ARTICLE

Check for updates

A combined GWAS approach reveals key loci for socially-affected traits in Yorkshire pigs

Pingxian Wu [1,4], Kai Wang[1,4], Jie Zhou[1], Dejuan Chen[1], Anan Jiang[1], Yanzhi Jiang[2], Li Zhu[1], Xiaotian Qiu[3], Xuewei Li[1] & Guoqing Tang [1✉]

Socially affected traits in pigs are controlled by direct genetic effects and social genetic effects, which can make elucidation of their genetic architecture challenging. We evaluated the genetic basis of direct genetic effects and social genetic effects by combining single-locus and haplotype-based GWAS on imputed whole-genome sequences. Nineteen SNPs and 25 haplotype loci are identified for direct genetic effects on four traits: average daily feed intake, average daily gain, days to 100 kg and time in feeder per day. Nineteen SNPs and 11 haplotype loci are identified for social genetic effects on average daily feed intake, average daily gain, days to 100 kg and feeding speed. Two significant SNPs from single-locus GWAS (SSC6:18,635,874 and SSC6:18,635,895) are shared by a significant haplotype locus with haplotype alleles 'GGG' for both direct genetic effects and social genetic effects in average daily feed intake. A candidate gene, *MT3*, which is involved in growth, nervous, and immune processes, is identified. We demonstrate the genetic differences between direct genetic effects and social genetic effects and provide an anchor for investigating the genetic architecture underlying direct genetic effects and social genetic effects on socially affected traits in pigs.

[1] Farm Animal Genetic Resources Exploration and Innovation Key Laboratory of Sichuan Province, Sichuan Agricultural University, Chengdu, Sichuan, China. [2] College of Life Science, Sichuan Agricultural University, Yaan, Sichuan, China. [3] National Animal Husbandry Service, Beijing, Beijing, China. [4]These authors contributed equally: Pingxian Wu, Kai Wang. ✉email: tyq003@163.com

Social interactions inevitably occur among groups of individuals and these interactions influence the genetic variation of a trait in livestock[1]. In current pig production systems, pigs are penned together into contemporary groups, social interactions, and observed phenotypes genetically contribute to socially affected traits of individuals and their group mates[2]. Socially affected traits are influenced both by the genes of the individual itself (direct genetic effects, DGE) and by the genotype of other individuals in the same group (social genetic effects, SGE, also called indirect genetic effects)[2–5]. Previous research has proven that many phenotypes are affected by both DGE and SGE and the contribution of SGE to total genetic variance is large in pigs, such as growth rate, body weight (BW)[6,7], feed intake, backfat thickness (BFT), muscle depth[2], behavior[8,9], and average daily gain (ADG)[10,11]. The existence of SGE affect the estimation of DGE and therefore result in unfavorable consequence on genetic improvement of complex traits. Until recently, the genetic architecture of socially affected traits that involve both DGE and SGE is still poorly understood in pigs.

With the availability of genomic information, GWAS have increased the understanding of the contributions of genetic variants toward various complex traits. However, previous studies have mainly focused on DGE and have not considered the contributions of SGE. For many phenotypes, DGE and SGE can be directly quantified using a social genetic model[2]. Based on the social genetic model, the classical (direct) genetic model is expanded with SGE[12]. This model simultaneously contains both DGE and SGE that enables researchers to simultaneously estimate DGE and SGE values.

Detecting the genetic architecture of DGE and SGE may provide deeper insight into a better understanding of complex traits. To date, only a few studies have identified some SNPs associated with SGE. Using the Illumina 60 K SNP BeadChip, an important SNP was identified to be associated with DGE and SGE in chickens[13]. In pigs, five SNPs on chromosome 6 associated with social ADG were detected with an Illumina panel[14]. Additionally, in our previous study, a total of 27 SNPs associated with six growth traits were identified using whole-genome sequences (WGS) of 40 Yorkshire pigs (Zhejiang Tianpeng Group Co., Ltd. Zhejiang, China)[15]. However, because of the limited sample size, the study was not efficient in detecting credible associations with DGE and SGE.

Assessing the impact of SGE on socially affected traits is challenging. Haplotype-based GWAS can capture those genetic variants that are not detected by single-locus GWAS[16]. Therefore, to better investigate the genetic architecture of socially affected traits in pigs, the present study used both GWAS approaches to estimate the DGE and SGE on each trait by employing a social genetic model, WGS of 60 pigs, and genotyping of 1204 Yorkshire pigs with Illumina 50 K SNP chips. Therefore, the main objectives of this study were (i) to construct haplotype loci based on imputed WGS data and identify associations between haplotype-DGE and haplotype-SGE by GWAS; (ii) to perform single-locus analysis to explore SNPs associated with socially affected traits by considering both DGE and SGE in Yorkshire pigs; (iii) to compare the results of GWAS based on haplotypes and SNPs for DGE and SGE; and (iv) to combine single-locus and haplotype-based GWAS to reveal important SNPs and genes for DGE and SGE in pigs.

Results

Phenotype statistics. Summary statistics for all phenotypic data from 1204 Yorkshire pigs were shown in Table 1. Based on the social genetic model, the estimated accuracies and deregressed EBVs of DGE and SGE for each trait were presented in Table 2. All the DGE and SGE values were used for further analysis.

Table 1 Descriptive statistics of phenotypes for eight socially-affected traits in Yorkshire pigs.

Trait	Unit	Number	Mean ± SD	Max	Min
ADFI	kg	1112	1.87 ± 0.34	2.84	0.53
ADG	kg	1112	0.74 ± 0.14	1.26	0.15
B100	mm	1026	9.27 ± 2.34	19.98	4.23
D100	day	1112	183.61 ± 22.37	391.93	100
FCR	kg/kg	1112	2.53 ± 0.31	3.74	1.51
RFI	g	1112	7.89 ± 216.07	905.05	−923.5
TPD	min/day	1107	64.51 ± 15.91	123.24	28.09
FS	g/min	1107	30.79 ± 8.01	57.76	13.51

ADFI average daily feed intake, *ADG* average daily gain, *B100* backfat thickness to 100 kg, *D100* days to 100 kg, *FCR* feed conversion ratio, *RFI* residual feed intake, *TPD* time in feeder per day, *FS* feeding speed, *Number* number of phenotypic records, *Mean* arithmetic mean, *SD* standard deviation, *Max* maximum, *Min* minimum.

Table 2 Summary of the estimated accuracies and deregressed EBVs for direct genetic (DGE) and effects social genetic effects (SGE) in Yorkshire pigs.

| Trait		Number	Estimated accuracies (Mean ± SD)	Deregressed EBVs (Mean ± SD)
DGE	ADFI	1112	0.80 ± 0.17	1.77E-02 ± 0.22
	ADG	1112	0.80 ± 0.17	5.40E-04 ± 0.18
	B100	1026	0.40 ± 0.22	6.00E-02 ± 1.38
	D100	1112	0.55 ± 0.22	7.85E-02 ± 19.49
	FCR	1112	0.30 ± 0.21	7.64E-03 ± 0.24
	RFI	1112	0.44 ± 0.24	14.43 ± 194.37
	TPD	1107	0.55 ± 0.22	−6.09E-03 ± 8.07
	FS	1107	0.44 ± 0.23	0.11 ± 3.87
SGE	ADFI	1112	0.68 ± 0.19	2.99E-05 ± 0.07
	ADG	1112	0.60 ± 0.21	3.95E-05 ± 0.07
	B100	1026	0.40 ± 0.22	−1.37E-03 ± 0.62
	D100	1112	0.23 ± 0.18	−8.67E-03 ± 7.99
	FCR	1112	0.30 ± 0.21	8.16E-05 ± 0.16
	RFI	1112	0.12 ± 0.14	0.08 ± 119.57
	TPD	1107	0.51 ± 0.22	0.01 ± 2.55
	FS	1107	0.44 ± 0.23	−1.02E-03 ± 1.63

ADFI average daily feed intake, *ADG* average daily gain, *B100* backfat thickness to 100 kg, *D100* days to 100 kg, *FCR* feed conversion ratio, *RFI* residual feed intake, *TPD* time in feeder per day, *FS* feeding speed, *Number* number of phenotypic records, *Mean* arithmetic mean, *SD* standard deviation.

Evaluation of sequencing data and genotype imputation. A total of 60 pigs were sequenced and the statistic results were list in Supplementary Data 1. After initial quality filtering, 3.0-TB clean data with the average mapped read depth of 17.11 was retained. The mean reads were 336,176,275 for each sample. After mapping to the pig reference genome, the mean mapped reads and the mean uniquely mapped reads were 312,374,099 and 293,904,451 for each sample, respectively. The average mapping and unmapping ratios were 92.92 and 1.28% per individual, respectively.

This study conducted genotype imputation from a 50 K SNP chip to WGS data in Yorkshire pigs using a two-breed reference population (20 Landrace and 40 Yorkshire pigs). In total, 36,969 and 14,539,997 SNPs were identified in the target and reference populations, respectively. Beagle 5.1 software was used to impute the WGS reference panel across the 1204 genotyped pigs. After genotype imputation, a total of 14,539,997 SNPs with an average accuracy of 0.63 were available for this study. For each chromosome, the average imputation accuracy was in the range
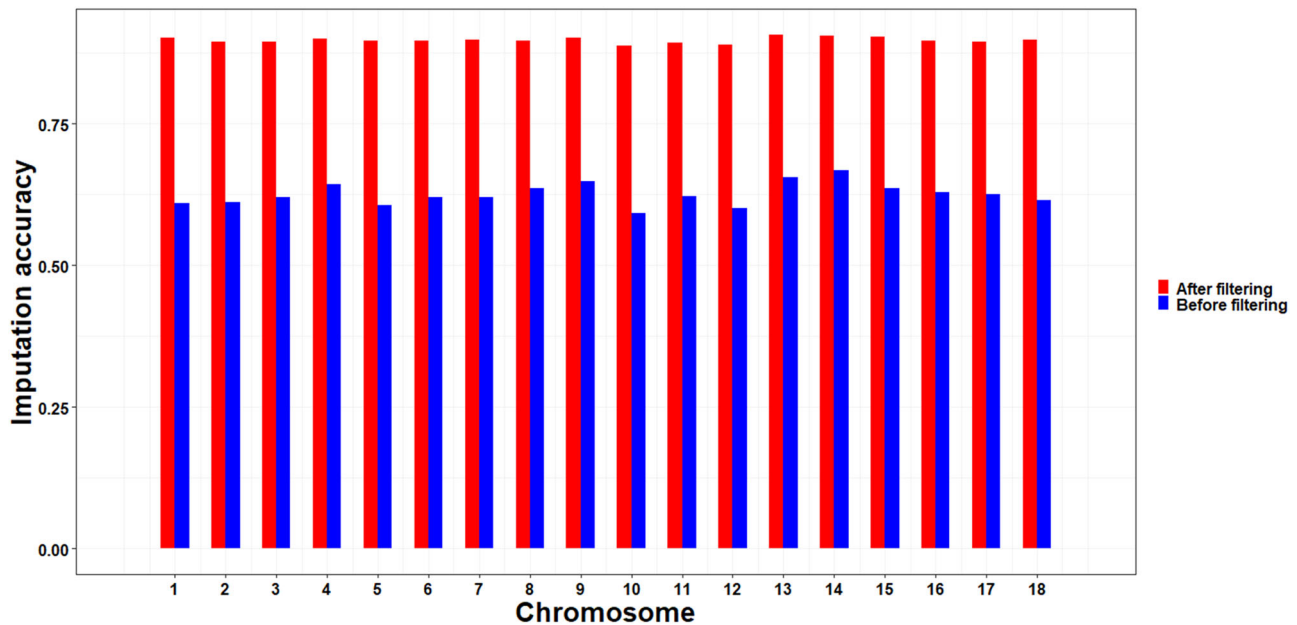
Fig. 1 Imputation accuracy for each chromosome in Yorkshire pigs. Accuracy of imputation based on Beagle R^2 before and after filtering (Beagle R^2 < 0.8). "Before filtering" for the blue part, "After filtering" for the red part.
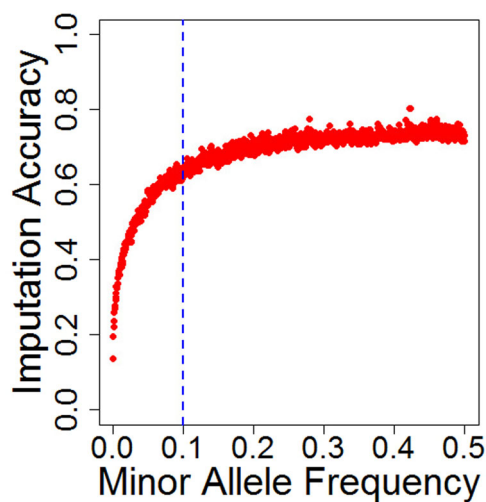
Fig. 2 Imputation accuracy against minor allele frequency (MAF). SNPs were divided into bins of SNPs with common MAF.

of 0.59–0.67 (Fig. 1). The SSC14 had the highest imputation accuracy (0.67), while SSC10 had the lowest imputation accuracy (0.59). After quality control, a total of 3,072,572 SNPs with an average imputation accuracy of 0.90 were retained for further analyses.

The average imputation accuracies of imputed SNPs with different minor allele frequency (MAF) were calculated and are shown in Fig. 2. The imputation accuracies were comparatively stable for different MAFs where the MAF was higher than 0.1. However, the imputation accuracy decreased sharply when the MAF was lower than 0.1.

Single-locus GWAS for eight socially affected traits

For average daily feed intake (ADFI). The imputed-GWAS results for DGE in ADFI are shown in Table 3, Supplementary Data 2, and Fig. 3a. The imputed-GWAS identified 413 SNPs

($P < 3.25 \times 10^{-7}$) associated with DGE at the suggestive threshold. Of these 413 SNPs, the 17 SNPs ($P < 1.63 \times 10^{-8}$) located on SSC6 and SSC8 showed a significant signal. The most significant SNP ($P = 1.07 \times 10^{-10}$) was located on SSC6: 46,448,592. In a narrow 9.75 kb region (18.64–18.65 Mb) on chromosome 6, there were ten consecutive genome-wide significant SNPs that showed significant peaks. In the region of SSC8: 79.51–79.53 Mb, five genome-wide significant SNPs showed high signals. A λ of less than 1.05 is considered to indicate a lack of population stratification[17]. The λ calculated for the DGE on ADFI was 1.01; implying no population stratification (Supplementary Fig. 1a). For chip-based GWAS, the significant threshold was set at $P = 2.70 \times 10^{-5}$. On the basis of the 50-K chip data, a total of 97 SNPs were identified (Supplementary Fig. 2). Three SNPs from the chip-based GWAS were also in the imputed GWAS ($P < 2.70 \times 10^{-5}$, Supplementary Table 1).

For SGE on ADFI, a total of 26 SNPs reached the suggestive threshold using imputed GWAS (Table 4, Supplementary Data 2, and Fig. 3b). Among these, 14 genome-wide significant SNPs were detected on chromosome 6, which had the highest signals (Fig. 3b). The λ was equal to 1.00, which indicates no population stratification (Supplementary Fig. 1b). A total of 89 SNPs were found from the chip-based GWAS (Supplementary Fig. 2). Among them, two overlapped with the imputed GWAS ($P < 2.70 \times 10^{-5}$, Supplementary Table 1).

A comparison between the imputed GWAS of DGE and SGE on ADFI revealed a total of 17 common SNPs shared between DGE and SGE (Fig. 4). Notably, the use of SGE identified associations for ADFI: two SNPs (SSC6: 25,649,764, $P = 1.67 \times 10^{-9}$ and SSC6: 25,651,261, $P = 2.68 \times 10^{-9}$) were significant for SGE but were not significant for DGE ($P > 1.00 \times 10^{-7}$).

For ADG. On the basis of imputed GWAS, seven common SNPs reached the genome-wide significant threshold ($P = 1.63 \times 10^{-8}$) and were identified for both DGE (Table 3 and Figs. 4, 5a) and SGE in ADG (Table 4 and Figs. 4, 5b). Of these SNPs, five consecutive loci located within SSC8: 79.51–79.53 Mb were identified with a P value of 1.85×10^{-9} for DGE and 3.17×10^{-9} for SGE. Ten consecutive SNPs were detected at a suggestive

Table 3 Single-locus GWAS for direct genetic effects (DGE) in Yorkshire pigs.

Trait	SSC	Position (bp)	Range (Mb)	Allele	P_value	Candidate Gene
DGE_ADFI	6	18640605	18.14–19.14	A/G	4.09E−09	*CDH1, CDH3, ZFP90, BBS2, MT4, MT3, MT1A, NUP93, MIR138-2, HERPUD1,*
	6	18635874	18.14–19.14	C/G	4.40E−09	*SLC12A3, NLRC5, CPNE2, FAM192A*
	6	18635895	18.14–19.14	A/G	4.40E−09	
	6	18639708	18.14–19.14	C/T	4.40E−09	
	6	18641480	18.14–19.14	C/T	4.40E−09	
	6	18642431	18.14–19.14	C/G	4.40E−09	
	6	18643649	18.14–19.14	A/G	4.40E−09	
	6	18644213	18.14–19.14	T/C	4.40E−09	
	6	18645373	18.15–19.15	A/T	4.40E−09	
	6	18645626	18.15–19.15	A/G	4.40E−09	
	6	46451435	45.95–46.95	A/G	5.89E−09	*ZNF566, ZFP82, ZFP14, ZNF793, ZNF527, ZNF569, U2, ZNF570, ZFP30, WDR87*
	6	46448592	45.95–46.95	T/G	1.07E−10	
	8	79506792	79.01–80.01	C/T	5.98E−10	*IQCM*
	8	79525267	79.03–80.03	A/T	5.98E−10	
	8	79529184	79.03–80.03	G/A	5.98E−10	
	8	79529190	79.03–80.03	T/C	5.98E−10	
	8	79529891	79.03–80.03	C/T	5.98E−10	
DGE_ADG	6	25649764	25.15–26.15	C/T	2.21E−09	*CDH11*
	6	25651261	25.15–26.15	T/C	2.43E−09	
	8	79506792	79.01–80.01	C/T	1.85E−09	*IQCM*
	8	79525267	79.03–80.03	A/T	1.85E−09	
	8	79529184	79.03–80.03	G/A	1.85E−09	
	8	79529190	79.03–80.03	T/C	1.85E−09	
	8	79529891	79.03–80.03	C/T	1.85E−09	

DGE_ADFI direct genetic effects of average daily feed intake, DGE_ADG direct genetic effects of average daily gain, SSC chromosome, Range range of significant chromosome region.
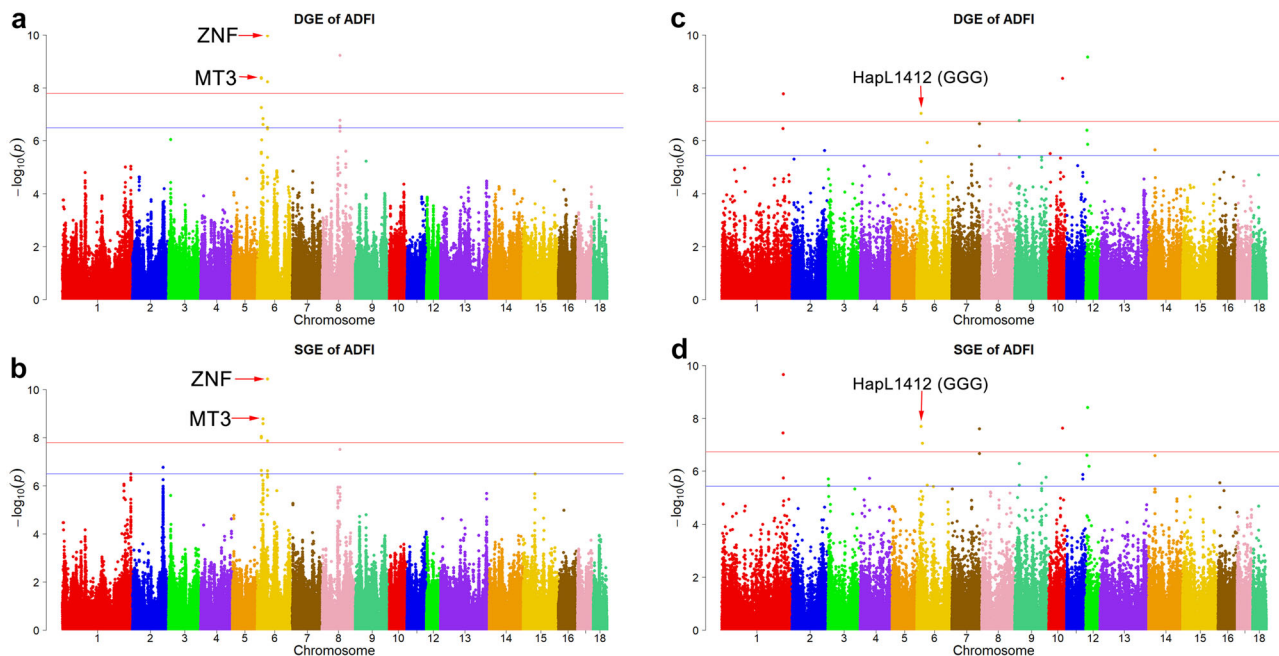
Fig. 3 GWAS results for direct genetic effects (DGE) and social genetic effects (SGE) in average daily feed intake (ADFI). a for DGE using imputed-GWAS; **b** for SGE using imputed-GWAS; **c** for DGE using haplotype-based GWAS; **d** for SGE using haplotype-based GWAS. For imputed-GWAS, the horizontal red and blue lines indicate the genome-wide (1.63×10^{-8}) and suggestive (3.25×10^{-7}) level, respectively. For haplotype-based GWAS, the horizontal red and blue lines indicate the genome-wide (1.82×10^{-7}) and suggestive (3.64×10^{-6}) level, respectively.

threshold for DGE in ADG in the region between 18.64 and 18.65 MB on chromosome 6 (Supplementary Data 2). The λ was 1.02 for DGE (Supplementary Fig. 1c) and 1.03 for SGE (Supplementary Fig. 1d). These λ values were similar to the values found for DGE and SGE in ADFI. Moreover, in the chip-based

GWAS, 87 SNPs were identified for DGE and 85 SNPs were identified for SGE ($P < 2.70 \times 10^{-5}$, Supplementary Fig. 2).

For the other six traits. Using imputed GWAS, no significant SNPs were identified for either DGE or SGE for the backfat

Table 4 Single-locus GWAS for social genetic effects (SGE) in Yorkshire pigs.

Trait	SSC	Position (bp)	Range (Mb)	Allele	P_value	Candidate Gene
SGE_ADFI	6	18635874	18.14–19.14	C/G	8.94E−09	*CDH1, CDH3, ZFP90, BBS2, MT4, MT3, MT1A, NUP93, MIR138-2, HERPUD1,*
	6	18635895	18.14–19.14	A/G	8.94E−09	*SLC12A3, NLRC5, CPNE2, FAM192A*
	6	18639708	18.14–19.14	C/T	8.94E−09	
	6	18641480	18.14–19.14	C/T	8.94E−09	
	6	18642431	18.14–19.14	C/G	8.94E−09	
	6	18643649	18.14–19.14	A/G	8.94E−09	
	6	18644213	18.14–19.14	T/C	8.94E−09	
	6	18645373	18.15–19.15	A/T	8.94E−09	
	6	18645626	18.15–19.15	A/G	8.94E−09	
	6	18640605	18.14–19.14	A/G	1.00E−08	
	6	25649764	25.15–26.15	C/T	1.67E−09	*CDH11*
	6	25651261	25.15–26.15	T/C	2.68E−09	
	6	46451435	45.95–46.95	A/G	1.39E−08	*OVOL3, POLR2I, TBCB, CAPNS1, COX7A1, ZNF146, ZNF565, ZNF567, ZNF461,*
	6	46448592	45.95–46.95	T/G	3.63E−11	*ZNF382, ZNF260, ZNF566, ZFP82, ZFP14, ZNF793, ZNF527, ZNF569, U2, ZNF570,*
						ZFP30, WDR87
	8	79506792	79.01–80.01	C/T	3.16E−08	*IQCM*
	8	79525267	79.03–80.03	A/T	3.16E−08	
	8	79529184	79.03–80.03	G/A	3.16E−08	
	8	79529190	79.03–80.03	T/C	3.16E−08	
	8	79529891	79.03–80.03	C/T	3.16E−08	
SGE_ADG	6	25649764	25.15–26.15	C/T	3.00E−09	*CDH11*
	6	25651261	25.15–26.15	T/C	3.68E−09	
	8	79506792	79.01–80.01	C/T	3.17E−09	*IQCM*
	8	79525267	79.03–80.03	A/T	3.17E−09	
	8	79529184	79.03–80.03	G/A	3.17E−09	
	8	79529190	79.03–80.03	T/C	3.17E−09	
	8	79529891	79.03–80.03	C/T	3.17E−09	

SGE_ADFI social genetic effects of average daily feed intake, *SGE_ADG* social genetic effects of average daily gain, *SSC* chromosome, *Range* range of significant chromosome region.
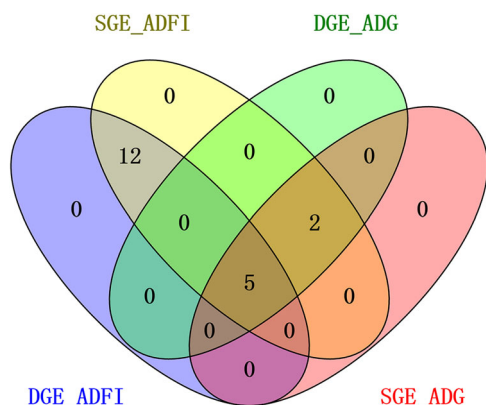
Fig. 4 The common SNPs were shared by both direct genetic effects (DGE) and social genetic effects (SGE). Venn diagram of common significant SNPs of DGE and SGE in average daily feed intake (ADFI) and average daily gain (ADG).

thickness to 100 kg (B100), days to 100 kg (D100), feed conversion ratio (FCR), residual feed intake (RFI), time in feeder per day (TPD), and feeding speed (FS) traits (Supplementary Fig. 3). However, two peaks were found for D100; one of which was located on SSC6 for DGE and the other on SSC5 for SGE. For RFI, a distinct peak located on SSC10 was found for both DGE and SGE.

In the chip-based GWAS, the total number of SNPs detected ($P < 2.70 \times 10^{-5}$) was 4 for B100, 15 for D100, 20 for RFI, 3 for FS, and 5 for TPD. For SGE, the total number of SNPs detected ($P < 2.70 \times 10^{-5}$) was 1 for B100, 5 for RFI, 1 for FS, and 1 for TPD (Supplementary Fig. 2). Among them, three SNPs were also

detected by imputed GWAS for the B100, D100, and FS traits ($P < 2.70 \times 10^{-5}$, Supplementary Table 1).

**Haplotype-based GWAS for eight socially affected traits**
For ADFI. Four significant haplotype loci were identified for DGE (Supplementary Table 2 and Fig. 3c) and seven for SGE (Supplementary Table 3 and Fig. 3d). These were distributed on SSC1, SSC6, SSC7, SSC10, and SSC12 ($P < 1.82 \times 10^{-7}$). Among these, four common haplotype loci (HapL1834, HapL910, HapL1412, and HapL1193) were shared by DGE and SGE. Only the *MT3* gene was located within the haplotype HapL1412 "GGG", with a haplotype frequency of 0.05. Furthermore, 12 suggestive haplotype loci were detected for DGE and 17 for SGE ($1.82 \times 10^{-7} < P < 3.64 \times 10^{-6}$, Supplementary Table 4). The λ was 1.06 for DGE (Supplementary Fig. 4a) and 1.08 for SGE (Supplementary Fig. 4b); implying no population stratification.

For ADG. Three haplotype loci surpassing the significance threshold ($P = 1.82 \times 10^{-7}$) were detected for DGE (Supplementary Table 2 and Fig. 5c) and one was detected for SGE (Supplementary Table 3 and Fig. 5d). The haplotype loci (HapL910) were shared by DGE and SGE. In addition, this haplotype was shared by ADFI and ADG with a haplotype frequency of 0.03. A total of ten suggestive haplotype loci were identified for DGE and SGE (Supplementary Table 4). The λ was 1.04 for DGE (Supplementary Fig. 4c) and 1.03 for SGE (Supplementary Fig. 4d); implying no population stratification.

For D100 and B100. When considering the DGE and SGE, we identified 20 different haplotype loci that reached the significance threshold ($P = 1.82 \times 10^{-7}$) for D100 (Supplementary Tables 2, 3 and Figs. 6a, b). Among these, three common haplotype loci
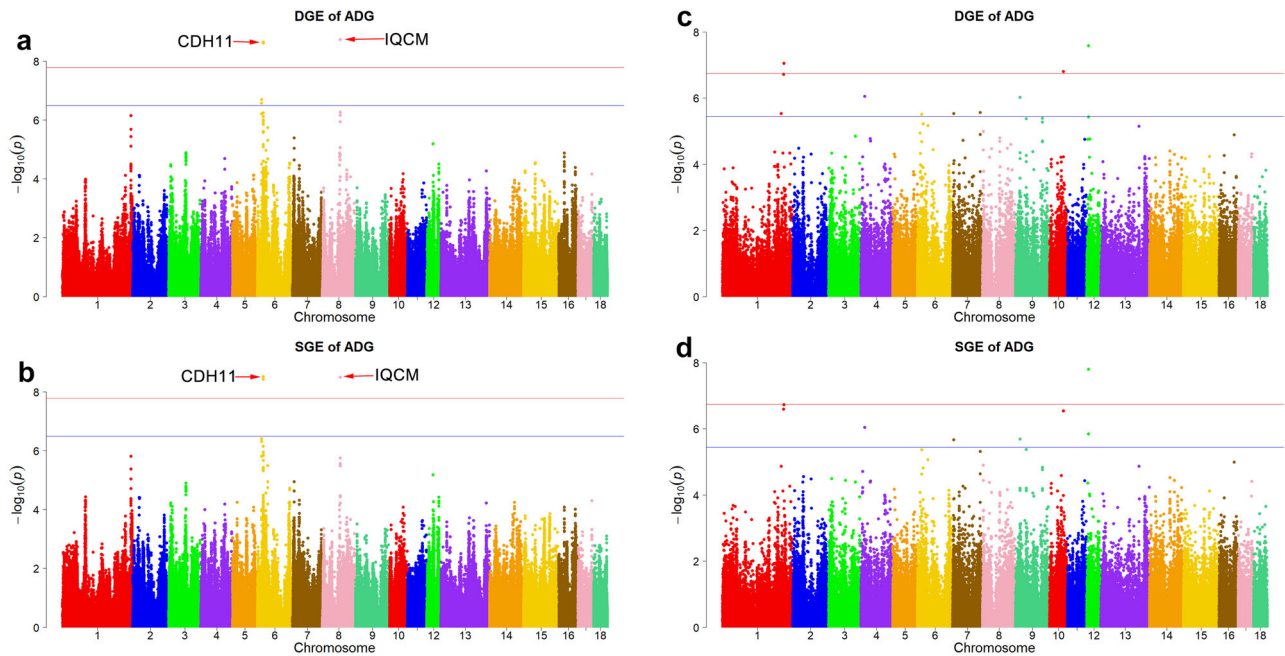
Fig. 5 GWAS results for direct genetic effects (DGE) and social genetic effects (SGE) in average daily gain (ADG). a for DGE using imputed-GWAS; **b** for SGE using imputed-GWAS; **c** for DGE using haplotype-based GWAS; **d** for SGE using haplotype-based GWAS. For imputed-GWAS, the horizontal red and blue lines indicate the genome-wide (1.63×10^{-8}) and suggestive (3.25×10^{-7}) level, respectively. For haplotype-based GWAS, the horizontal red and blue lines indicate the genome-wide (1.82×10^{-7}) and suggestive (3.64×10^{-6}) level, respectively.
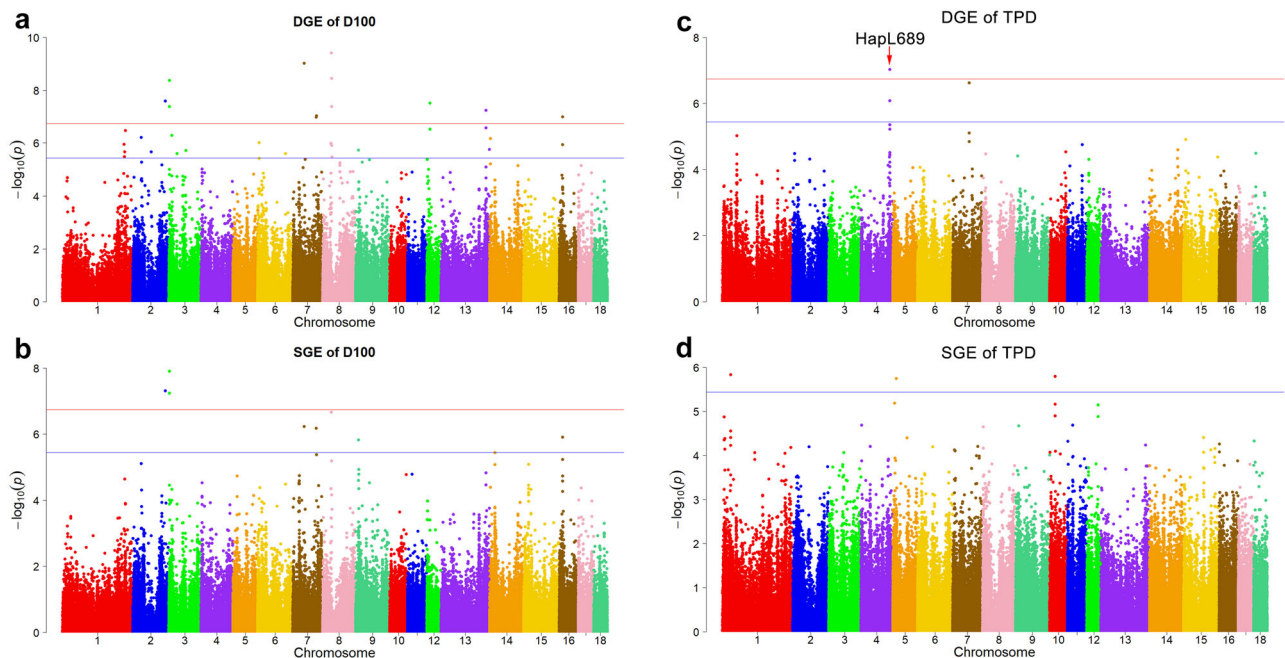
Fig. 6 Haplotype-based GWAS results for direct genetic effects (DGE) and social genetic effects (SGE). a for DGE in days to 100 kg (D100); **b** for SGE in D100; **c** for DGE of the time in feeder per day (TPD); **d** for SGE of TPD. The horizontal red and blue lines indicate the genome-wide (1.82×10^{-7}) and suggestive (3.64×10^{-6}) level, respectively.

(HapL36 HapL1601, and HapL213) were detected for DGE and SGE. Eighteen candidate genes were located within these significant regions. The *SDK1* gene within the significant haplotype HapL36 was shared by DGE and SGE. On basis of the suggestive level ($P = 3.64 \times 10^{-6}$), a total of 21 haplotype loci were detected for DGE and 13 for SGE (Supplementary Table 4). For B100, no haplotype loci were found to be associated with DGE (Supplementary Fig. 5a) or SGE (Supplementary Fig. 5b).

For FCR and RFI. No significant haplotype loci were detected for DGE or SGE (Supplementary Fig. 6); however, five suggestive haplotype loci ($1.82 \times 10^{-7} < P < 3.64 \times 10^{-6}$) were identified for DGE (three for FCR and two for RFI) and one for SGE on RFI (Supplementary Table 4 and Supplementary Fig. 6).

For TPD and FS. Significant haplotype loci located on SSC4: 130,340,342–130,390,531 were identified for the DGE on TPD

(Fig. 6c) and significant haplotype loci were detected for the SGE on FS (Supplementary Fig. 7). Furthermore, 15 different haplotype loci that surpassed the suggestive level ($P = 3.64 \times 10^{-6}$) were identified for DGE and SGE (Supplementary Table 4).

Comparing single-locus and haplotype-based GWAS. Haplotype-based GWAS identified more associations than single-locus GWAS for the DGE and SGE on the eight socially affected traits in pigs. Importantly, the significant HapL1412 with haplotype allele "GGG" overlapped with two significant SNPs (SSC6: 18,635,874 and SSC6: 18,635,895) for both DGE and SGE on ADFI. A candidate gene, *MT3*, was found in these regions. Furthermore, multiple significant SNPs and haplotype loci were shared by different traits, which indicates the pleiotropism.

Discussion

The socially affected traits are influenced by multiple genes. This study is designed to investigate DGE and SGE for eight socially affected traits using imputed WGS, as there is strong evidence that these traits are socially affected in pigs[2,8–10,18]. Using the social genetic model[12], we estimated the DGE and SGE of each socially affected trait in Yorkshire pigs. Then the single-locus and haplotype-based GWAS were performed to investigate associations for DGE and SGE using imputed data.

Different imputation strategies can result in different levels of accuracy[19]. The use of multi-breed reference populations is an effective way to improve the accuracy of imputation[19,20]. However, a large genetic distance between the reference and target populations can result in extremely low accuracies (less than 0.49)[21]. Landrace and Yorkshire pigs are bred according to the same breeding goals and are therefore likely genetically similar[22]. Additionally, because of the small effective population size for pigs[23], the imputation parameter of effective population size was set at 100 in the Beagle 5.1 software. After imputation, the average accuracy at the whole-genome level was 0.63, and 3,072,572 SNPs had a Beagle R^2 higher than 0.80 in the imputed WGS dataset. The imputation accuracy observed in the current study was higher than that reported in other studies[19–21], for example, a study using a ten-breed reference population of 168 pigs and using imputation from low- or high-density SNP data to WGS resulted in relatively low accuracy (0.39–0.49)[21]. Our results were more similar to the accuracy obtained from a multi-breed reference population of cattle (>0.70)[19,20]. Thus, the present study indicates that using mixed reference populations with similar genetic backgrounds can result in reasonable imputation accuracy in pigs. The imputation accuracy increased with reference population size[24]. In this study, a total of 40 sequenced pigs was used. Although the reference population size is smaller than that used in previous studies (>100) for pigs[25], this study also obtained a high imputation accuracy, and therefore, the imputed data can provide important information regarding socially affected traits in pigs.

Imputation accuracy is heavily affected by the MAF[26]. Rare variants with a low MAF result in poor imputation accuracy and therefore decrease the average imputation accuracy. In our study, the accuracy decreased sharply when the MAF was lower than 0.10. Filtering out the SNPs with MAF < 0.10 from the imputed data increased the average imputation accuracy from 0.63 to 0.71. In our study, the imputation accuracy increased with the MAF, which is in agreement with previous studies in livestock[21,24,27,28]. Rare variants are expected to have larger effects on complex traits than common variants[29]. It is also harder to construct the haplotype background using rare SNPs and they decrease the set of template haplotypes for genotype imputation because they are observed only a few times in reference genotype datasets[26].

Thus, genotype imputation of rare SNPs does not provide high accuracy levels. It is important to investigate imputation methods for rare variants in further studies.

In the processes of natural selection and the artificial selection of pigs, complex traits have developed high levels of genetic variation. The single-locus analysis focuses on identifying SNPs and mining functional genes for complex traits[13], but this method has only identified a small fraction of genetic variants in pigs. Haplotype-based GWAS is a complementary method that intensifies the benefits from linkage disequilibrium (LD) and enables the evaluation of the genetic determinants of complex traits. The use of haplotype information in GWAS would likely be beneficial for detecting further genetic variants[30]. However, haplotype-based GWAS using imputed WGS has not yet been reported in pigs. The rare variants may contribute large effects to complex traits[29]. Our analysis were done on imputed sequence data, which may contain imputation errors[31]. To alleviate this issue, further quality control was also conducted on the imputed sequence genotypes, rare variants with a MAF lower than 0.01 were moved. After quality control, the average MAF is 0.30 for chip-based data, 0.27 for imputed data, and 0.09 for haplotype alleles. Our results show that haplotype-based GWAS can be effective and can provide more genetic information than single-locus GWAS. These results are in agreement with previous studies; for example, haplotype-based GWAS in plants[32] and cattle[16] captured genetic variants that were not identified by single-locus analysis. Importantly, by combining single-locus and haplotype-based GWAS, we confirmed the presence of two important SNPs (SSC6: 18,635,874 and SSC6: 18,635,895) for DGE and SGE in ADFI. Therefore, our results highlight the advantages of haplotype-based GWAS for detecting genetic variants that influence complex traits in pigs. The power of GWAS is limited by sample size and SNP density. Using the imputed WGS instead of the 50 K SNP chip is powerful to investigate socially affected traits in pigs. Although only 1204 imputed WGS were used in this study, our results also provided important genetic variation for socially affected traits. Further studies are recommended to validate these results using a larger sample size.

Socially affected traits are known to be affected by SGE, but the mechanisms of SGE are not fully known. Many studies have been conducted to understand the genetic architecture of complex traits and they have identified many important SNPs and genes associated with these traits[33]. However, these studies have mainly focused on investigations of DGE, despite the majority of complex traits in pigs being affected by social interactions among individuals[2]. Pigs are highly social animals, so ignoring SGE could incorrectly estimate genetic variance and restrict genetic improvement.

Recently, numerous studies have shown that using social genetic models during genetic evaluation can enhance the genetic improvement of socially affected traits in pigs[1,10]. However, few studies have investigated the genetic architecture of DGE and SGE in pigs[14,15,34]. On the basis of our GWAS results, we found that the significant loci and candidate genes of DGE are different from SGE for each trait in Yorkshire pigs. These results are in agreement with previous studies, for example, a distinct difference was found between DGE and SGE for socially affected traits in laying hens[13], and a study based on single-step GWAS found only one QTL (SSC6: 19.9–20.9 Mb) for DGE and SGE on ADG in pigs[34]. The population composition may affect the estimation of DGE and SGE. An experimental design using groups comprising two families is optimal for estimating SGE[35]. Our group's composed of numerous families because all pigs in our study were from a commercial pig performance testing station. The group design in this study may have affected the estimation of DGE and SGE; however, it is difficult to conduct an experiment containing

only two families from a commercial breeding program. Moreover, in our study, all pigs were derived from the same herd and there were 20 pigs in each group, which made it easier to estimate DGE and SGE[5].

Our results provide further evidence for the genetic differences in DGE and SGE with regard to socially affected traits. Economically important traits are selected by artificial selection when based on the classical model that only considers DGE. SGE is, therefore, less frequently selected when compared with DGE. The complex traits in pigs are highly polygenic, and more research is required to understand the genetic architecture of the SGE on the complex traits of pigs.

To assess whether the identified SNPs associated with socially affected traits in our study replicate previously identified QTL, we compared our results with the PigQTL database based on the genome position of each SNP and QTL. A total of 78 important QTL were found to overlap with previously identified QTL for DGE of complex traits in pigs (Supplementary Table 5). No previous QTL has been reported to be associated with SGE. SGE has been evaluated to play an important role in livestock[12]. Thus, the investigation of the genetic architecture of SGE has important implications for pig breeding programs.

In our study, a total of 56 candidate genes associated with SGE and DGE were detected. Gene enrichment analysis showed that enriched pathways were associated with nucleic acid, calcium, and metal ion binding (Supplementary Table 6). In particular, two candidate genes, *IQCM* and *CDH11*, were associated with both DGE and SGE on the ADG and ADFI traits. *CDH11* gene encodes a calcium-dependent glycoprotein and mediates calcium-dependent cell-cell adhesion[36]. The *CDH11* gene is essential for tissue development, regulation of cell proliferation, and survival[37,38]. This gene has also been associated with postnatal bone development in mice[39]. Previous studies have shown that the *CDH11* gene is associated with growth traits in cattle including weaning weight[40] and RFI[41]. In pigs, this gene was found to be associated with fat and meat quality traits[42] and to regulate neural development and cell motility[43]. In this important chromosome region, six QTL for growth traits were overlapped, including ADG[44] and BW[45] in pigs. Additionally, these significant SNPs were located on the QTL for health traits[46].

In the region of 45.95–46.95 Mb on SSC6, numerous zinc-finger protein family genes were found to be associated with both DGE and SGE. The zinc-finger protein is one of the most abundant classes of transcription factors and regulates cell growth and differentiation. According to our gene ontology (GO) analysis, these genes regulate transcription, nucleic acid binding, and metal ion binding. Furthermore, this chromosome region overlapped with four reported QTL affecting ADG[44,47,48] and eating behavior[49] traits in pigs. Thus, daily interactions (SGE) between group pigs would influence their health. Interestingly, the significant SNPs within the region SSC6: 45.95–46.95 Mb were located within 20 QTL known to be associated with health traits[46,50] in pigs (https://www.animalgenome.org/).

In the region of SSC6: 18.14–19.14 Mb, *MT3* gene was found in both single-locus and haplotype-based GWAS for socially affected traits. This gene encodes a growth inhibitory factor that regulates many biological processes, particularly within the nervous and immune systems, and thereby influences health. *MT3* plays an important role in zinc homeostasis and cell death[51], and inhibits cell growth under zinc-deficient conditions[52]. Interestingly, our study also identified numerous zinc-finger protein family genes to be associated with DGE and SGE in SSC6: 45.95–46.95 Mb.

The previous studies identified several QTL on SSC6 for DGE and SGE in pigs[14,15,34]. The QTL (SSC6: 18.14–19.14 Mb) in the current study was close to the reported QTL (SSC6: 19.9–20.9 Mb) for both DGE and SGE on ADG in pigs[34]. Furthermore,

there were eight reported QTL associated with growth traits in the region of SSC6: 18.14–19.14 Mb. Among them, five QTL were related to ADG[53] and feeding intake[54]. The significant SNPs associated with DGE and SGE are located in nine reported QTL for health traits[46]. And, it was reported that three QTL located on SSC6: 18.14–19.14 Mb in pigs were associated with social interaction traits, and health traits[49,55].

In summary, we report the study employing combined single-locus and haplotype-based GWAS to identify the genetic architecture of socially affected traits that are influenced by both DGE and SGE in pigs. Our study shows the feasibility of mapping genomic variants that underlie SGE and provides genomic information for socially affected traits in pigs. These results provide evidence that the genomic architecture of SGE and DGE is different for socially affected traits. Hence, we recommend the use of both DGE and SGE to evaluate genetic architecture for socially affected traits in pigs.

Methods

Ethics declarations. All experimental procedures were performed in accordance with the Institutional Review Board (IRB14044) and the Institutional Animal Care and Use Committee of the Sichuan Agricultural University under permit number DKY-B20140302.

Animals and housing. During the period of 2017–2019, phenotypic data were collected for Yorkshire pigs from the national nucleus pig breeding farm of New Hope Group, Co., Ltd. (Sichuan, China) using the Osborne FIRE Pig Performance Testing System (Osborne, KS, United States). As target population, a total of 1204 Yorkshire pigs were placed in a temperature-controlled room at $25 \pm 2\,^{\circ}$C and relative humidity of 65–80% during the period of performance test from 30 to 110 kg. As reference population, a total of 60 pigs (20 Landrace and 40 Yorkshire pigs) were randomly selected from core populations. These pigs were group-housed in cement-floor pens (20 pigs in each pen) and the nutrient requirements were met as recommended by the National Research Council (NRC 2012).

Phenotypic data. A total of 1112 pigs' phenotypic data were collected, including ADG (kg/d), D100, B100, ADFI (kg/d), RFI, FCR, TPD (min/d), and FS (g/min). In this study, each pig was labeled with a unique electric identification tag on ear that was detected by the Osborne FIRE Pig Performance Testing System. The feed time, feed consumption, and BW were recorded at each visit to the feeder for each pig. At the end of the performance test, BFT between the third and fourth last ribs of each pig was calculated by PIGLOG 105B ultrasound machine (SFK Technology, Søborg, Denmark). ADG was calculated as linear regressions of BW from 30 to 110 kg on the performance test days. ADFI was calculated based on the total amount of recorded total feed intake (TFI) divided by the number of corresponding feed days at the feeder. TPD was calculated based on the total amount of recorded total time divided by the number of corresponding feed days (min/d), and FS = ADFI/TPD (g/min)[56,57]. FCR, D100, B100, and RFI were calculated as follows[58]:

$$FCR = \frac{TFI}{Weight_2 - Weight_1} \tag{1}$$

$$D100 = tested\ days + (100 - Weight_2) \times \frac{tested\ days - A}{Weight_2} \tag{2}$$

$$B100 = BFT + (100 - Weight_2) \times \frac{BFT}{Weight_2 - B} \tag{3}$$

$$RFI = ADFI - 14.1ADG - 2.83BFT - 110.9AMW \tag{4}$$

$$AMW = \frac{\left(Weight_2^{1.6} - Weight_1^{1.6}\right)}{1.6 \times \left(Weight_2 - Weight_1\right)} \tag{5}$$

where **Weight$_1$** and **Weight$_2$** are weights at the start and end of the performance test, respectively; **tested days** is the duration of the performance test in days. **A** is 50.775 for males and 46.415 for females; **B** is −7.277 for males and −9.440 for females. A and B were calculated based on an actual dataset of performance tests involving 5000 pigs[59]. The true days and B100 were firstly obtained based on two data that were the closest to 100 kg using linear interpolation. Then the nonlinear models of D100 and B100 were constructed based on linear interpolation as models (2) and (3). Finally, the A and B were calculated based on models (2) and (3) using the NLIN procedure in SAS software, respectively. BFT is the tested backfat thickness at the end of a performance test. AMW is an average metabolic BW.

The deregressed EBVs of DGE and SGE. The DGE and SGE were estimated for eight socially affected traits using a social genetic effect model[12], as follows:

$$y = Xb + Z_D a_D + Z_S a_S + Wl + Vg + e \qquad (6)$$

where y is the vector of phenotypic values; b is the vector of fixed effects, including sex, test year and month, birth year and month effects; a_D and a_S are vectors of DGE and SGE, respectively; l is the vector of random litter effects; g is the vector of random group effects in which the pigs were penned during the performance test; e is the random residual vector; X, Z_D, Z_S, W, and V are the incidence matrices of b, a_D, a_S, l, and g, respectively. This model was implemented by AI-REML in DMU software[60]. The estimated accuracies of DGE and SGE were calculated based on the formula $r = \sqrt{1 - s_e^2/\sigma_a^2}$, s_e^2 is the error variance of DGE and SGE, σ_a^2 is the corresponding genetic variance. Based on the estimated accuracies of DGE and SGE, their deregressed breeding values (EBVs) were obtained based on the formula *Deregressed* $EBVs = g_i/r_i^2$, g_i is the EBVs of ith individual, r_i^2 is the square of estimated accuracies for ith individual[61]. Then, the deregressed EBVs were used to implement for association analysis.

Genomic DNA extraction. The ear tissue samples were collected and stored in 75% alcohol. Genomic DNA from 1264 ear tissues was extracted using the Tissues Genomic DNA (Omega Bio-Tek, Norcross, GA, USA) kit according to the manufacturer's instructions, and then the quality and quantity were measured using a Nanodrop-2000 spectrophotometer. The genomic DNA with the ratio of light absorption (A260/280) between 1.8 and 2.0, concentration ≥50 ng/μL, and total volume ≤50 μL were eligible.

Genotyping by Illumina Porcine SNP50K BeadChip. A total of 1204 Yorkshire pigs, 442 boars and 762 gilts, were genotyped by the Illumina Porcine 50K SNP Chip (Neogen, Lincoln, NE, USA), which contained 50,697 SNPs. Firstly, the SNPs with no position information and located on sex chromosomes were removed from the genotype data, which contains 11,874 SNPs. Then, quality control of the genotype data was performed using PLINK software[62]. The SNPs with call rate < 0.90, MAF < 0.05, and Hardy–Weinberg equilibrium test (HWE) $<10^{-6}$, were excluded from the dataset. After quality control, a total of 1854 SNPs were removed. Finally, a total of 36,969 SNPs were used as target genotype data.

Whole-genome sequencing and SNP calling. The WGS reference data were obtained for 60 pigs (20 Landrace and 40 Yorkshire pigs), with 150 bp paired-end reads on the Illumina HiSeq PE150 platform. The sequencing was performed by BGI Co., Ltd. (Wuhan, China). After sequencing, the quality of raw reads was checked with a Phred score of 20 as the minimum to filter the adapter polluted reads and multiple N reads (where N > 10% of one read) to produce clean reads by FastQC (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/). Then the clean reads were mapped to the pig reference genome (Sscrofa11.1) using BWA (version 0.7.15) software with the parameters mem -t 10 -k 32 -M[63]. The SAM files produced from BWA were converted to BAM files using SAMtools (version 1.19)[64]. The potential PCR duplicates were removed by MarkerDuplicates utility in Picard release 1.119 (https://sourceforge.net/projects/picard/files/picardtools/1.119/). After that, BAM files were used to call SNPs using GATK (version 3.5) software[65] with multi-sample approaches. The raw SNPs generated from GATK were filtered with QualByDepth (QD) < 2.0, FisherStrand (FS) < 60.0, RMSMappingQuality (MQ) < 40.0, MappingQualityRankSumTest < −12.5, and ReadPosRankSumTest < −8.0. After the initial filtering, a total of 21,104,245 SNPs were removed. For further analyses, the SNPs with MAF >0.05, missing rate <0.1, HWE $<10^{-6}$, read depth (dp) >6, and the SNPs located on autosomes were considered as reference genotype data.

Imputation from 50 K chip to WGS. Genotype imputation between target and reference genotype data were performed by Beagle (version 5.1)[66] with default parameter settings, except for setting the effective population size to 100[22]. The imputation accuracy of each SNP was assessed using the Beagle R^2, which is the estimated squared correlation between the estimated allele dosage and the true allele dosage[67]. A two-breed reference population (20 Landrace and 40 Yorkshire pigs) with a small genetic distance[22] was used to perform genotype imputation from 50 K SNPs to WGS. After imputation, to maintain a balance between the average accuracy and the number of SNPs, SNPs with a Beagle $R^2 < 0.8$ were excluded. Furthermore, SNPs with MAF less than 0.01 and HWE less than 10^{-6} were removed from imputed data. Finally, a total of 3,072,572 SNPs were retained for further analyses.

Constructing haplotype loci. Haplotype loci were identified from the imputed WGS data in 1204 pigs. Haplotype loci were detected for each chromosome as proposed by Gabriel et al.[68] using PLINK v1.90 software[62]. The parameters were set to "-blocks no-pheno-req -blocks-max-kb 1000 -blocks-strong-lowci 0.8 -geno 0.1". A haplotype block containing two or more SNPs with high LD was defined, and loci with a confidence interval of r^2 higher than 0.8 were considered into one block. Then, the haplotype calling and identification of haplotype alleles were performed by GHap package[69]. Furthermore, haplotype loci were transformed into multi-allelic markers, and the haplotype genotype matrix was used to perform GWAS. Using the imputed data with 3,072,572 SNPs, a total of 33,708 haplotype loci were constructed and each haplotype loci contained 34.8 alleles. Finally, 274,741 haplotype alleles with MAF >0.01 were retained.

Association analysis. GWAS were performed independently for eight socially affected traits by considering both DGE and SGE using GEMMA software[65]. Before association analysis, the centered genotypes were used to estimate the $n \times n$ genomic relationship matrix between the individuals. The genomic relatedness matrix was calculated as follows:

$$G = \frac{1}{p}\sum_{i=1}^{p}\left(X_i - 1_n \bar{x}_i\right)\left(X_i - 1_n \bar{x}_i\right)^T \qquad (7)$$

where G is the genomic relatedness matrix between the individuals; n is the number of individuals; p is the number of genotypes; i represents the ith SNP; X denote $n \times p$ matrix of genotypes; X_i as its ith column representing genotypes of ith SNP; $\bar{x}_i$ as the sample mean of ith SNP; 1_n as a $n \times 1$ vector of 1's.

The deregressed EBVs were used to perform single-locus GWAS (including chip-based GWAS and imputed-GWAS) and haplotype-based GWAS, the following unified mixed linear model was used:

$$y = Xm + Wa + e \qquad (8)$$

$$a \sim MVN\left(0, G\sigma_a^2\right)$$

$$e \sim MVN\left(0, I\sigma_e^2\right)$$

where y is the vector of the deregressed EBVs of DGE or SGE; m is the SNP effects; a is the vector of residual polygenic effects; e is the vector of random residuals; X and W are the incidence matrices for m and a, respectively; G is a genomic relationship matrix, I is an identity matrix; MVN denotes multivariate normal distribution.

The genome-wide significant threshold value was determined using the Bonferroni correction method[70]. For chip-based GWAS, the genome-wide significant and suggestive levels were set as $P = 0.05/N_1 = 1.35 \times 10^{-6}$ and $P = 1/N_1 = 2.70 \times 10^{-5}$, respectively, where N_1 is the number of analyzed SNPs. For imputed-GWAS, the genome-wide significant and suggestive levels were set as $P = 0.05/N_2 = 1.63 \times 10^{-8}$ and $P = 1/N_2 = 3.25 \times 10^{-7}$, respectively, where N_2 is the number of analyzed SNPs. For haplotype-based GWAS, the genome-wide significant and suggestive levels were calculated as $P = 0.05/N_3 = 1.82 \times 10^{-7}$ and $P = 1/N_3 = 3.64 \times 10^{-6}$, respectively, where N_3 is the number of studied haplotype alleles.

The Manhattan plots were drawn using qqman package[71]. Genomic inflation factor (λ) was calculated to judge the extent of false-positive signals with estlambda function in GenABEL package ($\lambda = \frac{\text{the observed } P \text{ values}}{\text{the expected } P \text{ values}}$)[72].

Candidate genes and functional analysis. The region within a 1 Mb region centering each significant SNP was defined as the QTL region. The candidate functional genes were searched within the identified QTL regions on the pig genome assembly 11.1 (https://asia.ensembl.org/Sus_scrofa/Info/Index). Then, the biological functions of these candidate genes were investigated on PubMed (https://www.ncbi.nlm.nih.gov/pubmed) and the reported literature. For functional annotation, GO analysis was performed on DAVID Bioinformatics Resources (https://david.ncifcrf.gov). The Fisher's test was used to assess the significance of the determined enriched terms[73,74]. Enriched GO terms (P < 0.05) were selected to investigate the genes involved in biological processes[75].

Data availability

The datasets supporting the results of this article are included within the article. All genotypic, phenotype data, and supplemental material were deposited at the figshare repository (https://doi.org/10.6084/m9.figshare.12611786). All other data are available from the corresponding authors upon reasonable request.

References

1. Ellen, E. D. et al. The prospects of selection for social genetic effects to improve welfare and productivity in livestock. *Front. Genet.* 5, 377 (2014).
2. Bergsma, R., Kanis, E., Knol, E. F. & Bijma, P. The contribution of social effects to heritable variation in finishing traits of domestic pigs (Sus scrofa). *Genetics* 178, 1559–1570 (2008).

3. Moore, A. J., Brodie, E. D. 3rd & Wolf, J. B. Interacting phenotypes and the evolutionary process:I. Direct and indirect genetic effects of social interactions. *Evolution* **51**, 1352–1362 (1997).

4. Canario, L., Lundeheim, N. & Bijma, P. Pig growth is affected by social genetic effects and social litter effects that depend on group size. In *Proc. 9th World Congress on Genetics Applied to Livestock Production* (ed. Giessen) 1–6 (German Society of Animal Science, 2010).

5. Bijma, P. The quantitative genetics of indirect genetic effects: a selective review of modelling issues. *Heredity* **112**, 61–69 (2014).

6. Rostellato, R., Sartori, C., Bonfatti, V., Chiarot, G. & Carnier, P. Direct and social genetic effects on body weight at 270 days and carcass and ham quality traits in heavy pigs. *J. Anim. Sci.* **93**, 1–10 (2015).

7. Kim, Y. et al. Social genetic effects on days to 90 kg in Duroc and Yorkshire pigs. Korean. *J. Agric. Sci.* **43**, 595–602 (2016).

8. Camerlink, I., Turner, S. P., Bijma, P. & Bolhuis, J. E. Indirect genetic effects and housing conditions in relation to aggressive behaviour in pigs. *PLoS ONE* **8**, e65136 (2013).

9. Camerlink, I., Ursinus, W. W., Bartels, A. C., Bijma, P. & Bolhuis, J. E. Indirect genetic effects for growth in pigs affect behaviour and weight around weaning. *Behav. Genet.* **48**, 413–420 (2018).

10. Nielsen, H. M., Ask, B. & Madsen, P. Social genetic effects for growth in pigs differ between boars and gilts. *Genet. Sel. Evol.* **50**, 4 (2018).

11. Bouwman, A. C., Bergsma, R., Duijvesteijn, N. & Bijma, P. Maternal and social genetic effects on average daily gain of piglets from birth until weaning. *J. Anim. Sci.* **88**, 2883–2892 (2010).

12. Muir, W. M. Incorporation of competitive effects in forest tree or animal breeding programs. *Genetics* **170**, 1247–1259 (2005).

13. Brinker, T., Bijma, P., Vereijken, A. & Ellen, E. D. The genetic architecture of socially-affected traits: a GWAS for direct and indirect genetic effects on survival time in laying hens showing cannibalism. *Genet. Sel. Evol.* **50**, 38 (2018).

14. Hong, J. K. et al. A genome-wide association study of social genetic effects in Landrace pigs. *Asian-Australas. J. Anim. Sci.* **31**, 784–790 (2018).

15. Wu, P. et al. Whole-genome re-sequencing association study for direct genetic effects and social genetic effects of six growth traits in Large White pigs. *Sci. Rep.* **9**, 9667 (2019).

16. Chen, Z., Yao, Y., Ma, P., Wang, Q. & Pan, Y. Haplotype-based genome-wide association study identifies loci and candidate genes for milk yield in Holsteins. *PLoS ONE* **13**, e0192695 (2018).

17. Price, A. L., Zaitlen, N. A., Reich, D. & Patterson, N. New approaches to population stratification in genome-wide association studies. *Nat. Rev. Genet.* **11**, 459–463 (2010).

18. Reimert, I. et al. Backtest and novelty behavior of female and castrated male piglets, with diverging social breeding values for growth. *J. Anim. Sci.* **91**, 4589–4597 (2013).

19. Brøndum, R. F., Guldbrandtsen, B., Sahana, G., Lund, M. S. & Su, G. Strategies for imputation to whole genome sequence using a single or multi-breed reference population in cattle. *BMC Genomics* **15**, 1–8 (2014).

20. Bouwman, A. C. & Veerkamp, R. F. Consequences of splitting whole-genome sequencing effort over multiple breeds on imputation accuracy. *BMC Genet.* **15**, 1–9 (2014).

21. van den Berg, S. et al. Imputation to whole-genome sequence using multiple pig populations and its use in genome-wide association studies. *Genet. Sel. Evol.* **51**, 2 (2019).

22. Grossi, D. A. et al. Genetic diversity, extent of linkage disequilibrium and persistence of gametic phase in Canadian pigs. *BMC Genet.* **18**, 6 (2017).

23. Hall, S. J. Effective population sizes in cattle, sheep, horses, pigs and goats estimated from census and herdbook data. *Animal* **10**, 1778–1785 (2016).

24. Ye, S. et al. Imputation from SNP chip to sequence: a case study in a Chinese indigenous chicken population. *J. Anim. Sci. Biotechnol.* **9**, 30 (2018).

25. van den Berg, S. et al. Imputation to whole-genome sequence using multiple pig populations and its use in genome-wide association studies. *Genet. Sel. Evol.* **51**, 1–13 (2019).

26. Das, S., Abecasis, G. R. & Browning, B. L. Genotype imputation from large reference panels. *Annu. Rev. Genomics Hum. Genet.* **19**, 73–96 (2018).

27. Song, H. et al. Using imputation-based whole-genome sequencing data to improve the accuracy of genomic prediction for combined populations in pigs. *Genet. Sel. Evol.* **51**, 58 (2019).

28. Hayes, B. J., Bowman, P. J., Daetwyler, H. D., Kijas, J. W. & van der Werf, J. H. Accuracy of genotype imputation in sheep breeds. *Anim. Genet.* **43**, 72–80 (2012).

29. Morris, A. P. & Zeggini, E. An evaluation of statistical approaches to rare variant analysis in genetic association studies. *Genet. Epidemiol.* **34**, 188–193 (2010).

30. Lorenz, A. J., Hamblin, M. T. & Jannink, J. L. Performance of single nucleotide polymorphisms versus haplotypes for genome-wide association analysis in barley. *PLoS ONE* **5**, e14079 (2010).

31. Brøndum, R. F., Ma, P., Lund, M. S. & Su, G. Short communication: genotype imputation within and across Nordic cattle breeds. *J. Dairy Sci.* **95**, 6795–6800 (2012).

32. Contreras-Soto, R. I. et al. A genome-wide association study for agronomic traits in soybean using SNP markers and SNP-based haplotype analysis. *PLoS ONE* **12**, e0171105 (2017).

33. Hu, Z. L., Park, C. A., Wu, X. L. & Reecy, J. M. Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era. *Nucleic Acids Res.* **41**, D871–D879 (2013).

34. Hong, J. K. et al. Single-step genome-wide association study for social genetic effects and direct genetic effects on growth in Landrace pigs. *Sci. Rep.* **10**, 14958 (2020).

35. Bijma, P. Estimating indirect genetic effects: precision of estimates and optimum designs. *Genetics* **186**, 1013–1028 (2010).

36. Cheng, S. L. et al. Human osteoblasts express a repertoire of cadherins, which are critical for BMP-2-induced osteogenic differentiation. *J. Bone Min. Res.* **13**, 633–644 (1998).

37. Larue, L. et al. A role for cadherins in tissue formation. *Development* **122**, 3185–3194 (1996).

38. Gumbiner, B. M. Regulation of cadherin-mediated adhesion in morphogenesis. *Nat. Rev. Mol. Cell Biol.* **6**, 622–634 (2005).

39. Farber, C. R. et al. Identification of quantitative trait loci influencing skeletal architecture in mice: emergence of Cdh11 as a primary candidate gene regulating femoral morphology. *J. Bone Min. Res.* **26**, 2174–2183 (2011).

40. McClure, M. C. et al. A genome scan for quantitative trait loci influencing carcass, post-natal growth and reproductive traits in commercial Angus cattle. *Anim. Genet.* **41**, 597–607 (2010).

41. Sherman, E. L. et al. Fine mapping quantitative trait loci for feed intake and feed efficiency in beef cattle. *J. Anim. Sci.* **87**, 37–45 (2009).

42. Piórkowska, K., Żukowski, K., Ropka-Molik, K., Tyra, M. & Gurgul, A. A comprehensive transcriptome analysis of skeletal muscles in two Polish pig breeds differing in fat and meat quality traits. *Genet. Mol. Biol.* **41**, 125–136 (2018).

43. Schulte, J. D. et al. Cadherin-11 regulates motility in normal cortical neural precursors and glioblastoma. *PLoS ONE* **8**, e70962 (2013).

44. Harmegnies, N. et al. Results of a whole-genome quantitative trait locus scan for growth, carcass composition and meat quality in a porcine four-way cross. *Anim. Genet.* **37**, 543–553 (2006).

45. Rothammer, S. et al. Genome-wide QTL mapping of nine body composition and bone mineral density traits in pigs. *Genet. Sel. Evol.* **46**, 68 (2014).

46. Reiner, G. et al. Identification of QTL affecting resistance/susceptibility to acute Actinobacillus pleuropneumoniae infection in swine. *Mamm. Genome* **25**, 180–191 (2014).

47. Bidanel, J. P. et al. Detection of quantitative trait loci for growth and fatness in pigs. *Genet. Sel. Evol.* **33**, 289–309 (2001).

48. Rückert, C. & Bennewitz, J. Joint QTL analysis of three connected F2-crosses in pigs. *Genet. Sel. Evol.* **42**, 40 (2010).

49. Reiner, G. et al. Mapping of quantitative trait loci affecting behaviour in swine. *Anim. Genet.* **40**, 366–376 (2009).

50. Wang, J. Y. et al. Genome-wide association studies for hematological traits in swine. *Anim. Genet.* **44**, 34–43 (2013).

51. Vašák, M. & Meloni, G. Mammalian metallothionein-3: new functional and structural insights. *Int. J. Mol. Sci.* **18**, 1117 (2017).

52. Palmiter, R. D. Constitutive expression of metallothionein-III (MT-III), but not MT-I, inhibits growth when cells become zinc deficient. *Toxicol. Appl. Pharm.* **135**, 139–146 (1995).

53. Liu, G. et al. A genome scan reveals QTL for growth, fatness, leanness and meat quality in a Duroc-Pietrain resource population. *Anim. Genet.* **38**, 241–252 (2007).

54. Houston, R. D., Haley, C. S., Archibald, A. L. & Rance, K. A. A QTL affecting daily feed intake maps to chromosome 2 in pigs. *Mamm. Genome* **16**, 464–470 (2005).

55. Laenoi, W. et al. Quantitative trait loci analysis for leg weakness-related traits in a Duroc × Pietrain crossbred population. *Genet. Sel. Evol.* **43**, 13 (2011).

56. Ding, R. et al. Genome-wide association analysis reveals genetic loci and candidate genes for feeding behavior and eating efficiency in Duroc boars. *PLoS ONE* **12**, e0183244 (2017).

57. Do, D. N., Strathe, A. B., Jensen, J., Mark, T. & Kadarmideen, H. N. Genetic parameters for different measures of feed efficiency and related traits in boars of three pig breeds. *J. Anim. Sci.* **91**, 4069–4079 (2013).

58. Liu, X. The ministry of agriculture has formulated the "national pig genetic improvement program (2009-2020)". *Agri. Tech. Equip.* **19**, 10–13 (2011).

59. Zhang, Z., Zhang, H., Chen, Z. M. & Jia-Qi, L. I. Development of correction formula for production traits in swine breeding. Chinese. *J. Anim. Sci.* **51**, 49–54 (2015).

60. Madsen, O. & Jensen, J. A user's guide to DMU. A package for analysing multivariate mixed models. in Version 6, release 5.1 (2012).

61. Garrick, D. J., Taylor, J. F. & Fernando, R. L. Deregressing estimated breeding values and weighting information for genomic regression analyses. *Genet .Sel. Evol.* **41**, 55 (2009).

62. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

63. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

64. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

65. DePristo, M. A. et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).

66. Browning, B. L., Zhou, Y. & Browning, S. R. A one-penny imputed genome from next-generation reference panels. *Am. J. Hum. Genet.* **103**, 338–348 (2018).

67. Browning, B. L. & Browning, S. R. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am. J. Hum. Genet.* **84**, 210–223 (2009).

68. Gabriel, S. B. et al. The structure of haplotype blocks in the human genome. *Science* **296**, 2225–2229 (2002).

69. Utsunomiya, Y. T., Milanesi, M., Utsunomiya, A. T., Ajmone-Marsan, P. & Garcia, J. F. GHap: an R package for genome-wide haplotyping. *Bioinformatics* **32**, 2861–2862 (2016).

70. Bland, J. M. & Altman, D. G. Multiple significance tests: the Bonferroni method. *BMJ* **310**, 170 (1995).

71. Turner, S. D. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. Preprint at *bioRxiv* https://doi.org/10.1101/005165 (2014).

72. Aulchenko, Y. S., Ripke, S., Isaacs, A. & van Duijn, C. M. GenABEL: an R library for genome-wide association analysis. *Bioinformatics* **23**, 1294–1296 (2007).

73. Dennis, G. Jr. et al. DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol.* **4**, P3 (2003).

74. Rivals, I., Personnaz, L., Taing, L. & Potier, M. C. Enrichment or depletion of a GO category within a class of genes: which test? *Bioinformatics* **23**, 401–407 (2007).

75. Wang, K. et al. Genome wide association analysis reveals new production trait genes in a male Duroc population. *PLoS ONE* **10**, e0139207 (2015).

Author contributions
P.W., G.T., and K.W. performed experiments; P.W and K.W. analyzed data and prepared figures and tables; G.T. and P.W. edited and revised the manuscript; G.T., P.W., J.Z., D.C., A.J., Y.J., L.Z., X.Q., and X.L. conceived, designed research, and wrote this paper. All authors read and approved the final manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42003-021-02416-3.

Correspondence and requests for materials should be addressed to G.T.

Peer review information *Communications Biology* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editors: Brooke LaFlamme and George Inglis.

Reprints and permission information is available at http://www.nature.com/reprints

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