Investigating the genetic architecture of disease resilience in pigs by genome-wide association studies of complete blood count traits collected from a natural disease challenge model

Xuechun Bai[1], Tianfu Yang[1], Austin M. Putz[2], Zhiquan Wang[1], Changxi Li[1,3], Frédéric Fortin[4], John C. S. Harding[5], Michael K. Dyck[1], PigGen Canada[†], Jack C. M. Dekkers[2], Catherine J. Field[1], and Graham S. Plastow[1*]

[1] Livestock Gentec, Department of Agricultural, Food and Nutritional Science, University of Alberta, Edmonton, AB, Canada

[2] Department of Animal Science, Iowa State University, Ames, IA, United States

[3] Lacombe Research and Development Centre, Agriculture and Agri-Food Canada, Lacombe, AB, Canada

[4] Centre de Développement du Porc du Québec, Inc., Quebec City, QC, Canada

[5] Department of Large Animal Clinical Sciences, University of Saskatchewan, Saskatoon, SK, Canada

[*]Correspondence:
Graham S. Plastow
plastow@ualberta.ca

[†]PigGen Canada authors are listed in the acknowledgements of the article

Abstract

Background: Genetic improvement for disease resilience is anticipated to be a practical method to improve efficiency and profitability of the pig industry, as resilient pigs maintain a relatively undepressed level of performance in the face of infection. However, multiple biological functions are known to be involved in disease resilience and this complexity means that the genetic architecture of disease resilience remains largely unknown. Here, we conducted genome-wide association studies (GWAS) of 465,910 autosomal SNPs for complete blood count (CBC)

traits that are important in an animal's disease response. The aim was to identify the genetic control of disease resilience.

Results: Univariate and multivariate single-step GWAS were performed on fifteen CBC traits measured from the blood samples of 2743 crossbred (Landrace × Yorkshire) barrows drawn at 2-weeks before, and at 2 and 6-weeks after exposure to a polymicrobial infectious challenge. Overall, at a genome-wise false discovery rate of 0.05, five genomic regions located on *Sus scrofa* chromosome (SSC) 2, SSC4, SSC9, SSC10, and SSC12, were significantly associated with white blood cell traits in response to the polymicrobial challenge, and nine genomic regions on multiple chromosomes (SSC1, SSC4, SSC5, SSC6, SSC8, SSC9, SSC11, SSC12, SSC17) were significantly associated with red blood cell and platelet traits collected before and after exposure to the challenge. By functional enrichment analyses using Ingenuity Pathway Analysis (IPA) and literature review of previous CBC studies, candidate genes located nearby significant single-nucleotide polymorphisms were found to be involved in immune response, hematopoiesis, red blood cell morphology, and platelet aggregation.

Conclusions: This study helps to improve our understanding of the genetic basis of CBC traits collected before and after exposure to a polymicrobial infectious challenge and provides a step forward to improve disease resilience.

Keywords

Genome-wide association studies, disease resilience, complete blood count, pigs, natural disease challenge model.

1 Background

The prevalence of infectious diseases caused by a multitude of pathogens results in high economic losses for the pig industry [1, 2]. Genetic improvement for disease resilience is a practical option to help address the problem of infectious disease as it can ensure production efficiency, because resilient animals are defined as maintaining a relatively undepressed performance in the face of disturbances caused by infection [3, 4]. Disease resilience is a complex trait composed of multiple biological functions, such as production, health, nutrient status, and other dynamic elements, including the efficiency of immune response and the rate of recovery from infection [5]. This complexity makes disease resilience hard to properly characterize and little is known about the genetic architecture that drives disease resilience. Alternatively, indirect selection of disease resilience based on immune-related traits may be a feasible breeding strategy, because the disease response of an animal largely depends upon its immunity [6, 7].

Blood cells comprise white blood cells, red blood cells, and platelets that are important elements of an animal's immune status [8]. Complete blood count (CBC) is one of the most common clinical tests performed to evaluate concentrations and relative proportions of these circulating blood cells, which may help to uncover the layers of immune system complexity [9]. Our previous study [10] found that CBC traits collected from blood samples of pigs in both healthy and challenged conditions at 2-weeks before, and 2 and 6-weeks after exposure to a polymicrobial challenge were moderately to highly heritable (0.08 ± 0.04 to 0.53 ± 0.05). Changes of each CBC trait between blood samples collected at different time points (e.g. the change of a CBC level from 2-weeks before to 2-weeks after exposure to the challenge) were also found to be heritable, with estimates ranging from 0.06 ± 0.04 to 0.24 ± 0.04 [10]. These

heritability estimates indicate the importance of the genetic component of CBC traits. Moreover, significant genetic correlations (either positive or negative) were found for several CBC traits collected after exposure to the challenge with the economically important production traits of grow-to-finish growth rate (GFGR) and treatment rate (TR) in response to the polymicrobial challenge (-0.82 ± 0.47 to 0.89 ± 0.26) [10], which may further indicate the potential of developing those CBC traits as indicator traits of disease resilience. In addition to these significant genetic correlations for CBC with GFGR and TR, our previous study [10] also found high genetic correlations ($\geq$ 0.40 ± 0.04) between the CBC traits. Changes in CBC traits between each time point were also found to be genetically correlated, with significant estimates ranging from -0.42 ± 0.21 to -0.92 ± 0.11 to 0.44 ± 0.22 to 0.98 ± 0.03 [10]. This allows multivariate models to be used for joint analyses of these genetically correlated traits, which provides the potential to improve statistical power and explore pleiotropy [19–22].

To date, some quantitative trait loci (QTL) have been identified for some blood cell traits in pigs under either healthy or disease challenged status by linkage and association analyses [11–18]. However, due to the use of a pathogen-specific challenge or a relatively low density of genetic markers, the genetic components of blood cell traits in pigs under typical commercial environments, where multiple disease-causing pathogens are present, remains largely unknown.

In this study, CBC traits were collected from pigs in a natural polymicrobial disease challenge model, as described by Bai et al. [10]. Standard univariate genome-wide association studies (GWAS) and multivariate GWAS based on a relatively high-density panel of 465,910 autosomal single-nucleotide polymorphisms (SNPs) were conducted for these CBC traits. The objectives were: (1) to reveal the genomic regions associated with the CBC traits and with their

changes in response to the polymicrobial challenge; and (2) to explore the underlying genetic architecture for disease resilience of pigs in the face of a polymicrobial infectious challenge.

2 Results

2.1 Descriptive statistics and genetic parameters

Descriptive statistics for the CBC data of 2743 animals, including both genotyped (n = 2593) and non-genotype (n = 150) animals, are shown in **Additional file 1: Figure S1 to S3**. Details about genetic parameters for the evaluated CBC traits, including heritabilities and genetic correlations, can be found in our previous study [10], which used the same 2593 genotyped animals. In addition to the genetic correlations with resilience already reported for these data by Bai et al. [10], we also found significant genetic correlations for platelet concentration in Blood 3 collected at 2-weeks after exposure to a polymicrobial infectious challenge with GFGR (0.40 ± 0.22) and TR (-0.46 ± 0.26), and for the change of monocyte concentration from Blood 1 to Blood 3 (MONOΔ13) collected at 2-weeks before and 2-weeks after exposure to the challenge with GFGR (0.63 ± 0.21).

2.2 Population structure

As false positive results can be introduced in GWAS by confounding effects due to population stratification, multidimensional scaling (MDS) plots (**Additional file 2: Figure S4**) were generated to provide a visualization of the population structure in the first three dimensions (C1, C2, and C3). Animals tended to cluster by farm of origin, as they shared a similar genetic background when they came from the same farm. Since batches were nested within farms and coded uniquely, population stratification associated with the farm effect was accounted for in the association analysis model by fitting the fixed effect of batch. The genomic inflation factors of

single-step (SSGWAS) for the CBC traits ranged from 0.98 to 1.06, suggesting that there was no population stratification that confounded the GWAS results.

2.3 Association results and estimates for SNP effects

2.3.1 White blood cell traits

Five genomic regions were found to be significantly associated with white blood cell traits at a genome-wise false discovery rate (*FDR*) of 0.05. Of note, SNPs located on *Sus scrofa* chromosome (SSC) 4, SSC10, and SSC12 were found to be associated with eosinophil concentration in Blood 3, which was collected 2 weeks after exposure to the challenge (EOSB3). Meanwhile, SNPs on SSC2 and two adjacent floating SNPs (significant SNPs without a group of supportive SNPs) on SSC9 were identified to be associated with MONOΔ13. The Manhattan and Q-Q plots for EOSB3 and MONOΔ13 are shown in **Additional file 2: Figures S5 and S6**. Top lead SNPs (the most significant SNP with a group of supportive SNPs) for significant associations (genome-wise $FDR < 0.05$) with EOSB3 and MONOΔ13 are shown in **Table 1**. For EOSB3, the additive genetic variances explained by the 1 Mb window of the top lead SNPs (SNP1, SNP2, SNP3) and their adjacent SNPs on SSC4, SSC10, and SSC12 were estimated to be 0.46, 0.35, and 0.53% of the additive genetic variance for EOSB3, respectively. SNP4 was a floating SNP on SSC2 and its 1 Mb window only explained 0.12% of the additive genetic variance for MONOΔ13. The 1 Mb window for SNP5, the top lead SNP on SSC9, was estimated to explain about 1.23% of the additive genetic variance for MONOΔ13.

Estimates of additive and dominance effects for the top significant SNPs (genome-wise $FDR < 0.05$, including both top lead and top floating SNPs) associated with EOSB3 and MONOΔ13 are summarized in **Table 1**. A significant dominance effect ($p < 0.05$) was only

identified for SNP2, which was associated with EOSB3. Estimates of additive effects were found to be significant ($p < 0.05$) for all SNPs that were associated with EOSB3 and MONOΔ13. For EOSB3, estimates of additive effects were -0.05 ± 0.01, 0.14 ± 0.04, and -0.06 ± 0.01 for SNP1, 2 and 3, respectively. Estimates of additive effects for MONOΔ13 were 0.08 ± 0.02 and -0.08 ± 0.02 for SNP 4 and 5, respectively.

2.3.2 Red blood cell and platelet traits

Nine genomic regions were found to be significantly associated with red blood cell and platelet traits at the genome-wise *FDR* of 0.05. The Manhattan plots and Q-Q plots are shown in **Additional file 2: Figures S7 to S12**. The four top lead SNPs for significant associations (genome-wise $FDR < 0.05$) and estimates of additive genetic variances explained by these top lead SNPs and their adjacent SNPs in a 1 Mb window are summarized in **Table 2**. Five floating SNPs (genome-wise $FDR < 0.05$) that explained small amounts of the additive genetic variance (0.05% to 0.21%) for associated red blood cell and platelet traits were found and are summarized in **Table 3**. Of note, several pleiotropic SNPs associated with red blood cell or platelet traits were identified by multivariate SSGWAS of CBC traits in Blood 1, 3, and 4 (collected at 2-weeks before, 2- and 6-weeks after the challenge, respectively). High genetic correlations were found between mean corpuscular hemoglobin (MCH), mean corpuscular volume (MCV), and red blood cell concentration (RBC) traits (**Additional file 1: Table S1**), and also between all three sampling time points for each of these traits ($\geq 0.77 \pm 0.08$) [10]. Therefore, pleiotropic SNP7 on SSC6 was identified as the top lead and pleiotropic SNP for MCH in Blood 1 and for both MCV and RBC traits in all three blood samples (**Table 2**). The percentage of additive genetic variance explained by the 1 Mb window of SNP7 and its adjacent SNPs ranged from 0.29 to

0.57% for its associated traits. Moreover, SNP8 was the top lead and pleiotropic SNP on SSC8, which was associated with MCH, MCV, and RBC traits in all three blood samples. The percentages of additive genetic variance explained by SNP8 and its adjacent SNPs in a 1 Mb window were estimated to range from 0.28 to 0.35% for its associated traits. SNP9 on SSC17 was the top lead and pleiotropic SNP for mean platelet volume (MPV) in Blood 1 and 4. Together with adjacent SNPs in a 1 Mb window, SNP7 was estimated to explain about 0.49 and 0.40% of the additive genetic variances for MPV in Blood 1 and 4, respectively. Significant associations (genome-wise *FDR* < *0.05*) for SNP7 with MCV in Blood 1 (genome-wise *FDR* = 0.003) and for SNP8 with MCV in Blood 4 (genome-wise *FDR* = 0.04) were also found by univariate SSGWAS but at a lower significance level compared to the multivariate SSGWAS. Meanwhile, univariate SSGWAS only indicated suggestive associations (genome-wise *FDR* of 0.10) for SNP8 with RBC in Blood 1 (genome-wise *FDR = 0.09*) and with MCV in Blood 1 (genome-wise *FDR = 0.08*).

For red blood cell and platelet traits, the estimates of additive and dominance effects for the top lead SNPs are summarized in **Table 2** and for the top floating SNPs in **Table 3**. Of note, the additive effects for pleiotropic SNPs showed a tendency of affecting each CBC trait in the three blood samples in the same way, including SNP7 for MCV, SNP8 for MCH and MCV, SNP9 for MPV, SNP10 for MCV, SNP11 for PLT, and SNP12 for MCH. For pleiotropic SNP8, no significant additive effect was found for RBC traits.

2.4 Candidate genes and functional enrichment results

Candidate genes that were located within a maximum distance of 1 Mb on either side of the lead SNPs based on a genome-wise *FDR* < *0.1* for the associated CBC traits are summarized

in **Additional file 3: Table S2**. Candidate gene functions for each CBC trait were explored with the Ingenuity Pathway Analysis (IPA) database. Functions in the molecular and cellular functions category and the physiological system development and function category that were significantly enriched ($p < 0.05$) are summarized in **Additional file 4**. For the category of molecular and cellular functions, enriched functions such as cell morphology, cell-to-cell signaling and interaction, cellular development, cellular growth and proliferation, and cell movement were commonly identified for the candidate gene lists for white blood cell traits collected after exposure to the challenge, and the pleiotropic candidate gene lists for red blood cell and platelet traits collected before and after exposure to the challenge. Moreover, hematological system development and function, hematopoiesis, immune cell trafficking, and immune response were common enriched functions in the category of physiological system development and function that were identified for the gene lists for these CBC traits.

Candidate genes that have been reported by previous studies of pigs, human, mice, or rats to be functionally and biologically related to the same category of blood cells, as explored here, are summarized in **Table 4**. A group of immunity genes on SSC2 has been reported to be functionally and biologically related to monocytes, including *TICAM2*, *TMED7*, and *CDO1*, which were located proximal to SNP4, and *COMMD10*, which harbored SNP4 (**Table 4**). An overview of the location of these candidate genes and the distribution of all the SNPs in this region on SSC2 is shown in the LD haplotype map in **Additional file 3: Figure S13**. SNP6 on SSC4 is intronic within candidate gene *SPTA1* and the LD haplotype map for this region is shown in **Additional file 3: Figure S14**. In **Table 4**, a group of candidate genes, including *THAP11*, *PSMB10*, *LCAT*, and *SLC12A4*, was reported to be functionally related to red blood

cells and were located close to SNP7 in the same haplotype block on SSC6 (**Additional file 3: Figure S15**). SNP8 on SSC8 was found to be in LD ($r^2 > 0.30$) with SNPs in the *PDGFRA* gene (**Additional file 3: Figure S16**).

3 Discussion

3.1 Potential roles of candidate genes

Functional enrichment analyses for the candidate gene lists for CBC traits indicated multiple enriched functions that can be considered as functionally and biologically relevant to white blood cell traits in response to a polymicrobial infectious challenge, and red blood cell and platelet traits that were collected before and after exposure to the challenge, such as cell growth and proliferation, cell-to-cell signaling and interaction, cell movement, cellular development, hematological system development and function, hematopoiesis, immune cell trafficking, and immune response.

The candidate genes in **Table 4** have been reported to be relevant to particular types of CBC traits by studies in pigs, human, mice, and rat, which may help us to further understand the functions of these candidate genes related to CBC traits in response to the polymicrobial challenge. Of note, candidate genes *ARHGEF2* (*Rho/Rac guanine nucleotide exchange factor*), *TGFB2* (*transforming growth factor beta 2*), and *MIR21* (*microRNA miR-21*) were identified to be functionally and biologically relevant to eosinophils. The product of *ARHGEF2* regulates the activity of GTPases and has been identified to be highly expressed in eosinophils. GTPases are known to be involved in mediator release from granulocytes, which is a crucial event in the activation of eosinophils and neutrophils during inflammation [23, 24]. *TGFB2* has also been found to be expressed mainly in eosinophils, and greater expression of *TGFB2* has been

identified to be associated with persistent eosinophilic inflammation (severe asthma) in human [25]. However, in the polymicrobial challenge, an increase in the number of eosinophils may be associated with parasitic infection (e.g. *Ascaris suum*) rather than respiratory disease. Eosinophils play an important role of killing larvae by releasing the toxic content of their granules as part of the immune response [26]. Thus, further investigations are warranted to investigate the functional relationships between the expression of *TGFB2* and response to the challenge. Expression of *MIR21* has not been identified in eosinophils but in other white blood cells, including lymphocytes, monocytes, macrophages, and dendritic cells, which work collaboratively with eosinophils in the immune response [27–29]. Although the mRNA targets for *MIR21* are complex and remain an area of active investigation, it has been demonstrated that *MIR21* acts as a key signal mediating the balance of the inflammatory reaction to promote healing, resolution, and a return to homeostasis [27].

For the candidate genes on SSC2, the product of *COMMD10* (*COMM domain containing 10*) has been found to be related to the function of phagosomes in murine macrophages, which promotes phagolysosome maturation and facilitates the timely killing of pathogens [30, 31]. The product of *ATG12* (*autophagy related 12*) is involved in autophagy of circulating monocytes for degradation and recycling of cellular components, which prevents apoptosis (programmed cell death) of monocytes and is essential for monocyte-macrophage differentiation and cytokine production in the innate immune response [32, 33]. The product of *CDO1*, cysteine dioxygenase type 1, catalyzes taurine synthesis and it is commonly accepted that taurine plays an important role in the immune system as an antioxidant to protect phagocytes, including macrophages, from oxidative stress caused by the generation of reactive oxygen species at the site of inflammation

[34–37]. Both *TMED7* (*transmembrane emp24 domain-containing protein 7 precursor*) and *TICAM2* (*toll-like receptor adaptor molecule 2*) are immunity genes and their products are involved in the function of toll-like receptors (TLRs), which are expressed on macrophages and monocytes and are responsible for the sensing of pathogen-associated-molecular-patterns in the extracellular environment and in endosomes [38-40]. Of note, overexpression of *TMED7* has been found to be associated with inhibition of MyD88-independent TLR4 signaling and the protein encoded by *TICAM2* has been identified as a bridge adaptor recruiting TLRs to mediate innate immune responses [38–40]. In addition, *NAMPT* (*nicotinamide phosphoribosyl transferase*) on SSC9 has been found to be functionally and biologically related to monocytes, and its gene product has been found to play an important role in governing monocyte recruitment and in monocyte-macrophage differentiation [41, 42].

For red blood cells, the majority of candidate genes reported here have been identified as key components involved in hematopoiesis and erythropoiesis responsible for the differentiation and development of red blood cells, including *MNDA* (*myeloid cell nuclear differentiation antigen*) on SSC4, *CBFB* (*core-binding factor subunit beta*) and *THAP11* (*THAP domain containing protein 11*) on SSC6, *PDGFRA* (*platelet derived growth factor receptor alpha*) and *KIT* (*KIT proto-oncogene, receptor tyrosine kinase*) on SSC8, and *RARA* (*retinoic receptor alpha*) and *THRA* (*thyroid hormone receptor alpha*) on SSC12 [43–52]. In addition, *SPTA1* (*spectrin alpha*) on SSC4 encodes a protein in the red blood cell membrane, the products of *LCAT* (*lecithin-cholesterol acyltransferase*) and *SLC12A4* (*Potassium/Chloride Cotransporter 1*) on SSC6 regulate the lipid composition in the red blood cell membrane and cell swelling, respectively, and all these gene products work together to maintain the normal volume and

biconcave shape of red blood cells, which helps to ensure the biological and biomechanical functions of the cells [53–57]. *ACKR1* (*atypical chemokine receptor 1*) on SSC4 and *PSMB10* (*proteasome subunit beta type 10*) on SSC6 are candidate genes that have been shown to be involved in the immune response of red blood cells. The receptor ACKR1 expressed in red blood cells was found to regulate immune responses by interacting with chemokines, and which works as a blood-based chemokine buffer involved with the uptake and degradation of chemokines [58]. Meanwhile, ACKR1 has also been identified as an essential regulator of hematopoiesis and erythropoiesis promoting interactions between nuclear progenitor red blood cells and hematopoietic stem cells in the bone marrow [58, 59]. PSMB10 is found to be responsible for intracellular protein degradation and generation of peptides that bind to class I major histocompatibility complex (MHC) molecules [60]. The MHC molecules display these peptides to cytotoxic $CD8^+$ T cells to support their activity of immune surveillance [61]. Further, through a study of anemia caused by congenital red blood cell aplasia in human, PSMB10 has been suggested to be functional in the MHC class I machinery in mature red blood cells in response to inflammatory signaling [62].

Candidate genes for platelet traits were annotated into two major functions, platelet aggregation and megakaryopoiesis. Platelet aggregation involves platelet-to-platelet adhesion, which is essential for effective hemostasis following injury and bleeding, and megakaryopoiesis is the process of differentiation and development of platelets [63]. Among them, *CD9* (*CD9 antigen*) on SSC5 encodes a major platelet cell surface glycoprotein and plays dual roles in megakarypoiesis and platelet aggregation [64–67]. The products of *VWF* (*von Willebrand factor*), *PHB2* (*prohibitin 2*), and *GNB3* (*G protein subunit beta 3*) on SSC5 and *GNAS* (*guanine*

nucleotide binding protein) on SSC17 were found to be involved in platelet aggregation [68–73]. In addition to megakaryopoiesis, tubulin beta class VI coded by *TUBB1* on SSC17 has been reported to play a role in maintaining platelet morphology [74–76].

3.2 Overlap with previously discovered QTL

In addition to the novel QTL for CBC traits identified in this study, some of the QTL identified have been previously reported. QTL on SSC8 located nearby the *KIT* gene were found to be associated with MCH, MCV, and RBC in this study. In addition, this region has also been identified to show a significant effect on the levels of NEU and HCT in the crossbreds of European Wild Boar × Yorkshire and Landrace × Yorkshire subsequent to stress and disease challenges [11, 18]. For QTL on SSC5 that associated with PLT traits here, Reiner et al. [13] found them to be associated with red blood cell traits in Pietrain × Meishan pigs including HCT, HGB, and RBC traits. These results may be caused by the common myeloid progenitors for all cells mentioned above. Moreover, it may also further indicate the pleiotropic roles of QTL involved in the functions of different blood cells. Apart from studies in pigs, the candidate gene *SPTA1* associated with MCHC has also been identified by GWAS for red blood cell traits in human, which also functions in maintaining the shape and deformability of human red blood cells [77].

3.3 Potential links with disease resilience

Although the QTL uncovered for blood cell traits have small effects in this study, which has also been found in previous GWAS for blood cell traits of pigs and human [17, 77], the genes involved in these QTL are suggested to be involved in hematopoiesis and immune responses in the face of a polymicrobial infectious challenge. In turn, they may contribute to

disease resilience, as hematopoiesis and immune response are collaborative mechanisms that play essential roles in defending against pathogens, maintaining homeostasis, and preventing death from the infection [6, 78, 79]. None of the QTL identified for the CBC traits were pleiotropic with GFGR or TR in response to the challenge. However, some candidate genes are known to have pleiotropic effects among different CBC traits and play roles in both hematopoiesis and immune response. For example, *KIT* may be a pleiotropic gene for multiple blood cell populations in response to stress and disease challenge, and *ACKR1* exhibits pleiotropic effects on hematopoiesis and immune responses, as discussed above [11, 18, 58, 59]. Accordingly, these results highlight the importance of further investigating and validating the function of such pleiotropic genes in disease resilience.

4 Conclusions

In this study, we identified fourteen genomic regions that were significantly associated (genome wise $FDR < 0.05$) with CBC traits collected from the natural polymicrobial challenge model, including five for white blood cell traits and nine for red blood cell and platelet traits. Candidate genes or regions located nearby significant SNPs were found to have potential roles in immune response pathways, red blood cell morphology, platelet aggregation, and hematopoiesis, including granulopoiesis and granulocytic differentiation, erythropoiesis, and megakarypoiesis. These results complement previous GWAS for blood cell traits in pigs and contribute to improving our understanding of the genetic basis of blood cell composition before and after exposure to a polymicrobial infectious challenge. This study also advances understanding of the genetic control of disease resilience, as blood cells are key players in an animal's immune response and are recruited by hematopoiesis. Validation and identification of the candidate genes

and causal mutations are necessary to further investigate and develop the use of CBC traits to enhance genetic improvement of disease resilience for the pig industry.

5 Methods

5.1 Natural disease challenge model and phenotypic traits

Details of the natural disease challenge model (NDCM) and the collection of phenotypic traits are described in Bai et al. [10] and Putz et al. [80]. Briefly, the NDCM was established to simulate a polymicrobial infectious challenge and severe disease pressure often found at the commercial level of pig production. A 3-week healthy quarantine nursery and a test station that consisted of a 4-week second-stage nursery and an approximately 16-week grow-to-finish stage were the two main facilities in the NDCM. A total of 2743 healthy F1 crossbred (Landrace × Yorkshire) barrows provided by company Centre de Développement du Porc du Québec, Inc were introduced into the NDCM in 42 batches at 3-week intervals after weaning. Pigs that went through the NDCM were first exposed to the polymicrobial infectious challenge in the second-stage nursery. The challenge was established by co-introducing commercial seeder pigs with known diseases together with the first four batches of healthy barrows into the NDCM. Common disease-causing pathogens found in commercial farms were the major target pathogens in the NDCM, including multiple strains of porcine reproductive and respiratory syndrome virus and swine influenza A virus, various respiratory and enteric bacterial pathogens (such as *Mycoplasma hyopneumoniae*, *Haemophilus parasuis*, *Brachyspira hampsonii*, *Salmonella enterica* serovar typhimurium, and *Streptococcus suis*), and two parasites (*Cystoisospora suis* and *Ascaris suum*). Subsequently, the challenge model was maintained as a continuous flow system by nose-to-nose contact between the new batch of healthy barrows and the preceding

challenged group during the first week of challenge nursery period. The disease pressure varied by batch and on a seasonal basis; not all pigs were exposed to all the same pathogens, mimicking what occurs naturally on a commercial farm. A part of animals died due to infectious diseases after exposure to the challenge, the other animals reaching the target slaughter weight at approximately 181 days old were slaughtered commercially and entered the food chain after the study.

In total, four sets of blood samples (Blood 1, Blood 2, Blood 3, and Blood 4) were collected from the jugular vein. Blood 2 was collected immediately before entry into the challenge nursery and polymicrobial infectious challenge at 40 days of age. Whole EDTA-anticoagulated blood was collected for CBC analyses at Blood 1, Blood 3, and Blood 4 (2-weeks before, and at 2- and 6-weeks after exposure to the challenge) at an average age of 26 days, 54 days, and 82 days, respectively, using the ADVIA® 2120i Hematology System (Siemens Healthineer, Erlangen, Germany) within 24 to 48 hours of collection. CBC traits used for this study were described previously [10] and consisted of three categories: (1) six white blood cell traits, including total white blood cell concentration (WBC, 10^3/μL), neutrophil concentration (NEU, 10^3/μL), lymphocyte concentration (LYM, 10^3/μL), monocyte concentration (MONO, 10^3/μL), eosinophil concentration (EOS, 10^3/μL), and basophil concentration (BASO, 10^3/μL); (2) seven red blood cell traits, consisting of red blood cell concentration (RBC, 10^6/μL), hemoglobin concentration (HGB, g/L), hematocrit (HCT, %), mean corpuscular volume (MCV, fL), mean corpuscular hemoglobin (MCH, pg), mean corpuscular hemoglobin concentration (MCHC, g/L), and red blood cell distribution width (RDW, %); and (3) two platelet traits, including platelet concentration (PLT, 10^3/μL) and mean platelet volume (MPV, fL). CBC traits

were assessed at individual time points (Blood 1, Blood 3, Blood 4) as well as calculating and testing the changes between time points: Blood 1 to Blood 3 (Δ13, calculated as Blood 3−Blood 1), Blood 3 to Blood 4 (Δ34, Blood 4−Blood 3), and Blood 1 to Blood 4 (Δ14, Blood 4−Blood 1). All white blood cell traits in Blood 1, Blood 3, and Blood 4 were $\log_{10}$-transformed to reduce skewness of the distribution.

Body weights and veterinary treatments were recorded on an individual pig basis and used to calculate production traits, grow-to-finish growth rate (GFGR), and treatment rate (TR), which were regarded as economically important traits related to disease resilience. The GFGR for each animal was estimated using linear regression of body weights measured every 3 weeks in the grow-to-finish phase on ages from an average age of 69 days of age to the endpoint, i.e., either mortality due to infectious diseases or reaching the target slaughter weight at approximately 181 days old. The TR for each animal was the number of individual treatment events standardized by the number of days the animal spent in the NDCM (TR = number of treatment events/days × 100%).

5.2 SNP array genotyping and quality control

The genotyping using the 650K Affymetrix Axiom® Porcine Genotyping Array was performed at Delta Genomics (Edmonton AB, Canada). Missing genotypes were imputed with the reference genome (Sscrofa 11.1) using FImpute [81]. The preGSf90 software in the BLUPF90 suite of programs was used for quality control to remove SNPs with a minor allele frequency lower than 0.01 and call rates lower than 0.90 [82]. After the quality control, there were 2593 genotyped animals with 465,910 autosomal SNPs remained for the subsequent analyses.

5.3 Population stratification and linkage disequilibrium estimation

Population stratification among genotyped animals was investigated using PLINK 1.90 based on pairwise identity-by-state (IBS) distance, which was estimated using SNP genotypes [83]. A multidimensional scaling (MDS) plot was drawn by the 'ggplot2' package in R [83] to show the first three dimensions of the population structure. The genomic inflation factor and quantile–quantile (Q–Q) plots were applied to assess genomic inflation of the test statistics using the R packages of 'GenABEL' and 'qqman' [84-86]. The linkage disequilibrium (LD) of pairwise SNPs was measured as the squared correlation (r^2) of allele counts for the two SNPs and haplotype blocks were built using the Haploview software [87, 88].

5.4 Single-step GWAS and Models

Univariate and multivariate single-step GWAS (SSGWAS) for CBC traits were implemented in the BLUPF90 suite of programs with the joint pedigree-genomic relationship matrix (**H**) for single-marker associations, accommodating both genotyped (n = 2593) and non-genotyped (n = 150) animals [82, 89]. Details for algorithms employed for these analyses have been described by Aguilar et al. [89]. Briefly, BLUPF90 combines the algorithms for single-step GBLUP and for back-solving to obtain estimates and p-values for SNP associations from estimates of breeding values. The genomic relationship matrix (***G***) for genotyped animals was constructed as $\mathbf{ZZ}'/2\sum p_i(1-p_i)$, where the **Z** matrix contains centered SNP genotype codes and p_i is the minor allele frequency for SNP *i* [90]. The p-values for SNP associations were adjusted for multiple testing by the Benjamini and Hochberg correction (false discovery rate, *FDR*) [84, 91]. An *FDR* threshold of 0.05 was used to control false positive results and to declare significant associations. The most significant SNP above the genome-wise *FDR* of 0.05 in each

genomic region were referred to as the top significant SNP, which were further separated into top lead and top floating SNPs, which referred to top significant SNPs in a genomic region with or without a group of supportive SNPs, respectively.

The univariate mixed linear model used for GWAS can be described as follows:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{Wc} + \mathbf{e}$$

where $\boldsymbol{y}$ is a vector of observations on a CBC trait for all individuals, **b** is a vector of fixed effects, including the effect of batch and the covariate of bleeding age, **X** is a design matrix relating observations to the fixed effects, **a** is a vector of breeding values, **Z** is a design matrix that relates observations to breeding values, including genotyped and ungenotyped animals, and $\boldsymbol{e}$ is a vector of residual effects. Vector **c** represents a stack of vectors ($\mathbf{c}_{\mathbf{Litter}}$, $\mathbf{c}_{\mathbf{Pen1}}$, $\mathbf{c}_{\mathbf{Pen2}}$, and $\mathbf{c}_{\mathbf{Pen3}}$) of independent and uncorrelated random environmental effects, including litter ($\mathbf{c}_{\mathbf{Litter}}$) and pen effects in the quarantine unit ($\mathbf{c}_{\mathbf{Pen1}}$), in the test station second-stage nursery ($\mathbf{c}_{\mathbf{Pen2}}$), and in the test station grow-to-finish stage ($\mathbf{c}_{\mathbf{Pen3}}$). These random environmental effects were tested and fitted in the model for each CBC trait when they were significant ($p < 0.05$). Matrix **W** ($\mathbf{W}_{\mathbf{Litter}}$, $\mathbf{W}_{\mathbf{Pen1}}$, $\mathbf{W}_{\mathbf{Pen2}}$, and $\mathbf{W}_{\mathbf{Pen3}}$) is a stack of incidence matrices that relate observations to the corresponding random environmental effects. The random effects fitted for each of CBC traits were the same as Bai et al. [10].

Assuming the random effects **c** and **e** are uncorrelated and identically distributed, the (co-)variances of random effects for univariate models are:

$$\mathrm{var}\begin{bmatrix}\mathbf{a}\\ \mathbf{c}\\ \mathbf{e}\end{bmatrix} = \begin{bmatrix}\mathbf{H}\sigma_a^2 & 0 & 0\\ 0 & \mathbf{I}\boldsymbol{\sigma}_{\mathbf{c}}^{\mathbf{2}} & 0\\ 0 & 0 & \mathbf{I}\sigma_e^2\end{bmatrix}$$

where **H** is the joint pedigree-genomic relationship matrix for genotyped and non-genotyped animals as mentioned above, **I** is the identity matrix, σ_a^2 is the additive genetic variance, $\boldsymbol{\sigma}_{\mathbf{c}}^{\mathbf{2}}$ represents a stack of random effect variances (e.g. $\boldsymbol{\sigma}_{\mathbf{c}}^{\mathbf{2}} = \begin{bmatrix} \sigma^2_{c_{Litter}} & 0 \\ 0 & \sigma^2_{c_{Pen1}} \end{bmatrix}$, when the random effects $\mathbf{c_{Litter}}$ and $\mathbf{c_{Pen1}}$ are significant and fitted in the model for a trait), and σ_e^2 is the residual variance.

The model for multivariate analyses resembles a stack of univariate models for each of the traits that were found to be highly genetic correlated in [10], which can be written as [19, 92]:

$$\begin{bmatrix} \mathbf{y_1} \\ \mathbf{y_2} \\ \mathbf{y_3} \end{bmatrix} = \begin{bmatrix} \mathbf{X_1 b_1 + Z_1 a_1 + W_1 c_1 + e_1} \\ \mathbf{X_2 b_2 + Z_2 a_2 + W_2 c_2 + e_2} \\ \mathbf{X_3 b_3 + Z_3 a_3 + W_3 c_3 + e_3} \end{bmatrix}$$

For each trait in the multivariate model, the same effects were fitted as in the univariate models. For multivariate models, assuming random effects $\mathbf{c_n}$ and residual effects $\mathbf{e_n}$ for the n^{th} trait (n = 1, 2, 3) are uncorrelated and identically distributed, the (co-) variances of random effects are:

$$\text{Var}\begin{bmatrix} \mathbf{a_1} \\ \mathbf{a_2} \\ \mathbf{a_3} \\ \mathbf{c_1} \\ \mathbf{c_2} \\ \mathbf{c_3} \\ \mathbf{e_1} \\ \mathbf{e_2} \\ \mathbf{e_3} \end{bmatrix} = \begin{bmatrix} \mathbf{H}\sigma^2_{a_1} & \mathbf{H}\sigma_{a_{12}} & \mathbf{H}\sigma_{a_{13}} & 0 & 0 & 0 & 0 & 0 & 0 \\ \mathbf{H}\sigma_{a_{21}} & \mathbf{H}\sigma^2_{a_2} & \mathbf{H}\sigma_{a_{23}} & 0 & 0 & 0 & 0 & 0 & 0 \\ \mathbf{H}\sigma_{a_{31}} & \mathbf{H}\sigma_{a_{32}} & \mathbf{H}\sigma^2_{a_3} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbf{I}\boldsymbol{\sigma}^2_{\mathbf{c_1}} & \mathbf{I}\boldsymbol{\sigma}_{\mathbf{c_{12}}} & \mathbf{I}\boldsymbol{\sigma}_{\mathbf{c_{13}}} & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbf{I}\boldsymbol{\sigma}_{\mathbf{c_{21}}} & \mathbf{I}\boldsymbol{\sigma}^2_{\mathbf{c_2}} & \mathbf{I}\boldsymbol{\sigma}_{\mathbf{c_{23}}} & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbf{I}\boldsymbol{\sigma}_{\mathbf{c_{31}}} & \mathbf{I}\boldsymbol{\sigma}_{\mathbf{c_{32}}} & \mathbf{I}\boldsymbol{\sigma}^2_{\mathbf{c_3}} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{I}\sigma^2_{e_1} & \mathbf{I}\sigma_{e_{12}} & \mathbf{I}\sigma_{e_{13}} \\ 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{I}\sigma_{e_{21}} & \mathbf{I}\sigma^2_{e_2} & \mathbf{I}\sigma_{e_{23}} \\ 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{I}\sigma_{e_{31}} & \mathbf{I}\sigma_{e_{32}} & \mathbf{I}\sigma^2_{e_3} \end{bmatrix}$$

where $\sigma_{a_{12}} = \sigma_{a_{21}}$, $\sigma_{a_{13}} = \sigma_{a_{31}}$, and $\sigma_{a_{23}} = \sigma_{a_{32}}$ are additive genetic covariances between traits, $\boldsymbol{\sigma}_{\mathbf{c_{12}}} = \boldsymbol{\sigma}_{\mathbf{c_{21}}}$, $\boldsymbol{\sigma}_{\mathbf{c_{13}}} = \boldsymbol{\sigma}_{\mathbf{c_{31}}}$, and $\boldsymbol{\sigma}_{\mathbf{c_{23}}} = \boldsymbol{\sigma}_{\mathbf{c_{32}}}$ are covariances for common random effects

between two traits, $\sigma_{e_{12}} = \sigma_{e_{21}}$, $\sigma_{e_{13}} = \sigma_{e_{31}}$, and $\sigma_{e_{23}} = \sigma_{e_{32}}$ are covariances for residual effects between two traits.

5.5 Post-GWAS marker effect analyses

The percentage of additive genetic variance explained by a 1 Mb window (with a median of 224 adjacent SNPs) was estimated by conducting window-based inferences for additive genetic variance in the BLUPF90 suit of programs [82]. Each chromosome was evaluated by using a sliding (moving) 1 Mb window by using every SNP on the chromosome as a starting SNP for a window segment [82]. Therefore, a top significant SNP was contained within multiple windows and among them, the largest percentage of additive genetic variance explained by a window that contained that top significant SNP was reported for each trait.

Additive and dominance effects of each top significant SNP were estimated using the BLUPF90 suite of programs [82] based on the following model:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{Wc} + \mathbf{v}\alpha + \mathbf{d}\delta + \mathbf{e}$$

where **y**, **X**, **b**, **Z**, **a**, **W**, **c**, and **e** are the same as for the univariate model described above; **v** is a vector of the top significant SNP genotypes coded as -1, 0, and 1 for the AA, AB, and BB, respectively; α is the additive effect; **d** is a vector of dominance coded as 1 for heterozygous genotype (AB) and 0 for homozygous genotypes (AA and BB); δ is the dominance effect. Vectors **v** and **d** were fitted as covariates and the top significant SNPs were fitted one by one in the model. The likelihood ratio test was used to test the significance of the additive and dominance effects for each of the top significant SNPs by comparing full models to restricted models that constrained additive or dominance effects to zero using the REMLF90 program of

BLUPF90 [82]. When the dominance effect was not significant ($p > 0.05$), the additive effect for a SNP was re-estimated by removing the dominance effect from the model.

5.6 Post-GWAS bioinformatics analyses

Ingenuity Pathway Analysis (IPA) (Ingenuity® Systems, Redwood City, CA; https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis/, IPA Spring 2020 release) was used for functional enrichment analyses of candidate genes in the significant genomic region. A maximum distance of 1 Mb on either side of the lead SNPs based on a genome-wise $FDR < 0.10$ was used to search for candidate genes for each of the associated CBC traits. The lead pleiotropic SNPs were used to search for common candidate genes for red blood cell and platelet traits in different time points before and after exposure to the challenge. A relaxed *FDR* threshold was used here to increase identification of true positives for the significance of biological and functional relevance of candidate genes [93]. Identification of positional candidate genes was conducted using the UCSC Genome Browser for the Ensembl annotation of the Sscrofa11.1 build of the swine genome (https://genome.ucsc.edu). Human, mouse, and rat genes in the IPA knowledge base database were used as background for the core analyses, including function analyses (molecular and cellular functions, and physiological system development and function), pathway investigation, and gene network construction. A function was considered significantly enriched in the significant windows if the p-value for the overlap comparison test between the input list of candidate genes and the IPA database was less than 0.05 [94-96].

List of abbreviations

BASO: basophil concentration; CBC: complete blood count; DNA: deoxyribonucleic acid; EOS: eosinophil concentration; FDR: Benjamini and Hochberg correction for false discovery rate; GFGR: grow-to-finish growth rate; GWAS: genome-wide association studies; HCT: hematocrit; HGB: hemoglobin concentration; IBS: identity-by-state; IPA: Ingenuity Pathway Analysis; LD: linkage disequilibrium; LYM: lymphocyte concentration; LSMs: least-squares means; MCH: mean corpuscular hemoglobin; MCHC: mean corpuscular hemoglobin concentration; MCV: mean corpuscular volume; MDS; multidimensional scaling; MHC: major histocompatibility complex; MONO: monocyte concentration; MPV: mean platelet volume; NDCM: natural disease challenge model; NEU: neutrophil concentration; PCA: principle component analysis; PLT: platelet concentration; Q-Q plot: quantile-quantile plot; QTL: quantitative trait loci; RBC: red blood cell concentration; RDW: red blood cell distribution width; SSC: *Sus scrofa* chromosome SSGWAS: single-step genome-wide association studies; SNP: single-nucleotide polymorphisms; TLRs: toll-like receptors; TR: treatment rate; WBC: total white blood cell concentration.

Declarations

Ethics approval and consent to participate

The animal study was reviewed and approved by the Animal Protection Committee of the Centre de Recherche en Sciences Animales de Deschambault (15PO283) and the Animal Care and Use Committee at the University of Alberta (AUP00002227).

Consent for publication

Not applicable.

Availability of data and materials

Because the data were generated on samples from commercially owned animals, the data analyzed in the current study are not publicly available, but they can be made available for non-commercial use by the corresponding author GP on reasonable request.

Competing interests

FF is employed by company Centre de Développement du Porc du Québec, Inc. who manage the challenge model developed by the team. We have also included the PC consortium as co-authors as they are intimately involved in the research studies as identified in the manuscript and this is custom and practice in our publication to date. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Funding

This project was funded by the Genome Canada, Genome Alberta, PigGen Canada, Swine Innovation Porc, and Alberta Agriculture and Forestry. PigGen Canada also intimately involved in the design of the project and the development of protocols for the natural disease challenge model. This research is also part of the AMR – One Health Consortium, funded by the Major Innovation Fund program of the Alberta Ministry of Economic Development, Trade and Tourism.

Authors' contributions

XB analyzed the data and wrote the manuscript with GP and help from TY, ZW and CL. FF, JH, MD, PC, JD, and GP designed the project and developed protocols for the natural disease challenge model. FF oversaw the sample collection and scheduling. JH was in charge of veterinary oversight on the project. CF provided support on CBC data measurement and

interpretation. GP was in charge of the database and genotyping for the project. AP and JD further processed the genotype data and provided the genomic relationship matrix for the project. All authors helped with the interpretation of results and reviewed and approved the final manuscript.

Acknowledgements

The following individuals served as collaborators and representatives for member companies of the PigGen Canada Consortium: Mr. D. Vandenbroek and Mr. B. DeVries, Alliance Genetics Canada, St. Thomas, ON, Canada; Dr. N. Dion and Ms. S. Blanchette, AlphaGene, St.-Hyacinthe, QU, Canada; Dr. T. Rathje, DNA Genetics, Columbus, NE, United States; Mr. M. Duggan, FastGenetics, Saskatoon, SK, Canada; Dr. R. Kemp, Genesus, Oakville, MB, Canada; Dr. P. Charagu, Hypor, Regina, SK, Canada; Dr. P. Mathur, Topigs Norsvin, Helvoirt, Netherlands; JH, University of Saskatchewan, Saskatoon, SK, Canada. They participated in the development and implementation of the project, coordinated the sources of piglets and the collection of associated data, and discussed with the research team regularly during the execution of the natural challenge model to contribute to the project. Thanks to Mr. Michael Lowing at Delta Genomics (Edmonton, AB, Canada) for processing the raw genotypes from the Affymetrix SNP chips. Thanks to Dr. Jason Grant at Livestock Gentec, University of Alberta for managing the project database and helping with questions concerning genotyping and SNP map file questions. Thanks to Ms. Yan Meng, Ms. Janelle Jiminez, Ms. Ziqi Yang, and Mr. Jiehan Lim at the University of Alberta for helping with the processing of blood samples and managing the database of samples and CBC measures. Thanks to Ms. Susan Goruk and Ms. Marnie Newell at the University of Alberta for

helping with the processing of blood samples and problem solving of technical problems with the CBC instrument.

References

1. Davies G, Genini S, Bishop SC, Giuffra E. An assessment of opportunities to dissect host genetic variation in resistance to infectious diseases in livestock. Animal. 2008;3:415–36.

2. Tomley FM, Shirley MW. Livestock infectious diseases and zoonoses. Philosophical Transactions Royal Soc B Biological Sci. 2009;364:2637–42.

3. Albers GA, Gray GD, Piper LR, Barker JS, Le Jambre LF, and Barger IA. The genetics of resistance and resilience to *Haemonchus contortus* infection in young merino sheep. Int. J. Parasitol. 1987;17:1355–1363.

4. Mulder HA and Rashidi H. Selection on resilience improves disease resistance and tolerance to infections. J. Anim. Sci. 2017;95, 3346–3358.

5. Friggens NC, Blanc F, Berry DP, Puillet L. Review: deciphering animal robustness. A synthesis to facilitate its use in livestock breeding and management. Animal. 2017;11:2237–51.

6. Calder PC. Feeding the immune system. P Nutr Soc. 2013;72:299–309.

7. Knap PW, Bishop SC. Relationships between genetic change and infectious disease in domestic livestock. BSAP Occas Publ. 2000;27:65–80.

8. Abbas, A. K., Lichtman, A. H., and Pillai, S. Basic immunology: functions and disorders of the immune system. Amsterdam: Elsevier; 2015.

9. George-Gay B, Parker K. Understanding the complete blood count with differential. J Perianesth Nurs. 2003;18:96–117.

10. Bai X, Putz AM, Wang Z, Fortin F, Harding JCS, Dyck MK, et al. Exploring phenotypes for disease resilience in pigs using complete blood count data from a natural disease challenge model. Frontiers Genetics. 2020;11:216.

11. Edfors- Lilja I, Wattrang E, Andersson L, Fossum C. Mapping quantitative trait loci for stress induced alterations in porcine leukocyte numbers and functions. Anim Genet. 2000;31:186–93.

12. Edfors-Lilja I, Wattrang E, Marklund L, Moller M, Andersson-Eklund L, Andersson L, et al. Mapping quantitative trait loci for immune capacity in the pig. J Immunol. 1998;161:829–35.

13. Reiner G, Fischer R, Hepp S, Berge T, Köhler F, Willems H. Quantitative trait loci for red blood cell traits in swine. Anim Genet. 2007;38:447–52.

14. Reiner G, Fischer R, Hepp S, Berge T, Köhler F, Willems H. Quantitative trait loci for white blood cell numbers in swine. Anim Genet. 2008;39:163–8.

589 15. Ponsuksili S, Reyer H, Trakooljul N, Murani E, Wimmers K. Single- and Bayesian multi-
590 marker genome-wide association for haematological parameters in pigs. PLoS One.
591 2016;11:e0159212.

592 16. Gong Y-F, Lu X, Wang Z-P, Hu F, Luo Y-R, Cai S-Q, et al. Detection of quantitative trait
593 loci affecting haematological traits in swine via genome scanning. BMC Genet. 2010;11:56.

594 17. Zou Z, Ren J, Yan X, Huang X, Yang S, Zhang Z, et al. Quantitative trait loci for porcine
595 baseline erythroid traits at three growth ages in a White Duroc × Erhualian F2 resource
596 population. Mamm Genome. 2008;19:640–6.

597 18. Wattrang E, Almqvist M, Johansson A, Fossum C, Wallgren P, Pielberg G, et al.
598 Confirmation of QTL on porcine chromosomes 1 and 8 influencing leukocyte numbers,
599 haematological parameters and leukocyte function. Anim Genet. 2005;36:337–45.

600 19. Lu Y, Vandehaar MJ, Spurlock DM, Weigel KA, Armentano LE, Connor EE, et al. Genome-
601 wide association analyses based on a multiple-trait approach for modeling feed efficiency. J
602 Dairy Sci. 2018;101 J. Dairy Sci. 93 2010:3140–54.

603 20. Fatumo S, Carstensen T, Nashiru O, Gurdasani D, Sandhu M, Kaleebu P. Complimentary
604 methods for multivariate genome-wide association study identify new susceptibility genes for
605 blood cell traits. Frontiers Genetics. 2019;10:334.

606 21. Park SH, Lee JY, Kim S. A methodology for multivariate phenotype-based genome-wide
607 association studies to mine pleiotropic genes. BMC Syst Biol. 2011;5 Suppl 2:S13.

608 22. Shriner D. Moving toward system genetics through multiple trait analysis in genome-wide
609 association studies. Frontiers Genetics. 2012;3:1.

610 23. Lacy, P. The role of Rho GTPases and SNAREs in mediator release from granulocytes.
611 Pharmacol Therapeut. 2005;107, 358–376.

612 24. Turton, KB. Case studies of cryptic proteins contributing to shape change in eosinophils.
613 https://search.proquest.com/docview/2089435247?pq-origsite=gscholar&fromopenview=true
614 (2018). Accessed 5 May 2020.

615 25. Balzar S, Chu HW, Silkoff P, Cundall M, Trudeau JB, Strand M, et al. Increased TGF-β2 in
616 severe asthma with eosinophilia. J Allergy Clin Immun. 2005;115:110–7.

617 26. Masure D, Vlaminck J, Wang T, Chiers K, Van den Broeck W, Vercruysse J, et al. A role for
618 eosinophils in the intestinal immunity against infective Ascaris suum larvae. PLoS Negl Trop
619 Dis. 2013;7:e2138.

620 27. Sheedy FJ. Turning 21: Induction of miR-21 as a key switch in the inflammatory response.
621 Front Immunol. 2015;6:19.

622 28. Cobb BS, Hertweck A, Smith J, O'Connor E, Graf D, Cook T, et al. A role for Dicer in
623 immune regulation. J Exp Medicine. 2006;203:2519–27.

624 29. Wu H, Neilson JR, Kumar P, Manocha M, Shankar P, Sharp PA, et al. miRNA profiling of
625 naïve, effector and memory CD8 T cells. PLoS One. 2007;2:e1020.

30. Dill BD, Gierlinski M, Härtlova A, Arandilla AG, Guo M, Clarke RG, et al. Quantitative proteome analysis of temporally resolved phagosomes following uptake via key phagocytic receptors. Mol Cell Proteomics. 2015;14:1334–49.

31. Shlomo SB, Mouhadeb O, Cohen K, Varol C, Gluck N. COMMD10-guided phagolysosomal maturation promotes clearance of *Staphylococcus aureus* in macrophages. Iscience. 2019;14:147–63.

32. Zhang Y, Morgan MJ, Chen K, Choksi S, Liu Z. Induction of autophagy is essential for monocyte-macrophage differentiation. Blood. 2012;119:2895–905.

33. Geng J, Klionsky DJ. The Atg8 and Atg12 ubiquitin- like conjugation systems in macroautophagy. Embo Rep. 2008;9:859–64.

34. Wang L, Zhao N, Zhang F, Yue W, Liang M. Effect of taurine on leucocyte function. Eur J Pharmacol. 2009;616:275–80.

35. Schaffer SW, Azuma J, Mozaffari M. Role of antioxidant activity of taurine in diabetes. Can J Physiol Pharm. 2009;87:91–9.

36. Booken N, Gratchev A, Utikal J, Weiß C, Yu X, Qadoumi M, et al. Sézary syndrome is a unique cutaneous T-cell lymphoma as identified by an expanded gene signature including diagnostic marker molecules CDO1 and DNM3. Leukemia. 2008;22:393–9.

37. Marcinkiewicz J, Kontny E. Taurine and inflammatory diseases. Amino Acids. 2012;46:7–20.

38. Oshiumi H, Sasai M, Shida K, Fujita T, Matsumoto M, Seya T. TIR-containing adapter molecule (TICAM)-2, a bridging adapter recruiting to toll-like receptor 4 TICAM-1 that induces interferon-β. J Biol Chem. 2003;278:49751–62.

39. Mekonnen E, Bekele E, Stein CM. Novel polymorphisms in TICAM2 and NOD1 associated with tuberculosis progression phenotypes in Ethiopian populations. Global Heal Epidemiology Genom. 2018;3:e1.

40. Doyle SL, Husebye H, Connolly DJ, Espevik T, O'Neill LAJ, McGettrick AF. The GOLD domain-containing protein TMED7 inhibits TLR4 signalling from the endosome upon LPS stimulation. Nat Commun. 2012;3:707.

41. Schilling E, Wehrhahn J, Klein C, Raulien N, Ceglarek U, Hauschildt S. Inhibition of nicotinamide phosphoribosyltransferase modifies LPS-induced inflammatory responses of human monocytes. Innate Immun. 2011;18:518–30.

42. Travelli C, Colombo G, Mola S, Genazzani AA, Porta C. NAMPT: a pleiotropic modulator of monocytes and macrophages. Pharmacol Res. 2018;135:25–36.

43. Xie J, Briggs JA, Briggs RC. Human hematopoietic cell specific nuclear protein MNDA interacts with the multifunctional transcription factor YY1 and stimulates YY1 DNA binding. J Cell Biochem. 1998;70:489–506.

44. Kong X-Z, Yin R-H, Ning H-M, Zheng W-W, Dong X-M, Yang Y, et al. Effects of THAP11 on erythroid differentiation and megakaryocytic differentiation of K562 Cells. PLoS One. 2014;9:e91557.

45. Cools J, DeAngelo DJ, Gotlib J, Stover EH, Legare RD, Cortes J, et al. A tyrosine kinase created by fusion of the *PDGFRA* and *FIP1L1* genes as a therapeutic target of imatinib in idiopathic hypereosinophilic syndrome. New Engl J Medicine. 2003;348:1201–14.

46. Zhai P-F, Wang F, Su R, Lin H-S, Jiang C-L, Yang G-H, et al. The Regulatory roles of microRNA-146b-5p and its target platelet-derived growth factor receptor α (PDGFRA) in erythropoiesis and megakaryocytopoiesis. J Biol Chem. 2014;289:22600–13.

47. Moller MJ, Chaudhary R, Hellmén E, Höyheim B, Chowdhary B, Andersson L. Pigs with the dominant white coat color phenotype carry a duplication of the *KIT* gene encoding the mast/stem cell growth factor receptor. Mamm Genome. 1996;7:822–30.

48. Thorén LA, Liuba K, Bryder D, Nygren JM, Jensen CT, Qian H, et al. Kit regulates maintenance of quiescent hematopoietic stem cells. J Immunol. 2008;180:2045–53.

49. Kastner P, Chan S. Function of RARα during the maturation of neutrophils. Oncogene. 2001;20:7178–85.

50. Kastner P, Lawrence HJ, Waltzinger C, Ghyselinck NB, Chambon P, Chan S. Positive and negative regulation of granulopoiesis by endogenous RARα. Blood. 2001;97:1314–20.

51. Kendrick TS, Payne CJ, Epis MR, Schneider JR, Leedman PJ, Klinken SP, et al. Erythroid defects in TRα−/− mice. Blood. 2008;111:3245–8.

52. Zhu J, Heyworth CM, Glasow A, Huang Q-H, Petrie K, Lanotte M, et al. Lineage restriction of the RARα gene expression in myeloid differentiation. Blood. 2001;98:2563–7.

53. Diez-Silva M, Dao M, Han J, Lim C-T, Suresh S. Shape and biomechanical characteristics of human red blood cells in health and disease. Mrs Bull. 2010;35:382–8.

54. Godin DV, Gray GR, Frohlich J. Erythrocyte membrane alterations in lecithin:cholesterol acyltransferase deficiency. Scand J Clin Laboratory Investigation Suppl. 1978;150:162–7.

55. Karai I, Fukumoto K, Horiguchi S. Alterations of lipids of the erythrocyte membranes in workers exposed to lead. Int Arch Occ Env Hea. 1982;50:11–6.

56. Bize I, Güvenç B, Buchbinder G, Brugnara C. Stimulation of human erythrocyte K-Cl cotransport and protein phosphatase type 2A by n-ethylmaleimide: role of intracellular Mg++. J Membr Biology. 2000;177:159–68.

57. Rust MB, Alper SL, Rudhard Y, Shmukler BE, Vicente R, Brugnara C, et al. Disruption of erythroid K-Cl cotransporters alters erythrocyte volume and partially rescues erythrocyte dehydration in SAD mice. J Clin Invest. 2007;117:1708–17.

58. Permanyer M, Bošnjak B, Förster R. Dual role for atypical chemokine receptor 1 in myeloid cell hematopoiesis and distribution. Cell Mol Immunol. 2018;15:399–401.

59. Bonavita O, Poeta VM, Massara M, Mantovani A, Bonecchi R. Regulation of hematopoiesis by the chemokine system. Cytokine. 2018;109 Nature 404 2000:76–80.

60. Wu X, Yang SL, Yerle M, Zhu ZM, Wang HL, Wang H, et al. Genomic organization, localization and polymorphism of porcine *PSMB10*, a gene encoding the third beta- type proteasome subunit of 26S proteasome complex. J Anim Breed Genet. 2006;123:331–6.

61. Leone P, Shin E-C, Perosa F, Vacca A, Dammacco F, Racanelli V. MHC class I antigen processing and presenting machinery: organization, function, and defects in tumor cells. Jnci J National Cancer Inst. 2013;105:1172–87.

62. Pesciotta EN, Lam H-S, Kossenkov A, Ge J, Showe LC, Mason PJ, et al. In-depth, label-free analysis of the erythrocyte cytoplasmic proteome in diamond blackfan anemia identifies a unique inflammatory signature. PLoS One. 2015;10:e0140036.

63. Rolando RE, Thiagarajan P. Platelet-vessel wall interactions in hemostasis and thrombosis. Morgan & Claypool Life Sciences. 2011. doi:10.4199/C00007ED1V01Y201002ISP004.

64. Kaprielian Z, Cho K, Hadjiargyrou M, Patterson P. CD9, a major platelet cell surface glycoprotein, is a ROCA antigen and is expressed in the nervous system. J Neurosci. 1995;15:562–73.

65. Boucheix C, Benoit P, Frachet P, Billard M, Worthington RE, Gagnon J, et al. Molecular cloning of the CD9 antigen. A new family of cell surface proteins. J Biological Chem. 1991;266:117–22.

66. Clay D, Rubinstein E, Mishal Z, Anjo A, Prenant M, Jasmin C, et al. CD9 and megakaryocyte differentiation. Blood. 2001;97:1982–9.

67. Worthington RE, Carroll RC, Boucheix C. Platelet activation by CD9 monoclonal antibodies is mediated by the FCγII receptor. Brit J Haematol. 1990;74:216–22.

68. Kanaji S, Fahs SA, Shi Q, Haberichter SL, Montgomery RR. Contribution of platelet vs. endothelial VWF to platelet adhesion and hemostasis. J Thromb Haemost. 2012;10:1646–52.

69. Zhang Y, Wang Y, Xiang Y, Lee W, Zhang Y. Prohibitins are involved in protease-activated receptor 1- mediated platelet aggregation. J Thromb Haemost. 2012;10:411–8.

70. Frey UH, Aral N, Müller N, Siffert W. Cooperative effect of GNB3 825C>T and GPIIIa PI(A) polymorphisms in enhanced platelet aggregation. Thromb Res. 2003;109:279–86.

71. Dusse F, Frey UH, Bilalic A, Dirkmann D, Görlinger K, Siffert W, et al. The GNB3 C825T polymorphism influences platelet aggregation in human whole blood. Pharmacogenet Genom. 2012;22:43–9.

72. Freson K, Izzi B, Geet CV. From genetics to epigenetics in platelet research. Thromb Res. 2012;129:325–9.

73. Freson K, Izzi B, Labarque V, Helvoirt MV, Thys C, Wittevrongel C, et al. GNAS defects identified by stimulatory G protein α-subunit signalling studies in platelets. J Clin Endocrinol Metabolism. 2008;93:4851–9.

74. Kunishima S, Nishimura S, Suzuki H, Imaizumi M, Saito H. TUBB1 mutation disrupting microtubule assembly impairs proplatelet formation and results in congenital macrothrombocytopenia. Eur J Haematol. 2014;92:276–82.

737 75. Schwer HD, Lecine P, Tiwari S, Italiano JE, Hartwig JH, Shivdasani RA. A lineage-
738 restricted and divergent β-tubulin isoform is essential for the biogenesis, structure and function
739 of blood platelets. Curr Biol. 2001;11:579–86.

740 76. Burley K, Westbury SK, Mumford AD. TUBB1 variants and human platelet traits. Platelets.
741 2018;29:209–11.

742 77. Ganesh SK, Zakai NA, Rooij FJA van, Soranzo N, Smith AV, Nalls MA, et al. Multiple loci
743 influence erythrocyte phenotypes in the CHARGE Consortium. Nat Genet. 2009;41:1191–8.

744 78. Boettcher S, Manz MG. Regulation of inflammation- and infection-driven hematopoiesis.
745 Trends Immunol. 2017;38:345–57.

746 79. Baldridge MT, King KY, Goodell MA. Inflammatory signals regulate hematopoietic stem
747 cells. Trends Immunol. 2011;32:57–65.

748 80. Putz AM, Harding JCS, Dyck MK, Fortin F, Plastow GS, Dekkers JCM, et al. Novel
749 resilience phenotypes using feed intake data from a natural disease challenge model in wean-to-
750 finish pigs. Front Genet. 2019;9:660.

751 81. Sargolzaei, M., Chesnais, J. P., and Schenkel, F. S. A new approach for efficient genotype
752 imputation using information from relatives. BMC Genomics. 2014;15:478.

753 82. Misztal, I., Tsuruta, S., Strabel, T., Auvray, B., Druet, T., and Lee, D. H. "BLUPF90 and
754 related programs (BGF90)," in Proceedings of 7th World Congress of Genetics Applied to
755 Livestock Production, (Montpellier) (2002). Accessed 5 May 2020.

756 83. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A
757 tool set for whole-genome association and population-based linkage analyses. Am J Hum
758 Genetics. 2007;81:559–75.

759 84. R core team. R: A Language and Environment for Statistical Computing. R foundation for
760 statistical computing, Vienna. Available online at: https:// www.R-project.org/. Accessed 1 Jan
761 2019.

762 85. Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. GenABEL: an R library for genome-wide
763 association analysis. Bioinformatics. 2007;23:1294–6.

764 86. Turner SD. qqman: an R package for visualizing GWAS results using Q-Q and manhattan
765 plots. 2014. doi:10.1101/005165.

766 87. VanLiere JM, Rosenberg NA. Mathematical properties of the r2 measure of linkage
767 disequilibrium. Theor Popul Biol. 2008;74:130–7.

768 88. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and
769 haplotype maps. Bioinformatics. 2005;21:263–5.

770 89. Aguilar I, Legarra A, Cardoso F, Masuda Y, Lourenco D, Misztal I. Frequentist p-values for
771 large-scale-single step genome-wide association, with an application to birth weight in American
772 Angus cattle. Genet Sel Evol. 2019;51:28.

773 90. VanRaden, P. Efficient methods to compute genomic prediction. J. Dairy Sci. 2008;91,
774 4414–4423.

775 91. Benjamini, Y., and Hochberg, Y. Controlling the false discovery rate: a practical and
776 powerful approach to multiple testing. J Royal Statistical Soc Ser B Methodol 1995;57, 289–300.

777 92. Mrode, R. A. Linear Models for the Prediction of Animal Breeding Values. Boston, MA:
778 CABI; 2013.

779 93. Waide EH, Tuggle CK, Serão NVL, Schroyen M, Hess A, Rowland RRR, et al. Genomewide
780 association of piglet responses to infection with one of two porcine reproductive and respiratory
781 syndrome virus isolates1. J. Anim. Sci. 2017;95:16–38.

782 94. Zhang F, Wang Y, Mukiibi R, Chen L, Vinsky M, Plastow G, et al. Genetic architecture of
783 quantitative traits in beef cattle revealed by genome wide association studies of imputed whole
784 genome sequence variants: I: feed efficiency and component traits. BMC Genomics. 2020;21:36.

785 95. Wang Y, Zhang F, Mukiibi R, Chen L, Vinsky M, Plastow G, et al. Genetic architecture of
786 quantitative traits in beef cattle revealed by genome wide association studies of imputed whole
787 genome sequence variants: II: carcass merit traits. BMC Genomics. 2020;21.
788 doi:10.1186/s12864-019-6273-1.

789 96. Wysocki M, Chen H, Steibel JP, Kuhar D, Petry D, Bates J, et al. Identifying putative
790 candidate genes and pathways involved in immune responses to porcine reproductive and
791 respiratory syndrome virus (PRRSV) infection. Anim Genet. 2012;43:328–32.

Additional files

Additional file 1: This file contains additional **Figure S1.** Violin plots for descriptive statistics for white blood cell traits in Blood 1, Blood 3, and Blood 4 collected at 2-weeks before, and at 2- and 6-weeks after the challenge, respectively. **Figure S2.** Violin plots for descriptive statistics for red blood cell traits in Blood 1, Blood 3, and Blood 4 collected at 2-weeks before, and at 2- and 6-weeks after the challenge, respectively. **Figure S3.** Violin plots for descriptive statistics for white blood cell traits in Blood 1, Blood 3, and Blood 4 collected at 2-weeks before, and at 2- and 6-weeks after the challenge, respectively. **Table S1.** Genetic correlations between red blood cell traits of mean corpuscular hemoglobin (MCH), mean corpuscular volume (MCV), and red blood cell concentration (RBC) traits in Blood 1, Blood 3, and Blood 4.

Additional file 2: This file contains **Figure S4.** Multidimensional scaling (MDS) plots showing the first three dimensions (C1, C2, and C3) of the population structure for genotyped animals based on pairwise identity-by-state distance. **Figure S5**. **(A)** Manhattan plot for the eosinophil concentration in Blood 3 (EOSB3). **(B)** Quantile-Quantile plot for EOSB3; **Figure S6**. **(A)** Manhattan plot for the change of monocyte concentration from Blood 1 to Blood 3 (MONOΔ13). **(B)** Quantile-Quantile plot for MONOΔ13; **Figure S7.** Manhattan plots (**A, C, E**) and Quantile-Quantile plots (**B, D, F**) for the mean corpuscular hemoglobin (MCH) in Blood 1, Blood 3, and Blood 4, respectively; **Figure S8.** Manhattan plots (**A, C**) and Quantile-Quantile plots (**B, D**) for the mean corpuscular hemoglobin concentration (MCHC) in Blood 3 and for the change of MCHC from Blood 1 to Blood 4, respectively; **Figure S9.** Manhattan plots (**A, C, E**) and Quantile-Quantile plots (**B, D, F**) for the mean corpuscular volume (MCV) in Blood 1, Blood 3, and Blood 4, respectively; **Figure S10.** Manhattan plots (**A, C, E**) and Quantile-Quantile plots (**B, D, F**) for the red blood cell concentration (RBC) in Blood 1, Blood 3, and Blood 4, respectively; **Figure S11.** Manhattan plots (**A, C**) and Quantile-Quantile plots (**B, D**) for the mean platelet volume (MPV) in Blood 1 and Blood 4, respectively; **Figure S12.** Manhattan plots (**A, C, E**) and Quantile-Quantile plots (**B, D, F**) for the platelet concentration (PLT) in Blood 1, Blood 3, and Blood 4, respectively.

Additional file 3: This file contains additional **Table S2**. Candidate genes located within 1 Mb on either side of the top significant SNPs; **Figure S13.** Haplotype block pattern (r^2-scheme) for the region of candidate genes on SSC2 located within the maximum distance of 1 Mb on either side of the top lead SNP4 (SSC2, 120,341,201bp); **Figure S14.** Haplotype block pattern (r^2-scheme) for the region of candidate genes on SSC4 located within the maximum distance of 1 Mb on either side of the top lead SNP6 (SSC4, 91,591,493bp); **Figure S15.** Haplotype block pattern (r^2-scheme) for the region of candidate genes on SSC6 located within the maximum distance of 1 Mb on either side of the top lead SNP7 (SSC6, 28,511,423bp); **Figure S16.** Haplotype block pattern (r^2-scheme) for SNPs (40,946,144bp to 41,198,574bp) on SSC8 located within the maximum distance of 1 Mb on either side of the top lead SNP8 (SSC8, 41,156,538bp).

Additional file 4: This file contains a list of enriched functions (p-value < 0.05) for the candidate gene lists for complete blood count (CBC) traits.

Table 1. Top significant SNPs for significant associations with white blood cell traits at a genome-wise false discovery rate (*FDR*) of 0.05.

Trait[1]	Blood[2]	SNP ID	SNP status[3]	SSC[4]	SNP position (bp)	MAF[5]	FDR	GVar (%)[6]	1-Mb window start SNP position[7] (bp)	Dominance effect ± standard error	Additive effect[8] ± standard error
EOS	Blood 3	SNP1	Top lead	4	93,647,202	0.31	0.006	0.46	93,331,316	0.001 ± 0.01	**-0.05 ± 0.01**[9]
EOS	Blood 3	SNP2	Top lead	10	8,186,695	0.08	0.03	0.35	7,396,201	**-0.09 ± 0.04**	**0.14 ± 0.04**
EOS	Blood 3	SNP3	Top lead	12	36,308,994	0.13	0.003	0.53	35,450,868	0.002 ± 0.02	**-0.06 ± 0.01**
MONO	Δ13	SNP4	Top floating	2	120,341,201	0.47	0.049	0.12	120,219,793	-0.03 ± 0.02	**0.08 ± 0.02**
MONO	Δ13	SNP5	Top lead	9	105,461,701	0.43	0.049	1.23	105,461,701	-0.02 ± 0.02	**-0.08 ± 0.02**

[1]EOS: eosinophil concentration; MONO: monocyte concentration.
[2]Blood 3: the CBC measures in Blood 3 collected at 2-weeks after the challenge; Δ13: the change of CBC measures from Blood 1 collected at 2-weeks before the challenge to Blood 3 collected at 2-weeks after a polymicrobial infectious challenge.
[3]Top lead: the most significant SNP with a group of supportive SNPs; Top floating: the most significant SNP without a group of supportive SNPs.
[4]*Sus scrofa* chromosome.
[5]Minor allele frequency.
[6]The largest percentage of additive genetic variance explained by the top significant SNP and its adjacent SNPs in a 1 Mb window.
[7]Positions of the start SNP for the 1 Mb window segment with the largest amount of additive genetic variance.
[8]Estimates of additive effects per additional copy of the "B" allele. When the dominance effect was not significant ($p > 0.05$) the estimate of the additive effect was based on a model without the dominance effect.
[9]Significant estimates of additive and dominance effects are highlighted in bold ($p < 0.05$).

846 **Table 2.** Top lead SNPs[1] for significant associations with red blood cell and platelet traits at a genome-wise false discovery rate (*FDR*)
847 of 0.05.

SNP ID	SSC[2]	SNP position (bp)	MAF[3]	Trait[4]	Blood[5]	FDR	GVar (%)[6]	1-Mb window start SNP[7] position (bp)	Dominance effect ± standard error	Additive effect[8] ± standard error
SNP6	4	91,591,493	0.38	MCHC	Blood 3	0.04	1.15	91,291,800	**1.08 ± 0.02**[9]	**-2.13 ± 0.02**
SNP7	6	28,511,423	0.41	MCH	Blood 1	0.04	0.29	28,110,554	0.10 ± 0.07	**-0.25 ± 0.06**
				MCV	Blood 1	0.001	0.57	28,110,554	**0.29 ± 0.07**	**-0.73 ± 0.06**
					Blood 3	0.002	0.49	28,096,004	0.05 ± 0.04	**-0.45 ± 0.04**
					Blood 4	0.002	0.48	28,096,004	**-0.12 ± 0.04**	**-0.53 ± 0.04**
				RBC	Blood 1	0.01	0.44	28,110,554	-0.02 ± 0.07	0.08 ± 0.06
					Blood 3	0.01	0.44	28,110,554	0.001 ± 0.04	**0.08 ± 0.04**
					Blood 4	0.03	0.40	28,110,554	-0.04 ± 0.05	0.06 ± 0.04
SNP8	8	41,156,538	0.45	MCH	Blood 1	0.01	0.36	40,257,441	**0.15 ± 0.06**	**0.20 ± 0.05**
					Blood 3	0.04	0.36	40,257,441	**0.11 ± 0.04**	**0.16 ± 0.04**
					Blood 4	0.04	0.35	40,257,441	-0.006 ± 0.04	**0.18 ± 0.04**
				MCV	Blood 1	0.03	0.33	40,219,864	**0.35 ± 0.06**	**0.59 ± 0.05**
					Blood 3	0.006	0.27	40,219,864	**0.19 ± 0.04**	**0.38 ± 0.04**
					Blood 4	0.002	0.33	40,219,864	0.04 ± 0.04	**0.51 ± 0.04**
				RBC	Blood 1	0.007	0.31	40,219,864	0.01 ± 0.06	-0.08 ± 0.05
					Blood 3	0.02	0.31	40,219,864	-0.02 ± 0.04	-0.06 ± 0.04
					Blood 4	0.03	0.30	40,219,864	0.01 ± 0.04	-0.06 ± 0.04
SNP9	17	59,739,745	0.34	MPV	Blood 1	0.02	0.49	59,053,639	0.002 ± 0.10	**0.28 ± 0.08**
					Blood 4	0.04	0.40	59,053,639	-0.09 ± 0.07	**0.25 ± 0.05**

[1]The most significant SNP with a group of supportive SNPs.
[2]*Sus scrofa* chromosome.
[3]Minor allele frequency.
[4]MCHC: mean corpuscular hemoglobin concentration; MCH: mean corpuscular hemoglobin; MCV: mean corpuscular volume; RBC: red blood cell concentration; MPV: mean platelet volume.
[5]Blood 1, Blood 3, and Blood 4: CBC measures in blood samples collected at 2-weeks before, and 2- and 6-weeks after a polymicrobial infectious challenge.
[6]The largest percentage of additive genetic variance explained by the top lead SNP and its adjacent SNPs in a 1 Mb window.
[7]Positions of the start SNP for the 1 Mb window segment with the largest amount of additive genetic variance.
[8]Estimates of additive effects per additional copy of the "B" allele. When the dominance effect was not significant ($p > 0.05$) the estimate of the additive effect was based on a model without the dominance effect.
[9]Significant estimates of additive and dominance effects are highlighted in bold ($p < 0.05$).

Table 3. Top floating SNPs[1] for significant associations with red blood cell and platelet traits at a genome-wise false discovery rate (*FDR*) of 0.05.

SNP ID	SSC[2]	SNP Position (bp)	MAF[3]	Trait[4]	Blood[5]	FDR	GVar (%)[6]	1-Mb window start SNP[7] position (bp)	Dominance effect ± standard error	Additive effect[8] ± standard error
SNP10	1	18,792,764	0.37	MCV	Blood 1	0.003	0.18	18,536,535	-0.16 ± 0.18	**0.52 ± 0.14**[9]
					Blood 3	0.004	0.21	18,546,024	0.03 ± 0.13	**0.52 ± 0.10**
					Blood 4	0.02	0.15	18,536,535	-0.02 ± 0.14	**0.37 ± 0.11**
SNP11	5	64,520,638	0.31	PLT	Blood 1	0.001	0.09	63,861,170	-3.80 ± 6.83	**26.78 ± 5.18**
					Blood 3	0.001	0.09	63,861,170	-9.28 ± 7.44	**23.92 ± 5.39**
					Blood 4	0.03	0.05	63,861,170	-10.57 ± 7.40	**18.33 ± 5.46**
SNP12	9	40,919,049	0.45	MCH	Blood 1	0.03	0.05	39,919,771	0.07 ± 0.08	**0.21 ± 0.07**
					Blood 3	0.04	0.07	40,490,005	-0.04 ± 0.05	**0.19 ± 0.05**
					Blood 4	0.04	0.06	40,490,005	0.03 ± 0.05	**0.21 ± 0.05**
SNP13	11	13,749,336	0.12	MCHC	Δ14	0.02	0.07	13,011,748	1.89 ± 2.42	2.38 ± 2.33
SNP14	12	22,234,265	0.3	MCV	Blood 3	0.04	0.08	21,749,390	-0.06 ± 0.14	**-0.40 ± 0.10**

[1]The most significant SNP without a group of supportive SNPs.
[2]*Sus scrofa* chromosome.
[3]Minor allele frequency.
[4]MCV: mean corpuscular volume; PLT: platelet concentration; MCH: mean corpuscular hemoglobin; MCHC: mean corpuscular hemoglobin concentration.
[5]Blood 1, Blood 3, and Blood4: CBC measured in blood samples collected at 2-weeks before, and 2- and 6-weeks after a polymicrobial infectious challenge; Δ14: the change of CBC measures from Blood 1 collected at 2-weeks before the challenge to Blood 4 collected at 6-weeks after the challenge.
[6]The largest percentage of additive genetic variance explained by the top significant SNP and its adjacent SNPs in a 1 Mb window.

[7]Positions of the start SNP for the 1 Mb window segment with the largest amount of additive genetic variance.
[8]Estimates of additive effects per additional copy of the "B" allele. When the dominance effect was not significant ($p > 0.05$) the estimate of the additive effect was based on a model without the dominance effect.
[9]Significant estimates of additive and dominance effects are highlighted in bold ($p < 0.05$).

875 **Table 4.** Candidate genes located within 1 Mb on either side of the top significant SNPs[1] that have been reported by previous
876 studies of pigs, human, mice, and rats to be functionally and biologically related to CBC traits.

SNPID	Traits[2]	Browsing region	Candidate genes and locations[3]
SNP1	Eosinophils	SSC4: 92,647,202bp – 94,647,202bp	*ARHGEF2* (401,994bp R)
SNP2	Eosinophils	SSC10: 7,186,695bp – 9,186,695bp	*TGFB2* (118,679bp R)
SNP3	Eosinophils	SSC12: 35,308,994bp – 37,308,994bp	*MIR21* (243,636bp L)
SNP4	Monocytes	SSC2: 119,341,201bp – 121,341,201bp	*COMMD10* (within, intron 5), *ATG12* (375,499bp L), *CDO1* (400,776bp L), *TMED7* (536,795bp L), *TICAM2* (580,442bp L)
SNP5	Monocytes	SSC9: 104,461,701bp – 106,461,701bp	*NAMPT* (660,208 bp R)
SNP6	Red blood cells	SSC4: 90,591,493bp – 92,591,493bp	*SPTA1* (within, intron 15), *MNDA* (158,250bp L), *ACKR1* (356,841bp L)
SNP7	Red blood cells	SSC6: 27,511,423bp – 29,511,423bp	*CBFB* (734,672bp L), *THAP11* (38,940 bp L), *PSMB10* (33,487 bp R), *LCAT* (45,036bp R), *SLC12A4* (42,739bp R)
SNP8	Red blood cells	SSC8: 40,156,538bp – 42,156,538bp	*PDGFRA* (135,096bp L), *KIT* (245,796bp R)
SNP9	Platelets	SSC17: 58,739,745bp – 60,739,745bp	*GNAS* (686,723bp L), *TUBB1* (571,360bp R)
SNP10	Red blood cells	SSC1: 17,792,764bp – 19,792,764bp	*STXBP5* (279,512bp L), *RAB32* (78,111bp R)
SNP11	Platelets	SSC5: 63,520,638bp – 65,520,638bp	*VWF* (within, intron 2), *CD9* (60,802bp L), *GNB3* (650,242bp L), *PHB2* (764,158bp L)
SNP12	Red blood cells	SSC9: 39,919,049bp – 41,919,049bp	*ZBTB16* (720,652bp R)
SNP13	Red blood cells	SSC11: 12,749,336bp – 14,749,336bp	*TRPC4* (231,768bp L), *FREM2* (210,529bp R)
SNP14	Red blood cells	SSC12: 21,234,265bp – 23,234,265bp	*RARA* (148,591bp L), *THRA* (35,797bp R)

[1]The most significant SNP above the genome-wise *FDR* of 0.05 in each genomic region.
[2]The category of traits that associated with candidate genes.
[3]The gene location from the top significant SNP, described as "within" or to the left (L) or right (R) side of the gene as found on the reference genome sequence.