

RESEARCH

Open Access



# Comparison of two multi-trait association testing methods and sequence-based fine mapping of six additive QTL in Swiss Large White pigs

A. Nosková<sup>1\*</sup>, A. Mehrotra<sup>1</sup>, N. K. Kadri<sup>1</sup>, A. Lloret-Villas<sup>1</sup>, S. Neuenschwander<sup>2</sup>, A. Hofer<sup>3</sup> and H. Pausch<sup>1</sup>

## Abstract

**Background** Genetic correlations between complex traits suggest that pleiotropic variants contribute to trait variation. Genome-wide association studies (GWAS) aim to uncover the genetic underpinnings of traits. Multivariate association testing and the meta-analysis of summary statistics from single-trait GWAS enable detecting variants associated with multiple phenotypes. In this study, we used array-derived genotypes and phenotypes for 24 reproduction, production, and conformation traits to explore differences between the two methods and used imputed sequence variant genotypes to fine-map six quantitative trait loci (QTL).

**Results** We considered genotypes at 44,733 SNPs for 5,753 pigs from the Swiss Large White breed that had dergressed breeding values for 24 traits. Single-trait association analyses revealed eleven QTL that affected 15 traits. Multi-trait association testing and the meta-analysis of the single-trait GWAS revealed between 3 and 6 QTL, respectively, in three groups of traits. The multi-trait methods revealed three loci that were not detected in the single-trait GWAS. Four QTL that were identified in the single-trait GWAS, remained undetected in the multi-trait analyses. To pinpoint candidate causal variants for the QTL, we imputed the array-derived genotypes to the sequence level using a sequenced reference panel consisting of 421 pigs. This approach provided genotypes at 16 million imputed sequence variants with a mean accuracy of imputation of 0.94. The fine-mapping of six QTL with imputed sequence variant genotypes revealed four previously proposed causal mutations among the top variants.

**Conclusions** Our findings in a medium-size cohort of pigs suggest that multivariate association testing and the meta-analysis of summary statistics from single-trait GWAS provide very similar results. Although multi-trait association methods provide a useful overview of pleiotropic loci segregating in mapping populations, the investigation of single-trait association studies is still advised, as multi-trait methods may miss QTL that are uncovered in single-trait GWAS.

**Keywords** Genome-wide association study, Multivariate analyses, Meta-analyses, Pleiotropy, Imputation

## Background

Genome-wide association studies (GWAS) combine genotype and phenotype information to identify trait-associated variants. Genotypes at polymorphic loci are tested for association with phenotypes to determine their impact on traits of interest. Multi-trait GWAS can increase the statistical power over single-trait GWAS

\*Correspondence:

A. Nosková  
anoskova@ethz.ch

<sup>1</sup> ETH Zürich, Universitätsstrasse 2, 8092 Zürich, Switzerland

<sup>2</sup> ETH Zürich, Tannenstrasse 1, 8092 Zürich, Switzerland

<sup>3</sup> SUISAG, Allmend 10, 6204 Sempach, Switzerland



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

because they exploit cross-phenotype associations at pleiotropic loci [1–3].

Several methods have been developed to detect pleiotropic variants. These methods can be divided into two groups based on their underlying statistical framework [4, 5]. First, multivariate methods jointly model all traits of interest. This group of methods requires that all individuals included in the study have phenotypic records for all traits analysed, although there are exceptions (e.g., single step GWAS [6], imputation of phenotypes [7]). These methods exploit the genetic covariance between traits, thereby increasing statistical power over their univariate counterparts [5, 8, 9], unless all traits are highly correlated [9, 10]. Second, the meta-analysis of summary statistics enables to combine results from single trait GWAS, which means that the analyses can be carried out with different sets of individuals for each trait [1, 11–13].

The power to detect trait-associated variants increases as the marker density increases [14–17]. Low-pass sequencing is a cost-effective approach to provide high marker density [18–21]. The imputation from medium-density arrays to the whole-genome sequence level using a sequenced reference panel is another approach to provide sequence variant genotypes for large cohorts. Large and diverse porcine haplotype reference panels facilitate imputing sequence variant genotypes at high accuracy for animals from various breeds [22]. Medium-sized breed-specific reference panels may enable similar accuracy while reducing computational costs [23]. Imputation from medium-density genotypes to the whole-genome sequence level has been explored when high-density array-derived genotypes were not available [16, 24–26].

Only few genome-wide association studies have been conducted in the Swiss Large White (SLW) population. Becker et al. [27] performed association tests between 26 complex traits and 60 K SNPs genotyped in 192 breeding boars. This effort revealed only 4 QTL likely because the sample size was too small. Large-scale association testing had been conducted in other pig breeds (e.g., [27–29]). Fat deposition and weight gain-related traits have been considered frequently in these GWAS as they are economically relevant and highly heritable. Previous GWAS led to tens of proposed candidate genes affecting these traits, including *MCAR*, *BMP2*, *IGF2*, and *CCND2* [30–35].

In this paper, we compare single-trait, multivariate and meta-GWAS in 5,753 genotyped pigs from a Swiss breed to investigate the genetic architecture of 24 traits. We inferred sequence variant genotypes from a sequenced reference panel to identify candidate causal variants for six pleiotropic QTL.

## Results

### Single-trait association studies between array genotypes and 24 traits

A detailed description of the 24 traits considered in our study including their grouping into four categories (reproduction, production, conformation, all) is shown in Table 1. Marked pairwise correlations exist between the drEBV of the 24 traits (Additional file 1). The SNP-based heritability estimates of the drEBV (Table 1) were between 0.04 and 0.67.

Mixed model-based single-trait genome-wide association studies (stGWAS) between 40,382 SNPs and deregressed estimated breeding values (drEBV) for 24 traits in 5,753 Swiss Large White (SLW) pigs revealed 237 significantly associated variants ( $P < 1.24 \times 10^{-6}$ ). Fifteen out of 24 traits had at least one significantly associated variant (Table 1; Additional files 2 and 3).

The number of variants that exceeded the Bonferroni-corrected significance threshold was between 1 for GL, NIT, MAS and MTF, and 49 for ADFI (Table 1). The inflation factors of the stGWAS were between 0.85 for NT and 1.03 for SCH with an average value of  $0.95 \pm 0.04$  across all 24 stGWAS indicating that population stratification was properly considered.

The 237 associations were detected at 99 unique SNPs located at 11 QTL on SSC1, 5, 11, 15, 17 and 18. The two strongest associations were detected between the number of teats (NT) and a variant on SSC7 (MARC0038565 at 97,652,632 bp,  $P: 3.35 \times 10^{-35}$ ), and between lifetime daily weight gain (LDWG) and a variant on SSC1 (ASGA0008077 at 270,968,825 bp,  $P: 3.28 \times 10^{-28}$ ). 57 SNPs were significantly associated with more than one trait (two SNPs, ALGA0123414 and ASGA0008077, were associated with six traits) suggesting that pleiotropic effects are present and detectable in our dataset.

### Comparison of multi-trait studies using array genotypes

In order to exploit genetic correlations among the traits to detect pleiotropic loci, we conducted multivariate linear mixed model-based (mtGWAS) association testing and performed a multi-trait meta-analysis of the single-trait GWAS (metaGWAS<sup>1</sup>) for traits within the four trait categories. For unbiased investigation of differences between both methods, we considered between 1,074 and 2,689 individuals with complete phenotypic records for all traits within a trait category (Table 2) for the multi-trait analyses.

Both methods yielded similar results, but the metaGWAS<sup>1</sup> revealed more significantly associated variants as well as QTL (Table 2; Fig. 1; Additional files 4, 5 and 6). Across the four trait categories, the metaGWAS<sup>1</sup> revealed slightly more significant SNPs (Fig. 1A) resulting in a 18% smaller FDR than mtGWAS. The metaGWAS<sup>1</sup>

**Table 1** Traits with their abbreviations, full descriptions, corresponding trait group and descriptive statistics of the drEBV

Trait group	Acronym	Full name [unit of raw records]	N <sup>a</sup>	N <sup>b</sup>	Mean ± SD <sup>b</sup>	Min, Max <sup>b</sup>	h <sup>2</sup> ± SE <sup>a</sup>	FDR
Reproduction	PSP	Stillborn piglets [%]	2,886	2,610	0.4 ± 3.57	-9.46, 19.63	0.07 ± 0.01	-
	PUP	Underweight (< 1 kg) piglets [%]	2,886	2,554	0.88 ± 3.20	-13.83, 18.29	0.09 ± 0.01	2.50
	NBA	Piglets born alive [number]	2,886	2,697	2.46 ± 1.46	-4.04, 8.35	0.13 ± 0.01	-
	GL	Gestation length [days]	2,886	2,866	0.55 ± 0.88	-2.77, 5.18	0.29 ± 0.02	5.00
Production	MAS	Meat surface in <i>longissimus dorsi</i> [cm <sup>2</sup> ]	5,457	3,829	-3.71 ± 3.35	-15.17, 12.13	0.53 ± 0.02	5.00
	IMF	Intramuscular fat content in MAS [%]	5,422	3,335	0.23 ± 0.53	-1.99, 2.92	0.52 ± 0.01	1.67
	DRL	Drip loss [%]	5,515	3,595	0.86 ± 1.57	-6.12, 7.32	0.43 ± 0.02	-
	LMC	Lean meat content [%]	5,468	5,155	0.19 ± 1.87	-10.34, 8.53	0.46 ± 0.02	0.18
	PH24	pH 24 h postmortem in the loin	4,615	2,729	0 ± 0.05	-0.20, 0.20	0.06 ± 0.01	-
	ADFI	Average daily feed intake [kg/day]	5,467	5,109	0 ± 0.15	-0.67, 0.62	0.37 ± 0.01	0.10
	DWG	Daily weight gain on test [g/day]	5,467	5,008	10.72 ± 68.15	-256.11, 347.54	0.24 ± 0.02	0.15
	LDWG	Lifetime daily weight gain [g/day]	5,468	5,399	14.27 ± 31.06	-113.33, 168.03	0.23 ± 0.02	0.11
	MT	Loin muscle thickness [mm]	5,467	3,944	-2.25 ± 1.74	-8.07, 5.58	0.09 ± 0.01	5.00
	BFT	Back fat thickness [mm]	5,468	5,120	-0.87 ± 1.95	-8.96, 7.7	0.47 ± 0.01	0.25
Conformation	WSFH	Week to steep fetlock, hind legs <sup>c</sup>	5,434	2,274	0.06 ± 0.19	-0.82, 0.81	0.05 ± 0.01	-
	GAIT	Gait	5,434	2,025	0.39 ± 0.22	-0.83, 1.45	0.08 ± 0.01	-
	BFL	Bent to pre-bent curve of forelegs <sup>c</sup>	5,434	2,581	0.14 ± 0.14	-0.55, 0.87	0.41 ± 0.01	2.50
	NEIH	Narrowed (reduced) to enlarged inner hoof, hind claws <sup>c</sup>	5,430	2,564	0.18 ± 0.13	-0.45, 0.64	0.06 ± 0.01	-
	SCH	Sword- to chair-legged, hind legs <sup>c</sup>	5,434	2,361	0.17 ± 0.17	-0.65, 0.91	0.06 ± 0.01	-
	XOH	X- to O-legged (knocked-kneed to bow-legged), hind <sup>c</sup>	5,434	2,516	0.04 ± 0.12	-0.6, 0.67	0.04 ± 0.01	-
	CL	Carcass length [cm]	5,433	3,638	4.11 ± 2.66	-5.25, 14.64	0.67 ± 0.01	0.17
	NT	Number of teats (both sides)	5,433	5,419	2.19 ± 0.78	-0.72, 5.03	0.41 ± 0.01	0.25
	NIT	Number of inverted teats	5,427	2,721	-0.28 ± 0.21	-1, 0.96	0.09 ± 0.01	5.00
NUT	Number of underdeveloped teats	5,433	2,234	0.01 ± 0.09	-0.27, 0.45	0.09 ± 0.01	2.50	

The false discovery rate (FDR) and genomic heritability based on drEBV (h<sup>2</sup>) were based on array genotypes

<sup>a</sup> before filters

<sup>b</sup> after filters

<sup>c</sup> raw phenotypes for conformation traits are recorded on a scale from 1 to 7, where 4 represents the optimum value, but phenotypes for genetic evaluation and hence (deregressed) breeding values reflect the deviation from the optimum

**Table 2** Number of QTL in trait groups revealed by each of the methods

Group	Reproduction	Production	Conformation	All
Number of traits	4	10	10	24
Number of animals with records for all traits	2,553	1,927	2,689	1,074
Mean correlation between drEBV (± SD)	0.14 ± 0.18	0.32 ± 0.24	0.17 ± 0.19	0.10 ± 0.16
QTL—stGWAS <sup>a,c</sup> (in N traits)	2 (2)	5 (8)	4 (5)	11 (15)
QTL—mtGWAS <sup>a,c</sup>	0	5	3	4
QTL—metaGWAS <sup>a,c</sup>	0	7	3	4
QTL—metaGWAS <sup>b,c</sup>	0	6	3	7
QTL—metaGWAS <sup>b</sup> (imputed WGS)	0	6	3	6

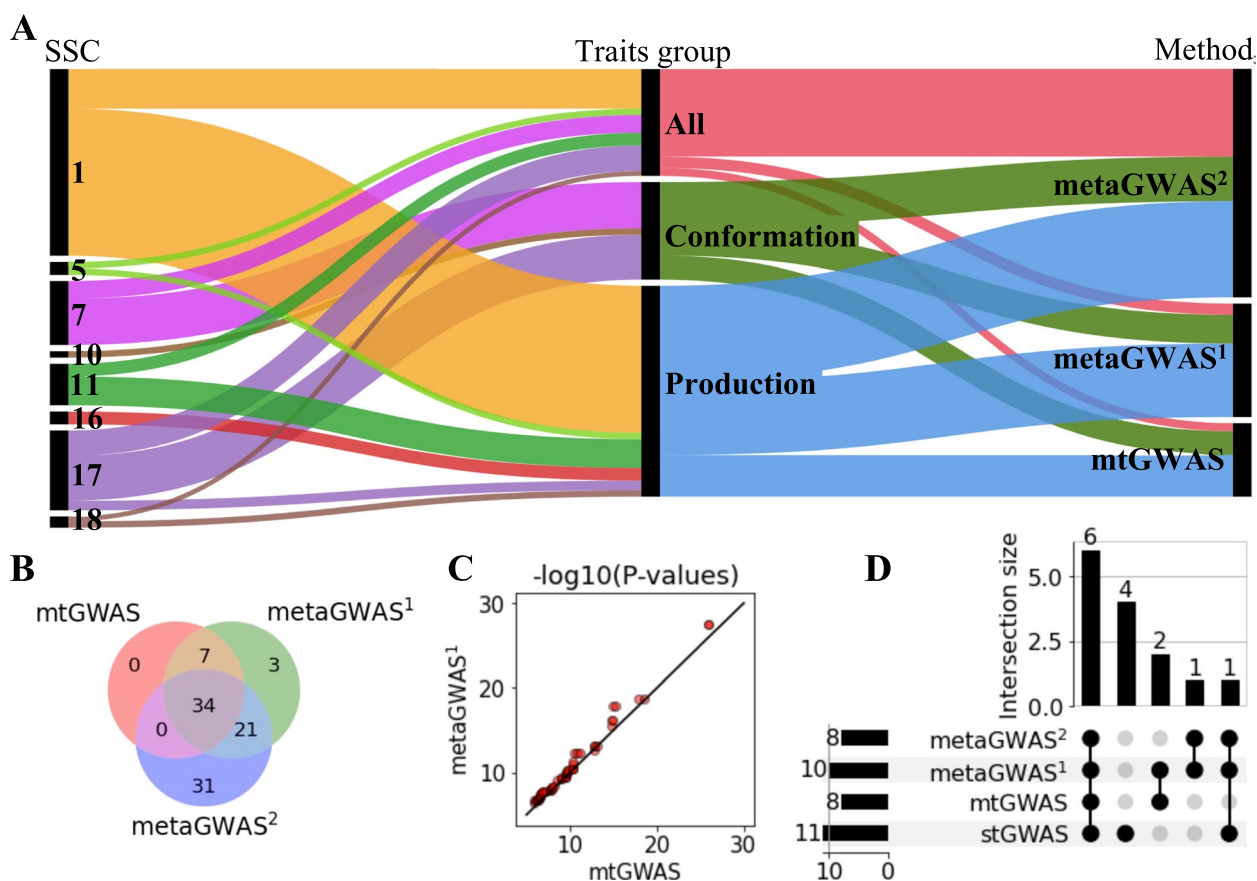
<sup>a</sup> performed in pigs with complete records in the trait groups

<sup>b</sup> performed in GWAS conducted in between 2,025 and 5,419 pigs

<sup>c</sup> based on array-derived genotypes at 40,382 SNPs

revealed 65 unique variants that were significantly associated with at least one trait category, of which 41 were also detected using mtGWAS. The mtGWAS revealed only associations that were also detected by metaGWAS<sup>1</sup>

(Fig. 1B). The *P*-values of lead SNPs were highly correlated ( $r=0.8$ ), but slightly lower (i.e., more significant) in the metaGWAS<sup>1</sup> than the mtGWAS (Fig. 1C).



**Fig. 1** Comparison of variants associated with 24 traits from 3 multi-trait GWAS methods. Multivariate (mtGWAS), meta-analyses with complete dataset (metaGWAS<sup>1</sup>), and meta-analyses including samples with missing trait records (metaGWAS<sup>2</sup>) were based on array-derived genotypes. **A** Proportions of significantly associated variants discovered across chromosomes, groups of traits and multi-trait methods. **B** Overlaps between the associated variants revealed by each of the methods. Sum across all four trait-groups. **C** Q-Q plot between  $-\log_{10}(P)$  of variants ( $N = 41$ ) associated in both mtGWAS and metaGWAS<sup>2</sup>. The line denotes a correlation of 1. **D** QTL detected by different methods across all trait groups

Neither of the multi-trait methods detected significantly associated SNP for the reproduction trait category. For the conformation trait category, mtGWAS and metaGWAS<sup>1</sup> revealed 15 and 18 associated SNPs, respectively. The associated SNPs defined three QTL on SSC7, 10, and 17. For the production trait category, the mtGWAS and metaGWAS<sup>1</sup> revealed 26 and 46 associations, respectively. Both methods revealed QTL at SSC1, 16, two QTL at SSC17, and SSC18. The metaGWAS<sup>1</sup> revealed two additional QTL at SSC1 and SSC11. When all 24 traits were combined for 1,074 pigs, the mtGWAS and metaGWAS<sup>1</sup> revealed only 5 and 7 associated SNPs, respectively. These SNPs spanned four QTL on SSC5, 7, 17, and 18.

Seven QTL detected by mtGWAS and metaGWAS<sup>1</sup>, were also detected by stGWAS (Fig. 1D). Both multi-trait methods detected three QTL on SSC11, 16 and 17 that were not detected in the stGWAS. Four QTL detected in the stGWAS were not revealed by either of the multi-trait

methods: these were QTL on SSC7, 11, and 15 that were slightly above the Bonferroni-corrected significance threshold for one trait (GL with  $P = 5.03 \times 10^{-7}$ , NIT with  $P = 1.06 \times 10^{-6}$ , PUP with  $P = 2.66 \times 10^{-7}$ , respectively), and one QTL on SSC10 that was associated with MT ( $P = 5.60 \times 10^{-7}$ ) and MES ( $P = 8.44 \times 10^{-7}$ ). From the seven QTL detected by both multi-trait and single-trait methods, six were associated with more than one trait in stGWAS. For the six pleiotropic QTL, at least one single trait analyses revealed more associated variants, and smaller  $P$ -value of the top SNP, than the mtGWAS or metaGWAS<sup>1</sup> (Additional file 7).

Using summary statistics from stGWAS for a multi-trait metaGWAS facilitates including data from animals with partially missing phenotypes. In order to maximize the power to identify trait-associated pleiotropic variants, we reran the metaGWAS using summary statistics from stGWAS with all available animals per trait (Additional file 8), denoted as metaGWAS<sup>2</sup>.

Compared to the previous metaGWAS<sup>1</sup> that was based on fewer individuals that had phenotypes for all traits, the number of SNPs exceeding the Bonferroni-corrected significance threshold ( $P < 1.24 \times 10^{-6}$ ) increased by 10, 14 and 48 to 28, 60 and 55 in the conformation, production, and all groups, respectively (Additional file 9). No significant markers were detected for the reproduction trait category. Across the four trait groups, the metaGWAS<sup>2</sup> revealed 86 variants, from which 34 were also detected by both other methods, while 21 were detected only by the metaGWAS<sup>1</sup> (Fig. 1B). Including all available samples into the metaGWAS<sup>2</sup> did not reveal any additional QTL (Fig. 2A).

#### Reference panel and imputation of array genotypes to sequence level

Whole-genome sequence-variant genotypes from 421 pigs were used to impute the medium-density genotypes of 5,753 pigs to the whole-genome sequence level. The principal components analysis (PCA; Additional file 10) of a GRM built from 16 million biallelic SNP genotypes confirmed that the sequenced reference panel is representative for our GWAS cohort. Five-fold cross-validation indicated high accuracy of imputation (Additional file 11) with values of 0.92, 0.97 and 0.95 for the squared Pearson's correlation ( $R^2$ ) between true and imputed allele dosages, cross-validated proportion of correctly imputed genotypes (concordance ratio – CR) and model-based accuracies from Beagle5.2 (Beagle DR2), respectively. Although the model-based estimate from Beagle was highly correlated with the Pearson  $R^2$  (0.81), the Beagle DR2 values were consistently higher.

#### Imputed sequence-based association studies

The stGWAS between 24 traits and the 16,051,635 imputed variants revealed 45,288 variants exceeding the Bonferroni-corrected significance threshold in 7 QTL regions which largely agreed with the results from the array-based GWAS (Additional file 12). Two QTL on SSC5 and SSC12 were detected in stGWAS for BFT and NT, respectively, while the other five QTL were associated with multiple traits. The sequence-based GWAS revealed one additional QTL (on SSC12) but did not detect significant association at 5 previously detected QTL likely due to a more stringent Bonferroni-corrected significance threshold resulting from a 350-fold denser marker panel.

A metaGWAS using the summary statistics from stGWAS between the 16,051,635 imputed whole-genome sequence variants and 24 traits using all animals revealed six QTL on SSC1, 5, 7, 17 and 18 (Fig. 2, Table 3) with a total of 9,774 variants exceeding the Bonferroni-corrected significance threshold. When the six lead imputed

SNPs were fitted as fixed effects in stGWAS, the peaks in the metaGWAS Manhattan plot disappeared (Additional file 13), indicating that the lead SNP accounted for the QTL variance. Four QTL revealed by metaGWAS were significantly associated in stGWAS exclusively with production traits, whereas two QTL (on SSC7 and SSC17) were associated with traits from both the production and conformation categories (Fig. 2C). The total trait variance explained per QTL ranged from 0.18 to 11.02% (Fig. 2C).

#### QTL 1 with lead SNP 1\_159637589

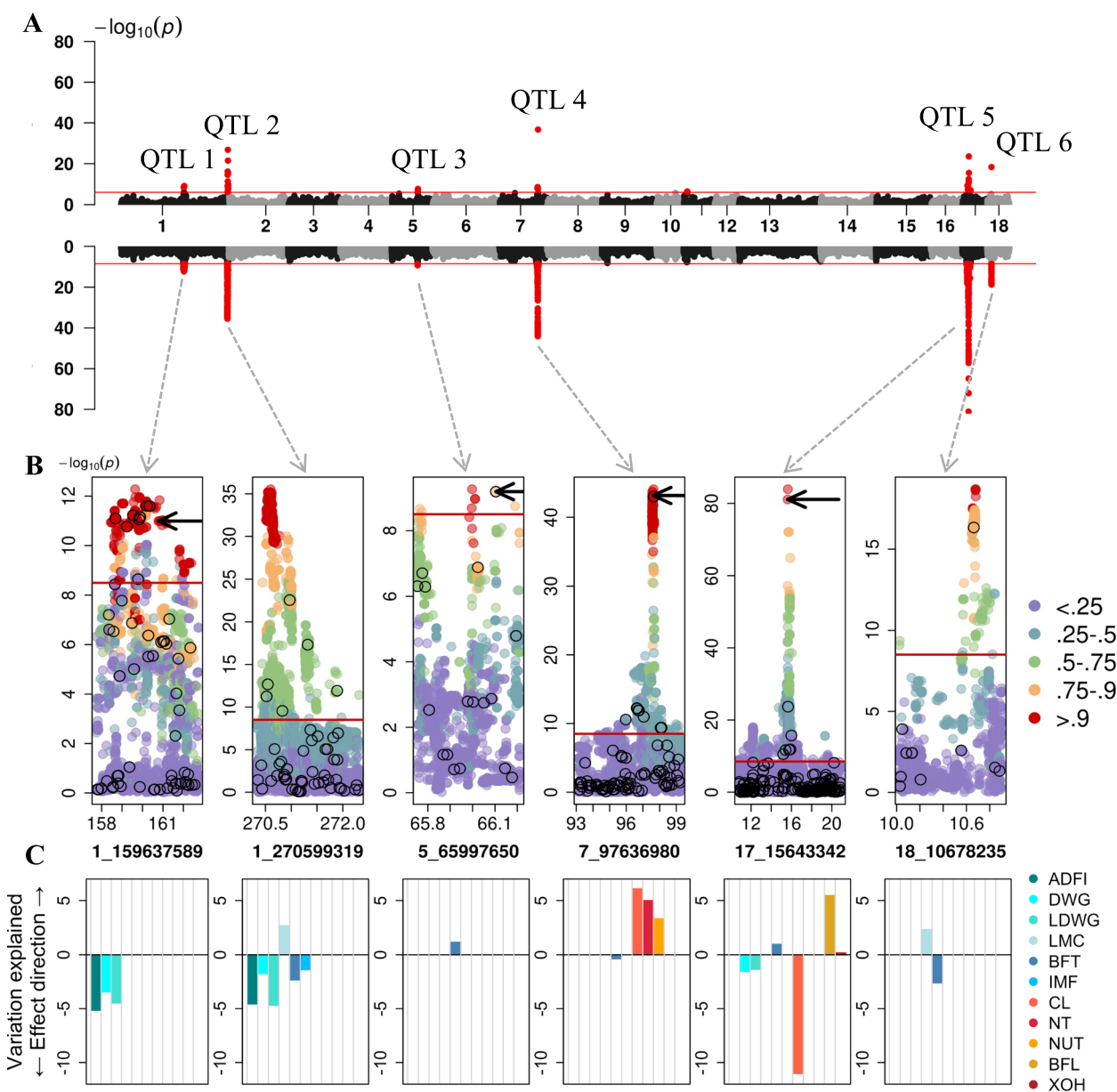
A QTL on SSC1 was between 157.73 and 162.74 Mb and encompassed 34,312 imputed sequence variants including 1,157 that were significantly associated in the sequence-based metaGWAS ( $-\log_{10}(P) > 8.5$ ). The QTL was associated with ADFI, DWG, and LDWG, and explained 5.16, 3.45, and 4.49% of the trait variance, respectively. The lead SNP at this QTL was an imputed intergenic sequence variant (rs692827816) located at 159,637,589 bp ( $P = 5.26 \times 10^{-13}$ ). rs692827816 was in high LD ( $R^2 > 0.90$ ) with 935 variants that had similar  $P$  values. One of the variants in high LD was a missense variant (rs81219178) at 160,773,437 bp within the *MC4R* gene, which has previously been proposed as candidate causative variant for growth and fatness traits [36]. rs81219178 segregated at MAF of 0.36 in the SLW population and was imputed from the reference panel with high accuracy (DR2 = 1).

#### QTL 2 with lead SNP 1\_270599319

Another QTL on SSC1 was between 270.33 and 272.32 Mb and encompassed 21,683 imputed sequence variants including 3,765 that were significant in the sequence-based metaGWAS (Additional file 14). The QTL was associated with ADFI, DWG, LDWG, LMC, BFT, and IMF, and explained 4.57, 1.77, 4.69, 2.70, 2.35, and 1.40% of the trait variance, respectively. The lead variant was an imputed insertion polymorphism (C > CT) located at 270,599,319 bp ( $P = 2.85 \times 10^{-36}$ ), approximately 13 kb downstream from *ASS1* and 53 kb upstream from *FUBP3*. The lead variant was in high LD ( $R^2 > 0.90$ ) with 747 variants. The MAF of the lead variant was 0.14 and its genotypes were imputed from the reference panel with high accuracy (DR2 = 0.98).

#### QTL 3 with lead SNP 5\_65997650

A QTL on SSC5 was located between 65.75 and 66.21 Mb and encompassed 5,065 imputed sequence variants, from which eight were significantly associated with BFT (Additional file 15). The QTL explained 1.16% of the phenotypic variance of BFT. The lead SNP was an imputed sequence variant located in an intergenic region at 65,997,650 ( $P = 5.40 \times 10^{-10}$ ;



**Fig. 2** Fine mapping of six QTL detected by metaGWAS. **A** Manhattan plots from array (upper) and imputed sequence (bottom) variants in metaGWAS with 24 traits. **B** Linkage disequilibrium between the lead SNPs and all other variants. Black circles mark array SNPs, arrows point to previously proposed causal variants. The red line indicates the genome-wide Bonferroni-corrected significance threshold. **C** Variation explained (in % of the drEBV variance) by alternative alleles of the lead SNPs in the single traits. Production traits are in blue scale (ADFI—Average daily feed intake; DWG—Daily weight gain on test; LDWG—Lifetime daily weight gain; LMC—Lean meat content; BFT—Back fat thickness; IMF—Intramuscular fat content in loin), and conformation traits in red scale (CL—Carcass length; NT—Number of teats—both sides; NUT—Number of underdeveloped teats; BFL—Bent to pre-bent, front legs; XOH—X- to O-legged)

rs346219461) which was in high LD ( $R^2 > 0.90$ ) with seven other variants. rs346219461 was 7.60 kb downstream the fibroblast growth factor 6 (*FGF6*) encoding gene and it had MAF of 0.41, and it was imputed from the reference panel with high accuracy ( $DR2 = 0.99$ ).

The rs346219461 was in LD ( $R^2 = 0.82$ ) with a non-coding variant (rs80985094 at 66,103,958 bp,  $P = 6.41 \times 10^{-10}$ ) in the third intron of *CCND2*, that was previously proposed as putative causal variant for backfat thickness [35].

**Table 3** Imputed lead variants in pleiotropic QTL revealed by a multi-trait meta-analyses

QTL	SSC	Start (bp)	Stop (bp)	Lead SNP	P-value	MAF	Ref   Alt	#traits	Candidate gene(s)
1	1	157,732,172	162,736,434	1_159637589	$5.26 \times 10^{-13}$	0.31	G <sup>a</sup>   C	3	<i>MC4R</i>
2	1	270,331,332	272,315,496	1_270599319	$2.85 \times 10^{-36}$	0.14	C   CT <sup>a</sup>	6	na
3	5	65,753,637	66,210,538	5_65997650	$5.40 \times 10^{-10}$	0.41	G   A <sup>a</sup>	1	<i>CCND2</i>
4	7	93,119,538	99,261,691	7_97636980	$8.09 \times 10^{-45}$	0.34	A <sup>a</sup>   C	4	<i>VRTN</i> ; <i>ABCD4</i>
5	17	10,801,927	20,928,904	17_15643342	$1.19 \times 10^{-84}$	0.20	C <sup>a</sup>   T	5	<i>BMP2</i>
6	18	10,024,756	10,904,242	18_10678235 18_10678293	$2.00 \times 10^{-19}$	0.32	C   A <sup>a</sup> A   G <sup>a</sup>	2	na

<sup>a</sup> Minor allele in the genotyped dataset; Reference alleles were determined according to the Sscrofa11.1 genome assembly

#### QTL 4 with lead SNP 7\_97636980

A pleiotropic QTL on SSC7 was between 93.12 and 99.26 Mb and it encompassed 31,013 imputed sequence variants including 2,341 that were significant in the sequence-based metaGWAS. The QTL was associated with BFT from the production group (0.37% variance explained), and CL, NT, and NUT from the conformation group, where it explained 6.11, 5.00, and 3.32% of the trait variance, respectively. The lead SNP was an imputed variant (rs333375257 at 97,636,980 bp,  $P=8.09 \times 10^{-45}$ ) located 12.7 kb downstream *VRTN*. The rs333375257 had MAF of 0.34 and was imputed from the reference panel with high accuracy (DR2=0.99). The rs333375257 was in high LD ( $R^2>0.90$ ) with 424 sequence variants. A previously described candidate causal variant (rs709317845 at 97,614,602 bp,  $P=6.71 \times 10^{-44}$ ) for the number of thoracic vertebrae [37] was in LD ( $R^2>0.99$ ) with the rs333375257. In addition, 334 significant variants in the ATP binding cassette subfamily D member 4 (*ABCD4*) gene were detected. This gene was proposed to impact NT in a Duroc population [38] with top SNP rs692640845 at position 97,568,284. In our study, the rs692640845 was highly significantly associated ( $P=1.49 \times 10^{-42}$ ) with CL, NT and NUT, and in almost complete LD ( $R^2=0.98$ ) with the lead SNP.

#### QTL 5 with lead SNP 17\_15643342

A pleiotropic QTL on SSC17 encompassed 82,132 imputed sequence variants including 2,112 that were significantly associated residing between 10.80 and 20.93 Mb. The QTL was associated with DWG, LDWG, and BFT from the production group (explaining 1.56, 1.36, and 0.97% of the trait variance, respectively), and with BL, BFL, and XOH from the conformation group (explaining 11.02, 5.48, 0.18% of the trait variance, respectively). The strongest association was from an imputed sequence variant (rs342044514 at 15,643,342,  $P=1.19 \times 10^{-84}$ ) in an intergenic region 106 kb upstream the *BMP2* gene. The variant had MAF of 0.2 and it was imputed from the reference panel with high accuracy (DR2=0.97). The lead SNP was in high LD ( $R^2>0.90$ )

with two other variants. One of them was a previously proposed candidate causative variant for carcass length [34] (rs320706814 at 15,626,425,  $P=8 \times 10^{-82}$ ) in an intergenic region upstream of the *BMP2* gene, 17 kb away from the lead SNP.

#### QTL 6 with lead SNP 18\_10678235

A QTL on SSC18 between 10.03 and 10.90 Mb encompassed 5,790 imputed sequence variants including 408 that were significant. The QTL was associated with LMC and BFT, explaining 2.34 and 2.62% of the phenotypic variance, respectively. The QTL had two lead SNPs ( $P=2 \times 10^{-19}$ ) in complete LD, which were imputed sequence variants located at 10,678,235 bp (rs338817164) and 10,678,293 bp (rs334203353). Both variants had MAF of 0.32. The imputation accuracy was 0.80. They were in high LD ( $R^2>0.90$ ) with other 11 intergenic variants.

## Discussion

Single- and multi-trait genome-wide association studies involving array-derived and imputed sequence variant genotypes from 5,753 SLW pigs enabled us to investigate the genetic architecture of 24 complex traits from three trait groups. The response variables for the association tests were deregressed breeding values because the genotyped pigs had progeny-derived phenotypes. Progeny-derived phenotypes have been frequently used to perform association studies in animals that lack own performance records for the traits of interest. To avoid false-positive associations arising from the accumulation of family information in the progeny-derived phenotypes [39], we used the deregressed breeding values and weighed them according to equivalent relatives' contributions.

The single-trait association analyses revealed 26 trait  $\times$  QTL associations at eleven QTL of which seven were associated with at least two traits. Exploiting genetic correlations among the traits in a multi-trait framework revealed association for six out of the seven pleiotropic QTL detected in the stGWAS. Despite considering up to

10 phenotypes in the mtGWAS and up to 24 phenotypes in the metaGWAS, the multivariate methods applied in our study revealed three QTL, that were not detected by the single-trait association studies. The multi-trait methods did not reveal association at four QTL that were revealed by stGWAS. These QTL had low  $P$ -values, and perhaps because of higher penalty for multiple testing their effects were too small to be detected in our medium-sized cohort. The phenomenon that associations detected by stGWAS might disappear in multi-trait analyses has been reported earlier [12, 40].

The mtGWAS and metaGWAS<sup>1</sup> detected largely the same associated SNPs for almost all trait-groups. However, the metaGWAS<sup>1</sup> revealed more associated SNPs, and lower  $P$ -values for the lead SNPs. Combining all 24 traits in the mtGWAS revealed five associated SNPs, and the metaGWAS<sup>1</sup> conducted with the same individuals detected association of seven SNPs. Multivariate linear mixed models may suffer from over-parametrisation and loss of power when more than ten traits are considered [41]. The low number of detected QTL might also result from low genetic correlations between the 24 traits, or from a small sample size ( $N=1,074$  pigs with non-missing records for all 24 traits).

The metaGWAS approach enabled us to establish a larger sample size by considering summary statistics from stGWAS that were conducted with a various number of individuals (i.e., some pigs had missing records for some of the traits). In this setting, the number of associated SNPs detected by the metaGWAS<sup>2</sup> increased to 55 (tenfold higher than before). According to Bolormaa et al. [1], in situations where each stGWAS is performed on partially different set of individuals, the metaGWAS approach still appropriately considers variances and covariances among the  $t$ -values. It is worth mentioning that there are also frameworks that enable considering samples with partially missing phenotypes in multi-trait GWAS [42–44], but these avenues were not explored in the current study.

Our comparisons between GWAS approaches considered microarray-derived SNP genotypes. The imputation of array-derived genotypes up to the sequence level provides more statistical power to identify associated loci because causal variants are in the data and directly tested for association with traits of interest [24, 26]. The pigs in our study were genotyped at 44,733 SNPs. No samples were genotyped with denser (e.g., 600 K) arrays precluding the stepwise imputation of genotypes up to the sequence level. The imputation of sequence variant genotypes into sparsely genotyped samples may be inaccurate particularly for rare alleles [16]. However, imputation in our mapping cohort with a haplotype reference panel of 421 sequenced animals

was accurate. This is likely because the haplotype reference panel mainly contained animals from the target breed.

The imputed sequence variants were used in metaGWAS to fine-map QTL and prioritise candidate causal variants. The top associated variants in four QTL were variants that had been previously identified as candidate causal variants [34–38], indicating that they might also underpin these QTL in the SLW breed. However, none of the proposed candidate causal variants was the top variant in our association studies possibly indicating sampling bias [45], presence of multiple trait-associated variants in linkage disequilibrium [46, 47], or that the top variants were inaccurately imputed [48]. It is also possible that the previously reported candidate causal variants are not causal. Further in-depth functional investigations are required to determine and validate the molecular mechanisms underpinning the QTL identified in our study.

Our meta-analyses approach using imputed sequence variants revealed six QTL of which five were associated with multiple traits. Several porcine pleiotropic loci are underpinned by heterozygous loss-of-function alleles that may have fatal consequences in the homozygous state [49–51]. Pleiotropic QTL have also been described in pigs for highly correlated traits [31, 52]. The meta-analyses of 24 traits conducted in our study revealed six QTL, that were significantly associated in one to six stGWAS. These results emphasize the importance of the single-trait analyses for dissecting the pleiotropic effects. A QTL on SSC17 which is associated with traits belonging to distinct trait categories. This QTL is associated with carcass length ( $P=1 \times 10^{-62}$ ) and daily weight gain ( $P=6 \times 10^{-13}$ ). These two traits are barely correlated with each other ( $r=0.02$ – $0.05$ ). Another pleiotropic QTL on SSC1 at ~270 Mb was associated with six traits ADFI, DWG, LDWG, LMC, BFT, and IMF, that were moderately to highly correlated (mean  $r \pm SD=0.37 \pm 0.27$ ). This chromosomal region harbours QTL for backfat thickness and feed efficiency-related traits in other pig populations [33, 53]. However, candidate causal variants underpinning this QTL had not been proposed so far. The lead SNP in our study was at position 270,599,319 bp, which is 13 kb downstream from the *ASS1* gene. Expression of *ASS1* has been associated with digestive tract development, cell adhesion, response to lipopolysaccharide, and arginine and proline metabolism in pigs [54, 55]. Considering its putative role in energy metabolism, we propose *ASS1* as a positional and functional candidate gene for a pleiotropic QTL at ~270 Mb. Further functional annotations of the trait-associated variants in the non-coding regions might help elucidating the genetic mechanism underpinning this pleiotropic QTL.



## Conclusions

Multi-trait associations analyses provide strength of evidence for the presence or absence of a QTL segregating in populations. Here, we compared the multivariate linear model with meta-analyses of single-trait summary statistics using real data. Both approaches performed similarly in correlated groups of traits with complete datasets. The ability of meta-analyses to include different sets of individuals and unrestricted number of traits promoted the detection power. Thus, we recommend using the meta-analyses for getting overview of pleiotropic QTL in cohorts with more than 10 traits. For analyses of reduced and correlated groups of traits, the choice of the method seems to provide indifferent results. Yet, we stress the importance of the single-trait analyses for accurate interpretation of the pleiotropic effects and for the assignment of the affected traits.

The reference-guided imputation to whole-genome sequence level assigned genotypes to 22 million variants with high accuracy. Putative causal variants found in literature were among the top variants in the fine-mapped QTL. Our analyses provide overview of the QTL affecting economically important traits in Swiss Large White and might serve as catalogue for future research examining the causal variants for complex traits.

## Methods

### Animals and phenotypes

Deregressed estimated breeding values (drEBV) with their corresponding degrees of determination ( $r^2_{\text{drEBV}}$ ) and weights ( $w_{\text{drEBV}}$ ) for 24 traits were provided by the Swiss breeding company SUISAG for 5,753 pigs of the SLW breed. Breeding values were estimated using BLUP multiple trait animal models neglecting genomic information and subsequently deregressed according to Garrić et al. [56]. For all our analyses, we considered drEBV which had  $r^2_{\text{drEBV}} > 0.3$  and were within five standard deviations from the mean values. We considered only traits for which at least 2,000 genotyped animals had records. The final number of animals with phenotypes was between 2,025 for gait (GAIT) and 5,419 for number of teats (NT). Up to 37% of the pigs had missing records for at least one trait. The traits were assembled in four trait groups (Table 1): reproduction (4 traits), conformation (10 traits), production (10 traits) and all (24 traits).

### Genotypes

Microarray-derived genotypes were available for 17,006 pigs from different breeds. The genotypes were obtained with five SNP panels with medium density. There were 2,970 pigs genotyped with the Illumina PorcineSNP60 Bead Chip comprising either 62,163 (v. 1) or 61,565 (v.

2) SNPs; 13,342 pigs were genotyped using customized 60 K Bead Chips comprising either 62,549 (v. 1) or 77,119 (v. 2) SNPs; and 546 pigs had genotypes at 68,528 SNPs obtained with the GeneSeek Genomic Profiler (GGP) Porcine 80 K array.

We used PLINK (v. 1.9; [57]) to merge the genotypes from the five SNP panels based on the physical positions of the SNPs according to the Sscrofa11.1 assembly [58] of the porcine genome. Then, we performed a quality control for the combined dataset. We retained unique autosomal SNPs that did not deviate from Hardy–Weinberg proportions ( $P < 0.00001$ ), had SNP- and individual-level genotyping rates above 80%, and minor allele frequency (MAF) greater than 0.5%. Finally, sporadically missing genotypes for the resulting 44,733 variants were imputed for 14,292 animals using Beagle (v. 5.0; [59]).

For the array-based GWAS and for the comparison of the multi-trait GWAS methods, we considered 40,382 SNPs that had MAF greater than 5% in 5,753 animals of the SLW breed.

### Genomic heritability

We used the Fisher-scoring algorithm implemented in the GREML module of GCTA (v. 1.92.1; [60]) to estimate variance components while considering the inversed weight of drEBV ( $w_{\text{drEBV}}$ ). The genomic relationship matrix was built for 14,292 individuals with 44,733 SNPs, but only up to 5,753 SLW animals were used to estimate genomic heritability.

### Imputation to whole-genome sequence level

Whole-genome sequence (WGS) data were available for 421 SLW pigs that had been sequenced at an average read depth of 7.1x, ranging between  $2.35 \times$  and  $37.5 \times$ . This panel also included 32 key ancestors of the genotyped SLW pigs that explained a large fraction of the genetic diversity of the current breeding population [19].

Raw sequence data were trimmed and pruned for low-quality bases and reads with default parameter settings of the fastp software (v. 0.20.0; [61]) and subsequently mapped to the Sscrofa11.1 reference genome using the mem-algorithm of the BWA software (v. 0.7.17; [62]). Duplicated reads were marked with the Picard tools software suite (v. 2.25.2; [63]), followed by sorting the alignments by coordinates with Sambamba tool (v. 0.6.6; [64]). The read depth at each genomic position was calculated with the mosdepth software (v. 0.2.2; [65]), considering reads with mapping quality  $> 10$ . Variant calling and filtering followed Genome Analysis Toolkit (GATK—v. 4.1.0; [66]) best practice recommendations. Base quality scores were adjusted using the BaseRecalibrator module while considering 63,881,592 unique positions from the porcine dbSNP (v. 150) as known variants. The discovery,

genotyping, and filtering of SNPs and INDELS in the 421 pigs was done using the HaplotypeCaller, GenomicsDBImport, GenotypeGVCFs and VariantFiltration modules of the GATK.

The filtered WGS dataset containing 421 pigs and 22,018,148 variants (18,839,630 SNPs and 3,178,518 INDELS) with MAF greater than 0.01 was used for the imputation of sequence variant genotypes into the array dataset. The reference panel was pre-phased using SHAPEIT4 (v. 4.2; [67]) using the `-sequencing` parameter. The target array dataset was pre-phased with SHAPEIT4 using the phased sequence data as the reference. Sequence variant genotypes were imputed with Beagle (v. 5.2; [59]) with an effective population size of 50. The effective population size was estimated using SneP [68].

The accuracy of imputation was assessed empirically by five-fold cross-validation in the 421 animals as follows: 40 animals which were sequenced at high coverage (> 10x), were used as target panel. The remaining 381 animals served as the reference panel. The SNP density in the target panel was reduced to 44,733 SNP chip genotypes and subsequently imputed to the sequence level based on 381 reference animals as described above. The imputed and actual genotypes of the target samples were compared to derive concordance ratio (CR; proportion of correctly imputed genotypes) and squared correlation ( $R^2$ ) between imputed and true genotypes.

Relationship between the animals included in the reference panel and the target animals was assessed with a principal components (PC) analysis. First, a genomic relationship matrix (GRM) was built among 14,629 pigs that had 16,387,582 (partially) imputed biallelic SNPs with MAF > 5% (421 reference animals and 14,208 animals with imputed sequence variant genotypes) using GCTA [60]. Then, the first 10 principal components (PC) of the GRM were obtained with PLINK (v. 1.9).

Post-imputation quality control excluded SNPs with MAF < 5%, model-based accuracy of imputation (Beagle DR2) < 0.6, and deviations from Hardy–Weinberg proportions ( $P < 10^{-8}$ ), resulting in a total of 16,051,635 biallelic variants (13,773,179 SNPs and 2,278,456 INDELS) which were used for association analyses and the fine-mapping of QTL in 5,753 SLW pigs.

#### Single-trait genome-wide association analysis (stGWAS)

Single marker-based GWAS were conducted between 24 traits (see Table 1 for more information about the traits and the number of individuals with records) and either 40,382 array-derived or 16,051,635 imputed sequence variant genotypes using the mixed model-based approach implemented in the GEMMA software (v.0.98.5; [69]).

The linear mixed model fitted to the data was in the following form:  $y = W\alpha + x\beta + u + \epsilon$ , where  $y$  is a

vector of phenotypes of  $n$  animals;  $W$  is a vector of ones;  $\alpha$  is a vector of corresponding coefficients;  $x$  is a vector of marker genotypes, coded as 0, 1 and 2 for genotype  $A_1A_1$ ,  $A_1A_2$  and  $A_2A_2$ ;  $\beta$  is the effect of the  $A_2$  allele;  $u \sim MVN_n(0, G\sigma_a^2)$  is a random polygenetic effect with  $G$  representing the  $n \times n$ -dimensional genomic relationship matrix (GRM);  $\sigma_a^2$  is the additive genetic variance;  $\epsilon \sim MVN_n(0, I\sigma_e^2)$  is a vector of errors, with  $I$  representing an identity matrix; and  $\sigma_e^2$  is the residual variance.  $MVN_n$  denotes the  $n$ -dimensional multivariate normal distribution.

The centred GRM was calculated with GEMMA (`-nk 1`) using either array-based or imputed sequence variant genotypes. The  $P$ -value of each SNP was estimated by the score test implemented in GEMMA (`-lmm 3`).

The stGWAS was run with either all available individuals or considering only preselected individuals that had non-missing phenotypes within a group of traits. The markers were separately filtered for MAF > 5%. Thus, in the latter run, the stGWAS for the reproduction traits included 41,242 variants typed in 2,553 samples; the stGWAS for the production traits included 40,557 variants typed in 2,689 samples; the stGWAS for the conformation traits included 41,168 variants typed in 1,927 samples, and the stGWAS for the 24 traits together included 41,152 variants typed in 1,074 samples with non-missing records.

#### Multi-trait genome-wide association analyses (mtGWAS)

Multi-trait association tests (mtGWAS) were conducted using a multivariate mixed model-based approach implemented in the GEMMA software (v.0.98.5; [41]). The multivariate linear mixed model was parameterised similar to the stGWAS model ( $y = W\alpha + x\beta + u + \epsilon$ ), except that  $y, \alpha, u, \epsilon$  are matrices with  $d$  (number of traits) columns, and  $\beta$  is a vector with length  $d$ .  $\sigma_a^2$  and  $\sigma_e^2$  are  $d \times d$  symmetric matrices of genetic and environmental variance components, respectively. Because multivariate association testing as implemented in GEMMA requires phenotype data for all individuals and traits, we considered only 2,553, 1,927, 2,689 and 1,074 individuals, respectively, for the reproduction (4 traits), conformation (10 traits), production (10 traits), and all-trait (24 traits) mtGWAS. The GRM, used during the mtGWAS, was the one from the stGWAS.

#### Meta-analyses multi-trait genome-wide association (metaGWAS)

A multi-trait meta-analysis (metaGWAS) was conducted with the summary statistics from stGWAS as suggested by Bolormaa et al. [1]. Briefly, the  $t$ -values for each marker-trait combination were calculated based on the allele substitution effect and corresponding standard

error obtained from the stGWAS. The multi-trait  $\chi^2$  statistic was subsequently calculated as  $\chi_{df=d}^2 = t_j^d V^{-1} t_j$ , where  $t$  is a  $j \times d$  matrix of signed  $t$ -values at the  $j^{\text{th}}$  marker across  $d$  traits, and  $V^{-1}$  is the inversed  $d \times d$  variance–covariance matrix.

$P$ -values for the  $j$  markers were calculated with `pchisq` function with  $d-1$  degrees of freedom, as implemented in R. We carried out the meta-analyses with the 24 traits classified into the same four trait categories as in mtGWAS (reproduction, production, conformation, and all). First, to enable unbiased comparison between the metaGWAS and the mtGWAS results, we considered summary statistics obtained from stGWAS based on individuals with complete records within the respective trait-group. Second, to increase the power of the association tests through maximizing the volume of entering information, hence exploiting the benefit of the metaGWAS approach, we used the stGWAS summary statistics based on all available individuals for the trait (the first run). For clarity we denote the first and second meta-analyses as metaGWAS<sup>1</sup> and metaGWAS<sup>2</sup>, respectively. The latter approach was repeated for the fine-mapping of the QTL with imputed sequence variant genotypes.

### Comparison of the association methods

We used a 5% Bonferroni-corrected significance threshold ( $1.24 \times 10^{-6}$  and  $3.11 \times 10^{-9}$  for array and imputed sequence variant genotypes, respectively) to consider multiple testing. Genomic inflation factors were calculated to compare the distributions of the expected and observed test statistics.

The statistical power was assessed using false discovery rate (FDR). Following Bolormaa et al. [70], the FDR was calculated as  $\frac{P*(1-\frac{A}{T})}{\frac{A}{T}*(1-P)}$ , where  $P$  is the significance threshold (e.g.,  $1.24 \times 10^{-6}$  or  $3.11 \times 10^{-9}$ ),  $A$  is the number of significant variants and  $T$  is the total number of variants tested.

### Fine mapping of detected QTL

We defined QTL as a region of 1 Mb non-overlapping windows, containing at least one significantly associated marker. The marker with the smallest  $P$ -value within a QTL was defined as lead variant. Linkage disequilibrium (LD) between the lead variant and all other variants was calculated with the PLINK (v. 1.9) `-r2` command. Variants within QTL were annotated with Ensembl's Variant Effect Predictor (VEP; [71]) tool using local cache files from the Ensembl (release 104) annotation of the porcine genome. The deleteriousness of missense variants was predicted with the SIFT scoring algorithm [72] implemented in VEP.

The proportion of drEBV variance explained by a QTL was estimated with  $\frac{2p(1-p)\beta^2}{\sigma^2}$ , where  $p$  is the frequency of the minor allele of the lead SNP and  $\sigma^2$  is the drEBV variance;  $\beta$  is the regression coefficient of the lead SNP. To avoid overestimating the variance explained by a lead variant, we followed the approach described in Kadri et al. [73] and estimated the regression coefficients jointly for all QTL from the stGWAS, i.e., the lead variants of those QTL, that were significantly associated in the stGWAS, were fitted as covariates.

### Abbreviations

EBV	Estimated breeding value
drEBV	Deregressed estimated breeding value
GWAS	Genome-wide association study
INDEL	Insertions and deletions
LD	Linkage disequilibrium
MAF	Minor allele frequency
metaGWAS	Meta-analyses GWAS
mtGWAS	Multi-trait GWAS
PCA	Principal component analysis
QTL	Quantitative trait loci
$r^2_{\text{drEBV}}$	Reliability of the drEBV
SD	Standard deviation
SLW	Swiss Large White
SNP	Single nucleotide polymorphism
SSC	<i>Sus scrofa</i> Chromosome
stGWAS	Single-trait GWAS
$W_{\text{drEBV}}$	Weight of the drEBV
WGS	Whole-genome sequencing

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-023-09295-4>.

**Additional file 1.** Correlations between the 24 traits. In the upper and lower triangles are correlations between signed  $t$  values, and between deregressed breeding values, respectively. Shades of purple and orange indicate positive and negative correlation coefficients, respectively. The three rectangulars are defining reproduction, production, and conformation groups of traits.

**Additional file 2.** Manhattan plots of single-trait GWAS based on array genotypes. Each page contains 4, 10 and 10 plots for traits from reproduction, conformation and production groups, respectively. Blue suggestive line is at 5.9.

**Additional file 3.** Results of single-trait GWAS based on array genotypes. Table contains numbered QTL regions, chromosome, start and stop positions of the QTL, lowest  $P$ -value, top SNP with MAF, number of significantly associated variants within the QTL and the single-trait abbreviation.

**Additional file 4.** Manhattan plots of multi-trait GWAS (upper) and meta-analyses GWAS (bottom) based on array genotypes. Suggestive line is at 5.9.

**Additional file 5.** Results of multi-trait GWAS based on array genotypes. Table contains numbered QTL regions, chromosome, start and stop positions of the QTL, lowest  $P$ -value, top SNP with MAF, number of significantly associated variants within the QTL and the trait-group.

**Additional file 6.** Results of meta-analyses GWAS based on array genotypes and complete dataset. Table contains numbered QTL regions, chromosome, start and stop positions of the QTL, lowest  $P$ -value, top SNP with MAF, number of significantly associated variants within the QTL and the trait-group.

**Additional file 7.** Comparison of pleiotropic QTL across methods based on array genotypes. Summary table of *P*-values and number of significantly associated variants within the 6 pleiotropic resulting from single-trait GWAS, meta-analyses GWAS with complete datasets and multi-trait GWAS.

**Additional file 8.** Results of meta-analyses GWAS based on array genotypes and dataset including missing phenotypic records. Table contains numbered QTL regions, chromosome, start and stop positions of the QTL, lowest *P*-value, top SNP with MAF, number of significantly associated variants within the QTL and the trait-group.

**Additional file 9.** Number of significantly associated pleiotropic variants in the meta-analyses GWAS within groups using all possible samples and stGWAS of individual traits. The groups are denoted as \$ (production group) and ¥ (conformation group). The single traits are: BFL - Bent to pre-bent curve of forelegs; NUT - Number of underdeveloped teats; IMF - Intramuscular fat content in MAS; BFT - Back fat thickness; NT - Number of teats (both sides); LMC - Lean meat content; CL - Carcass length; LDWG - Lifetime daily weight gain; DWG - Daily weight gain on test; ADFI - Average daily feed intake.

**Additional file 10.** PCA plot of reference panel samples (red and blue) and the target samples with array-genotypes (black +). First and second principal components captured 7.55 and 0.83% of the genomic variation.

**Additional file 11.** Accuracy of imputation to whole-genome sequence versus minor allele frequency. Dosage R<sup>2</sup> values from Beagle, and empirical measures of concordance and accuracy (R<sup>2</sup>) derived in 5 cross-validation sets, were assessed in 421 animals based on 44,733 downsampled chip genotypes.

**Additional file 12.** Results of single-trait GWAS based on imputed sequence variants. Table contains numbered QTL regions, chromosome, start and stop positions of the QTL, lowest *P*-value, top SNP with MAF, number of significantly associated variants within the QTL and the single-trait abbreviation.

**Additional file 13.** Manhattan plot of multi-trait meta-analyses GWAS of all 24 traits after fitting the 6 QTL as covariates, which resulted in loss of the peaks. Suggestive line is at 8.5.

**Additional file 14.** Results of meta-analyses GWAS based on imputed sequence variants and dataset including missing phenotypic records. Table contains numbered QTL regions, chromosome, start and stop positions of the QTL, lowest *P*-value, top SNP with MAF, number of significantly associated variants within the QTL and the trait-group.

**Additional file 15.** List of genes ID and genes symbols within the 6 pleiotropic QTL.

#### Acknowledgements

We thank Claudia Kasper-Vökl for granting access to low-pass sequence data for 294 SLW animals.

#### Authors' contributions

Analysis: AN AM HP; Preparation of sequence data: AN ALV HP; Conceived and designed the experiments: AN AM HP; Conceptualisation: HP; Secured funding: HP SN; Wrote the paper: AN HP; Critically revised the manuscript: all authors; Read and approved the final version of the manuscript: all authors.

#### Funding

Open access funding provided by Swiss Federal Institute of Technology Zurich. This study was financially supported by SUISAG, Micarna SA and the ETH Zürich Foundation. The funding bodies were not involved in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

#### Availability of data and materials

High-coverage sequencing read data are available at the European Nucleotide Archive (ENA) (<http://www.ebi.ac.uk/ena>) of the EMBL at BioProject PRJEB38156 and PRJEB39374.

## Declarations

#### Ethics approval and consent to participate

No animals were sampled for the present study. Thus, no ethical approval was required.

#### Consent for publication

Not applicable.

#### Competing interests

AH is employee of SUISAG (the Swiss pig breeding and competence centre). HP is a member of the editorial board of *BMC Genomics*. All other authors declare that they have no competing interests.

Received: 14 December 2022 Accepted: 4 April 2023

Published online: 10 April 2023

## References

- Bolormaa S, Pryce JE, Reverter A, Zhang Y, Barendse W, Kemper K, et al. A Multi-Trait, Meta-analysis for Detecting Pleiotropic Polymorphisms for Stature, Fatness and Reproduction in Beef Cattle. *PLoS Genet*. 2014;10:e1004198.
- Bonnemajjer PWM, van Leeuwen EM, Iglesias AI, Gharahkhani P, Vitart V, Khawaja AP, et al. Multi-trait genome-wide association study identifies new loci associated with optic disc parameters. *Commun Biol*. 2019;2:1–12.
- Yoshida GM, Yañez JM. Multi-trait GWAS using imputed high-density genotypes from whole-genome sequencing identifies genes associated with body traits in Nile tilapia. *BMC Genomics*. 2021;22:1–13.
- Hackinger S, Zeggini E. Statistical methods to detect pleiotropy in human complex traits. *Open Biol*. 2017;7:170125.
- Lee SH, Yang J, Goddard ME, Visscher PM, Wray NR. Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. *Bioinformatics*. 2012;28:2540.
- Legarra A, Vitezica ZG. Genetic evaluation with major genes and polygenic inheritance when some animals are not genotyped using gene content multiple-trait BLUP. *Genetics Sel Evol*. 2015;47:89.
- Wang H, Pei F, Vanyukov MM, Bahar I, Wu W, Xing EP. Coupled mixed model for joint genetic analysis of complex disorders with two independently collected data sets. *BMC Bioinformatics*. 2021;22:50.
- Galesloot TE, van Steen K, Kiemeneij LALM, Janss LL, Vermeulen SH. A Comparison of Multivariate Genome-Wide Association Methods. *PLoS ONE*. 2014;9:e95923.
- Porter HF, O'Reilly PF. Multivariate simulation framework reveals performance of multi-trait GWAS methods. *Sci Rep*. 2017;7:1–12.
- Korte A, Vilhjálmsson BJ, Segura V, Platt A, Long Q, Nordborg M. A mixed-model approach for genome-wide association studies of correlated traits in structured populations. *Nat Genet*. 2012;44:1066.
- Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics Appl Note*. 2010;26:2190–1.
- Turley P, Walters RK, Maghziyan O, Okbay A, Lee JJ, Fontana MA, et al. Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat Genet*. 2018;50:229–37.
- Zhu X, Feng T, Tayo BO, Liang J, Young JH, Franceschini N, et al. Meta-analysis of Correlated Traits via Summary Statistics from GWASs with an Application in Hypertension. *Am J Hum Genet*. 2015;96:21.
- Goddard ME, Hayes BJ. Genomic Selection Based on Dense Genotypes Inferred From Sparse Genotypes. *Proc Assoc Advmt Anim Breed Genet*. 2009;18:26–29.
- Daetwyler HD, Capitan A, Pausch H, Stothard P, Van Binsbergen R, Brøndum RF, et al. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nat Genet*. 2014;46:858–65.
- Van Den Berg S, Vandenplas J, Van Eeuwijk FA, Bouwman AC, Lopes MS, Veerkamp RF. Imputation to whole-genome sequence using multiple pig

- populations and its use in genome-wide association studies. *Genet Sel Evol.* 2019;51:1–13.
17. Iso-Touru T, Sahana G, Gulbrandsen B, Lund MS, Vilkki J. Genome-wide association analysis of milk yield traits in Nordic Red Cattle using imputed whole genome sequence variants. *BMC Genet.* 2016;17:55.
  18. Li JH, Mazur CA, Berisa T, Pickrell JK. Low-pass sequencing increases the power of GWAS and decreases measurement error of polygenic risk scores compared to genotyping arrays. *Genome Res.* 2021;26:6486–120.
  19. Nosková A, Bhati M, Kadri NK, Crysanto D, Neuenschwander S, Hofer A, et al. Characterization of a haplotype-reference panel for genotyping by low-pass sequencing in Swiss Large White pigs. *BMC Genomics.* 2021;22:290.
  20. Rubinacci S, Ribeiro DM, Hofmeister RJ, Delaneau O. Efficient phasing and imputation of low-coverage sequencing data using large reference panels. *Nat Genet.* 2021;53:120–6.
  21. Zhang Z, Ma P, Zhang Z, Wang Z, Wang Q, Pan Y. The construction of a haplotype reference panel using extremely low coverage whole genome sequences and its application in genome-wide association studies and genomic prediction in Duroc pigs. *Genomics.* 2022;114:340–50.
  22. Ding R, Savegnago R, Liu J, Long N, Tan C, Cai G, et al. Nucleotide resolution genetic mapping in pigs by publicly accessible whole genome imputation. *bioRxiv.* 2022;2022.05.18.492518.
  23. Lloret-Villas A, Pausch H, Leonard AS. Size and composition of haplotype reference panels impact the accuracy of imputation from low-pass sequencing in cattle. *bioRxiv.* 2023;2023.01.13.523894.
  24. Casu S, Usai MG, Sechi T, Salaris SL, Miari S, Mulas G, et al. Association analysis and functional annotation of imputed sequence data within genomic regions influencing resistance to gastro-intestinal parasites detected by an LDLA approach in a nucleus flock of Sarda dairy sheep. *Genet Sel Evol.* 2022;54:1–16.
  25. Ros-Freixedes R, Valente BD, Chen C-Y, Herring WO, Gorjanc G, Hickey JM, et al. Rare and population-specific functional variation across pig lines. *Genet Sel Evol.* 2022;54:1–16.
  26. Yan G, Liu X, Xiao S, Xin W, Xu W, Li Y, et al. An imputed whole-genome sequence-based GWAS approach pinpoints causal mutations for complex traits in a specific swine population. *Science China Life Sciences.* 2021;2021:1–14.
  27. Becker D, Wimmers K, Luther H, Hofer A, Leeb T, Moore S. A Genome-Wide Association Study to Detect QTL for Commercially Important Traits in Swiss Large White Boars. *PLoS One.* 2013;8:2.
  28. Broekema R v, Jonkers IH, Bakker OB. A practical view of fine-mapping and gene prioritization in the post-genome-wide association era. *Open Biol.* 2020;10(1):190221.
  29. Johnsson M, Jungnickel MK. Evidence for and localization of proposed causative variants in cattle and pig genomes. *Genet Sel Evol.* 2021;53:1–18.
  30. Cai Z, Christensen OF, Lund MS, Ostensen T, Sahana G. Large-scale association study on daily weight gain in pigs reveals overlap of genetic factors for growth in humans. *BMC Genomics.* 2022;23:1–13.
  31. Delpuech E, Aliakbari A, Labruno Y, Fève K, Billon Y, Gilbert H, et al. Identification of genomic regions affecting production traits in pigs divergently selected for feed efficiency. *Genet Sel Evol.* 2021;53:49.
  32. Ding R, Zhuang Z, Qiu Y, Ruan D, Wu J, Ye J, et al. Identify known and novel candidate genes associated with backfat thickness in Duroc pigs by large-scale genome-wide association analysis. *J Anim Sci.* 2022;100:1–8.
  33. Gozalo-Marcilla M, Buntjer J, Johnsson M, Batista L, Diez F, Werner CR, et al. Genetic architecture and major genes for backfat thickness in pig lines of diverse genetic backgrounds. *Genet Sel Evol.* 2021;53:1–14.
  34. Li J, Peng S, Zhong L, Zhou L, Yan G, Xiao S, et al. Identification and validation of a regulatory mutation upstream of the BMP2 gene associated with carcass length in pigs. *Genet Sel Evol.* 2021;53:1–13.
  35. Oliveira HC, Lopes MS, Derks MFL, Madsen O, Harlizius B, van Son M, et al. Fine Mapping of a Major Backfat QTL Reveals a Causal Regulatory Variant Affecting the CCND2 Gene. *Front Genet.* 2022;0:1241.
  36. Kim KS, Larsen N, Short T, Plastow G, Rothschild MF. A missense variant of the porcine melanocortin-4 receptor (MC4R) gene is associated with fatness, growth, and feed intake traits. *Mamm Genome.* 2000;11:131–5.
  37. Duan Y, Zhang H, Zhang Z, Gao J, Yang J, Wu Z, et al. VRTN is required for the development of thoracic vertebrae in mammals. *Int J Biol Sci.* 2018;14:667–81.
  38. Zhuang Z, Ding R, Peng L, Wu J, Ye Y, Zhou S, et al. Genome-wide association analyses identify known and novel loci for teat number in Duroc pigs using single-locus and multi-locus models. *BMC Genomics.* 2020;21:344.
  39. Ekine CC, Rowe SJ, Bishop SC, deKoning DJ. Why breeding values estimated using familial data should not be used for genome-wide association studies. *G3 Genes Genomes Genetics.* 2014;4:341–7.
  40. Pausch H, Emmerling R, Schwarzenbacher H, Fries R. A multi-trait meta-analysis with imputed sequence variants reveals twelve QTL for mammary gland morphology in Fleckvieh cattle. *Genet Sel Evol.* 2016;48:1–9.
  41. Zhou X, Stephens M. Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nat Methods.* 2014;11:407–9.
  42. Jiang J, Zhang Q, Ma L, Li J, Wang Z, Liu JF. Joint prediction of multiple quantitative traits using a Bayesian multivariate antedependence model. *Heredity.* 2015;115:29–36.
  43. Hormozdiari F, Kang EY, Bilow M, Ben-David E, Vulpe C, McLachlan S, et al. Imputing Phenotypes for Genome-wide Association Studies. *Am J Hum Genet.* 2016;99:89.
  44. Aguilar I, Legarra A, Cardoso F, Masuda Y, Lourenco D, Misztal I. Frequentist p-values for large-scale single step genome-wide association, with an application to birth weight in American Angus cattle. *Genetics Selection Evolution.* 2019;51:1–8.
  45. Barendse W. The effect of measurement error of phenotypes on genome wide association studies. *BMC Genomics.* 2011;12:1–12.
  46. Bickel RD, Kopp A, Nuzhdin SV. Composite Effects of Polymorphisms near Multiple Regulatory Elements Create a Major-Effect QTL. *PLoS Genet.* 2011;7:e1001275.
  47. Abell NS, DeGorter MK, Gloudemans MJ, Greenwald E, Smith KS, He Z, et al. Multiple causal variants underlie genetic associations in humans. *Science.* 1979;202(375):1247–54.
  48. Pausch H, MacLeod IM, Fries R, Emmerling R, Bowman PJ, Daetwyler HD, et al. Evaluation of the accuracy of imputed sequence variant genotypes and their utility for causal variant detection in cattle. *Genet Sel Evol.* 2017;49:1–14.
  49. Flossmann G, Wurmser C, Pausch H, Tenghe A, Dodenhoff J, Dahinten G, et al. A nonsense mutation of bone morphogenetic protein-15 (BMP15) causes both infertility and increased litter size in pigs. *BMC Genomics.* 2021;22:1–9.
  50. Derks MFL, Lopes MS, Bosse M, Madsen O, Dibbitts B, Harlizius B, et al. Balancing selection on a recessive lethal deletion with pleiotropic effects on two neighboring genes in the porcine genome. *PLoS Genet.* 2018;14:e1007661.
  51. Fujii J, Otsu K, Zorzato F, de Leon S, Khanna VK, Weiler JE, et al. Identification of a mutation in porcine ryanodine receptor associated with malignant hyperthermia. *Science.* 1979;199(1253):448–51.
  52. Bovo S, Mazzoni G, Bertolini F, Schiavo G, Galimberti G, Gallo M, et al. Genome-wide association studies for 30 haematological and blood clinical-biochemical traits in Large White pigs reveal genomic regions affecting intermediate phenotypes. *Scientific Reports.* 2019;9:1–17.
  53. Jiao S, Maltecca C, Gray KA, Cassady JP. Feed intake, average daily gain, feed efficiency, and real-time ultrasound traits in Duroc pigs: II Genomewide association. *J Anim Sci.* 2014;92:2846–60.
  54. 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.* 2018;41:125–36.
  55. Ponsuksili S, Murani E, Brand B, Schwerin M, Wimmers K. Integrating expression profiling and whole-genome association for dissection of fat traits in a porcine model. *J Lipid Res.* 2011;52:668.
  56. Garrick DJ, Taylor JF, Fernando RL. Deregressing estimated breeding values and weighting information for genomic regression analyses. *Genet Sel Evol.* 2009;41:55.
  57. Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience.* 2015;4:7.
  58. Warr A, Affara N, Aken B, Beiki H, Bickhart DM, Billis K, et al. An improved pig reference genome sequence to enable pig genetics and genomics research. *Gigascience.* 2020;9:1–14.
  59. Browning BL, Zhou Y, Browning SR. A One-Penny Imputed Genome from Next-Generation Reference Panels. *Am J Hum Genet.* 2018;103:338–48.
  60. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet.* 2011;88:76–82.

61. Chen S, Zhou Y, Chen Y, Gu J. Fastp: An ultra-fast all-in-one FASTQ pre-processor. *Bioinformatics*. 2018;34(17):i884–890.
62. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:13033997. 2013.
63. Picard Toolkit. Broad Institute, GitHub Repository. 2019.
64. Tarasov A, Vilella AJ, Cuppen E, Nijman IJ, Prins P. Sambamba: Fast processing of NGS alignment formats. *Bioinformatics*. 2015;31:2032–4.
65. Pedersen BS, Quinlan AR. Mosdepth: Quick coverage calculation for genomes and exomes. *Bioinformatics*. 2018;34:867–8.
66. DePristo MA, Banks E, Poplin R, Garimella Kv, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 2011;43:491–501.
67. Delaneau O, Zagury JF, Robinson MR, Marchini JL, Dermitzakis ET. Accurate, scalable and integrative haplotype estimation. *Nature Commun*. 2019;10:1–10.
68. Barbato M, Orozco-terWengel P, Tapio M, Bruford MW. SNeP: A tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. *Front Genet*. 2015;6 MAR:109.
69. Zhou X, Stephens M. Genome-wide Efficient Mixed Model Analysis for Association Studies. *Nat Genet*. 2012;44:821.
70. Bolormaa S, Hayes BJ, Savin K, Hawken R, Barendse W, Arthur PF, et al. Genome-wide association studies for feedlot and growth traits in cattle. *J Anim Sci*. 2011;89:1684–97.
71. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl Variant Effect Predictor. *Genome Biol*. 2016;17:122.
72. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc*. 2009;4:1073–82.
73. Kadri NK, Harland C, Faux P, Cambisano N, Karim L, Coppieters W, et al. Coding and noncoding variants in HFM1, MLH3, MSH4, MSH5, RNF212, and RNF212B affect recombination rate in cattle. *Genome Res*. 2016;26:1323–32.

### Publisher's Note

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

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

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

At BMC, research is always in progress.

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

