ANIMAL GENETICS Immunogenetics, Molecular Genetics and Functional Genomics

SHORT COMMUNICATION

doi: 10.1111/age.12959

Haplotype-based genome-wide association studies reveal new loci for haematological and clinical–biochemical parameters in Large White pigs

S. Bovo*, M. Ballan*, G. Schiavo*, M. Gallo†, S. Dall'Olio* and L. Fontanesi*

*Department of Agricultural and Food Sciences, Division of Animal Sciences, University of Bologna, Viale G. Fanin 46, Bologna 40127, Italy.
†Associazione Nazionale Allevatori Suini, Via Nizza 53, Roma 00198, Italy.

Summary

We report haplotype-based GWASs for 33 blood parameters measured in 843 Italian Large White pigs. In the single-trait analysis, a total of 30 QTL for number of basophils, six erythrocyte traits (haemoglobin, haematocrit, mean corpuscular haemoglobin, mean corpuscular haemoglobin concentration, mean corpuscular volume and red blood cell count) and two clinical–biochemical traits (alkaline phosphatase and Ca^{2+} contents) were identified. In the multiple-trait analysis, a total of five QTL affected three different clusters of traits. Only four of these QTL were already reported in the single-marker and multi-marker GWASs we previously carried out on the same pig population. QTL on SSC11 and SSC17 showed effects on multiple traits. These results further dissected the genetic architecture of parameters that could be used as proxies in breeding programmes for more complex traits. In addition, these results might help to better define the pig as an animal model for several blood-related biological functions.

Keywords alkaline phosphatase, blood parameter, creatine kinase, electrolyte, enzyme, erythrocyte, leukocyte, QTL, *Sus scrofa*

Haematological and clinical–biochemical parameters are important descriptors or proxies of the physiological state and functions of various organs and they are used as general indicators of the immunological status and health conditions of an organism (Etim *et al.* 2014). Many of these parameters have medium to high heritability (Bovo *et al.* 2019). The identification of the genetic determinants underlying their variability could help to clarify the biological mechanisms of different diseases directly or indirectly related to these proxies (Okada & Kamatani 2012).

GWASs have identified several loci linked to blood parameters in humans (reviewed in Vasquez *et al.* 2016) and in a few livestock species, including chickens (Sun *et al.* 2016), ducks (Zhu *et al.* 2020), cattle (Gan *et al.* 2019) and yak (Ma *et al.* 2019). Several QTL investigations involving F2 reference populations and GWAS within breeds have also been carried out in pigs. Studies in this species focused mainly on haematological traits, i.e. blood cell-related parameters (e.g. Reiner *et al.* 2007; Gong *et al.* 2010; Zhang *et al.* 2013; Zhang *et al.* 2014; Ponsuksili *et al.* 2016; Yan *et al.* 2018; Bovo *et al.* 2019), whereas fewer studies included clinical–biochemical parameters (e.g. Reiner *et al.* 2009; Bovo *et al.* 2016, 2019; Reyer *et al.* 2019). The genetic architecture of these traits could be useful (i) to define selection strategies to improve disease resistance in pigs and (ii) to further describe the pig as an animal model for several biological aspects related to blood cell and other blood indicators.

In our previous studies (Bovo *et al.* 2016, 2019) we performed GWASs in Italian Large White pigs and identified more than 50 QTL affecting 15 haematological and 11 clinical–biochemical parameters. Here, we examined the same pig population taking advantage of haplotype analyses to further exploit genetic variability related to a large number of blood parameters. A detailed description of the pig population, blood parameters and processing of data is reported in Bovo *et al.* (2016, 2019) and provided in Appendix S1 and Table S1. Briefly, a total of 843 performance-tested Italian Large White pigs (278 castrated males and 565 gilts) were blood sampled at the commercial abattoir just after slaughtering of the animals. A total of 15 haematological parameters (seven erythrocyte, six

Address for correspondence

Luca Fontanesi, Department of Agricultural and Food Sciences, Division of Animal Sciences, University of Bologna, Viale G. Fanin 46, 40127 Bologna, Italy.
E-mail: luca.fontanesi@unibo.it

Accepted for publication 02 May 2020

leukocyte and two platelet-related traits) and 18 clinical–biochemical parameters (electrolytes and lipid, metabolism and protein-related parameters) were determined (Bovo *et al*. 2016, 2019). A correlation network based on Pearson's correlation coefficients (r) was used to study the dependence among these parameters, as previously described (Bovo *et al*. 2019; Appendix S1).

DNA was extracted from blood and genotyped with the Illumina PorcineSNP60 BeadChip version 2. SNPs were mapped on the Sscrofa11 genome version. Haplotypes were called with the R package GHap 1.2.2 (Utsunomiya *et al*. 2016) on genotype data previously phased using SHAPEIT version 2 (Delaneau *et al*. 2012). A total of 280 787 haplotypes, corresponding to 21 296 haploblocks, were used in the genome-wide association analyses by examining single trait–haplotype pairs with linear mixed models fitted with GEMMA version 0.98.1 (Zhou & Stephens 2012). In addition, based on network analysis results, a few medium/highly correlated ($|r| > 0.4$; $P < 1.15 \times 10^{-4}$) haematological trait clusters were identified and subsequently used in the multivariate genome-wide association scans, as reported in Bovo *et al*. (2019). Details of the correlation network, quality control procedures, haplotype calling and genome-wide association analyses are given in Appendix S1. Table S1 reports the estimated genomic inflation factor (λ) and chip heritability (h^2_{SNP}).Table S2 reports statistics of correlation coefficients used to build the network. The suggestive and significant thresholds were set to $0.05/m$ and $0.05/n$, where m and n are the number of haploblocks and the number of haplotypes, respectively. Haplotypes presenting the lowest P in chromosome regions separated by at least 5 Mbp were considered as tag haplotypes. All Manhattan plots and quantile–quantile plots are provided in Figs S1 and S2, respectively.

Significantly associated tag haplotypes are reported in Table 1. The complete list of haplotypes below the suggestive threshold is given in Table S3. Over-imposed Manhattan plots obtained in the GWASs for haematological traits, clinical–biochemical traits and in the multivariate GWAS scans for correlated blood parameters are reported in Fig. 1. Heritability estimates for the analysed traits showed an average value of $h^2_{SNP} = 0.378$ whereas we previously reported a mean value of $h^2_{SNP} = 0.290$ using SNP-based analyses (Bovo *et al*. 2019). The highest value was $h^2_{SNP} = 0.572$ (SE = 0.074) for alanine aminotransferase activity (ALT), confirming its medium-high heritability (Bovo *et al*. 2019). In the single-trait analysis, a total of 30 tag haplotypes on 13 different autosomes (Table 1) were significantly associated with different traits, including one leukocyte trait (number of basophils, BASO), six erythrocyte traits (haemoglobin, HGB; haematocrit, HCT; mean corpuscular haemoglobin, MCH; mean corpuscular haemoglobin concentration, MCHC; mean corpuscular volume, MCV; and red blood cell count, RBC) and two clinical–biochemical traits (alkaline phosphatase, ALP; Ca^{2+} content). In the multiple-trait analysis, a total of five tag haplotypes were significantly associated with three different clusters of traits (Table 1). Five of these QTL (HCT on SSC10 and SSC11; HGB on SSC11; MCHC on SSC11; MCV on SSC15) have already been reported by other studies in a close chromosome region in F2 populations or Chinese breeds (Table 1). The haplotype on SSC15 includes the *methyl-CpG binding domain protein* 5 (*MBD5*) gene that has been associated with MCV in humans (Kichaev *et al*. 2019). Only four QTL (two for haematological traits – BASO on SSC14 and HCT on SSC18; one for a clinical–biochemical parameter – Ca^{2+} levels on SSC13; and one for the cluster aspartate aminotransferase-creatine kinase – AST–CK on SSC14, near the *glutamic-oxaloacetic transaminase 1* (*GOT1*) and the *carboxypeptidase N subunit 1* (*CPN1*) genes; Table 1) have already been reported in the single-marker GWASs we previously carried out on the same pig population (Bovo *et al*. 2016, 2019).

Two significant tag haplotypes had effects on multiple blood parameters: the haplotype spanning 69.30–69.77 Mbp on SSC11 affected HCT, HGB and RBC and this result was confirmed by the multiple-trait analysis that identified the same SSC11 with the cluster constituted by the same three parameters; and the haplotype spanning 41.50–41.97 Mbp on SSC17 affected both MCH and MCV. This latter SSC17 region was also identified as significant in the multiple-trait analysis for the RBC–HGB–HCT cluster. Other two QTL for another cluster of traits (MCH–MHCH–MCV–RDW) were located on SSC5 and SSC9 spanning the *protein phosphatase, Mg^{2+}/Mn^{2+} dependent 1H* (*PPM1H*) gene and the gene cluster *BUD13/SIK3/ZPR1/APOA5* which have been associated with RDW and other haematological traits in humans (Astle *et al*. 2016; Pilling *et al*. 2017; Kichaev *et al*. 2019).

Several QTL were identified for ALP that were not previously reported in any other studies in pigs. These QTL might reflect the ubiquitous and broad role of this enzyme complex, that might interact with several regulators and cellular mechanisms or could be involved in many different biological pathways. Moreover, serum ALP levels are due mainly to the isoforms from bone and liver, which is of particular interest for the evaluation of disease states (hepatobiliary disease, vitamin D deficiency, bone diseases and malignancy). Among these QTL, one on SSC8 (spanning positions 116.10–116.57 Mbp) harbours the *inorganic pyrophosphatase 2* (*PPA2*) gene, whose functions are related to ALP levels (Wei *et al*. 2011). Other suggestively associated haplotypes identified putative QTL regions in this study (Table S2). Among them, a suggestively significant QTL for ALP on SSC7 included the *glycosylphosphatidylinositol-specific phospholipase D1* (*GPLD1*) gene that is associated with ALP levels in humans (Chambers *et al*. 2011).

A few other suggestively associated haplotypes identified QTL encompassing putative causative genes already reported in the single-marker analysis (Bovo *et al*. 2019):

Table 1 Tag haplotypes identified at the genome-wide significant level

Blood parameters[1]	SSC[2]	Haploblock[3]	Start[4]	End[5]	*P*[6]	Previous studies[7]
Haematological parameters						
HGB	8	B964	96 300 001	96 770 001	9.95×10^{-8}	–
HCT	10	B87	8 600 001	9 070 001	2.77×10^{-8}	Zhang *et al*. (2014) (Pigs); Gong *et al*. (2010) (pigs)
HCT	11	B694	69 300 001	69 770 001	4.14×10^{-8}	Zhang *et al*. (2013) (pigs); Gong *et al*. (2010) (pigs); Yan *et al*. (2018) (pigs)
HGB	11	B694	69 300 001	69 770 001	8.43×10^{-8}	Gong *et al*. (2010) (pigs)
RBC	11	B694	69 300 001	69 770 001	1.43×10^{-7}	–
MCHC	11	B722	72 100 001	72 570 001	1.17×10^{-7}	Gong *et al*. (2010) (pigs); Yan *et al*. (2018) (pigs)
BASO	14	B711	710 00 001	71 470 001	1.44×10^{-8}	Bovo *et al*. (2019) (pigs); Kichaev *et al*. (2019) (humans)
MCV	15	B38	3 700 001	4 170 001	1.48×10^{-7}	Yan *et al*. (2018) (pigs)
MCH	17	B416	41 500 001	41 970 001	4.60×10^{-8}	–
MCV	17	B416	41 500 001	41 970 001	3.53×10^{-8}	–
HCT	18	B368	36 700 001	37 170 001	1.60×10^{-7}	Bovo *et al*. (2019) (pigs)
Clinical–biochemical parameters						
ALP	1	B42	4 100 001	4 570 001	1.68×10^{-7}	–
ALP	1	B412	41 100 001	41 570 001	1.48×10^{-11}	–
ALP	2	B79	7 800 001	8 270 001	4.00×10^{-8}	–
ALP	2	B273	27 200 001	27 670 001	1.10×10^{-8}	–
ALP	3	B1083	108 200 001	108 670 001	1.50×10^{-10}	–
ALP	4	B86	8 500 001	8 970 001	1.00×10^{-8}	–
ALP	8	B189	18 800 001	19 270 001	1.13×10^{-10}	–
ALP	8	B298	29 700 001	30 170 001	1.15×10^{-8}	–
ALP	8	B1162	11 610 0001	116 570 001	4.26×10^{-9}	–
ALP	11	B454	45 300 001	45 770 001	2.09×10^{-8}	–
ALP	12	B410	4 090 0001	41 370 001	2.75×10^{-8}	–
Ca^{2+}	13	B939	93 800 001	9 4270 001	5.07×10^{-8}	Bovo *et al*. (2016) (pigs)
ALP	14	B169	16 800 001	17 270 001	1.38×10^{-10}	–
ALP	15	B188	18 700 001	19 170 001	5.00×10^{-8}	–
ALP	17	B398	39 700 001	40 170 001	1.65×10^{-8}	–
ALP	17	B475	47 400 001	47 870 001	2.82×10^{-8}	–
ALP	17	B586	58 500 001	58 970 001	5.12×10^{-8}	–
ALP	18	B118	11 700 001	12 170 001	1.16×10^{-7}	–
ALP	18	B119	11 800 001	12 270 001	4.06×10^{-8}	–
Cluster of traits						
MCH–MCHC–MCV–RDW	5	B278	27 700 001	28 170 001	1.43×10^{-7}	Astle *et al*. (2016) (humans); Pilling *et al*. (2017) (humans); Kichaev *et al*. (2019) (humans)
MCH–MCHC–MCV–RDW	9	B442	44 100 001	44 570 001	7.17×10^{-8}	Astle *et al*. (2016) (humans); Pilling *et al*. (2017) (humans); Kichaev *et al*. (2019) (humans)
RBC–HGB–HCT	11	B694	69 300 001	69 770 001	1.90×10^{-8}	This study – single-trait analysis
AST–CK	14	B1097	109 600 001	110 070 001	2.89×10^{-11}	Bovo *et al*. (2019) (pigs)
RBC–HGB–HCT	17	B416	41 500 001	41 970 001	4.23×10^{-8}	This study – single-trait analysis

[1]Haematological parameters: erythrocyte traits – red blood cell count (RBC), haemoglobin (HGB), haematocrit (HCT), mean corpuscular volume (MCV), mean corpuscular haemoglobin (MCH), mean corpuscular haemoglobin concentration (MCHC), red cell distribution width (RDW); leukocyte traits – basophil count (BASO). Clinical–biochemical parameters: enzyme traits – alkaline phosphatase (ALP), creatine kinase (CK), aspartate aminotransferase (AST); electrolytes – calcium (Ca^{2+}).
[2]*Sus scrofa* chromosome.
[3]Haploblock identifier (chromosome specific).
[4]Position (bp) of the first SNP in the haplotype.
[5]Position (bp) of the last SNP in the haplotype.
[6]*P*-value from GEMMA (Wald test).
[7]References that showed QTL in the same chromosome region in pigs or in the corresponding homologous region in humans.

on SSC4 a region suggestively significant for ALT (also known as glutamic-pyruvic transaminase or GPT; routinely measured in serum as a biomarker of liver injury caused by drug toxicity, infection, alcohol and steatosis) was close to the *glutamic-pyruvic transaminase* (*GPT*) gene; other haplotypes suggestively associated with low-density lipoprotein cholesterol (LDL-Chol) and total cholesterol (Tot-Chol) were located on SSC3, close to the *apolipoprotein B* (*APOB*) gene

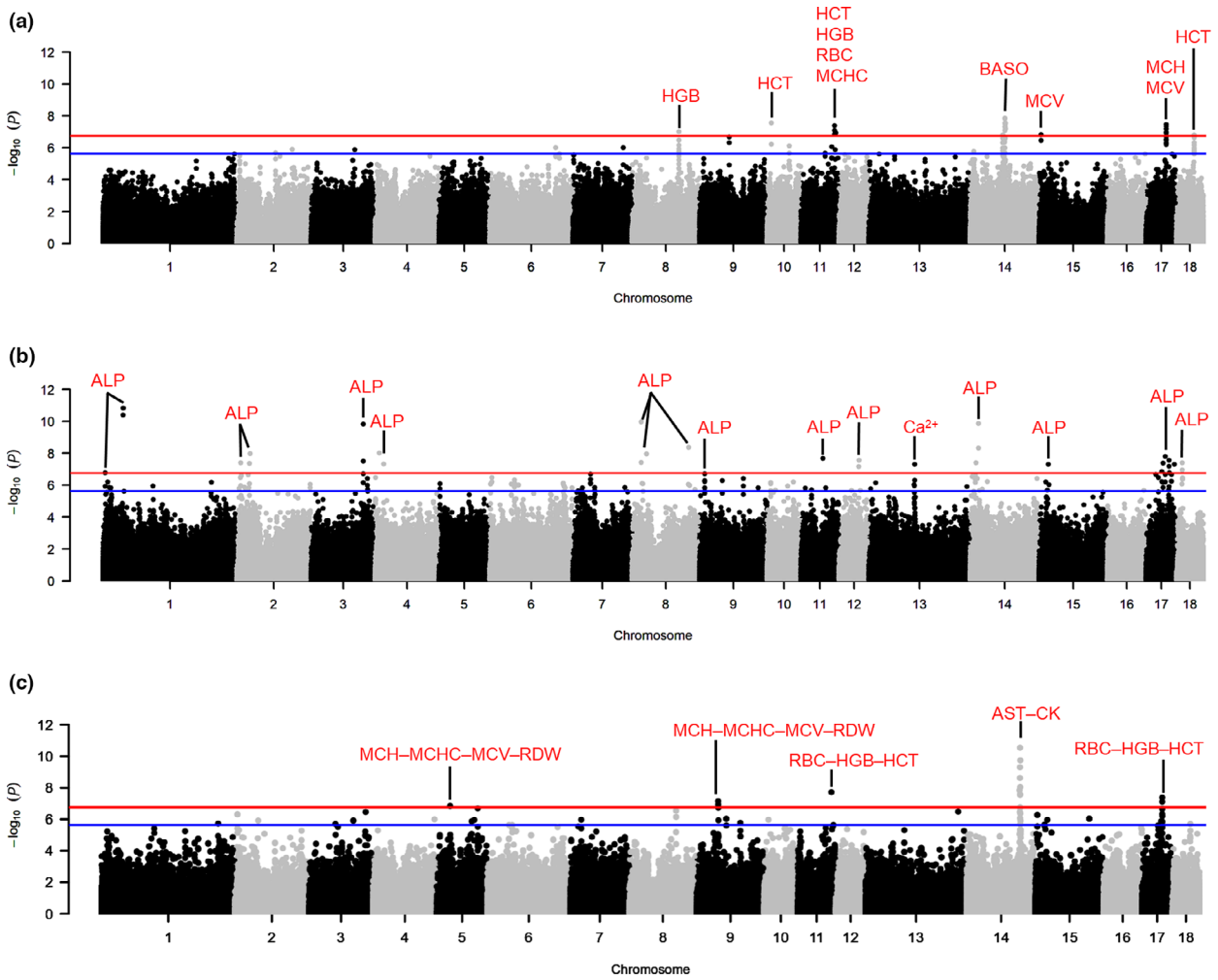

Figure 1 Over-imposed Manhattan plots displaying the results of the haplotype-based GWASs for the 33 measured blood parameters. (a) GWASs for haematological traits; (b) GWASs for clinical–biochemical traits; (c) multivariate GWASs for correlated blood parameters. The blue and red lines mark the suggestive and significance association thresholds, respectively. Full names of traits are given in the note to Table 1.

which is involved in its homeostasis (the association was confirmed also with the cluster T-Chol–LDL-Chol–HDL-Chol).

Other suggestively associated haplotypes involving putative causative genes were not detected in the previous single-marker analyses. Haplotypes on SSC2 spanning the *adenosine monophosphate deaminase 3* (*AMPD3*) gene were suggestively associated with the cluster T-Chol–LDL-Chol–HDL-Chol. Variants in the human *AMPD3* gene are associated with serum cholesterol level (Teslovich *et al.* 2010; Willer *et al.* 2013). Another suggestively associated haplotype with CK not previously detected in the single-marker analysis mapped on SSC6 near the *creatine kinase, muscle* (*CKM*) gene, which encodes a subunit of CK (Table S2).

Altogether, these results evidenced strong genetic components affecting and modulating blood composition and related enzyme activities. In summary, in this work we carried out haplotype-based GWASs for haematological and clinical–biochemical traits in pigs, refining the results of our previous studies (Bovo *et al.* 2016, 2019). The haplotype analysis highlighted several novel QTL that were not detected with the single-marker analysis and including, in several cases, obvious candidate genes, as deduced from their known biological functions or previous results in humans. These results further dissected the genetic architecture of parameters that could be used as proxies in breeding programmes for more complex traits, including disease resistance and resilience. In addition, they might help to better define the pig as animal model for several blood-related biological functions.

Acknowledgements

MB was supported by the Ministry of Foreign Affairs of Italy. This work was supported by fund from the MiPAAF

Innovagen project, the MIUR PRIN2017 PigPhenomics project and from the University of Bologna RFO 2018-2019 programmes.

Competing interests

The authors declare that hey do not have any competing interests.

Availability of data

The datasets used and analysed during the current study are available from the corresponding author on reasonable request.

References

Astle W.J., Elding H., Jiang T. *et al.* (2016) The allelic landscape of human blood cell trait variation and links to common complex disease. *Cell* **167**, 1415–29.

Bovo S., Schiavo G., Mazzoni G. *et al.* (2016) Genome-wide association study for the level of serum electrolytes in Italian Large White pigs. *Animal Genetics* **47**, 597–602.

Bovo S., Mazzoni G., Bertolini F. *et al.* (2019) Genome-wide association studies for 30 haematological and blood clinical–biochemical traits in Large White pigs reveal genomic regions affecting intermediate phenotypes. *Scientific Reports* **9**, 7003.

Chambers J.C., Zhang W., Sehmi J. *et al.* (2011) Genome-wide association study identifies loci influencing concentrations of liver enzymes in plasma. *Nature Genetics* **43**, 1131–8.

Delaneau O., Marchini J., Zagury J.-F. (2012) A linear complexity phasing method for thousands of genomes. *Nature Methods* **9**, 179–81.

Etim N.N., Williams M.E., Akpabio U., Offiong E.E.A. (2014) Haematological parameters and factors affecting their values. *Agricultural Science* **2**, 37–47.

Gan Q.F., Li Y.R., Lund M., Su G.S., Liang X.W. (2019) Genome-wide association study identifies loci linked to serum electrolyte traits in Chinese Holstein cattle. *Animal Genetics* **50**, 744–8.

Gong Y.-F., Lu X., Wang Z.-P. *et al.* (2010) Detection of quantitative trait loci affecting haematological traits in swine via genome scanning. *BMC Genetics* **11**, 56.

Kichaev G., Bhatia G., Loh P.-R. *et al.* (2019) Leveraging polygenic functional enrichment to improve GWAS power. *American Journal of Human Genetics* **104**, 65–75.

Ma X., Jia C., Fu D. *et al.* (2019) Analysis of hematological traits in polled yak by genome-wide association studies using individual SNPs and haplotypes. *Genes (Basel)* **10**, E463.

Okada Y., Kamatani Y. (2012) Common genetic factors for hematological traits in Humans. *Journal of Human Genetics* **57**, 161–9.

Pilling L.C., Atkins J.L., Duff M.O. *et al.* (2017) Red blood cell distribution width: Genetic evidence for aging pathways in 116,666 volunteers. *PLoS ONE* **12**, e0185083.

Ponsuksili S., Reyer H., Trakooljul N., Murani E., Wimmers K. (2016) Single- and Bayesian multi-marker genome-wide association for haematological parameters in pigs. *PLoS ONE* **11**, e0159212.

Reiner G., Fischer R., Hepp S., Berge T., Köhler F., Willems H. (2007) Quantitative trait loci for red blood cell traits in swine. *Animal Genetics* **38**, 447–52.

Reiner G., Clemens N., Fischer R. *et al.* (2009) Mapping of quantitative trait loci for clinical-chemical traits in swine. *Animal Genetics* **40**, 57–64.

Reyer H., Oster M., Wittenburg D., Murani E., Ponsuksili S., Wimmers K. (2019) Genetic contribution to variation in blood calcium, phosphorus, and alkaline phosphatase activity in pigs. *Frontiers in Genetics* **10**, 590.

Sun Y., Li Q., Hu Y. *et al.* (2016) Genomewide association study of immune traits in chicken F2 resource population. *Journal of Animal Breeding and Genetics* **133**, 197–206.

Teslovich T.M., Musunuru K., Smith A.V. *et al.* (2010) Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707–13.

Utsunomiya Y.T., Milanesi M., Utsunomiya A.T.H., Ajmone-Marsan P., Garcia J.F. (2016) GHap: an R package for genome-wide haplotyping. *Bioinformatics* **32**, 2861–2.

Vasquez L.J., Mann A.L., Chen L., Soranzo N. (2016) From GWAS to function: lessons from blood cells. *ISBT Science Series* **11**, 211–9.

Wei Y.J., Tsai K.S., Lin L.C. *et al.* (2011) Catechin stimulates osteogenesis by enhancing PP2A activity in human mesenchymal stem cells. *Osteoporosis International* **22**, 1469–79.

Willer C.J., Schmidt E.M., Sengupta S. *et al.* (2013) Discovery and refinement of loci associated with lipid levels. *Nature Genetics* **45**, 1274–83.

Yan G., Guo T., Xiao S. *et al.* (2018) Imputation-based whole-genome sequence association study reveals constant and novel loci for hematological traits in a large-scale swine F2 resource population. *Frontiers in Genetics* **9**, 401.

Zhang Z., Hong Y., Gao J. *et al.* (2013) Genome-wide association study reveals constant and specific loci for hematological traits at three time stages in a White Duroc × Erhualian F2 resource population. *PLoS ONE* **8**, e63665.

Zhang F., Zhang Z., Yan X. *et al.* (2014) Genome-wide association studies for hematological traits in Chinese Sutai pigs. *BMC Genetics* **15**, 41.

Zhou X., Stephens M. (2012) Genome-wide efficient mixed model analysis for association studies. *Nature Genetics* **44**, 821–4.

Zhu F., Cui Q.-Q., Yang Y., Hao J.-P., Yang F.-X., Hou Z.-C. (2020) Genome-wide association study of the level of blood components in Pekin ducks. *Genomics* **112**, 379–87.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Appendix S1 Supplementary methods: description of the animals, phenotypes, genomic data and genome-wide association analyses

Table S1 Descriptive statistics of the 33 measured blood parameters in the Italian Large White pig population

Table S2 Pearson's correlation coefficients (r) used to graph the network ($|r| > 0.4$; $P_r < 1.15 \times 10^{-4}$). For

comparison, Spearman's rank correlation coefficients (ρ) and the related significance (P_ρ) are provided. Short names are defined in Table S1.

Table S3 Haplotypes identified at the genome-wide suggestive level. Short names are defined in Table S1

Figure S1 Manhattan plots showing the results of the GWASs for the 33 blood parameters. Short names are defined in Table S1

Figure S2 Quantile–Quantile plots of the GWASs. Short names are defined in Table S1