Sequence-based GWAS meta-analyses for beef production traits

M P Sanchez, T Tribout, N Kadri, P K Chitneedi, S Maak, C Hozé, M Boussaha, P Croiseau, R Philippe, M Spengeler, et al.

Sequence-based GWAS meta-analyses for beef production traits

M.P. Sanchez[1*], T. Tribout[1], N. Kadri[2], P.K. Chitneedi[3], S. Maak[3], C. Hozé[1,4], M. Boussaha[1], P. Croiseau[1], R. Philippe[5], M. Spengeler[6], C. Kühn[3], Y. Wang[7], H. Pausch[2], D. Boichard[1]

[1]Université Paris Saclay, INRAE, AgroParisTech, GABI, 78350 Jouy-en-Josas, France; [2]ETH, 8092 Zürich, Switzerland; [3]Research Institute for Farm Animal Biology (FBN), 18196 Dummerstorf, Germany; [4]Allice, 75595 Paris, France; [5]INRAE, USC1061 GAMAA, Université de Limoges, 87060 Limoges, France; [6]QualitasAG, 6300 Zug, Switzerland; [7]University of Alberta, Edmonton, Canada. [*]marie-pierre.sanchez@inrae.fr

Abstract

Five partners of the H2020 BovReg project have performed sequence-based GWAS for 28 beef production traits (4 growth, 9 morphology, and/or 15 carcass traits) using 54,782 animals from 15 different purebred or crossbred cattle populations. These results were herein combined to conduct 16 different meta-analyses (MA) with both the z-score and the fixed effects MA methods. We identified QTL in 15 MA on BTA2, 5, 6, 7, 10, 11, 13, 14, 15, 17, and 20, most of them being common to several MA. Overall, the fixed effects method outperformed the z-score method in terms of significance level and number of QTL detected. Compared to within-population GWAS, MA found a higher number of QTL in which variants were more frequently located in genes (e.g. *MSTN*, *LCORL, ARRDC3, PLAG1, COL3A1*).

Introduction

When performed at the whole genome sequence (WGS) level, the meta-analysis (MA) of within-population GWAS results can be powerful and accurate to identify causal variants for complex traits (Bouwman et al., 2018). One of the objectives of the H2020 BovReg project is to perform MA at the sequence level for various dairy and beef cattle traits. For beef production, five partners from France (INRAE/ULIM), Switzerland (ETH), Germany (FBN), and Canada (UAL) contributed with 54,782 animals from 15 purebred populations Charolais (CHA), Montbéliarde (MON), Normande (NOR), Limousine (LIM), Blonde d'Aquitaine (BLA), Brown Swiss (BSW), Original Braunvieh (OBR) or crossbred Charolais x Holstein (CH), and Angus, Charolais and beef composite (BP). Each partner conducted sequence-based within-population GWAS for 4 growth, 9 morphology, and/or 15 carcass traits. Here, we combine these GWAS results to conduct 16 MA with fixed effects and z-score methods.

Materials & Methods

Within-population GWAS. Within-population GWAS were conducted on 54,782 steers, cows or bulls (19,656 females and 35,126 males) of 15 various populations. Depending on the population, 4 growth, 9 morphology, and/or 15 carcass traits were expressed as yield deviation (YD), daughter YD (DYD), deregressed proof (DRP) or adjusted performance (AP). All partners applied similar imputation and GWAS workflows. Genotypes, aligned on the ARS-UCD1.2 reference genome, were imputed using a stepwise approach: 1) 777K (HD) genotypes were imputed from 50k genotypes with Beagle (Browning and Browning, 2016) or FImpute (Sargolzaei et al., 2014) using animals with HD genotypes as a reference and 2) sequence variants were imputed using 372 to 3093 animals from the RUN7 or RUN8 reference panel of the 1000 Bull Genomes consortium (Bouwman et al, 2018) with Minimac (Howie et al., 2012) or Beagle. After filtering, allele dosages of between 12.9 and 20.6 million variants were tested

for association with different traits in each population separately using GCTA software (Yang et al., 2011) accounting for a polygenic effect estimated from 50k or HD SNPs. DRP and DYD were weighted to account for heterogeneous accuracy (Vandenplas, pers. comm.).

Table 1. Features of beef production meta-analyses (MA).

MA	Trait group	Traits [1]	Breeds (# populations)	#animals
G1	growth	W15/W18/ADG	CH/CHA/BP/LIM/BLA (7)	18,774
G2	growth	BW	CH/CHA/LIM/BLA (5)	2,720
M1	morphology	MS30/THIGHS/CC	CHA/MON/NOR/LIM/BLA (6)	17,418
M2	morphology	MS30/WITHER/CC	CHA/MON/NOR/LIM/BLA (6)	17,418
M3	morphology	LL	CH/CHA/LIM/BLA (5)	3,695
M4	morphology	WT	CH/CHA/LIM/BLA (5)	3,695
M5	morphology	SS30/SD	CHA/LIM/BLA (4)	12,140
C1	carcass	CW	CH/OBR/BSW/CHA/MON/NOR/BP (7)	19,989
C2	carcass	AS	INRAE/CHA/LIM/BLA (6)	12,208
C3	carcass	CY	CH/CHA/LIM/BLA (5)	3,694
C4	carcass	CG/LMY/MT/CC	CH/OBR/BSW/MON/NOR/BP/CHA/LIM/BLA(10)	25,367
C5	carcass	FS/ABT/FC6/FCU/CF	CH/OBR/BSW/NOR/BP/CHA/LIM/BLA (8)	14,622
C6	carcass	WS	CH/CHA/LIM/BLA (5)	2,636
C7	carcass	ALT	CH/CHA/LIM/BLA (5)	3,692
C8	carcass	IFW	CH/CHA/LIM/BLA (5)	3,686
C9	carcass	REA	CH/BP (3)	4,453

[1] Weight at birth (**BW**), at month 15 (**W15**), at month 18 (**W18**); average daily gain (**ADG**); muscularity score (**MS30**), skeletal (**SS30**) score, thickness of bones (**TB30**) at month 30; **THIGHS**; **WITHER**; fat score (**FS**); leg length (**LL**); maximum width of the thigh (**WT**); skeletal development (**SD**); carcass conformation (**CC**), weight (**CW**), yield (**CY**), grade (**CG**) and fat score (**CF**); age at slaughter (**AS**); weight at slaughter (**WS**); lean meat yield (**LMY**); meatiness (**MT**); average backfat thickness (**ABT**); fat content of the 6th rib (**FC6**); ultrasound fat content (**FCU**); area of *longissimus thoracis* (**ALT**); internal fat weight (**IFW**); rib eye area (**REA**).

Meta-analyses. We then conducted 16 MA – 2, 5 and 9 for growth, morphology and carcass traits, respectively – combining GWAS results of 1 to 5 traits measured in 3 to 10 different populations from 2 to 5 partners (Table 1). We considered 29.6 million variants imputed by at least two different partners with concordant REF/ALT alleles. For each partner x trait x population combination, we retained variants with a MAF $\geq$ 0.005 (0.02 for FBN which had a smaller sample size) and with an imputation $R^2 \geq 0.20$, i.e. 17.9 to 24.9 million variants depending on the MA. Variant effects were standardized by the genetic standard deviation of the trait. The z-score (ZSc) and fixed effects (FE) MA methods implemented in the METAL software (Willer et al., 2010) were applied. For each variant, the ZSc method converts the p-value in z-score $Z = \Sigma_i z_i w_i / (\Sigma_i w_i^2)$ where w_i is the square root of the sample size for study i and $z_i = \Phi^{-1}(1 - p_i/2)$*(*effect direction for study i*), with p_i the p-value of the i^{th} study. The FE method assumes that the true effect of each allele is the same across the different studies and combines effects by weighting them by the inverse of their error variance. Therefore, both MA methods weight the different studies by their sample size. For all GWAS or MA results, we considered an uniform threshold ($-\log_{10}(P) = 8.7$) corresponding to the 5% genome-wide threshold of significance after Bonferroni correction for ~25 million variants.

Results

For 15 of the 16 MA conducted in this study, we found variants with significant effects with both ZSc and FE methods (Table 2). These variants defined QTL regions on BTA2, 5, 6, 7, 10, 11, 13, 14, 15, 17, and 20. The most significant QTL were located on BTA2 and 6 (Figure 1). In 12 of the 15 MA, the FE method identified a higher number of variants with significant

effects than the ZSc method. In addition, QTL on BTA6 (C7), 13 (C5), 14 (M2), and 17 (C5) were detected only by the FE approach. In each of the 15 MA, variants with significant effects were also identified in 1 to 5 within-population GWAS. In 7 MA (G2, M1, M2, M5, C4, C5, and C8), the number of variants with significant effects was higher in at least one elementary GWAS than in any of the MA methods used. However, compared to within-population GWAS, MA revealed i) more significant effects, ii) novel QTL on BTA2 (C1), 5 (G1), 6 (C7), 7 (M1 and M2), 10 (C1 and C4), 14 (M2), and 17 (C5) and iii) variants with the most significant effects in the QTL peaks (TOP1) more frequently located in genes (72% in fixed effects MA vs 62% in GWAS). For the most significant QTL identified on BTA2 in M1, M2, M4, and C4 MA, the TOP1 variant is a stop gained variant (rs110344317, Q204x) within the *MSTN* gene. In 10 MA (G1, G2, M3, M4, M5, C1, C4, C5, C6, and C9), we found a QTL on BTA6 with 7 unique TOP1 variants located within or near to the *LCORL* gene. For other QTL, the variants with the most significant effects were located in *ANKAR*, *ASNSD1*, *SLC40A1*, and *COL3A1* (BTA2); *CCND2* (BTA5); *MLLT1*, *PIAS4*, and *ARRDC3* (BTA7); *GLDN* and *CYP19A1* (BTA10); *PLAG1* (BTA14); and *CHST1* and *SLC35C1* (BTA15).

Table 2. Within population GWAS and MA results.

	GWAS			Z-score MA			Fixed effects MA		
MA	# GWAS with sign. variants (min-max)	# BTA with sign. variants[1]	-log$_{10}$(P) max	# sign. variants	-log$_{10}$(P) max	# BTA with sign. variants	# sign. variants	-log$_{10}$(P) max	# BTA with sign. variants
G1	2 (300-1006)	6	35.7	1656	42.0	6	2456	45.5	6
G2	1 (1354)	1	18.7	314	17.8	1	1247	22.0	1
M1	4 (116-19,058)	1	154.0	6384	165.5	2	5667	165.1	2
M2	4 (58-9529)	1	154.0	5710	165.5	2	5197	165.1	3
M3	4 (88-1027)	1	15.8	2334	31.9	1	3198	37.1	1
M4	4 (36-508)	2	15.6	1953	23.4	2	2269	22.7	2
M5	3 (46-3527)	4	27.9	2957	28.2	4	3249	29.4	4
C1	1 (581)	3	23.5	461	29.4	4	740	31.4	4
C2	0	0	-	0	-	0	0	-	0
C3	4 (201-2575)	2	30.4	3977	48.5	1	4899	48.0	1
C4	5 (88-1120)	3	32.8	249	33.8	4	1004	42.0	4
C5	3 (21-446)	4	13.8	138	15.8	3	192	14.7	5
C6	3 (57-741)	1	13.4	1308	25.8	1	1852	29.2	1
C7	3 (9-110)	1	11.8	1275	24.9	1	1724	25.8	2
C8	1 (189)	1	9.7	28	11.2	1	8	9.3	1
C9	3 (24-436)	2	15.0	722	19.8	2	753	22.8	2

[1] In at least one GWAS

Discussion

This study, conducted on imputed WGS of 54,782 animals from 15 populations of various breeds, is the first meta-analysis of this magnitude dedicated to cattle beef production. We show the value of MA, in complement to within-population GWAS, in identifying i) a larger number of QTL, ii) a lower number of variants in QTL and iii) candidate variants located more frequently in genes. The use of more animals and the conservation of LD over shorter distances across breeds may explain the superiority of MA over GWAS in terms of power and mapping precision. By applying here the most commonly MA methods used for GWAS, i.e. ZSc and FE approaches, we confirm that the FE method appears more powerful in detecting QTL (Begum et al., 2012), although MA combine substantially different traits in the present study. In several regions, MA directly pointed out variants in genes, including *MSTN*, *LCORL*, *ARRDC3*, and

PLAG1, previously associated with morphology and carcass traits in various studies (e.g. Wang et al., 2020). For example, the Q204X mutation, ranked 1st in the QTL peaks at the proximal end of BTA2 and causing a premature stop codon in the gene encoding myostatin (*MSTN*), was reported as one of the polymorphisms responsible for the double-muscled phenotype in several cattle breeds (Grobet et al., 1998). We also identified dozens of other variants located in genes having a function that may be related to meat production traits (e.g. *COL3A1 collagen type III alpha 1 chain*). By better identifying genes and candidate causative variants associated with beef production traits in cattle, MA appears to be of great interest to decipher the biological mechanisms underlying these traits.

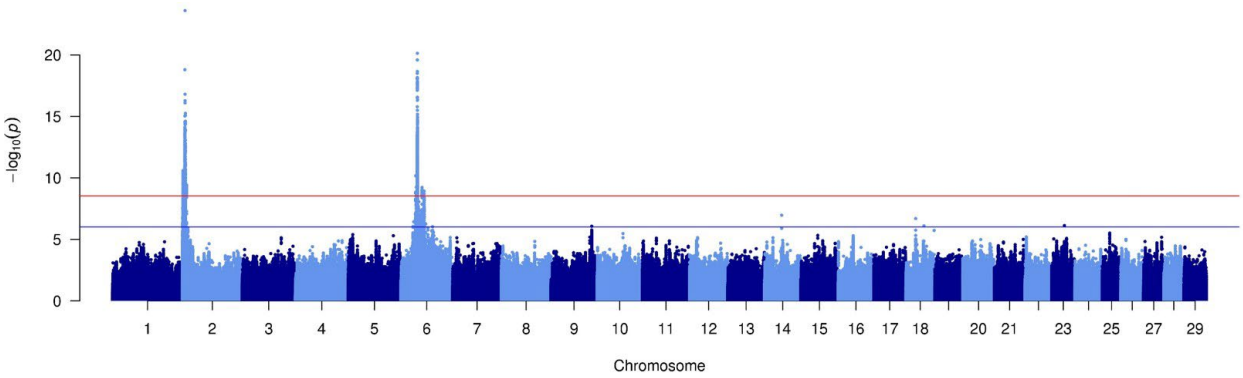

Figure 1. Manhattan plot obtained with M4 fixed effects MA results.

Acknowledgements.
The authors thank Dr J. Vandenplas (WUR) for the modification of GCTA to account for weights. The BovReg project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 815668.

References

Begum F., Ghosh D., Tseng G., and Feingold E. (2012) Nucleic Acids Res 40(9):3777-3784. https://doi.org/10.1093/nar/gkr1255
Bouwman A. C., Daetwyler H. D., Chamberlain A. J., Ponce C. H., Sargolzaei M. *et al.* (2018) Nat Genet 50(3):362-367. https://doi.org/10.1038/s41588-018-0056-5
Browning B. and Browning S. (2016) Am J Hum Genet 98(1):116-126. https://doi.org/10.1016/j.ajhg.2015.11.020
Grobet L., Poncelet D., Royo L., Brouwers B., Pirottin D. *et al.* (1998) Mamm Genome 9(3):210-213. https://doi.org/10.1007/s003359900727
Howie B., Fuchsberger C., Stephens M., Marchini J., and Abecasis G. R. (2012) Nat Genet 44(8):955-959. https://doi.org/10.1038/ng.2354
Sargolzaei M., Chesnais J., and Schenkel F. (2014) BMC Genomics 15(1):478. http://doi.org/10.1186/1471-2164-15-478
Wang Y., Zhang F., Mukiibi R., Chen L., Vinsky M. *et al.* (2020) BMC Genomics 21(1). https://doi.org/10.1186/s12864-019-6273-1
Willer C., Li Y., and Abecasis G. (2010) Bioinformatics 26(17):2190-2191. https://doi.org/10.1093/bioinformatics/btq340
Yang J., Lee S., Goddard M., and Visscher P. (2011) Am J Hum Genet 88(1):76-82. https://doi.org/10.1016/j.ajhg.2010.11.011