC, et al. A Genome-wide survey of transgenerational genetic effects in autism. *PLoS ONE*. 2013;8(10):e76978.
61. Hanai S, Kanai M, Ohashi S, Okamoto K, Yamada M, Takahashi H, et al. Loss of poly(ADP-ribose) glycohydrolase causes progressive neurodegeneration in *Drosophila melanogaster*. *Proc Natl Acad Sci U S A*. 2004;101(1):82–6.
62. Wang X, Gu J, Miyoshi E, Honke K, Taniguchi N. Phenotype changes of *Fut8* knockout mouse: Core fucosylation is crucial for the function of growth factor receptor(s). *Methods Enzymol*. 2006;417:11–22.
63. Lian W, Gao D, Huang C, Zhong Q, Hua R, Lei M. Heat stress impairs maternal endometrial integrity and results in embryo implantation failure by regulating transport-related gene expression in Tongcheng pigs. *Biomolecules*. 2022;12(3):388.
64. Martins TF, Braga Magalhães AF, Verardo LL, Santos GC, Silva Fernandes AA, Gomes Vieira JI, et al. Functional analysis of litter size and number of teats in pigs: From GWAS to post-GWAS. *Theriogenology*. 2022;193:157–66.
65. Song M, Yang X, Ren X, Maliskova L, Li B, Jones IR, et al. Mapping cis-regulatory chromatin contacts in neural cells links neuropsychiatric disorder risk variants to target genes. *Nat Genet*. 2019;51(8):1252–62.
66. Lu B, Jiao Y, Wang Y, Dong J, Wei M, Cui B, et al. A *FKBP5* mutation is associated with paget's disease of bone and enhances osteoclastogenesis. *Exp Mol Med*. 2017;49(5):e336.
67. Kimura M, Nagai T, Matsushita R, Hashimoto A, Miyashita T, Hirohata S. Role of *FK506* binding protein 5 (*FKBP5*) in osteoclast differentiation. *Mod Rheumatol*. 2013;23(6):1133–9.
68. Weedon MN, Lango H, Lindgren CM, Wallace C, Evans DM, Mangino M, et al. Genome-wide association analysis identifies 20 loci that influence adult height. *Nat Genet*. 2008;40(5):575–83.
69. Berndt SJ, Gustafsson S, Mägi R, Ganna A, Wheeler E, Feitosa MF, et al. Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture. *Nat Genet*. 2013;45(5):501–12.
70. Tachmazidou I, Süveges D, Min JL, Ritchie GRS, Steinberg J, Walter K, et al. Whole-genome sequencing coupled to imputation discovers genetic signals for anthropometric traits. *Am J Hum Genet*. 2017;100(6):865–84.
71. Cho HW, Jin HS, Eom YB. A genome-wide association study of novel genetic variants associated with anthropometric traits in Koreans. *Front Genet*. 2021;12:669215.
72. Gong H, Xiao S, Li W, Huang T, Huang X, Yan G, et al. Unravelling the genetic loci for growth and carcass traits in Chinese Bamaxiang pigs based on a 1.4 million SNP array. *J Anim Breed Genet*. 2019;136:3–14.
73. Falker-Gieske C, Blaj I, Preuß S, Bennewitz J, Thaller G, Tetens J. GWAS for meat and carcass traits using imputed sequence level genotypes in pooled F2-designs in pigs. *G3-Genes Genom Genet*. 2019;9(9):2823–34.
74. Yoneshima E, Okamoto K, Sakai E, Nishishita K, Yoshida N, Tsukuba T. The transcription factor EB (TFEB) regulates osteoblast differentiation through ATF4/CHOP-dependent pathway. *J Cell Physiol*. 2016;231(6):1321–33.
75. Bassett JHD, Logan JG, Boyde A, Cheung MS, Evans H, Croucher P, et al. Mice lacking the calcineurin inhibitor *Rcan2* have an isolated defect of osteoblast function. *Endocrinology*. 2012;153(7):3537–48.
76. Pei YF, Liu L, Le LT, Yang XL, Zhang H, Wei XT, et al. Joint association analysis identified 18 new loci for bone mineral density. *J Bone Miner Res*. 2019;34(6):1086–94.
77. Bellucci MM, Schoenmaker T, Rossa-Junior C, Orrico SR, de Vries TJ, Everts V. Magnesium deficiency results in an increased formation of osteoclasts. *J Nutr Biochem*. 2013;24(8):1488–98.
78. Mabileau G, Mieczkowska A, Irwin N, Flatt PR, Chappard D. Optimal bone mechanical and material properties require a functional glucagon-like peptide-1 receptor. *J Endocrinol*. 2013;219(1):59–68.
79. 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.
80. Mullin BH, Zhu K, Xu J, Brown SJ, Mullin S, Tickner J, et al. Expression quantitative trait locus study of bone mineral density GWAS variants in human osteoclasts. *J Bone Miner Res*. 2018;33(6):1044–51.
81. Sudmant PH, Rausch T, Gardner EJ, Handsaker RE, Abyzov A, Huddleston J, et al. An integrated map of structural variation in 2,504 human genomes. *Nature*. 2015;526(7571):75–81.
82. Quan C, Li Y, Liu X, Wang Y, Ping J, Lu Y, et al. Characterization of structural variation in Tibetans reveals new evidence of high-altitude adaptation and introgression. *Genome Biol*. 2021;22:159.
83. Shanta O, Noor A, Chaisson MJP, Sanders AD, Zhao X, Malhotra A, et al. The effects of common structural variants on 3D chromatin structure. *BMC Genomics*. 2020;21:95.
84. Fudenberg G, Pollard KS. Chromatin features constrain structural variation across evolutionary timescales. *Proc Natl Acad Sci U S A*. 2019;116(6):2175–80.
85. Jakubosky D, D'Antonio M, Bonder MJ, Smail C, Donovan MKR, Young Greenwald WW, et al. Properties of structural variants and short tandem repeats associated with gene expression and complex traits. *Nat Commun*. 2020;11:2927.
86. Alonge M, Wang X, Benoit M, Soyk S, Pereira L, Zhang L, et al. Major impacts of widespread structural variation on gene expression and crop improvement in tomato. *Cell*. 2020;182:145–61.
87. Matsuo K, Irie N. Osteoclast-osteoblast communication. *Arch Biochem Biophys*. 2008;473(2):201–9.
88. Okamoto K, Nakashima T, Shinohara M, Negishi-Koga T, Komatsu N, Terashima A, et al. Osteoimmunology: The conceptual framework unifying the immune and skeletal systems. *Physiol Rev*. 2017;97(4):1295–349.
89. Duncan Bassett JH, Williams GR. Role of thyroid hormones in skeletal development and bone maintenance. *Endocr Rev*. 2016;37(2):135–87.
90. Xia B, Zhang W, Wudzinska A, Huang E, Brosh R, Pour M, et al. The genetic basis of tail-loss evolution in humans and apes. *bioRxiv*. 2021. <https://doi.org/10.1101/2021.09.14.460388>.
91. Pastor T, Talotti G, Lewandowska MA, Pagani F. An Alu-derived intronic splicing enhancer facilitates intronic processing and modulates aberrant splicing in *ATM*. *Nucleic Acids Res*. 2009;37(21):7258–67.
92. Vacik T, Raska I. Alternative intronic promoters in development and disease. *Protoplasma*. 2017;254(3):1201–6.
93. Su M, Han D, Boyd-Kirkup J, Yu X, Han JDJ. Evolution of Alu elements toward enhancers. *Cell Rep*. 2014;7(2):376–85.

94. Li J, Kannan M, Trivett AL, Liao H, Wu X, Akagi K, et al. An antisense promoter in mouse L1 retrotransposon open reading frame-1 initiates expression of diverse fusion transcripts and limits retrotransposition. *Nucleic Acids Res.* 2014;42:4546–62.
95. Ding M, Liu Y, Liao X, Zhan H, Liu Y, Huang W. Enhancer RNAs (eRNAs): New insights into gene transcription and disease treatment. *J Cancer.* 2018;9(13):2334–40.
96. Román AC, González-Rico FJ, Moltó E, Hernando H, Neto A, Vicente-García C, et al. Dioxin receptor and SLUG transcription factors regulate the insulator activity of B1 SINE retrotransposons via an RNA polymerase switch. *Genome Res.* 2011;21(3):422–32.
97. Mastrangelo MF, Weinstock KG, Shafer BK, Hedge AM, Garfinkel DJ, Strathern JN. Disruption of a silencer domain by a retrotransposon. *Genetics.* 1992;131(3):519–29.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

