# Supplementary Information

## Supplementary Note

### 1. Samples for whole genome sequencing

#### 1.1 Environmental conditions of southern and northern Chinese pigs

#### 1.2 Selection of the Wuzhishan pig reference genome

### 2. Population genetics and evolutionary history of Chinese pigs

#### 2.1 Methodology

#### 2.2 Population diversity

#### 2.3 Population genetic differentiation

#### 2.4 Linkage disequilibrium extents

#### 2.5 Genomic landscape of introgression between Chinese and European domestic pigs

### 3. Coalescent simulation analysis supports that the 14 Mb sweep region is not likely caused by gene drift

Nature Genetics: doi:10.1038/ng.3199

## Overview List of Supplementary Figures

**Supplementary Fig. 1** Selection of Chinese representative pigs from a large sample of 678 animals based on the Illumina porcine 60K SNP data.

**Supplementary Fig. 2** The concordance between sequence-based SNPs and Illumina porcine 60K SNPs in the tested samples.

**Supplementary Fig. 3** Venn diagram showing novel variants absent from Build 138 of dbSNP in Chinese pigs.

**Supplementary Fig. 4** Circos plot showing global distribution of variants along the genome.

**Supplementary Fig. 5** Global distribution of structural variants along chromosomes in a whole population.

**Supplementary Fig. 6** Distribution profile of minor allele frequencies of SNPs called in this study.

**Supplementary Fig. 7** Global distribution of structural variants along chromosome in each Chinese breed and wild boars.

**Supplementary Fig. 8** Comparison of the size distribution patterns between structural variants and annotated transposon elements.

**Supplementary Fig. 9** Principal component analysis of whole genome SNP data on 111 Chinese and European pigs.

**Supplementary Fig. 10** Admixture analysis of 111 Chinese and European pigs.

**Supplementary Fig. 11** Illustration of introgression between Chinese and European domestic pig genomes.

**Supplementary Fig. 12** The large potential introgression region on chromosome 13 in Min pigs.

**Supplementary Fig. 13** The large potential introgression region on chromosome 15 in Min pigs.

**Supplementary Fig. 14** The linkage disequilibrium decay plots for Chinese and European pigs.

Nature Genetics: doi:10.1038/ng.3199

**Supplementary Fig. 15** Illustration of the X-linked sweep region by three statistics.

**Supplementary Fig. 16** Comparison of heterozygosity of the X-linked sweep, flanking regions and autosomes.

**Supplementary Fig. 17** The distribution of recombination rate on chromosome X.

**Supplementary Fig. 18** The distribution of GC content, repeat sequence and gene counts on chromosome X.

**Supplementary Fig. 19** Distribution of the 6K poly-T sequence on chromosome X.

## Overview List of Supplementary Tables

**Supplementary Table 1**. Samples, origin and environmental variables of Chinese pigs in this study.

**Supplementary Table 2.** Comparison of the Wuzhishan and Duroc reference genomes.

**Supplementary Table 3**. Statistics of 1 Mb reads of one individual from each breed mapped to the Wuzhishan and Duroc reference genomes.

**Supplementary Table 4**. Distribution of variants in the pig genome.

**Supplementary Table 5.** Statistics for putative loss-of-function variants.

**Supplementary Table 6**. Statistics of structural variants overlapped with annotated transposon elements.

**Supplementary Table 7**. Summary of 42 publicly available pig genomes.

**Supplementary Table 8**. Population genetics statistics for Chinese pigs in this study.

**Supplementary Table 9**. Pair-wise $F_{ST}$ values among Chinese pig populations.

**Supplementary Table 10.** Statistics of introgression between Chinese and European domestic pig genomes.

**Supplementary Table 11**. Putative sweep regions on autosomes for local adaptation to low or high latitude in Chinese pigs.

**Supplementary Table 12**. The 219 candidate genes on autosomes for local adaptation to low or high latitude in Chinese pigs.

Nature Genetics: doi:10.1038/ng.3199

**Supplementary Table 13**. Go enrichment analysis on the 219 candidate genes that likely contribute to local adaptation to high or low latitude in Chinese pigs.

**Supplementary Table 14.** Protein-altering SNPs outliers in 10 of 219 candidate genes under selection for high/low-latitude environments in Chinese pigs.

**Supplementary Table 15.** The genomic difference between the 48 Mb recombination-decreasing region and the outside region on chromosome X.

**Supplementary Table 16.** The spearman correlation between recombination rate and the count of 6K poly-T repeat sequence.

**Supplementary Table 17.** The allele frequency of 7 diagnostic chip SNPs within the 48 Mb region in Chinese and European pigs.

**Supplementary Table 18.** Number of segregating sites within and between northern and southern Chinese pigs.

**Supplementary Table 19.** Coalescent simulations showing that gene drift is not likely a cause of the sweep signal on chromosome X.

**Supplementary Table 20**. Z-test showing gene drift is not likely a cause of the sweep signal on chromosome X.

**Supplementary Table 21**. The time to most recent common ancestor ($T_{MRCA}$) within and outside the 48 Mb low-recombination region on chromosome X for southern and northern Chinese pigs.

**Supplementary Table 22.** The divergence time within and outside the 48 Mb low-recombination region on chromosome X among southern Chinese, northern Chinese and European pigs.

Nature Genetics: doi:10.1038/ng.3199

# Supplementary Note

## 1. Samples for whole genome sequencing

### 1.1 Environmental conditions of southern and northern Chinese pigs

The geographical origins of the sequenced 69 individuals can be roughly divided into two climate zones: the hot zone (south subtropical zone) at low latitudes, which accommodates the Bamaxiang, Luchuan, Wuzhishan pigs and southern Chinese wild boars, and that cold zone at high latitudes or altitude, which is home to the Erhualian, Laiwu, Min, Hetao and Tibetan pigs. The south subtropical zone is characterized for hot climate by long frost-free periods (338 - 365 days), and high average temperature (> 10 $^{o}$C) even in the coldest month **(Supplementary Table 1)**. No winter exists in these areas. Although Tibetan pigs from Sichuan, Yunnan and Tibet live at relatively low-latitude areas, they are explored to cold and hypoxia environments at high altitude. These Tibetan pigs together with another Tibetan geographic population at high latitude (Gansu) are thus classified into the cold-zone (high-latitude) group. The habitats of northern Chinese pigs and Tibetan pigs have variable climates in a year, however, they have a common climatic feature: a cold and dry winter with extreme temperature ranging from -50.1 to -15 $^{o}$C and short frost free period ranging from 50 - 191 days. Environmental temperature differs strikingly between the hot and cold climate zones, which differed by 30.6°C in minimal winter temperature and 220 frost-free days (**Supplementary Table 1**).

### 1.2 Selection of the Wuzhishan pig reference genome

The scaffolds and contigs of a previously sequenced Wuzhishan (WZSP) pig (Fang *et al.* 2012) were aligned to chromosomes of Duroc (Sscrofa10.2) (Groenen *et al.* 2012). LASTZ (Harris 2007) was used for alignment with the parameters of "M=254 K=4500 L=3000 Y=15000 C=2 T=2" (http://www.bx.psu.edu/miller_lab/). ChainNet that can accommodate inversions, translocations, duplications, large-scale deletions, and overlapping deletions (Kent *et al.* 2003) was used to combine traditional alignments

into larger structures. A total of 5,951 scaffolds and contigs were successfully mapped, resulting in a length of 2.47 Gb and 86% of genome coverage (**Supplementary Table 2**). Unaligned scaffolds and contigs were assembled into a pseudo-chromosome ChrUn with 100 bp gaps between adjacent sequence elements.

To evaluate which genome is more suitable for genomic analyses of Chinese pigs, we randomly chose 1 Mb reads each from 12 pigs representing the 11 Chinese breeds and wild boars, and mapped them to both WZSP and Duroc genomes. The alignment ratio of WZSP is higher than 96% while that of Duroc is only 89% (**Supplementary Table 3**). We also tested the mapping rate and SNP accuracy between the two reference genomes using the Mozaik aligner (Lee *et al.* 2014). To do so, we randomly selected sequencing reads from two individuals and mapped them against the WZSP and Duroc reference genomes, respectively. The mapping rate of WZSP was ~98% (98.24% and 97.96%) while that of Duroc was ~91% (91.23% and 90.84%). Then we compared the detected SNPs with the 60K chip SNP data. The consistent rate between sequence-based SNPs and 60K chip SNPs was 96% when using the Duroc SSC10.2 assembly, which was slightly less than the consistent rate of 98.2% when using the WZSP assembly. These results support that WZSP is better than Duroc as the reference genome of Chinese pigs. The associated gene set was also updated based on the WZSP genome by transferring the features from scaffolds to chromosome sequences.

## 2. Population genetics and evolutionary history of Chinese pigs

### 2.1 Methodology

To compare genomic similarity between Chinese and European pigs, we downloaded the publicly available whole-genome sequence data of 42 pigs (**Supplementary Table 7**) (Groenen *et al.* 2012). Most of these genomes were sequenced at less than 10-fold coverage. The 42 pig genome data were combined with our 69-samples to create a 111-samples data set. Population-based genotypes for the 111 pigs were created by GATK after BWA alignment and GATK preprocessing as mentioned above. To avoid the

potential bias between our data and the publicly available data, we called SNPs by comparing the genome sequence of each individual to the Wuzhishan reference genome, and then merge the called SNPs to form a common set of SNP data for all 111 individuals.

Genetic diversity at the genomic scale was measured for each individual by nucleotide diversity $\pi$ (Nei & Li 1979) and Watterson's estimator $\vartheta$(Watterson 1975) using the below formula.

$$\pi = \sum_{i=1}^{n} \sum_{j=1}^{i} x_i x_j \pi_{ij}$$

where $x_i$ and $x_j$ are the frequencies of the *i*th and *j*th sequences respectively, $\pi_{ij}$ is the number of nucleotide differences per nucleotide site between the *i*th and *j*th sequences and *n* is the number of sequences. The summation is taken over all distinct *i-j* pairs without repetition.

$$\vartheta = S/a_n$$

where *S* is the number of segregating sites and $a_n$ is the (*n*-1)th harmonic number calculated according to the following formula.

$$a_n = \sum_{i=1}^{n-1} \frac{1}{i}$$

To calculate derived allele frequency (DAF) in each population, only those alleles with frequencies of greater than 90% in 6 sequenced Chinese wild boars were defined as ancestral alleles. At most one heterozygote is allowed at each ancestral allele for the 6 wild boars. Loci with minor allele frequency (MAF) >=0.1 in Chinese wild boars were then excluded as derived alleles at these loci could not be defined.

To measure population differentiation, $F_{ST}$ (Akey *et al.* 2002) was calculated using the formula below.

$$F_{ST} = \frac{MSP - MSG}{MSP + (n_c - 1)MSG}$$

MSG represents the observed mean square errors for SNPs within populations,

$$MSG = \frac{1}{\sum_{i=1}^{s} n_i - 1} \sum_{i}^{s} n_i p_{Ai}(1 - p_{Ai})$$

MSP represents the observed mean square errors for SNPs between populations

$$MSP = \frac{1}{s-1} \sum_{i}^{s} n_i (p_{Ai} - \bar{p}_A)^2$$

Where s denotes the number of subpopulations, $p_{Ai}$ denotes the frequency of the SNP allele A in the ith subpopulation. $\bar{p}_A = \frac{n_i p_{Ai}}{\sum_i n_i}$ is weighted average of $\mathrm{p_{Ai}}$ across subpopulations, $n_c = \frac{1}{s-1}\sum_{i=1}^{s} n_i - \frac{\sum_i n_i^2}{\sum_i n_i}$ is the average sample size across samples that also incorporates and corrects for the variance in sample size over subpopulations.

Principal component analysis (PCA) was conducted using EIGENSOFT (Price *et al.* 2006). The neighbor-joining tree was constructed by MEGA (Tamura *et al.* 2011) using IBS distance matrix data of all individuals by Plink v.107. Linkage disequilibrium (LD) was calculated using PLINK (Purcell *et al.* 2007) with options "--r2 --ld-window-kb 1000 --ld-window-r2 0". LD extent was determined from LD decay plot at level 0.3.

**2.2 Population diversity**

In the 111-genome SNP dataset, in average, 10.4 million non-reference alleles were found for each individual, resulting in an individual SNP rate of 0.40% in Chinese pigs, which was much higher than the value (0.17%) reported for European pigs (Groenen *et al.* 2012) and was consistent with previous findings (Bosse *et al.* 2012; Groenen *et al.* 2012). Furthermore, the Chinese wild boars had a higher genetic diversity than the rest of populations, as reflected by the higher values of the $\pi$ and $\vartheta$ parameters (**Supplementary Table 8**). The frequencies of derived alleles and the observed heterozygosities were roughly comparable across the populations, suggesting similar genetic variability in all populations (**Supplementary Table 8**).

**2.3 Population genetic differentiation**

The PCA analysis (**Supplementary Fig. 9**) on the 111 sequenced individuals is consistent with the clear evolutionary split between Chinese and European pigs. The first principal component (PC1) distinguished European pigs from Chinese individuals, and the second principal component (PC2) illustrated the differentiation among the Chinese breeds (**Supplementary Fig. 9)**. The neighbour-joining tree also indicates the divergence – the European pigs defined their own separate clade, supporting the independent domestication origins of Chinese and European pigs (Larson *et al.* 2005; Larson *et al.* 2010). For Chinese domestic pigs, the clustering patterns reflected their geographical proximity. The southern Chinese breeds, including Bamaxiang, Luchuan, Wuzhishan and Xiang, were nested in a cluster, whereas the Laiwu, Min and Hetao breeds from northern China defined another group. All of the Tibetan pigs clustered together, as did the Erhualian, Jiangquhai and Meishan individuals, which are located in adjacent regions of the Yangtze River valley (**Fig. 1B**). Furthermore, the average genetic differentiation $F_{ST}$ values across all SNP studied between Chinese pig breeds were consistent with their geographical distributions (**Supplementary Table 9**). The Bamaxiang and Wuzhishan pigs, which are geographical neighbours, had the closest genetic relationship, as revealed by the lowest $F_{ST}$ (0.074) value. The highest $F_{ST}$ value (0.306) was found between the Min and Luchuan pigs, corresponding to their large geographical separation. Within Tibetan pigs, the Gansu population showed the highest average value of $F_{ST}$ against the other three Tibetan populations (Gansu: 0.129, Sichuan: 0.098, Tibet: 0.094, Yunnan: 0.094). This indicates the Gansu population is more distant to other Tibetan populations and is consistent with our previous conclusion based on the 60K chip SNP data (Ai *et al.* 2014).

**2.4 Linkage disequilibrium extents**

We investigated the extent of linkage disequilibrium (LD), as estimated by average distance between SNPs that correspond to linkage disequilibrium $r^2$ = 0.3, which is known as the "useful LD" (Aerts *et al.* 2007), in each breed (**Supplementary Table 8**).

As expected for an outbred and non-admixed population, the Chinese wild boars had the smallest LD extent (6 kb), in stark contrast to the Gansu Tibetan population, which had the largest LD extent (300 kb). High LD extents (>200 kb) were also found in the Laiwu, Luchuan and Min breeds, which were comparable to European domestic pigs (100–200 kb). The LD extended for shorter distances in the other Chinese breeds, from 50 kb in the Tibetan pigs to 92 kb in the Hetao pigs. The short-range LD pattern is different from that observed in dogs (Sutter *et al.* 2004) and cattle (de Roos *et al.* 2008). Notably, the inter-breed LD extent was much lower in Chinese domestic pigs (5 kb) than in European domestic pigs (17.5 kb) (**Supplementary Fig. 14**). This suggests that there is almost no LD between pairs of genes across Chinese diverse local breeds, agreeing with the observation by Amaral *et al* (2008) (Amaral *et al.* 2008). Chinese domestic pigs are thus well suited for an association-mapping strategy, as performed in humans.

**2.5 Genomic landscape of introgression between Chinese and European domestic pigs**

Introgression between species, subspecies, varieties or breeds is extensive and an important source of genetic diversity, which may contribute to adaptation in various organisms including animals (Yang *et al.* 2011) and even humans (Green *et al.* 2010). Introgression between Chinese and European domestic pigs has been documented. It is known that the introduction of southern Chinese pigs into Europe 200 years ago has contributed to the formation of European pigs (Giuffra *et al.* 2000). Here, the whole genome sequencing data enable us to identify genomic regions of potential introgression between Chinese and European pigs in a quantitative way.

We first conducted the ADMIXTURE analysis (Alexander *et al.* 2009) to roughly estimate the introgression fractions between Chinese and European pigs. As shown in **Supplementary Figure 10**, we found ~18.6% of Chinese haplotypes in European pigs when K=2. Almost all of the introgressed genomes in European pigs are of southern

Chinese pigs origin as demonstrated when K=3. The introgression origin of southern Chinese pigs is consistent with the historical records that pigs were imported from South China to in particular improve UK local breeds during the 18$^{th}$ century (Giuffra *et al.* 2000). From *K* = 2 to 4, the Min pigs showed signals of admixture with the European pigs (**Supplementary Fig. 10**), consistent with our previous observation of potential historical introgression in the Min pigs on the basis of the 60K SNP data (Ai *et al.* 2013; Ai *et al.* 2014). The fraction of introgression detected here is slightly smaller than the ~20% fraction estimated by Bosse *et al*. (2014) that used the same methodology but different Chinese breeds (Bosse *et al.* 2014). Bosse *et al*. (2014) used the whole-genome sequence information mainly from the Meishan pig, a breed quite similar to the Erhualian pig used in our study. Thus, we propose that the discrepancy between our study and the report by Bosse *et al*. (2014) is mainly attributable to the different Chinese breeds investigated.

We further applied a likelihood ratio test method (McNally *et al.* 2009) to identify potential ancestral introgression in genomic regions between Chinese and European domestic pigs. Briefly, all putative introgressions between source (Chinese domestic pigs or European domestic pigs) and destination groups (one European breed or one Chinese breed) were examined for every window of 50 kb with at least 10 SNPs. Major alleles of source and destination groups were defined as alleles with the frequency greater than 50%. If major alleles differed between source and destination groups and the alleles in the destination group were same as these in the source group, they were recognized as differential sites. If the number of differential sites of a certain individual in the destination group exceeds 80% of total number of SNPs, the window was defined as introgression for this individual. The predicted introgression loci of each breed in the destination group were plotted along each chromosome. Introgression length, ratio of the introgressed genome and gene number involved were also tabulated for each breed.

The proportion of the introgression region in the genome ranges from 0.79% and

0.92% in Yunnan and Sichuan Tibetan pigs, respectively, to 35.3% in Min from Europe, and from 7.61% in Hampshire to 18.46% in Large White from China (**Supplementary Table 10**). For comparison purpose, we also adopted the D- statistics described in the paper by Groenen et al (2012) to estimate the introgression fractions between Wuzhishan and Large White pigs as an example. We queried derived allele from Wuzhishan (WZS) into Netherlands Wild Boar (WBNL) and Large White (LW). The D-statistic for the group (P1:P2:P3; WBNL:LW:WZS) was estimated to be -0.2459 using the following formula:

$$\mathrm{D} = \frac{nBABA - nABBA}{nBABA + nABBA}$$

The negative D value indicates that the gene flow occurred from Wuzhishan to Large White. And we calculated the admixture fraction of WZS into LW using the below formula:

$$f_{WZS,LW} = \frac{S(WBNL, LW, WZS)}{S(WBNL, BMX, WZS)}$$

where BMX represents Bamaxiang, a sister breed to WZS. We detected 16.07% of Wuzhishan haplotypes in Large White pigs for this group. This result is comparable with our estimated 18.5% fraction (Supplementary Table 9) of Wuzhishan components in Large White pigs.

Consistent with the results from ADMIXTRUE, the Min pig stands out as an admixed breed having the largest proportion of genome introduced from European pigs. Surprisingly, we observed two unusual large introgression regions on chromosomes 13 (58 Mb, Supplementary Fig. 12) and 15 (38 Mb, Supplementary Fig. 13). The two regions indicate a recent introgression event or possible under a certain force of selection.

## 3. Coalescent simulation analysis supports that the 14 Mb sweep region is not likely caused by gene drift

First, we performed two coalescence simulations for the target regions in southern

and northern Chinese pigs using the ms software (Hudson 2002) with demographic model from the pairwise sequentially markovian coalescent (Li & Durbin 2011) results: one without recombination and one without recombination plus 2-fold reduction of mutation rate. For each group (southern Chinese pigs or northern Chinese pigs), we calculated the probabilities (P) of observing the number of segregating sites (S) and Tajima's D (D) values within the haplogroup without selection. Under the condition of no recombination, both the probabilities of observing the S and D statistics in southern and northern Chinese pigs within the 14 Mb region were 0. Under the condition of no recombination plus 2-fold reduction of mutation rate, the probabilities of observing the S and D values within the 14 Mb region in southern and northern Chinese pigs were 2.5% and 0.2%, respectively (**Supplementary Table 19**). We also compared segregating sites and Tajima's D values within the target regions with those within other regions on autosomes. We found that the probabilities of segregating sites and Tajima's D values within the 14 Mb sweep region against their distributions on autosomes were smaller than 5% in both southern and northern Chinese pig populations (**Supplementary Table 20**). These results thus support the assumption that the 14 Mb region is unlikely to be caused by a lack of recombination and coalescence variance, but must be explained by natural selection.

## References


Aerts J., Megens H.J., Veenendaal T., Ovcharenko I., Crooijmans R., Gordon L., Stubbs L. & Groenen M. (2007) Extent of linkage disequilibrium in chicken. *Cytogenet Genome Res* **117**, 338-45.

Ai H., Huang L. & Ren J. (2013) Genetic diversity, linkage disequilibrium and selection signatures in chinese and Western pigs revealed by genome-wide SNP markers. *PLoS One* **8**, e56001.

Ai H.S., Yang B., Li J., Xie X.H., Chen C. & Ren J. (2014) Inference of population history and genome-wide detection of signatures for high-altitude adaptation in Tibetan pigs. *BMC Genomics*.

Akey J.M., Zhang G., Zhang K., Jin L. & Shriver M.D. (2002) Interrogating a high-density SNP map for signatures of natural selection. *Genome Res* **12**, 1805-14.

Alexander D.H., Novembre J. & Lange K. (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* **19**, 1655-64.

Amaral A.J., Megens H.J., Crooijmans R.P., Heuven H.C. & Groenen M.A. (2008) Linkage disequilibrium decay and haplotype block structure in the pig. *Genetics* **179**, 569-79.

Bosse M., Megens H.J., Madsen O., Frantz L.A., Paudel Y., Crooijmans R.P. & Groenen M.A. (2014) Untangling the hybrid nature of modern pig genomes: a mosaic derived from biogeographically distinct and highly divergent Sus scrofa populations. *Mol Ecol* **23**, 4089-102.

Bosse M., Megens H.J., Madsen O., Paudel Y., Frantz L.A., Schook L.B., Crooijmans R.P. & Groenen M.A. (2012) Regions of homozygosity in the porcine genome: consequence of demography and the recombination landscape. *PLoS Genet* **8**, e1003100.

de Roos A.P., Hayes B.J., Spelman R.J. & Goddard M.E. (2008) Linkage disequilibrium and persistence of phase in Holstein-Friesian, Jersey and Angus cattle. *Genetics* **179**, 1503-12.

Fang X., Mou Y., Huang Z., Li Y., Han L., Zhang Y., Feng Y., Chen Y., Jiang X., Zhao W., Sun X., Xiong Z., Yang L., Liu H., Fan D., Mao L., Ren L., Liu C., Wang J., Li K., Wang G., Yang S., Lai L., Zhang G., Li Y., Wang J., Bolund L., Yang H., Wang J., Feng S., Li S. & Du Y. (2012) The sequence and analysis of a Chinese pig genome. *Gigascience* **1**, 16.

Frantz L.A., Schraiber J.G., Madsen O., Megens H.J., Bosse M., Paudel Y., Semiadi G., Meijaard E., Li N., Crooijmans R.P., Archibald A.L., Slatkin M., Schook L.B., Larson G. & Groenen M.A. (2013) Genome sequencing reveals fine scale diversification and reticulation history during speciation in Sus. *Genome Biol* **14**, R107.

Giuffra E., Kijas J.M., Amarger V., Carlborg O., Jeon J.T. & Andersson L. (2000) The origin of the domestic pig: independent domestication and subsequent introgression. *Genetics* **154**, 1785-91.

Green R.E., Krause J., Briggs A.W., Maricic T., Stenzel U., Kircher M., Patterson N., Li H., Zhai W., Fritz M.H., Hansen N.F., Durand E.Y., Malaspinas A.S., Jensen J.D., Marques-Bonet T., Alkan C., Prufer K., Meyer M., Burbano H.A., Good J.M., Schultz R., Aximu-Petri A., Butthof A., Hober B., Hoffner B., Siegemund M., Weihmann A., Nusbaum C., Lander E.S., Russ C., Novod N., Affourtit J., Egholm M., Verna C., Rudan P., Brajkovic D., Kucan Z., Gusic I., Doronichev V.B., Golovanova L.V., Lalueza-Fox C., de la Rasilla M., Fortea J., Rosas A., Schmitz R.W., Johnson P.L., Eichler E.E., Falush D., Birney E., Mullikin J.C., Slatkin M., Nielsen R., Kelso J., Lachmann M., Reich D. & Paabo S. (2010) A draft sequence of the Neandertal genome. *Science* **328**, 710-22.

Groenen M.A., Archibald A.L., Uenishi H., Tuggle C.K., Takeuchi Y., Rothschild M.F., Rogel-Gaillard C., Park C., Milan D., Megens H.J., Li S., Larkin D.M., Kim H., Frantz L.A., Caccamo M., Ahn H., Aken B.L., Anselmo A., Anthon C., Auvil L.,

Badaoui B., Beattie C.W., Bendixen C., Berman D., Blecha F., Blomberg J., Bolund L., Bosse M., Botti S., Bujie Z., Bystrom M., Capitanu B., Carvalho-Silva D., Chardon P., Chen C., Cheng R., Choi S.H., Chow W., Clark R.C., Clee C., Crooijmans R.P., Dawson H.D., Dehais P., De Sapio F., Dibbits B., Drou N., Du Z.Q., Eversole K., Fadista J., Fairley S., Faraut T., Faulkner G.J., Fowler K.E., Fredholm M., Fritz E., Gilbert J.G., Giuffra E., Gorodkin J., Griffin D.K., Harrow J.L., Hayward A., Howe K., Hu Z.L., Humphray S.J., Hunt T., Hornshoj H., Jeon J.T., Jern P., Jones M., Jurka J., Kanamori H., Kapetanovic R., Kim J., Kim J.H., Kim K.W., Kim T.H., Larson G., Lee K., Lee K.T., Leggett R., Lewin H.A., Li Y., Liu W., Loveland J.E., Lu Y., Lunney J.K., Ma J., Madsen O., Mann K., Matthews L., McLaren S., Morozumi T., Murtaugh M.P., Narayan J., Nguyen D.T., Ni P., Oh S.J., Onteru S., Panitz F., Park E.W., Park H.S., Pascal G., Paudel Y., Perez-Enciso M., Ramirez-Gonzalez R., Reecy J.M., Rodriguez-Zas S., Rohrer G.A., Rund L., Sang Y., Schachtschneider K., Schraiber J.G., Schwartz J., Scobie L., Scott C., Searle S., Servin B., Southey B.R., Sperber G., Stadler P., Sweedler J.V., Tafer H., Thomsen B., Wali R., Wang J., Wang J., White S., Xu X., Yerle M., Zhang G., Zhang J., Zhang J., Zhao S., Rogers J., Churcher C. & Schook L.B. (2012) Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* **491**, 393-8.

Harris R.S. (2007) Improved pairwise alignment of genomic DNA. *ProQuest*.

Hudson R.R. (2002) Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* **18**, 337-8.

Kent W.J., Baertsch R., Hinrichs A., Miller W. & Haussler D. (2003) Evolution's cauldron: duplication, deletion, and rearrangement in the mouse and human genomes. *Proc Natl Acad Sci U S A* **100**, 11484-9.

Larson G., Dobney K., Albarella U., Fang M., Matisoo-Smith E., Robins J., Lowden S., Finlayson H., Brand T., Willerslev E., Rowley-Conwy P., Andersson L. & Cooper A. (2005) Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. *Science* **307**, 1618-21.

Larson G., Liu R., Zhao X., Yuan J., Fuller D., Barton L., Dobney K., Fan Q., Gu Z., Liu X.H., Luo Y., Lv P., Andersson L. & Li N. (2010) Patterns of East Asian pig domestication, migration, and turnover revealed by modern and ancient DNA. *Proc Natl Acad Sci U S A* **107**, 7686-91.

Lee W.P., Stromberg M.P., Ward A., Stewart C., Garrison E.P. & Marth G.T. (2014) MOSAIK: a hash-based algorithm for accurate next-generation sequencing short-read mapping. *PLoS One* **9**, e90581.

Li H. & Durbin R. (2011) Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493-6.

McNally K.L., Childs K.L., Bohnert R., Davidson R.M., Zhao K., Ulat V.J., Zeller G., Clark R.M., Hoen D.R., Bureau T.E., Stokowski R., Ballinger D.G., Frazer K.A., Cox D.R., Padhukasahasram B., Bustamante C.D., Weigel D., Mackill D.J., Bruskiewich R.M., Ratsch G., Buell C.R., Leung H. & Leach J.E. (2009)

Genomewide SNP variation reveals relationships among landraces and modern varieties of rice. *Proc Natl Acad Sci U S A* **106**, 12273-8.

McVean G.A., Myers S.R., Hunt S., Deloukas P., Bentley D.R. & Donnelly P. (2004) The fine-scale structure of recombination rate variation in the human genome. *Science* **304**, 581-4.

Nei M. & Li W.H. (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc Natl Acad Sci U S A* **76**, 5269-73.

Price A.L., Patterson N.J., Plenge R.M., Weinblatt M.E., Shadick N.A. & Reich D. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904-9.

Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A., Bender D., Maller J., Sklar P., de Bakker P.I., Daly M.J. & Sham P.C. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-75.

Sutter N.B., Eberle M.A., Parker H.G., Pullar B.J., Kirkness E.F., Kruglyak L. & Ostrander E.A. (2004) Extensive and breed-specific linkage disequilibrium in Canis familiaris. *Genome Res* **14**, 2388-96.

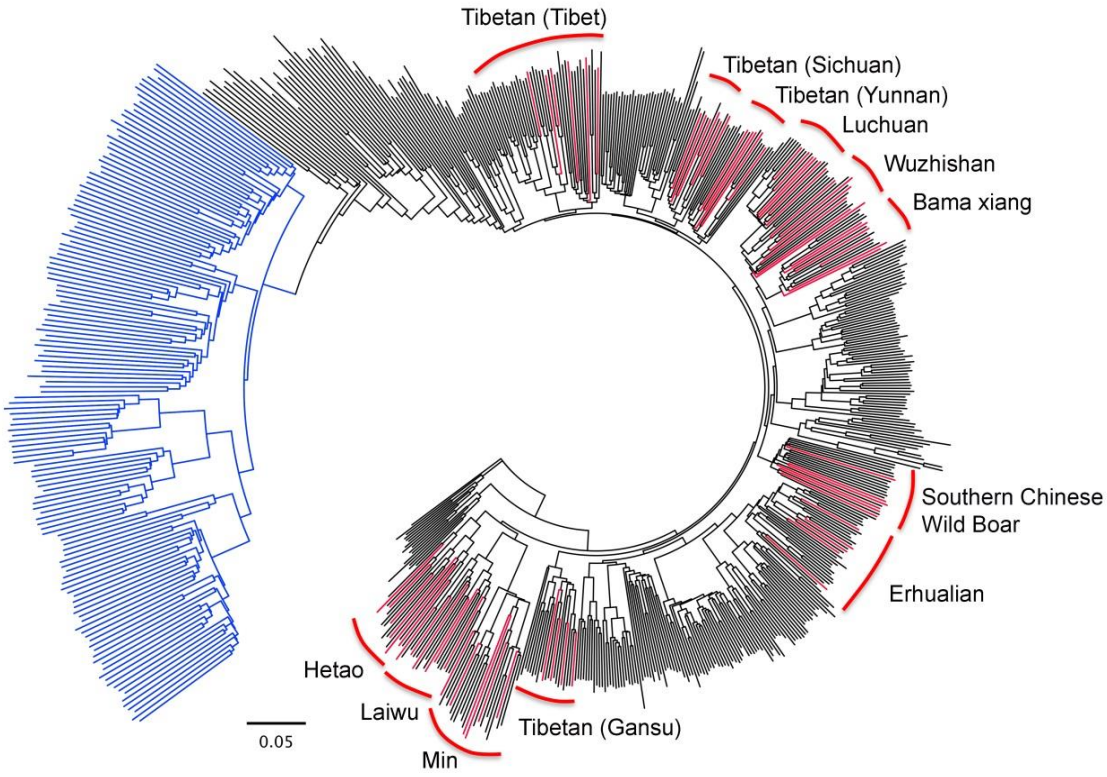
Tamura K., Peterson D., Peterson N., Stecher G., Nei M. & Kumar S. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28**, 2731-9.

Wang, L. (2011) Animal genetic resources in China: Pigs. *China Agricultural Press* 2-16.

Watterson G.A. (1975) On the number of segregating sites in genetical models without recombination. *Theor Popul Biol* **7**, 256-76.

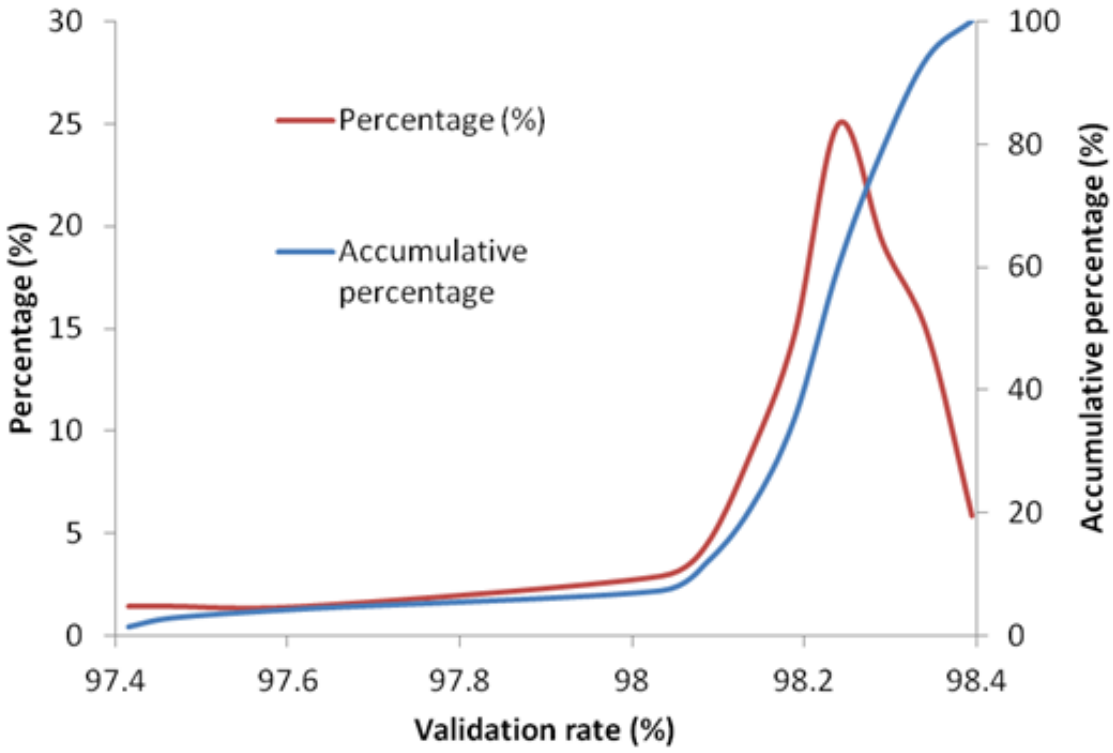
Yang H., Wang J.R., Didion J.P., Buus R.J., Bell T.A., Welsh C.E., Bonhomme F., Yu A.H., Nachman M.W., Pialek J., Tucker P., Boursot P., McMillan L., Churchill G.A. & de Villena F.P. (2011) Subspecific origin and haplotype diversity in the laboratory mouse. *Nat Genet* **43**, 648-55.

## Supplementary Figures




**Supplementary Fig. 1**

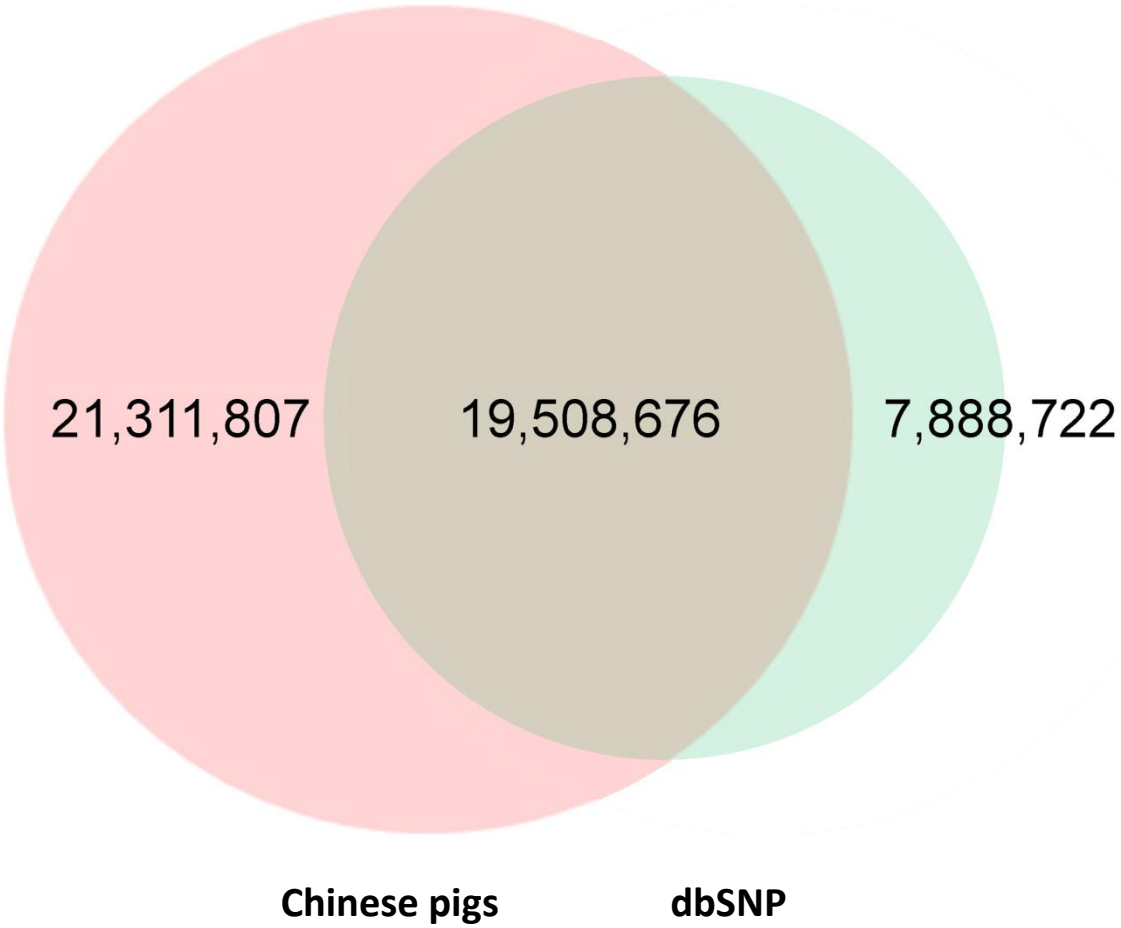
**Selection of Chinese representative pigs from a large sample of 678 animals based on the Illumina porcine 60K SNP data.** The selected 69 individuals are highlighted in red, and the corresponding breeds are indicated in the figure. The 137 European pigs as outgroups are indicated in blue.




**Supplementary Fig. 2**

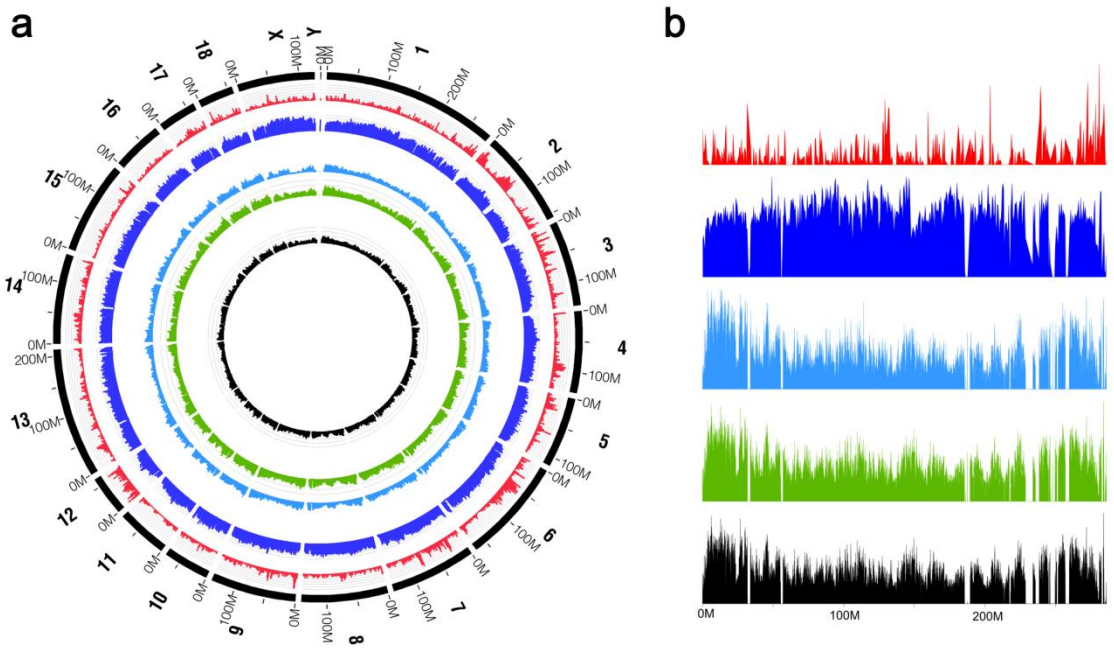
**The concordance between sequence-based SNPs and Illumina porcine 60K SNPs in the tested samples.** Informative chip SNPs were compared with sequenced-based SNPs for all 69 samples. Chip SNPs with the same base pairs as sequenced-based SNPs were defined as validated SNPs. The distribution of the percentages of validated SNPs for all samples were recorded and plotted in both frequency density (red) and accumulative ways (blue).

Nature Genetics: doi:10.1038/ng.3199




**Supplementary Fig. 3**

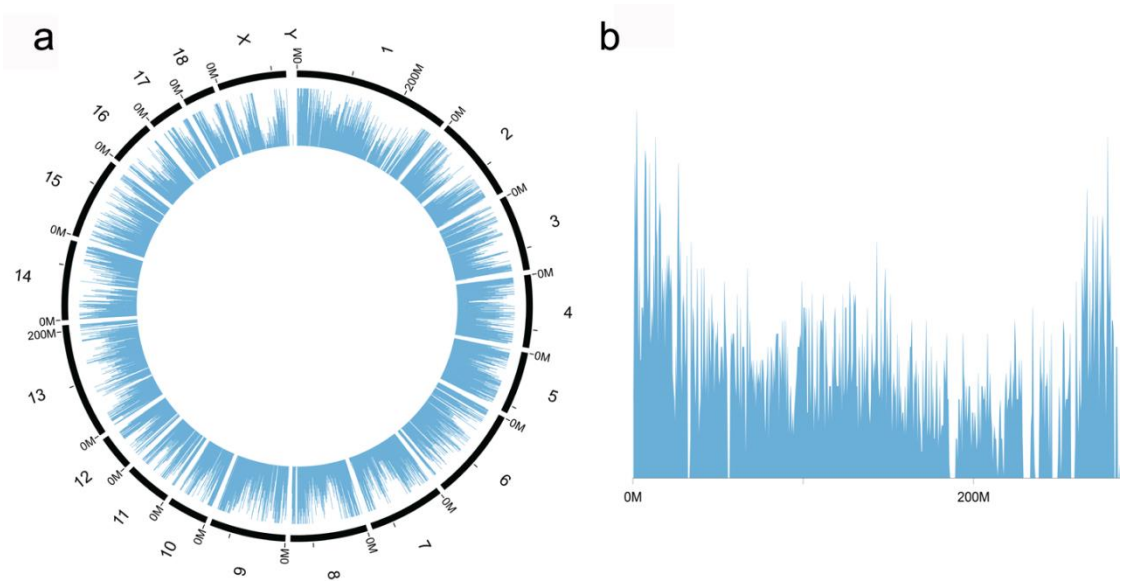
**Venn diagram showing novel variants absent from Build138 of dbSNP in Chinese pigs.**




**Supplementary Fig. 4**

**Circos plot showing global distribution of variants along the genome.** (a) The circles from outside to inside illustrate gene density, repeat density, $\pi$, $\vartheta$ and SNP density, respectively. Higher heterozygosity in telomere was observed by evaluating the statistical magnitude of $\pi$ in the 69 pig genomes. $Mean_{(terminus)}$ = 5.79e-3, $Variance_{(terminus)}$ = 6.42e-6; $Mean_{(inner)}$ = 3.51e-3, $Variance_{(inner)}$ = 4.20e-6. (b) Distribution of variants on chromosome 1.

Nature Genetics: doi:10.1038/ng.3199

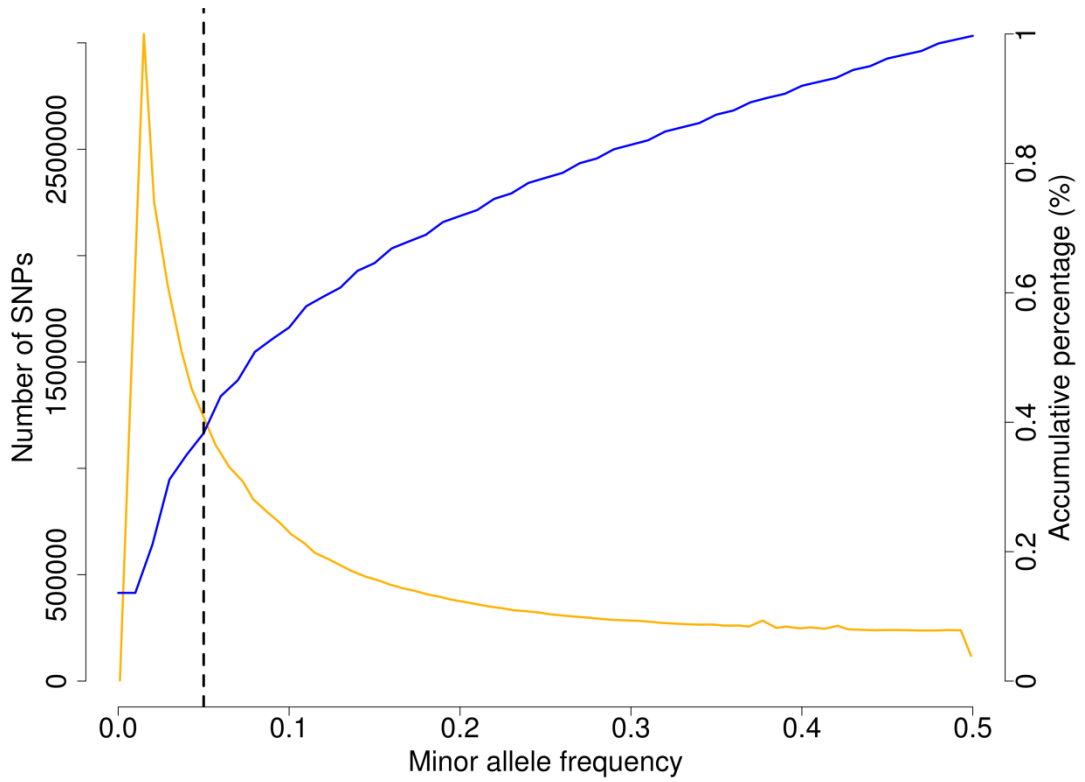



**Supplementary Fig. 5**

**Global distribution of structural variants along chromosomes in a whole population.**

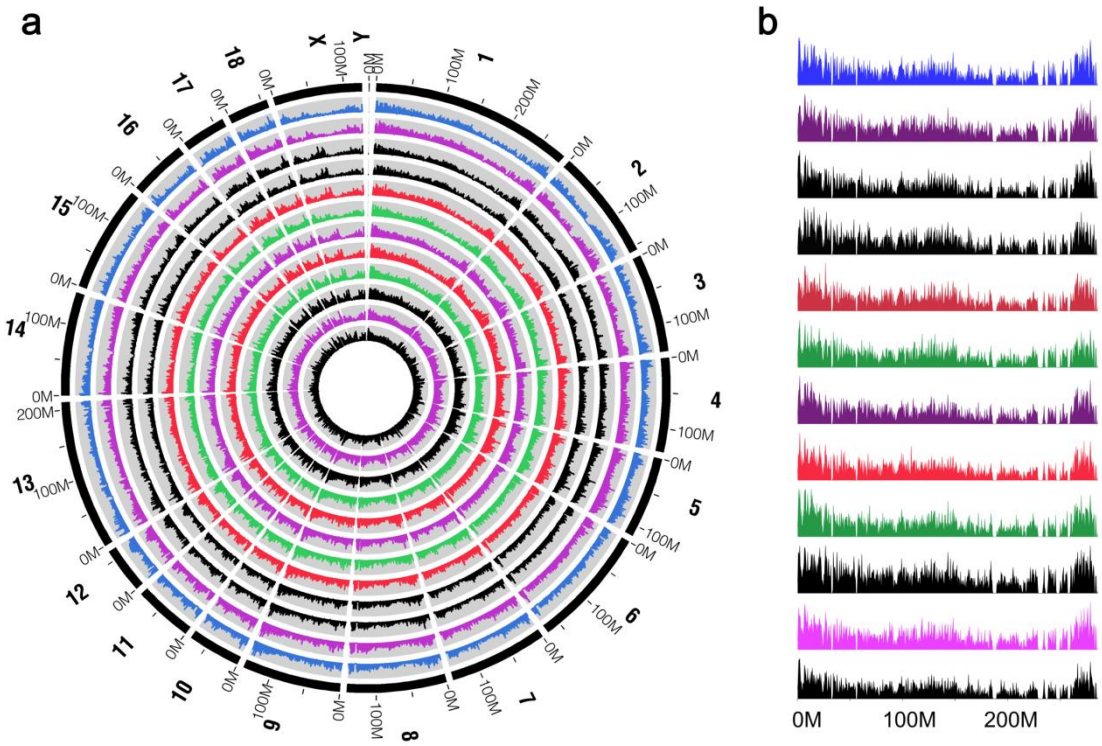
(a) The structural variants (SVs) number was slightly higher in the termini of chromosomes when counting the number of SVs in a 500 kb non-overlapping sliding window ($Mean_{(terminus)}$ = 11.6, $SD_{(terminus)}$ = 2.21; $Mean_{(inner)}$ = 8.20, $SD_{(inner)}$ = 1.04), which may also be caused by elevated recombination rates. (b) An illustration of distribution of structural variants on chromosome 1.

Nature Genetics: doi:10.1038/ng.3199



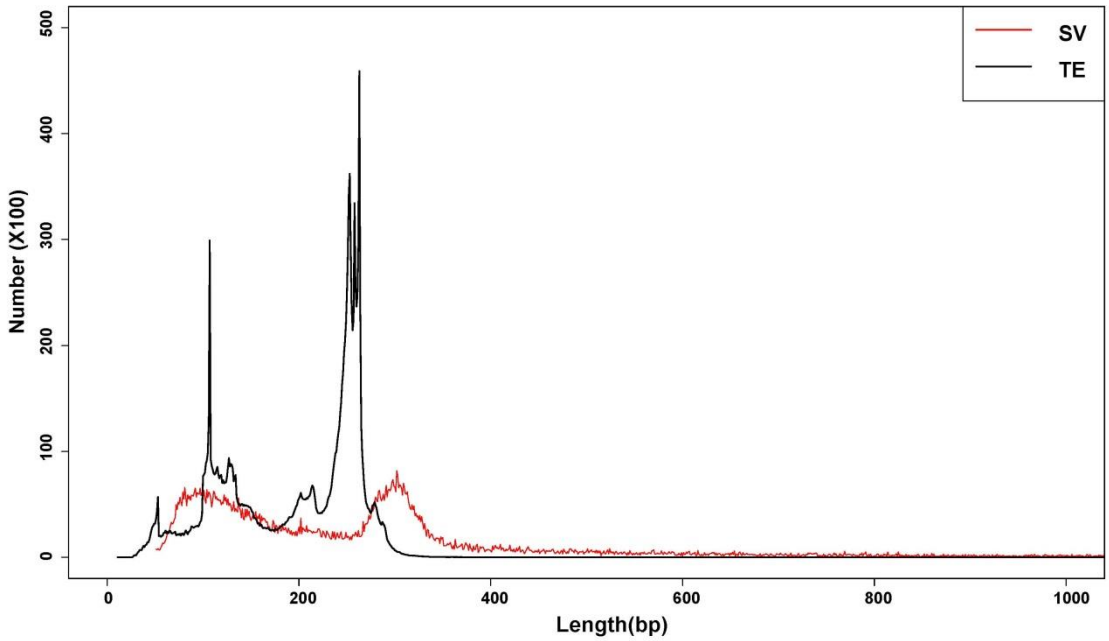

**Supplementary Fig. 6**

**Distribution profile of minor allele frequencies of SNPs called in this study.** Orange curve indicates density distribution and blue curve indicates accumulative distribution. Dashed vertical line denotes minor allele frequency of 0.05.




**Supplementary Fig. 7**

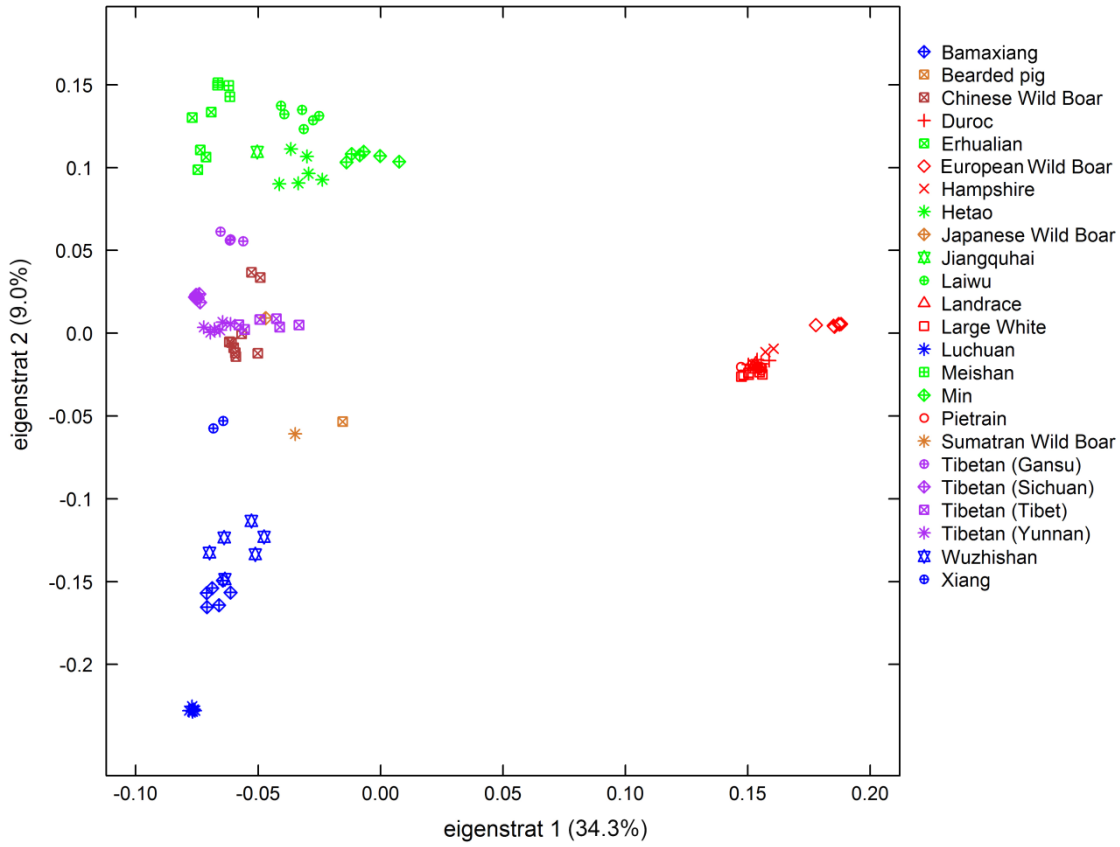
**Global distribution of structural variants along chromosome in each Chinese breed and wild boars.** (a) The circles from outside to inside illustrate different breeds: Bamaxiang, southern Chinese Wild Boar, Erhualian, Tibetan (Gansu), Hetao, Luchuan, Laiwu, Min, Tibetan (Sichuan), Tibetan (Tibet), Wuzhishan and Tibetan (Yunnan). (b) An illustration of distribution of structural variants on chromosome 1 in each breed.




**Supplementary Fig. 8**

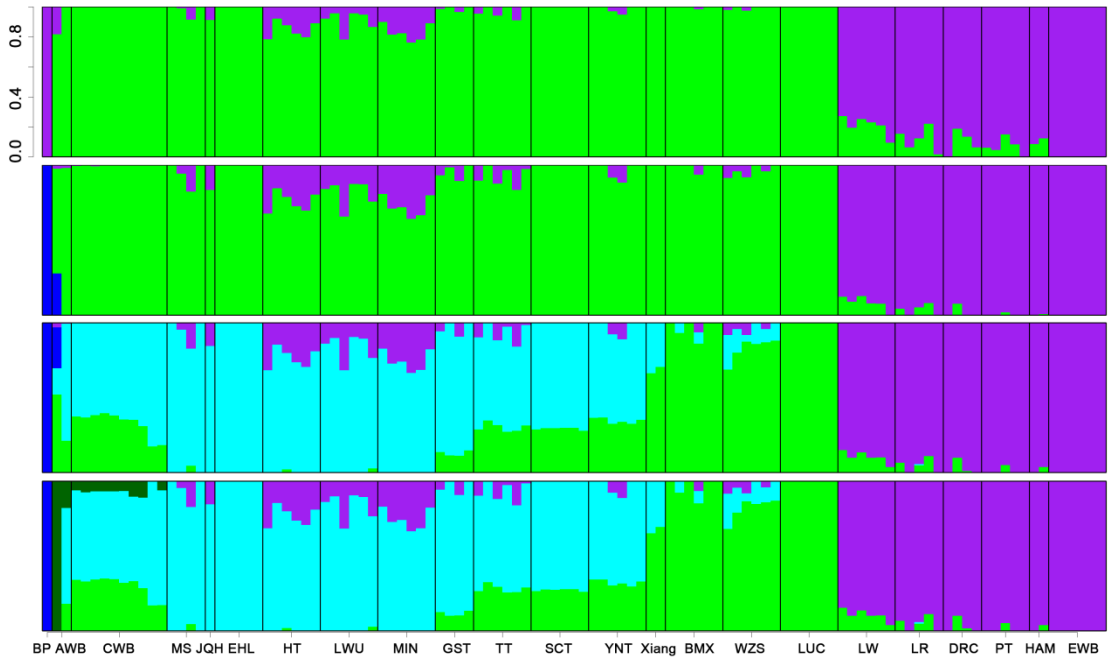
**Comparison of the size distribution patterns between structural variants and annotated transposon elements.** Both curves of SV and TE show bimodal distribution with peaks at length of around 80 and 300 bp.SV, Structural variants; TE, transposon element (SINE/tRNA).

Nature Genetics: doi:10.1038/ng.3199



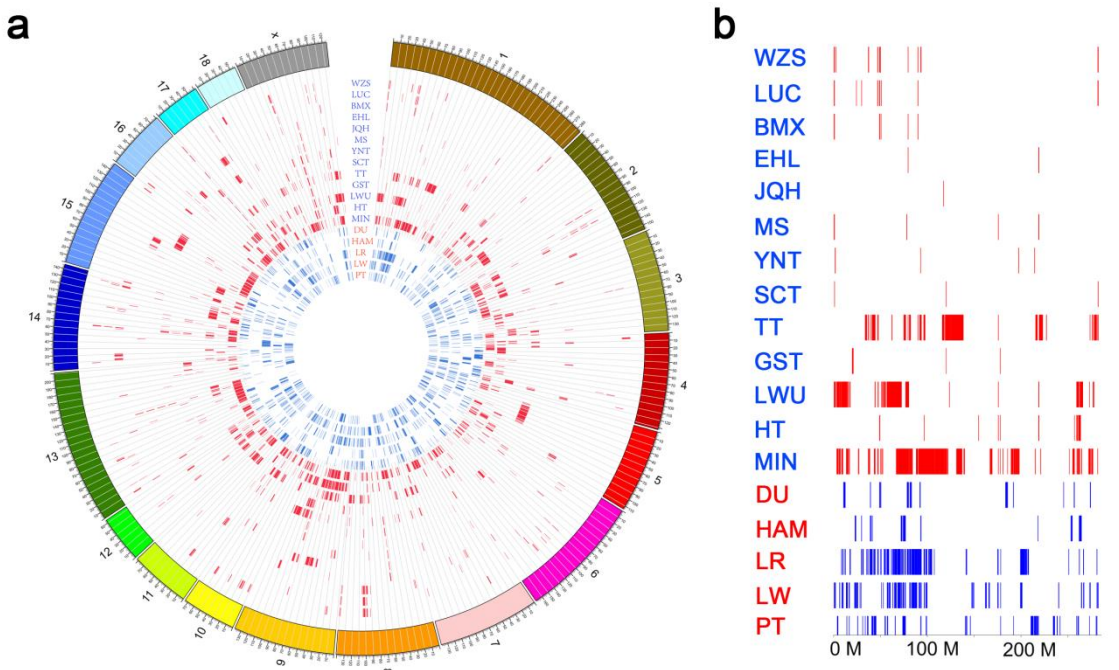

**Supplementary Fig. 9**

**Principal component analysis of whole-genome SNP data on 111 Chinese and European pigs.**




**Supplementary Fig. 10**

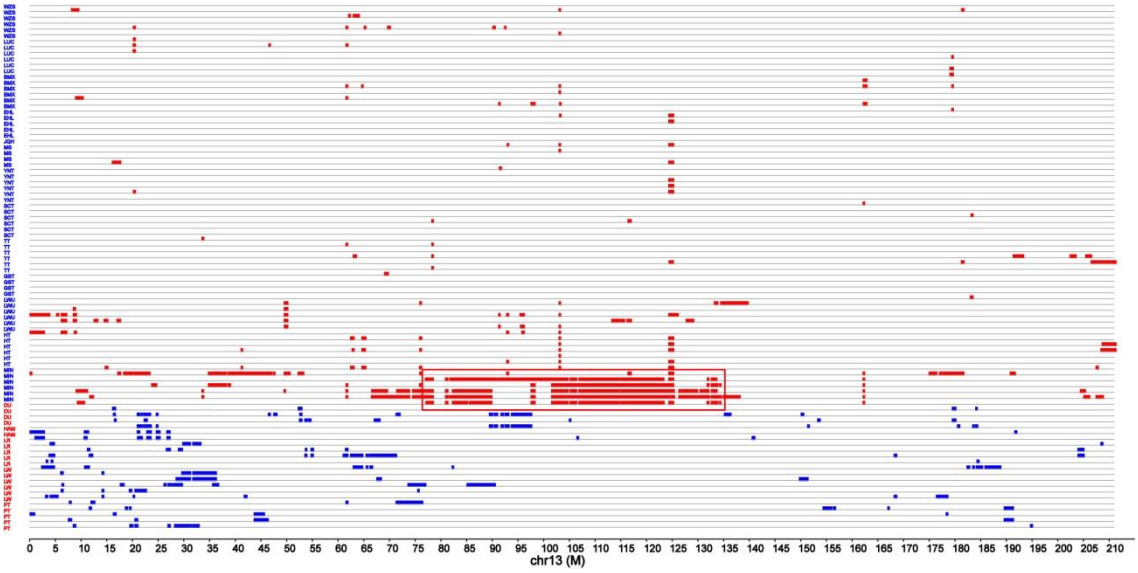
**Admixture analysis of 111 Chinese and European pigs.** AWB includes bearded pig, Sumatran wild boar and Japanese wild boar; CWB indicates southern and northern Chinese wild boars; GST, Tibetan (Gansu), TT, Tibetan (Tibet); SCT, Tibetan (Sichuan); YNT, Tibetan (Yunnan); LUC, Luchuan; WZS, Wuzhishan; BMX, Bamaxiang; MS, Meishan; JQH, Jiangquhai; EHL, Erhualian; HT, Hetao, LWU, Laiwu; MIN, Min; DRC, Duroc; LW, Large White; LR, Landrace; PT, Pietrain; HAM, Hampshire; EWB, European wild boars.




**Supplementary Fig. 11**

**Illustration of introgression between Chinese and European domestic pig genomes.** (a) The breeds are, from the periphery to the center, Wuzhishan (WZS), Luchuan (LUC), Bamaxiang (BMX), Erhualian (EHL), Jiangquhai (JQH), Meishan (MS), Tibetan (Yunnan, YNT), Tibetan (Sichuan, SCT), Tibetan (Tibet, TT), Tibetan (Gansu, GST), Laiwu (LWU), Hetao (HT), Min (MIN), Duroc (DU), Hampshire (HAM), Landrace (LR), Large White (LW) and Pietrain (PT). Red blocks refer to the introgression from European to Chinese lineages while blue blocks refer to the introgression from Chinese to European lineages. (b) Illustration of introgression between Chinese and European domestic pig on chromosome 1.

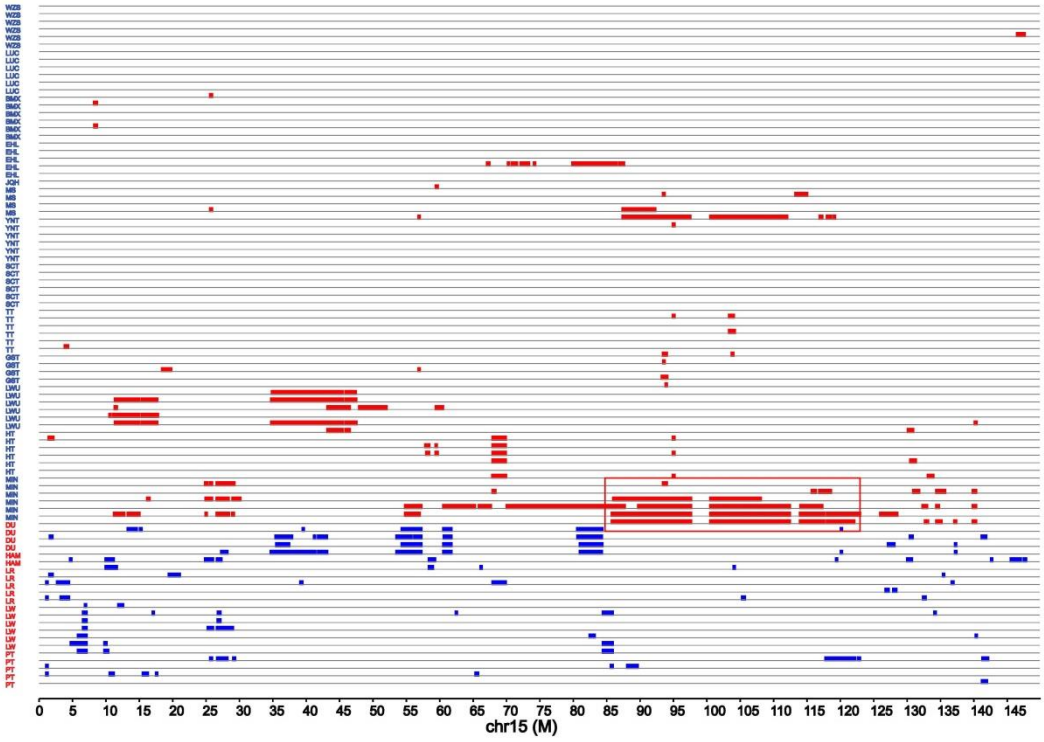
Nature Genetics: doi:10.1038/ng.3199




**Supplementary Fig. 12**

**The large potential introgression region on chromosome 13 in Min pigs.** Red blocks refer to the introgression from European into Chinese lineages and blue blocks refer to the introgression from Chinese into European lineages. The large introgression region in Min pigs is indicated by a red box. Genomic positions in Mb along this chromosome are given in the x-axis. The breed abbreviations in the y-axis are the same as those in Supplementary Fig. 11.

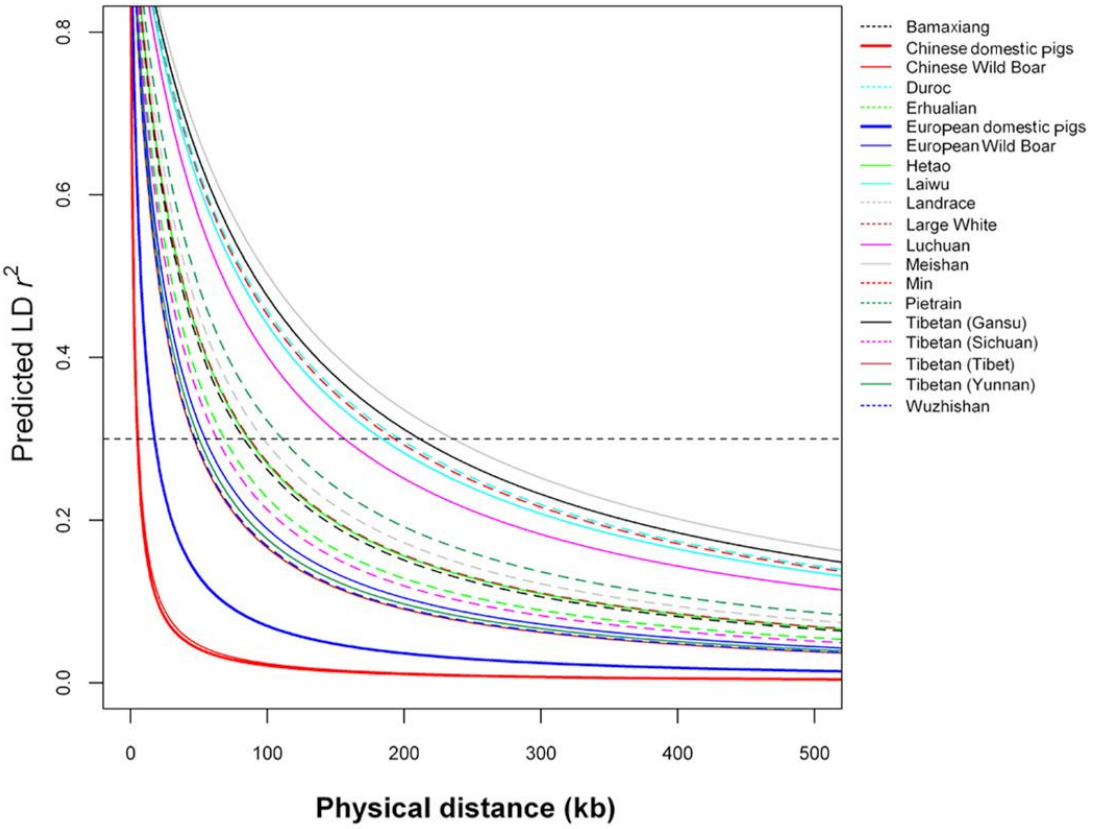
Nature Genetics: doi:10.1038/ng.3199




**Supplementary Fig. 13**

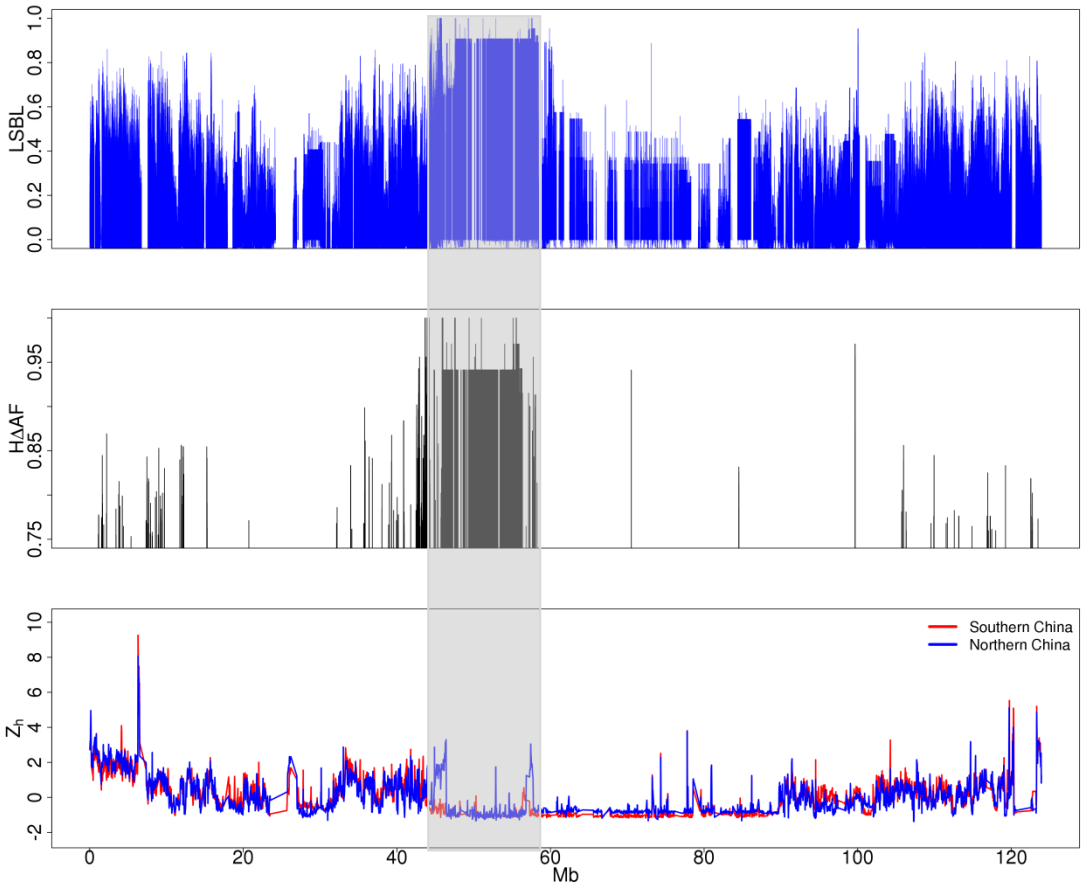
**The large potential introgression region on chromosome 15 in Min pigs.** Red blocks refer to the introgression from European to Chinese lineages and blue blocks refer to the introgression from Chinese to European lineages. The large introgression region in Min pigs is indicated by a red box. Genomic positions in Mb along this chromosome are given in the x-axis. The breed abbreviations in the y-axis are the same as those in Supplementary Fig. 11.

Nature Genetics: doi:10.1038/ng.3199



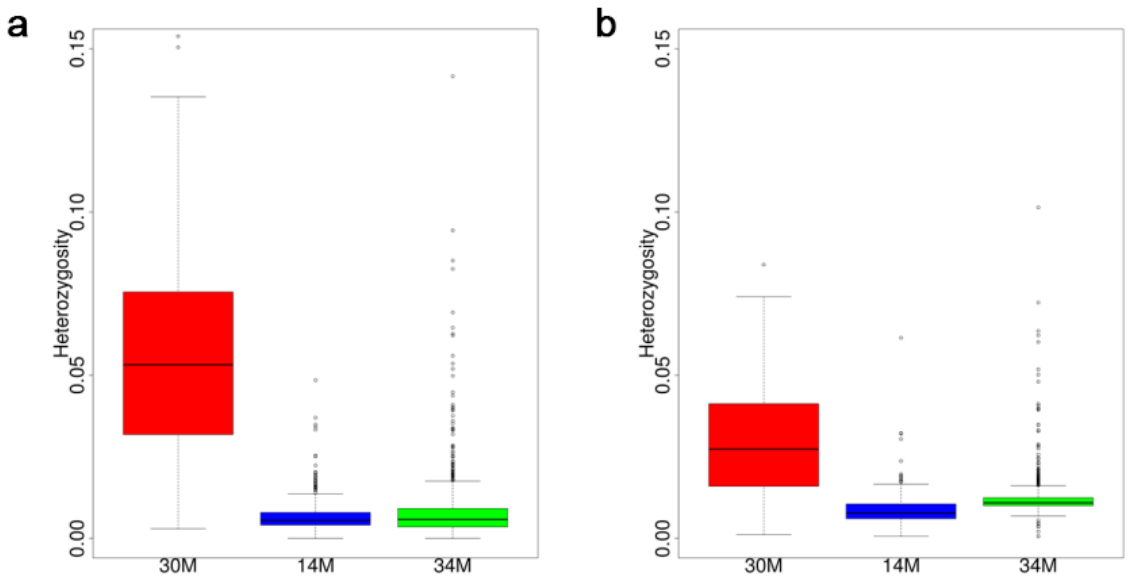

**Supplementary Fig. 14**

**The linkage disequilibrium decay plots for Chinese and European pigs.**




**Supplementary Fig. 15**

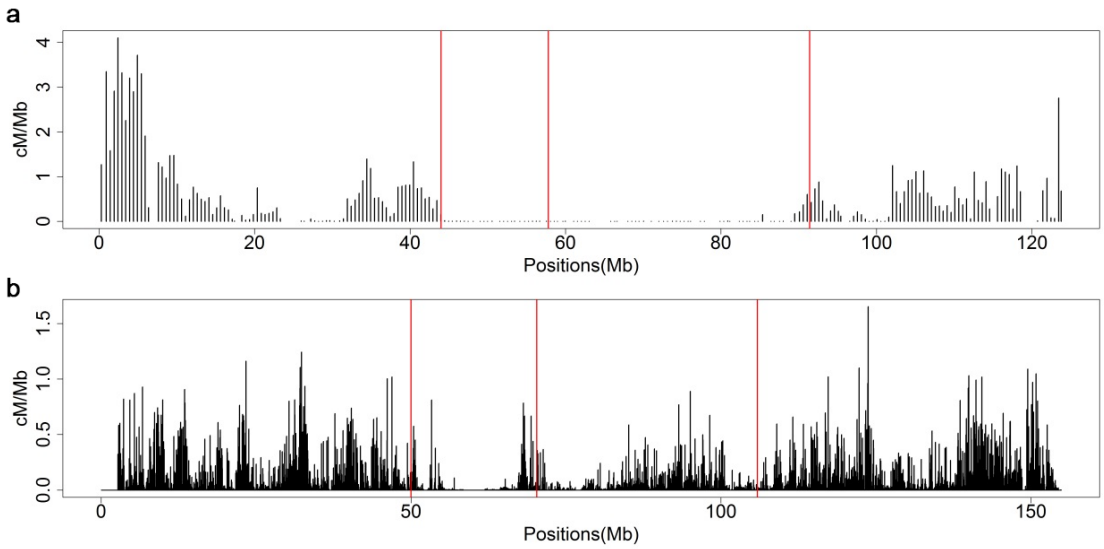
**Illustration of the X-linked sweep region by three statistics.** From the top down, the vertical axis indicates values of LSBL, HΔAF and $Z_h$ respectively. The horizontal axis shows positions in Mb along the X chromosome. Statistics were calculated separately for northern (high-latitude) and southern (low-latitude) Chinese pigs, and plotted for the whole region on the X chromosome. These statistics clearly indicate a strong sweep signal in a 14 Mb region (shaded) that exhibits strong LSBL and HΔAF scores between southern and northern Chinese pigs, reduced heterozygosity in both northern and southern Chinese pigs.




**Supplementary Fig. 16**

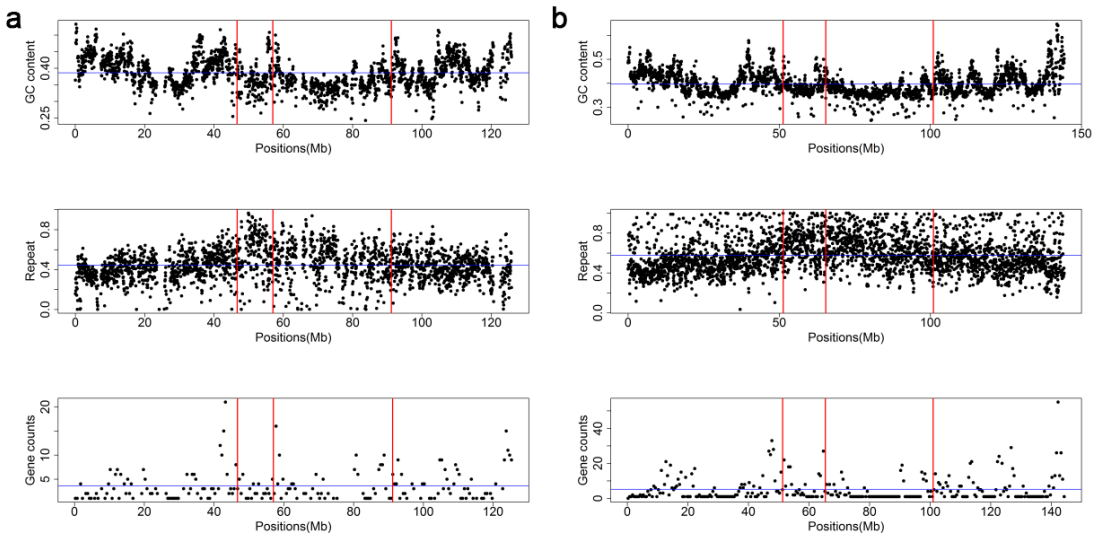
**Comparison of heterozygosity of the X-linked sweep, flanking regions and autosomes.** Heterozygosity was calculated for the 14 Mb sweep (44-58 Mb, blue) and its flanking 30 Mb (14-44 Mb, red) and 34 Mb regions (58-92 Mb, green) in southern Chinese pigs (A) and northern Chinese pigs (B). Each dot represents the heterozygosity calculated by each 10 kb non-overlapping window. From this figure, we can find that the 14 Mb and 34 Mb regions have extremely low heterozygosity in both southern and northern Chinese pigs.

Nature Genetics: doi:10.1038/ng.3199



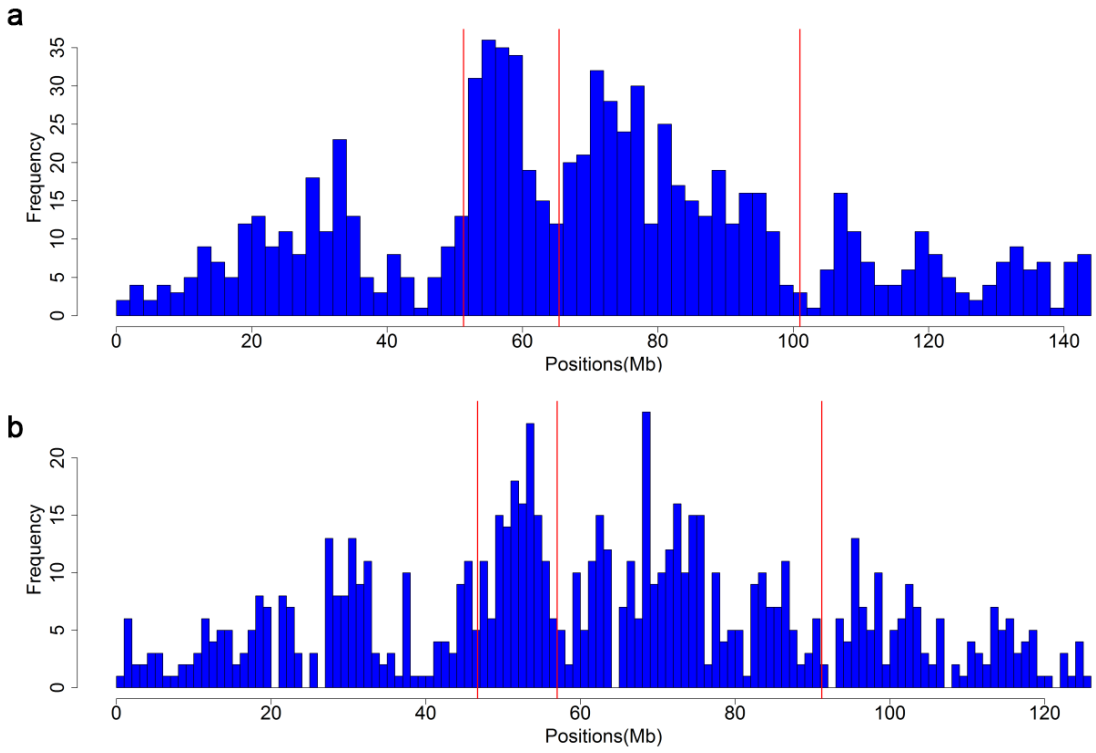

**Supplementary Fig. 17**

**The distribution of recombination rate on chromosome X.** The genetic distance (cM) was plotted against 500 kb non-overlapping windows on chromosome X in pig (upper panel) and human (lower panel) genomes. The vertical red line denotes the boundaries of the low recombination rate region corresponding to the 14 Mb sweep and the flanking 34 Mb segment. The boundaries in humans were determined by aligning chromosome X sequences of the Wuzhishan pig and human reference genome using the LastZ program (Harris 2007). The recombination rates in pigs were calculated by LDhat 2.1 (McVean *et al.* 2004) using whole genome sequence data generated in this study. The recombination rate data in humans were downloaded from database of HapMap project (http://hapmap.ncbi.nlm.nih.gov/).




**Supplementary Fig. 18**

**The distribution of GC content, repeat sequence and gene counts on chromosome X.** (A) the Wuzhishan reference genome. (B) the Duroc reference genome. The GC content (upper panels) and repeated sequence content (middle panels) were plotted by 50 kb non-overlapping windows, and gene counts (lower panels) were plotted by 500 kb non-overlapping windows. The blue horizontal lines represent the mean values, and red vertical lines denote boundaries of the low recombination region corresponding to the 14 Mb sweep and the flanking 34 Mb segment.




**Supplementary Fig. 19**

**Distribution of the 6K poly-T sequence on chromosome X.** (A) the Duroc reference genome.(B) the Wuzhishan reference genome. The 60K poly-T core sequence corresponds to 45,743,141 – 45,749,642 bp in the Wuzhishan reference genome. The counts of the 6K poly-T sequence were plotted by 50 kb non-overlapping windows. The red vertical lines represent boundaries of the low recombination region corresponding to the 14 Mb sweep and the flanking 34 Mb segment.

## Supplementary Tables

**Supplementary Table 1.** Samples, origin and environmental variables of Chinese pigs sequenced in this study [a]

| | | Environmental variable | | | | |
|---|---|---|---|---|---|---|
| **Population** | **Sample** | **Origin** | **Latitude (degree)** | **Altitude (m)** | **Frost-free period (d)** | **Extreme temp in winter (°C)** |
| **Low-latitude (Southern China) pigs** | | | | | | |
| Bamaxiang | 6 | Bama, Guangxi | 24.2 | 830 | 338 | -3.3 |
| Luchuan | 6 | Luchuan, Guangxi | 22.3 | 70 | 359 | 1 |
| Wuzhishan | 6 | Haikou, Hainan | 18.8 | 330 | 365 | 15 |
| Wild boar | 6 | Jiangxi, Zhejiang | 29.0 | 15 | 269 | -3 |
| **High-latitude (Northern China) pigs** | | | | | | |
| Erhualian | 5 | Wuxi, Jiangsu | 31.7 | 10 | 227 | -13 |
| Hetao | 6 | Wuyuan, Inner Mongolia | 40.8 | 1,200 | 120 | -30 |
| Laiwu | 6 | Laiwu, Shandong | 36.2 | 250 | 191 | -15 |
| Min | 6 | Lanxi, Heilongjiang | 46.3 | 240 | 135 | -50.1 |
| Tibetan (Gansu) | 4 | Hezuo, Gansu | 35.0 | 3,100 | 55 | -28.5 |
| Tibetan (Sichuan) | 6 | Litan, Sichuan | 30.0 | 4,000 | 50 | -30.6 |
| Tibetan (Tibet) | 6 | Gongbujiangda, Tibet | 29.6 | 3,600 | 130 | -16 |
| Tibetan (Yunnan) | 6 | Xianggelila, Yunnan | 27.8 | 3,300 | 113 | -27.4 |

[a] All data in this table were cited from Wang et al.(2011).

**Supplementary Table 2.** Comparison of the Wuzhishan and Duroc reference genomes[a]

| Chr[b] | Wuzhishan | | | | Duroc | | | |
|---|---|---|---|---|---|---|---|---|
| | Length (bp) | Ungapped length (bp) | Length (%) | Scaffold number | Length (bp) | Ungapped length (bp) | Length (%) | Scaffold number |
| 1 | 285,984,920 | 252,355,678 | 10.05 | 718 | 315,321,322 | 279,914,422 | 11.23 | 692 |
| 2 | 156,155,180 | 134,192,764 | 5.49 | 448 | 162,569,375 | 145,631,175 | 5.79 | 331 |
| 3 | 138,935,171 | 124,768,976 | 4.88 | 515 | 144,787,322 | 129,207,922 | 5.16 | 304 |
| 4 | 132,671,114 | 128,810,704 | 4.66 | 100 | 143,465,943 | 129,362,743 | 5.11 | 276 |
| 5 | 114,578,090 | 110,505,538 | 4.03 | 90 | 111,506,441 | 99,275,441 | 3.97 | 239 |
| 6 | 169,926,861 | 147,162,065 | 5.97 | 247 | 157,765,593 | 139,069,593 | 5.62 | 366 |
| 7 | 138,862,790 | 128,588,732 | 4.88 | 706 | 134,764,511 | 121,271,311 | 4.8 | 266 |
| 8 | 140,952,491 | 132,984,731 | 4.95 | 100 | 148,491,826 | 132,692,726 | 5.29 | 309 |
| 9 | 138,938,469 | 129,537,595 | 4.88 | 264 | 153,670,197 | 139,490,897 | 5.47 | 276 |
| 10 | 74,304,832 | 67,113,415 | 2.61 | 257 | 79,102,373 | 71,122,273 | 2.82 | 156 |
| 11 | 82,828,561 | 75,352,975 | 2.91 | 216 | 87,690,581 | 77,960,681 | 3.12 | 190 |
| 12 | 64,164,676 | 53,277,345 | 2.26 | 193 | 63,588,571 | 56,400,871 | 2.26 | 141 |
| 13 | 210,982,587 | 190,580,744 | 7.41 | 276 | 218,635,234 | 195,589,234 | 7.78 | 449 |
| 14 | 145,891,642 | 131,537,004 | 5.13 | 214 | 153,851,969 | 140,665,969 | 5.48 | 259 |
| 15 | 149,841,376 | 136,768,277 | 5.27 | 500 | 157,681,621 | 140,675,921 | 5.61 | 332 |
| 16 | 81,457,867 | 77,315,492 | 2.86 | 78 | 86,898,991 | 78,720,191 | 3.09 | 160 |
| 17 | 62,603,045 | 49,907,989 | 2.2 | 327 | 69,701,581 | 62,138,581 | 2.48 | 148 |
| 18 | 57,577,357 | 52,509,227 | 2.02 | 308 | 61,220,071 | 55,640,371 | 2.18 | 109 |
| X | 125,820,878 | 98,750,770 | 4.42 | 358 | 144,288,218 | 127,507,118 | 5.14 | 333 |
| Y | 638,869 | 359,753 | 0.02 | 36 | 1,637,716 | 1,333,916 | 0.06 | 7 |
| An | 2,473,116,776 | 2,222,379,774 | 86.92 | 5,951 | 2,596,639,456 | 2,323,671,356 | 92.46 | 5,343 |
| Nan | 372,305,695 | 364,257,811 | 13.08 | 1,132,340 | 211,869,922 | 195,490,322 | 7.54 | 4,562 |
| Total | 2,845,422,471 | 2,586,637,585 | 100 | 1,138,291 | 2,808,509,378 | 2,519,161,678 | 100 | 9,905 |

[a] The data of the Duroc reference genome (*Sscrofa* 10.2) are from Groenen et al. (2012)

[b] An: Anchored; Nan: Not Anchored.

**Supplementary Table 3**. Statistics of 1 Mb reads of one individual from each breed mapped to the Wuzhishan and Duroc reference genomes

| Breed | Sample | Mapped to WZSP [a] (%) | Mapped to Sscrofa10.2[b] (%) |
|---|---|---|---|
| Bamaxiang | BMX0001 | 97.50 | 89.25 |
| Erhualian | ER_CS0234 | 97.28 | 89.19 |
| Hetao | HTDE12 | 96.71 | 89.30 |
| Laiwu | LWH0F | 97.77 | 89.34 |
| Luchuan | LUC201 | 96.98 | 88.98 |
| Min | MZ-304-07 | 97.34 | 89.61 |
| Tibetan (Gansu) | hztR02 | 96.99 | 88.53 |
| Tibetan (Sichuan) | LTZ201 | 97.30 | 89.19 |
| Tibetan (Tibet) | B12 | 97.22 | 89.28 |
| Tibetan (Yunnan) | DQZ24 | 97.05 | 88.46 |
| Wuzhishan | WZS149 | 97.64 | 89.60 |
| Wild boar | NCYZ0010 | 97.56 | 89.28 |

[a] WZSP, the Wuzhishan reference genome.

[b] Sscrofa10.2, the Duroc reference genome.

**Supplementary Table 4**. Distribution of variants in the pig genome

| Variant | Overall | Intergenic | 10 kb Upstream | 5'UTR | Exon | Intron | 3'UTR | 10 kb Downstream |
|---|---|---|---|---|---|---|---|---|
| SNPs | 40,820,483 | 26,606,993 | 1,943,519 | 42,014 | 188,664 | 8,528,717 | 157,064 | 2,848,410 |
| Insertions | 2,927,933 | 1,887,820 | 146,114 | 2,837 | 1,846 | 627,706 | 13,220 | 210,711 |
| Deletions | 3,275,417 | 2,129,264 | 160,231 | 3,215 | 2,163 | 689,204 | 15,106 | 234,904 |
| SVs [a] | 44,170 | 31,765 | 3,026 | 109 | 484 | 12,211 | 260 | 3,037 |
| Total | 47,068,003 | 30,655,842 | 2,252,890 | 48,175 | 193,157 | 9,857,838 | 185,650 | 3,297,062 |

[a] SVs, Structural variants.

**Supplementary Table 5.** Statistics for putative loss-of-function variants

| Breed | Nonsense [a] | Frame shift | Splice [b] | SVs [c] | Overall | Fixed LOF [d] | Specific LOF [e] |
|---|---|---|---|---|---|---|---|
| Bamaxiang | 81 | 274 | 271 | 248 | 803 | 131 | 0 |
| Erhualian | 68 | 420 | 297 | 223 | 912 | 60 | 0 |
| Hetao | 65 | 248 | 264 | 211 | 738 | 96 | 0 |
| Laiwu | 66 | 258 | 263 | 252 | 779 | 153 | 1 |
| Luchuan | 63 | 211 | 231 | 245 | 702 | 84 | 1 |
| Min | 61 | 246 | 265 | 256 | 776 | 126 | 1 |
| Tibetan (Gansu) | 53 | 181 | 220 | 154 | 580 | 116 | 0 |
| Tibetan (Sichuan) | 81 | 284 | 292 | 237 | 813 | 125 | 0 |
| Tibetan (Tibet) | 89 | 271 | 285 | 257 | 834 | 88 | 0 |
| Tibetan (Yunnan) | 92 | 289 | 304 | 249 | 851 | 26 | 0 |
| Wuzhishan | 82 | 287 | 305 | 255 | 857 | 101 | 0 |
| Wild boar | 90 | 300 | 332 | 240 | 882 | 66 | 0 |
| Average | 74 | 272 | 277 | 236 | 794 | 98 | 0.25 |

[a] Premature stop mutations.

[b] Mutations within 2 bp flanking splice sites.

[c] Structural variants in the coding regions.

[d] Loss-of-function variants that are fixed in one breed.

[e] Loss-of-function variants that are fixed in one breed and rare (allele frequencies < 0.05) in all samples of the other breeds.

**Supplementary Table 6**. Statistics of structural variants overlapped with annotated transposon elements

| Breed | SV number | Number overlapped with TE [a] | % |
|---|---|---|---|
| Bamaxiang | 35,711 | 19,244 | 53.89 |
| Erhualian | 30,558 | 16,580 | 54.26 |
| Hetao | 36,343 | 19,641 | 54.04 |
| Luchuan | 34,717 | 18,653 | 53.73 |
| Laiwu | 38,101 | 20,455 | 53.69 |
| Min | 39,569 | 21,098 | 53.32 |
| Tibetan (Gansu) | 29,917 | 16,271 | 54.39 |
| Tibetan (Sichuan) | 37,554 | 20,305 | 54.07 |
| Tibetan (Tibet) | 40,916 | 21,834 | 53.36 |
| Tibetan (Yunnan) | 37,144 | 19,794 | 53.29 |
| Wuzhishan | 36,757 | 19,863 | 54.04 |
| Wild boar | 40,216 | 21,513 | 53.49 |
| Average | 36,459 | 19,604 | 53.80 |

[a] TE, transposon elements.

**Supplementary Table 7**. Summary of 42 publicly available pig genomes [a]

| Accession | Breed | Specie | Run Read Count | Run Base Count (Gb) | Coverage (X) |
|---|---|---|---|---|---|
| ERR173170 | Duroc (Europe) | *Sus scrofa* | 123,740,689 | 21 | 8.08 |
| ERR173171 | Duroc (Europe) | *Sus scrofa* | 132,568,899 | 23 | 8.85 |
| ERR173172 | Duroc (Europe) | *Sus scrofa* | 67,691,341 | 12 | 4.62 |
| ERR173173 | Duroc (Europe) | *Sus scrofa* | 84,342,308 | 15 | 5.77 |
| ERR173174 | Hampshire (Europe) | *Sus scrofa* | 130,177,332 | 23 | 8.85 |
| ERR173175 | Hampshire (Europe) | *Sus scrofa* | 114,097,075 | 20 | 7.69 |
| ERR173177 | Bearded pig | *Sus barbatus* | 84,799,356 | 14 | 5.38 |
| ERR173178 | Wild boar (Sumatra) | *Sus scrofa* | 127,336,765 | 22 | 8.46 |
| ERR173179 | Jiangquhai (China) | *Sus scrofa* | 124,020,627 | 21 | 8.08 |
| ERR173180 | Landrace (Europe) | *Sus scrofa* | 106,682,600 | 19 | 7.31 |
| ERR173181 | Landrace (Europe) | *Sus scrofa* | 154,718,233 | 28 | 10.77 |
| ERR173182 | Landrace (Europe) | *Sus scrofa* | 104,998,636 | 18 | 6.92 |
| ERR173183 | Landrace (Europe) | *Sus scrofa* | 83,521,719 | 15 | 5.77 |
| ERR173184 | Landrace(Europe) | *Sus scrofa* | 85,400,325 | 15 | 5.77 |
| ERR173186 | Large White (Europe) | *Sus scrofa* | 115,240,716 | 20 | 7.69 |
| ERR173188 | Large White(Europe) | *Sus scrofa* | 113,214,857 | 20 | 7.69 |
| ERR173190 | Large White (Europe) | *Sus scrofa* | 132,514,970 | 23 | 8.85 |
| ERR173192 | Large White (Europe) | *Sus scrofa* | 123,734,187 | 20 | 7.69 |
| ERR173193 | Large White (Europe) | *Sus scrofa* | 110,757,743 | 20 | 7.69 |
| ERR173196 | Large White (Europe) | *Sus scrofa* | 100,835,761 | 19 | 7.31 |
| ERR173199 | Meishan (China) | *Sus scrofa* | 105,491,123 | 18 | 6.92 |
| ERR173200 | Meishan (China) | *Sus scrofa* | 105,393,163 | 18 | 6.92 |
| ERR173201 | Meishan (China) | *Sus scrofa* | 101,672,777 | 17 | 6.54 |
| ERR173202 | Meishan (China) | *Sus scrofa* | 117,087,024 | 20 | 7.69 |
| ERR173204 | Meishan (China) | *Sus scrofa* | 118,752,799 | 21 | 8.08 |

| | | | | | |
|---|---|---|---|---|---|
| ERR173205 | Meishan (China) | *Sus scrofa* | 122,082,579 | 21 | 8.08 |
| ERR173206 | Pietrain (Europe) | *Sus scrofa* | 100,897,017 | 17 | 6.54 |
| ERR173207 | Pietrain (Europe) | *Sus scrofa* | 65,792,012 | 11 | 4.23 |
| ERR173208 | Pietrain (Europe) | *Sus scrofa* | 129,619,503 | 22 | 8.46 |
| ERR173212 | Wild Boar (Japan) | *Sus scrofa* | 130,215,795 | 22 | 8.46 |
| ERR173213 | Wild Boar (Netherlands Veluwe) | *Sus scrofa* | 101,703,565 | 18 | 6.92 |
| ERR173214 | Wild Boar (Netherlands Veluwe) | *Sus scrofa* | 132,693,822 | 23 | 8.85 |
| ERR173215 | Wild Boar (Netherlands Meinweg) | *Sus scrofa* | 61,858,289 | 10 | 3.85 |
| ERR173216 | Wild Boar (Netherlands Meinweg) | *Sus scrofa* | 88,850,546 | 16 | 6.15 |
| ERR173217 | Wild Boar (France) | *Sus scrofa* | 104,442,331 | 19 | 7.31 |
| ERR173218 | Wild Boar (Switzerland) | *Sus scrofa* | 162,047,007 | 29 | 11.15 |
| ERR173219 | Wild Boar (South China) | *Sus scrofa* | 61,520,753 | 10 | 3.85 |
| ERR173220 | Wild Boar (South China) | *Sus scrofa* | 117,783,048 | 20 | 7.69 |
| ERR173221 | Wild Boar (North China) | *Sus scrofa* | 59,165,306 | 9 | 3.46 |
| ERR173222 | Wild Boar (North China) | *Sus scrofa* | 117,185,548 | 20 | 7.69 |
| ERR173223 | Xiang (China) | *Sus scrofa* | 97,603,648 | 18 | 6.92 |
| ERR173224 | Xiang (China) | *Sus scrofa* | 96,373,669 | 18 | 6.92 |

[a] Genomic data of the 42 individuals are from Groenen et al. (2012). Note that the Duroc and Hampshire breeds are North American breeds of European origin.

**Supplementary Table 8**. Population genetic statistics for Chinese pigs in this study

| Breed | π | θ | Derived allele frequency [a] | Observed | LD extent (kb) [b] |
|---|---|---|---|---|---|
| Bamaxiang | 0.00430 | 0.00358 | 0.246 (7,560,499) | 0.342 | 90 |
| Erhualian | 0.00384 | 0.00331 | 0.294 (5,904,002) | 0.308 | 70 |
| Hetao | 0.00436 | 0.00359 | 0.252 (7,777,322) | 0.302 | 92 |
| Laiwu | 0.00379 | 0.00308 | 0.297 (6,538,969) | 0.369 | 250 |
| Luchuan | 0.00357 | 0.00292 | 0.308 (5,864,776) | 0.348 | 200 |
| Min | 0.00392 | 0.00313 | 0.309 (6,797,267) | 0.340 | 260 |
| Tibetan (Gansu) | 0.00404 | 0.00365 | 0.301 (6,042,185) | 0.310 | 300 |
| Tibetan (Sichuan) | 0.00448 | 0.00368 | 0.238 (7,593,022) | 0.324 | 60 |
| Tibetan (Tibet) | 0.00491 | 0.00410 | 0.211 (9,180,986) | 0.315 | 50 |
| Tibetan (Yunnan) | 0.00467 | 0.00390 | 0.219 (8,321,676) | 0.295 | 52 |
| Wuzhishan | 0.00481 | 0.00404 | 0.210 (8,907,522) | 0.332 | 50 |
| Wild boar | 0.00506 | 0.00423 | 0.083 (5,971,862) | 0.328 | 6 |

[a] The number of derived SNPs in each breed is given in brackets.

[b] Linkage disequilibrium (LD) extent was determined from LD decay plot at level 0.3.

**Supplementary Table 9**. Pair-wise $F_{ST}$ values among Chinese pig breeds[a]

| | BMX | EHL | HT | LWH | LUC | MIN | GST | SCT | TT | YNT | WZS | WB |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **BMX** | 0 | | | | | | | | | | | |
| **EHL** | 0.222 | 0 | | | | | | | | | | |
| **HT** | 0.190 | 0.149 | 0 | | | | | | | | | |
| **LWH** | 0.240 | 0.207 | 0.173 | 0 | | | | | | | | |
| **LUC** | 0.159 | 0.292 | 0.255 | 0.305 | 0 | | | | | | | |
| **MIN** | 0.241 | 0.225 | 0.166 | 0.207 | 0.306 | 0 | | | | | | |
| **GST** | 0.204 | 0.196 | 0.157 | 0.216 | 0.271 | 0.222 | 0 | | | | | |
| **SCT** | 0.164 | 0.172 | 0.145 | 0.200 | 0.227 | 0.208 | 0.134 | 0 | | | | |
| **TT** | 0.144 | 0.159 | 0.122 | 0.175 | 0.207 | 0.174 | 0.124 | 0.083 | 0 | | | |
| **YNT** | 0.146 | 0.163 | 0.134 | 0.189 | 0.210 | 0.195 | 0.129 | 0.078 | 0.075 | 0 | | |
| **WZS** | 0.074 | 0.175 | 0.144 | 0.192 | 0.116 | 0.194 | 0.161 | 0.124 | 0.103 | 0.107 | 0 | |
| **WB** | 0.142 | 0.16 | 0.135 | 0.183 | 0.200 | 0.183 | 0.148 | 0.113 | 0.096 | 0.101 | 0.101 | 0 |

[a] BMX, Bamaxiang; EHL, Erhualian; HT, Hetao; LWH, Laiwu; LUC, Luchuan; MIN, Min; GST, Tibetan (Gansu); SCT, Tibetan (Sichuan); TT, Tibetan (Tibet); YNT, Tibetan (Yunnan); WZS, Wuzhishan; WB, Wild boar from southern China.

**Supplementary Table 10.** Statistics of introgression between Chinese and European domestic pig genomes

| Breed [a] | Introgression length (bp) [b] | Introgression ratio(%) [c] | Gene number [d] |
|---|---|---|---|
| Chinese breed | | | |
| Wuzhishan | 61,100,000 | 2.47 | 375 |
| Luchuan | 29,250,000 | 1.18 | 162 |
| Bamaxiang | 40,900,000 | 1.65 | 230 |
| Erhualian | 41,000,000 | 1.66 | 429 |
| Jiangquhai | 49,700,000 | 2.01 | 334 |
| Meishan | 76,700,000 | 3.10 | 571 |
| Tibetan (Yunnan) | 22,700,000 | 0.92 | 144 |
| Tibetan (Sichuan) | 19,650,000 | 0.79 | 102 |
| Tibetan (Tibet) | 235,100,000 | 9.51 | 1,669 |
| Tibetan (Gansu) | 40,500,000 | 1.64 | 221 |
| Laiwu | 476,050,000 | 19.25 | 3,974 |
| Hetao | 299,450,000 | 12.11 | 2,425 |
| Min | 873,000,000 | 35.30 | 6,694 |
| European breed | | | |
| Duroc | 325,850,000 | 13.18 | 2,662 |
| Hampshire | 188,250,000 | 7.61 | 1,681 |
| Landrace | 418,150,000 | 16.91 | 3,492 |
| Large white | 456,500,000 | 18.46 | 3,676 |
| Pietrain | 322,300,000 | 13.03 | 2,629 |

[a] Whole genome data of Jiangquhai, Meishan and European breeds were from Groenen *et al*. (2012).

[b] The total length of the introgression genomic regions.

[c] The proportion of the introgression region in the genome.

[d] The total number of annotated genes in the introgression region

Nature Genetics: doi:10.1038/ng.3199

**Supplementary Table 11 see excel file “SupTable11.xls”.**

**Supplementary Table 12 see excel file “SupTable12.xls”.**

**Supplementary Table 13 see excel file “SupTable13.xls”.**

Nature Genetics: doi:10.1038/ng.3199

**Supplementary Table 14.** Protein-altering SNPs outliers in 10 of 219 candidate genes under selection for high/low-latitude environments in Chinese pigs

| Chr [a] | Position (bp) [b] | Mutation | Altered amino acid | Gene | LSBL Value |
|---|---|---|---|---|---|
| 1 | 232,120,787 | G/T | val/gly | *VPS13A* | 0.779 |
| 1 | 232,146,604 | T/A | phe/tyr | *VPS13A* | 0.728 |
| 1 | 250,274,295 | A/C | asn/ser | *ABCA1* | 0.787 |
| 1 | 250,174,231 | C/T | val/ala | *OR13D1* | 0.619 |
| 2 | 69,219,426 | A/G | met/val | *CREB3L3* | 0.646 |
| 2 | 69,536,280 | T/C | val/ala | *ZFR2* | 0.649 |
| 4 | 90,206,165 | A/G | his/arg | *CD84* | 0.801 |
| 4 | 90,221,189 | A/G | met/ile | *CD84* | 0.627 |
| 8 | 75,075,718 | G/T | ala/ser | *DCHS2* | 0.771 |
| 8 | 75,076,810 | C/T | his/tyr | *DCHS2* | 0.686 |
| 13 | 66,872,205 | G/T | ala/ser | *OGG1* | 0.721 |
| 14 | 125,210,642 | G/A | val/ile | *C10ORF81* | 0.647 |
| 14 | 109,993,950 | C/T | thr/ile | *LSM3* | 0.857 |
| 14 | 109,994,013 | C/T | ala/val | *LSM3* | 0.708 |
| 14 | 109,993,934 | C/A | asp/glu | *LSM3* | 0.678 |

[a] Chromosome.

[b] Positions refer to the Wuzhishan pig reference genome.

**Supplementary Table 15.** The genomic difference between the 48 Mb recombination-decreasing region and the outside region on chromosome X [a]

| | GC Content | | | Repeat | | | Gene counts | | |
|---|---|---|---|---|---|---|---|---|---|
| **Reference genome** | 48 Mb | Outside | *P*-value | 48 Mb | Outside | *P*-value | 48 Mb | Outside | *P*-value |
| Wuzhishan | 0.36 | 0.40 | 2.00E-77 | 0.50 | 0.42 | 4.02E-32 | 4.6 | 5.7 | 0.016 |
| Duroc | 0.37 | 0.41 | 1.26E-72 | 0.65 | 0.54 | 5.59E-68 | 8.2 | 11.5 | 8.52E-09 |

[a] 48 Mb, the 48 Mb region of low recombination; Outside, the outside region of the recombination cold spot on chromosome X. The *P* values were calculated by Mann Whitney U test using R function Wilcox.test.

**Supplementary Table 16.** The spearman correlation between recombination rate and the count of 6K poly-T repeat sequence on autosomes [a]

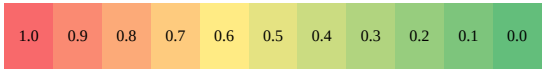
| **Window size** | **Spearman ($r^2$)** | ***P* value** |
|---|---|---|
| 50 kb | -0.039 | 3.07E-17 |
| 500 kb | -0.128 | 1.36E-18 |

[a] The 6 Kb poly-T core sequence corresponds to 45,743141 – 45,749,642 bp in the Wuzhishan reference genome.

**Supplementary Table 17.** The allele frequency of 7 diagnostic chip SNPs within the 48 Mb region in Chinese and European pigs [a]

| | | | INRA0056740(C) | INRA0056751(C) | ALGA0099769(A) | INRA0056771(C) | INRA0056813(G) | INRA0056883(G) | INRA0056920(G) |
|---|---|---|---|---|---|---|---|---|---|
| Population | Breed | No. [b] | 51,916,470 bp [c] | 62,454,725 bp | 62,650,052 bp | 66,093,757 bp | 75,594,444 bp | 97,152,327 bp | 100,478,965 bp |
| European | Duroc | 52 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| pig | White Duroc | 8 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | Large white | 48 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | Landrace | 51 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Northern | Kele | 16 | 0.00 | 0.00 | 0.00 | 0.75 | 0.75 | 0.75 | 0.75 |
| Chinese pig | Min | 28 | 0.00 | 0.00 | 0.00 | 0.79 | 0.79 | 0.79 | 0.79 |
| | Jinhua | 18 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | Erhualian | 62 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | Laiwu | 19 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | Hetao | 25 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | Bamei | 21 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | Neijiang | 22 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | Rongchang | 34 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | Tibetan (Gansu) | 31 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | Tibetan (Tibet 1) | 38 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | Tibetan (Tibet 2) | 17 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Tibetan (Sichuan) | 17 | 0.06 | 0.06 | 0.06 | 1.00 | 1.00 | 1.00 | 1.00 |
| | Tibetan (Yunnan) | 19 | 0.16 | 0.16 | 0.16 | 1.00 | 1.00 | 1.00 | 1.00 |
| Central | Tongcheng | 31 | 0.52 | 0.52 | 0.52 | 1.00 | 1.00 | 1.00 | 1.00 |
| Chinese pig | Shaziliang | 15 | 0.80 | 0.80 | 0.80 | 1.00 | 1.00 | 1.00 | 1.00 |
| Southern | Dahuabai | 20 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Chinese pig | Dongshan | 16 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | Congjiangxiang | 32 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | Diannanxiaoer | 16 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | Luchuan | 36 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | Bamaxiang | 31 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | Wuzhishan | 27 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | Wild boar | 31 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |




[a] The 7 diagnostic SNPs in the 48 Mb region on chromosome X show extremely allele imbalance between European pigs and Southern Chinese pigs. These SNPs were selected from the illumina porcine 60K chips. Kele and Min pigs show signatures of admixture with European pigs. Breeds from central China including Shaziling and Tongcheng are segregating for both southern and northern haplotypes.

[b] The total number of X chromosomes in each population.

[c] The position of each SNP on chromosome X in the *Sscrofa* build 10.2 genome assembly.

Nature Genetics: doi:10.1038/ng.3199

**Supplementary Table 18.** Number of segregating sites within and between northern and southern Chinese pigs [a]

| | Northern Chinese Pigs | | | Southern Chinese Pigs | | |
|---|---|---|---|---|---|---|
| | 14M [b] | 34M [c] | 30M [d] | 14M [b] | 34M [c] | 30M [d] |
| **Northern Chinese Pigs** | 14,318 | 13,295 | 175,116 | | | |
| **Southern Chinese Pigs** | 66,418 | 19,599 | 193,206 | 6,330 | 16,509 | 165,488 |

[a] Southern Chinese pigs included wild boar and domestic pigs from southern China. Northern China pigs included northern Chinese domestic pigs and Tibetan pigs, but excluded wild boars as their genetic similarity to European pigs in the target regions.

[b] 14M, the X-linked sweep of 14 Mb.

[c] 34M, the 34 Mb low-recombination region flanking the 14 Mb sweep on chromosome X.

[d] 30M, the 30 Mb region (14 - 44 Mb) flanking the 14 Mb sweep on chromosome X.

**Supplementary Table 19.** Coalescent simulations showing that gene drift is not likely a cause of the sweep signal on chromosome X

| | 14M [b] | | | 34M [c] | | | 30M [d] | | |
|---|---|---|---|---|---|---|---|---|---|
| Group [a] | Prob (%) | S_ob | D_ob | Prob (%) | S_ob | D_ob | Prob (%) | S_ob | D_ob |
| Coalescent simulation 1 | | | | | | | | | |
| SCP | 0 | 6,330 | -0.788 | 0 | 16,509 | -0.159 | 85.244 | 165,488 | 0.506 |
| NCP | 0 | 14,318 | -1.157 | 0 | 13,295 | -0.798 | 14.219 | 175,116 | 0.443 |
| Coalescent simulation 2 | | | | | | | | | |
| SCP | 2.469 | 6,330 | -0.788 | 0 | 16,509 | -0.159 | 85.22 | 165,488 | 0.506 |
| NCP | 0.206 | 14,318 | -1.157 | 0 | 13,295 | -0.798 | 38.75 | 175,116 | 0.443 |

[a] Coalescent simulation 1, Coalescent simulation with no recombination; Coalescent simulation 2, Coalescent simulation with no recombination and 2-fold reduced mutation rate. SCP: Southern Chinese domestic pigs and wild boars; NCP: Northern Chinese domestic pigs.

[b] 14M, the X-linked sweep of 14 Mb. Prob (%), the probability of both the simulated number of segregating sites and simulated Tajima's D values less than the observed values; S_ob, the observed number of segregating site in the target region; D_ob, the observed Tajima's D value in the target regions.

[c] 34M, the 34 Mb low-recombination region flanking the 14 Mb sweep on chromosome X.

[d] 30M, the 30 Mb region (14 - 44 Mb) flanking the 14 Mb sweep on chromosome X.

**Supplementary Table 20.** Z-test showing gene drift is not likely a cause of the sweep signal on chromosome X [a].

| | 14M [b] | | | | 34M [c] | | | | 30M [d] | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Group** | SS_ob [e] | Prob_SS(%) [f] | D_ob [g] | Prob_D(%) [h] | S_ob | Prob_SS(%) | D_ob | Prob_D(%) | S_ob | Prob_SS(%) | D_ob | Prob_D (%) |
| **SCP** | 33.75 | 0.75 | -0.258 | 1.41 | 30.59 | 0.60 | 0.420 | 6.34 | 267.25 | 24.92 | 1.694 | 68.70 |
| **NCP** | 77.24 | 2.38 | -0.542 | 0.73 | 25.46 | 0.36 | -0.501 | 0.78 | 301.44 | 26.51 | 0.943 | 13.66 |

[a] Z-test in the present table means testing the probabilities of observed segregating sites and Tajima's D values in the target regions against the distributions on whole autosomes

[b] 14M, the X-linked sweep of 14 Mb. In the table the interior no-recombination region from 46.4 – 56 M represents the 14 Mb region.

[c] 34M, the 34 Mb region of low recombination flanking the 14 Mb sweep on chromosome X.

[d] 30M, the 30 Mb region (14 - 44 Mb) of normal recombination flanking the 14 Mb sweep on chromosome X.

[e] SS_ob: the average segregating site observed in 50K windows in the target region.

[f] Prob_SS(%): the probability of the observed average segregating sites in 50K windows in the target region compared with the distribution of segregating sites on autosomes.

[g] D_ob: the average Tajima's D value observed in 50K windows in the target region.

[h] Prob_D(%): the probability of the observed average Tajima's D values in 50k windows in the target region compared with the distribution of Tajima's D values on autosomes.

Nature Genetics: doi:10.1038/ng.3199

**Supplementary Table 21**. The time to most recent common ancestor ($T_{MRCA}$) within and outside the 48 Mb low-recombination region on chromosome X for southern and northern Chinese pigs[a]

| $T_{MRCA}$ (Mya) | 14 M[b] | 34 M[c] | 30 M[d] | Chr2 [e] |
|---|---|---|---|---|
| Northern Chinese pigs | 0.13 | 0.18 | 1.43 | 0.99 |
| Southern Chinese pigs | 0.11 | 0.23 | 1.63 | 0.89 |

[a] Chinese wild boars were excluded in this analysis

[b] $T_{MRCA}$ of the X-linked sweep of 14 Mb.

[c] $T_{MRCA}$ of the 34 Mb region of low recombination flanking the 14 Mb sweep on chromosome X.

[d] $T_{MRCA}$ of the 30 Mb region (14 - 44 Mb) of normal recombination flanking the 14 Mb sweep on chromosome X.

[e] $T_{MRCA}$ of an equivalent autosomal region (44 - 57.8 Mb) on chromosome 2.

**Supplementary Table 22.** The divergence time within and outside the 48 Mb low-recombination region on chromosome X among southern Chinese, northern Chinese and European pigs [a]

| Dviergent time (Mya) | 14 M[b] | 34 M[c] | 30 M[d] | Chr2 [e] |
|---|---|---|---|---|
| Northern Chinese pigs vs Southern Chinese pigs | 8.48 | 0.04 | 0.24 | 0.24 |
| Northern Chinese pigs vs European pigs | 0.29 | 7.91 | 1.28 | 1.17 |
| Southern Chinese pigs vs European pigs | 8.42 | 7.87 | 1.40 | 1.61 |
| Sumatran pigs vs Southern Chinese pigs | 1.03 | 0.92 | 2.01 | 2.33 |
| Bearded pigs vs Southern Chinese pigs | 3.74 | 3.42 | 5.51 | 6.97 |

[a] Southern Chinese pigs included wild boar and domestic pigs from southern China. Northern China pigs included northern Chinese domestic pigs and Tibetan pigs, but excluded wild boars as their genetic similarity to European pigs in the target regions. The divergence time was calibrated by the split time of 9.9 Mya between *Sus scrofa* populations and Africa Warthog as reported recently (Frantz *et al.* 2013). For two non-introgression regions each on chromosome X and 2, the divergence time between groups is in agreement with the known evolution history of *Sus scrofa*. For example, the divergence time between northern Chinese pigs and European pigs is ~1.2 Mya, and that between Sumatran wild boars and southern Chinese pigs is ~2.1 Mya, which are consistent with the recently reported values between these two groups (Frantz *et al*. 2013).

[b] The X-linked 14 Mb sweep region.

[c] The 34 Mb low-recombination region flanking the 14 Mb sweep on chromosome X.

[d] The 30 Mb region (14 - 44 Mb) flanking the 14 Mb sweep on chromosome X.

[e] An equivalent autosomal region (44-57.8 Mb) on chromosome 2.