

Genetic and non-genetic factors distinctly shape the variation of the immune response in cattle

Received: 8 July 2024

Accepted: 18 December 2025

Published online: 15 January 2026

 Check for updates

Shifang Li¹, Françoise Myster^{1,6}, Célia Darimont¹, Lijing Tang², Justine Javaux¹, Rémy Sandor¹, Gabriel Costa Monteiro Moreira², José Luis Gualdrón Duarte^{2,3}, Philippe Crepin³, Marc Dive^{3,5}, Patrick Mayeres³, Tom Druet², Mihai G. Netea⁴, Michel Georges², Carole Charlier² & Laurent Gillet¹✉

Increasing agricultural production while reducing the reliance on synthetic inputs such as antibiotics is an important challenge. For cattle breeding, this implies better understanding the genetics underlying meat production and the immune response. Here, we use systems immunology to investigate the genetic and environmental drivers of immune variation in Belgian White Blue male cattle, a breed historically bred for meat production. While seasonality and other non-genetic factors account for much of the immune variation observed, genome-wide association studies identify loci with major effects on specific immunophenotypes. Genetics also emerges as the primary driver of cytokine production. Finally, we develop a predictive model linking genetic data to cytokine responses. Our findings support the selection of cattle with improved immunity and advance our understanding of mammalian immune variation.

Given the profound impact of immune responses on health, it is necessary to understand their variability in a given population and how this variability relates to susceptibility to infections and immune-mediated diseases^{1–5}. In humans and livestock species, this variation is determined by intrinsic (such as sex, age, and genetics), as well as extrinsic factors (such as diet, smoking, physical activity, microbial exposure)⁵. Despite recent advances, our understanding of the precise mechanisms and the extent to which these diverse factors contribute to the heterogeneity of immune responses among individuals and populations remains in its nascent stages⁶.

There is a growing consensus that between 20 and 40% of variation in immune traits is inherited in humans^{2,4,7}. However, the majority of the underlying genetic variants remain largely unidentified^{7,8}. Until recently, the study of genetic factors affecting immune responses has

mainly been performed through the study of “in-born errors of immunity”, i.e. monogenic immune deficiencies⁹. While these approaches have successfully revealed critical immune pathways, they are less effective at elucidating how genetic variation acts in more subtle ways to regulate transcription or translation of key immune mediators⁴. In the past decade, strategies and technologies, collectively termed “systems immunology”, have emerged to explore the immune system with unprecedented precision^{7,8,10–16}.

In general, systems immunology aims to identify the key components of immune networks, such as the immune cells that interact within specific systems and the molecular mediators they employ, and to determine how these components are altered when the system is perturbed^{12,17}. To this end, these approaches were usually coupled with high-throughput “omics” technologies that measure mRNA, proteins,

¹Immunology-Vaccinology, Faculty of veterinary medicine, FARAH, ULiège, Belgium. ²Unit of Animal Genomics, Faculty of veterinary medicine, GIGA, ULiège, Belgium. ³Walloon Breeders Association, Rue des Champs Elysées, 4, 5590 Ciney, Belgium. ⁴Department of Internal Medicine, Radboud Center for Infectious Diseases, Radboud University Medical Centre, Nijmegen, The Netherlands. ⁵Deceased: Marc Dive. ⁶These authors contributed equally: Shifang Li, Françoise Myster. ✉e-mail: L.gillet@uliege.be

metabolites and epigenetic modifications that orchestrate gene transcription^{18,19}. Systems immunology thus holds the potential to deepen our understanding of the mechanisms underlying disease and to provide insights into host-pathogen co-evolution^{18,19}. Some recent projects, such as the Human Immunology Project Consortium (HIPC, <https://immunespace.org>), the Human Functional Genomics Project (HFGP, <http://www.humanfunctionalgenomics.org>), and the Milieu Interieur Project (<https://www.milieuinterieur.fr>), collectively provide a comprehensive grasp of interindividual variation in human immune responses. Interestingly, when integrated with the exploration of the genomic landscape governing these responses, these studies, identified some of the most important genetic variants associated with immune cell phenotypes and cytokines production^{7,10,15,20–27}.

Strikingly, while systems immunology approaches have demonstrated their value in various contexts, they remain underutilized in livestock species. Although several pioneering studies have begun to explore systems-level immune responses in production animals, including pigs, cattle, and chicken^{28–36}, these remain relatively rare and focus primarily on transcriptomic data or examine a small number of specific immunophenotypes. This is particularly surprising given that resistance or susceptibility to infectious diseases is a major determinant of production efficiency³⁷. Moreover, the increased susceptibility of animals to infectious diseases is a threat to humans as it contributes to the emergence of zoonotic diseases and the propagation of antimicrobial resistance³⁸. The scarcity of systems immunology studies in animals becomes even more perplexing when considering the notable advancements witnessed in the past decade in animal genetics. Genomic selection, for instance, has been introduced in several major livestock species, leading to an impressive doubling of genetic advancements in some cases³⁹. These developments highlight the potential for a better understanding of immune responses in these populations through a formalized and integrative systems immunology framework.

The Belgian White Blue (BWB) cattle breed is the most muscular breed in the world and is attracting considerable interest to increase meat production in purebreds and industrial crossbreeding. Until now, selection programs in the BWB breed have focused on production traits and have not included health-related traits, particularly immune resistance to pathogens. Moreover, the respective effects of host and environmental factors, as well as the precise nature of the genetic variants that affect heterogeneity in the immune system in bovines, are largely unknown. The objectives of this study were to investigate the immune response in BWB cattle in a comprehensive manner by integrating both environmental and genetic factors that contribute to immune variation. Specifically, we aimed to identify genetic variants associated with immune traits that could be used to improve animal health through marker-assisted selection. Finally, we sought to demonstrate the potential of this genetic information to predict immune responses, thereby opening avenues for sustainable livestock production.

Results

Measurement of immune parameters in BWB cattle

Over the course of an 18-month period, we examined 246 BWB calves enrolled in the RESISTOMICS cohort, a bovine population cohort gathered to understand immune variation in cattle and raised in standardized conditions at the Center for Bovine Selection (CBS) in Ciney, Belgium. These calves, aged between 1 and 7 months upon entry in CBS, were evaluated for immune parameters at two distinct time points (28 days interval) by a broad immunophenotyping approach (Fig. 1a, b). Briefly, five unique flow cytometry staining panels were applied to immune cells, enabling the collection of 21 immune cell population percentages for each blood sampling (Fig. 1a and Supplementary Fig. 1–4, Supplementary Tables 1–3). The concentrations of 10 cytokines (IL-1 β , IL-6, IL-4, IFN- γ , IL-17, TNF- α , IP-10, IL-8, MCP1 and IL-

10) and of three antibody isotypes (IgM, IgG1, and IgG2) in serum were also measured (Supplementary Fig. 2a). In parallel, we measured the ex vivo production of 6 cytokines (IL-1 β , IL-4, IL-6, IL-10, IL-17 and IFN- γ) by peripheral blood mononuclear cells (PBMCs) and whole white blood cells (WWBCs) stimulated by 15 different stimuli (Supplementary Fig. 2b). 66 cytokine-stimulus measurement pairs were excluded from further analysis because they showed no difference in comparison to the non-stimulated states (Wilcoxon signed rank test, $p > 0.01$) or had mostly (over 50%) values below the detection limit (Supplementary Fig. 2b). In total, the quantitative values of 148 immune parameters were compared for each animal at each time point.

Correlations between immune compartments

We first analyzed correlations between immunophenotypes to reveal some underlying patterns of variation at the population level. Those analyses were performed while correcting for age and season effects as several studies highlighted the importance of these factors on immune parameters^{40,41}. At steady state in blood, concentrations of immunoglobulin isotypes were positively correlated with each other, while concentrations of IP-10 (CXCL10) was positively correlated with those of IL-10 and MCP-1 (CCL2) (Fig. 2a) (FDR < 0.05). Secondly, we explored the correlation between concentrations of circulating immunoglobulins and cytokines and the proportions of the different immune cell populations in blood (Fig. 2b). Notably, these analyses revealed a strong positive correlation between the percentage of a specific monocyte population circulating in the blood (Live CD172 + CD16 + CD14+), known as intermediate monocytes in human⁴², and the blood concentration of IL-10 ($r = 0.419$; FDR = 5.8e-08) (Fig. 2b). Other less strong and less significant correlations were also found, such as a positive correlation between CD21- IgM+ B lymphocytes and circulating IgM levels ($r = 0.315$, FDR = 0.000546).

We then performed unsupervised clustering of the concentrations of cytokines produced by immune cells in response to ex vivo stimulations (Fig. 2c). Four main clusters were observed. Cluster 1 encompassed IL-10 production in PBMCs and WWBCs. Cluster 2 mainly grouped monocyte-derived cytokines (IL-6 and IL-1 β) in WWBCs. Cluster 3 comprised IFN- γ and IL-1 β responses in PBMCs, while cluster 4 exhibited IL-17 and IL-6 responses in PBMCs. Although a distinct response to PMA/ionomycin stimulation was identified, it was surprising to note that the cytokine response primarily followed a cell-type (WWBCs vs PBMCs) and a cytokine-dependent pattern (Fig. 2c), not a stimulation defined one (Supplementary Fig. 5). These cell-type (WWBCs vs PBMCs) and cytokine-dependent patterns were confirmed when these parameters were integrated using Non-metric Multi-Dimensional Scaling (NMDS) as significant differences in cytokine productions were noted between PBMCs and WWBCs in both the first (FDR = 4.87e-09) and second (FDR = 2.27e-16) NMDS dimensions (Fig. 2d), suggesting the major influence of specific whole-blood components, such as neutrophils, on the variation in cytokine production. When comparisons were based on the cytokines, significant differences were also highlighted in both NMDS dimensions, especially when IL-10 production was compared to IFN- γ , IL-17, and IL-1 β productions (Fig. 2d).

Finally, to comprehensively understand the interaction between immune cell populations in blood and ex vivo cytokine production, their relationship was evaluated across the 15 stimulations (Fig. 2e). It first identified a major negative correlation between the percentages of neutrophils and the production of most of the measured cytokines by WWBCs. Strikingly, percentages of circulating CD14+ monocytes exhibited positive correlations with the ex vivo production of monocyte-derived cytokines (IL-1 β , IL-6 and particularly IL-10) in PBMCs and especially in WWBCs. In parallel, NK, T, and B cell percentages in blood displayed strong positive correlations with IFN- γ and IL-17 production after *A. fumigatus* stimulation of WWBCs (FDR < 0.05) (Fig. 2e). Conversely, negative correlations emerged between some

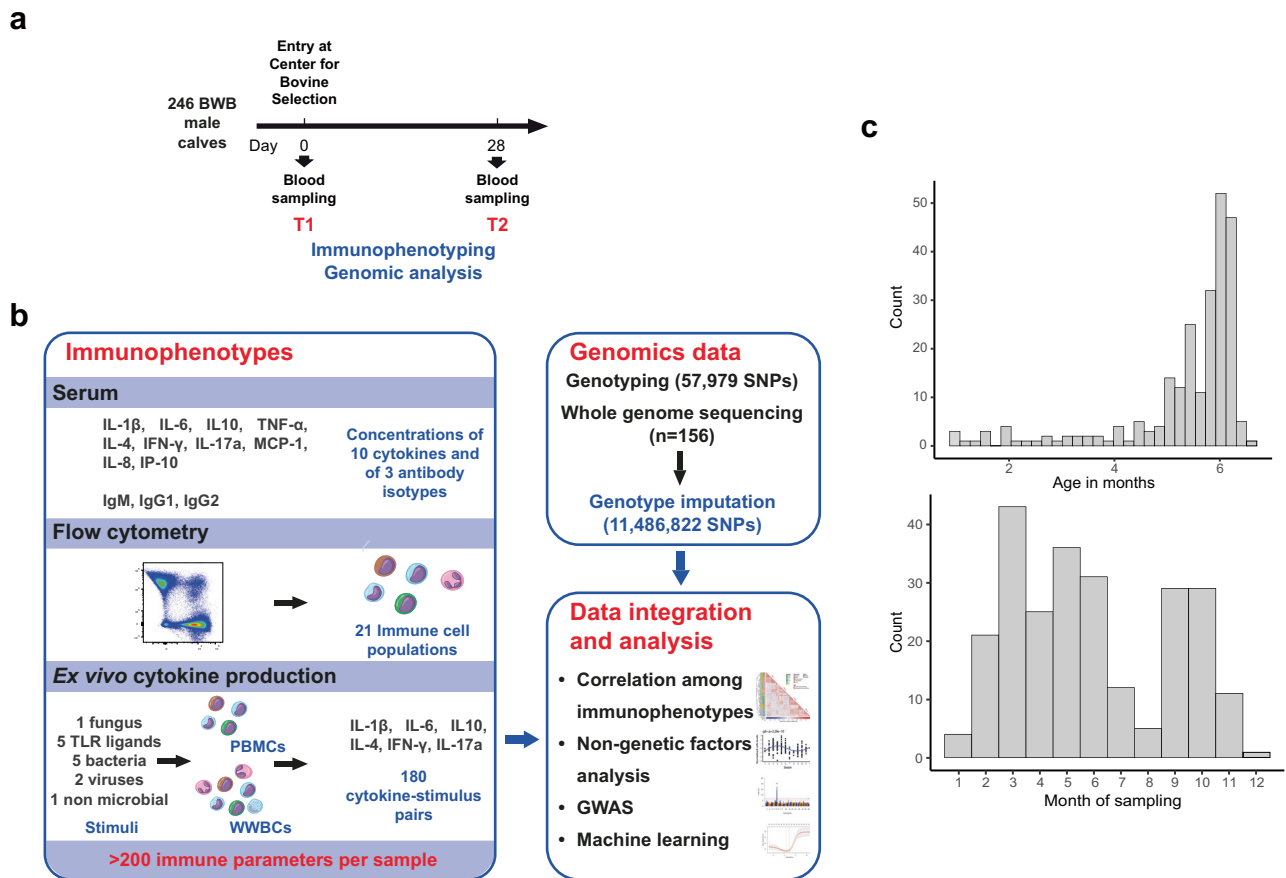


Fig. 1 | Study overview. **a** Sampling points. 246 male calves, entering the Center for Bovine Selection (CBS), underwent two blood samplings 28 days apart. Syringes represent treatment (vaccines, antiparasitic) administrations. **b** Immunophenotypes

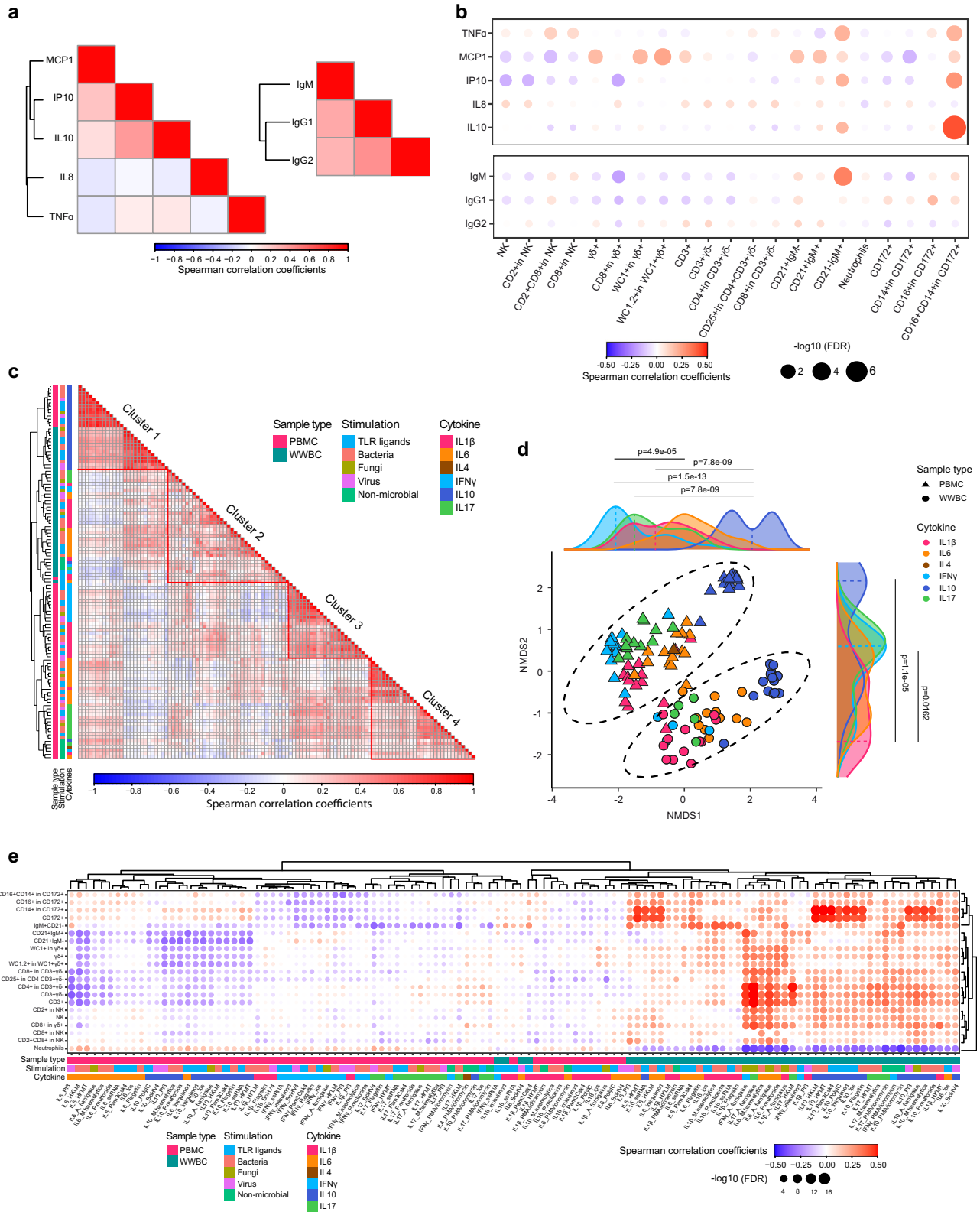
were measured on each blood sample and genomics data were collected for each animal before integration. **c** The distribution of ages at CBS entry (top) and month of sampling (bottom) for the animals in the cohort.

monocyte-derived cytokine productions and lymphocyte populations. Specifically, CD3+ $\gamma\delta^-$, CD3+ $\gamma\delta^+$ and B cell subpopulations were inversely linked with HKLM- and HKMT-stimulated IL-6 production by PBMCs (FDR < 0.005). Similarly, those populations of lymphocytes negatively correlated with the production of IL-10 from PBMCs after various stimulations (Fig. 2e).

Contribution of inter- and intra-individual factors to immune variation

Out of the 246 animals, 211 underwent two separate samplings (which enabled an investigation into the evolution of immunophenotypes over time). Initial sampling (T1) occurred when the animals entered the CBS, followed by standardized vaccines or treatments (Supplementary Table 1). Re-sampling took place 28 days after entry (T2) (Fig. 1a). When we compared correlations between data obtained at T1 or between data obtained at T2, we observed similar patterns (Supplementary Fig. 6a, b). Interestingly, when we correlated immunophenotypes measured at T1 with those measured at T2, the cytokine-dependent patterns described above were still observed suggesting that animals display functional immune properties that are conserved over time (Supplementary Fig. 6c). To assess how the immune landscape of calves evolved over time, a longitudinal ANOVA model was developed, incorporating two independent variables: individual calf ID and sample time point (Fig. 3a and Supplementary Fig. 7)⁴³. Each immune parameter was tested separately as response variable (Supplementary Fig. 7). The majority of the variance in each parameter was explained by a model depicting stable intraindividual immune profiles across longitudinal sampling (Fig. 3a). The median R^2 across the parameters

was 0.66, ranging from 0.50 to 0.89 (Supplementary Fig. 7). Notably, 65.97% of models maintained statistical significance (Bonferroni-corrected $p < 0.05$) even after correction for multiple testing (Supplementary Fig. 7a, b). Variation between repeated samples of the same individuals contributed minimally to the observed total variation, with a median proportion of R^2 of 0.0091 (range 1.21e-05–0.067). In contrast, interindividual variation explained a much larger effect, with a median proportion of R^2 of 0.663 (range 0.523–0.891) (Fig. 3a and Supplementary Fig. 7c, d), with some exceptions where sampling time (intraindividual variation) was significant (Supplementary Fig. 8). Consistent with this result, moderate but significant (FDR < 0.05) correlations between time points were found for most immunophenotypes, with a median correlation coefficient of 0.323 (range -0.0295 to 0.796). Interestingly, the temporal correlation between immunophenotypes measured in serum was significantly higher than for the production of cytokines after stimulation (FDR < 0.05; Fig. 3b). Furthermore, cytokine production following stimulation in WWBCs exhibited slightly higher interindividual variability, but lower intraindividual variability compared to PBMCs ($p < 0.05$) (Supplementary Fig. 7e). To further explore whether cumulative minor parameter changes could anyway lead to intraindividual variation over time, NMDS on paired samples was conducted. As expected, samples did not primarily cluster by sampling point (Supplementary Fig. 9a). However, NMDS2 revealed a difference between T1 and T2 ($p = 0.0034$) (Supplementary Fig. 9a). This difference originated from immune cell proportions ($p = 0.005$) and from the production of cytokines after stimulations ($p = 0.00041$) (Supplementary Fig. 9a). To analyze this sampling point variation further, an integrative analysis method called



multiomics factor analysis (MOFA) was used to identify shared sources of variation and correlation between T1 and T2 (Supplementary Fig. 9b)^{44–46}. This method enabled the detection of hidden parameters (namely latent factors (LFs)) that explain variance across the whole dataset and their potential relationship with differences between sampling points. For the immunoglobulins and cytokines in serum, percentages of immune cells and production of cytokines after stimulation, the MOFA model explained 2.11%, 66.05%, and 54.27% of the

variance, respectively. Seven LFs were identified to explain the variance in the combined dataset. LF1 explained most of the variance observed across the percentages of immune cells (Supplementary Fig. 9b) but did not explain differences between T1 and T2 (Supplementary Fig. 9c). LF1 was mainly impacted by the variation in the percentages of neutrophils and T cells among samples (Supplementary Fig. 9d). LF6, explaining 8.99% of the total variance in the model, was the most significant LF that could distinguish between T1 and T2

Fig. 2 | Correlation between individual immune parameters. **a** Unsupervised hierarchical clustering of levels of cytokines or immunoglobulins in serum using Spearman correlations as the measure of similarity. Red depicts positive correlation and blue indicates negative correlation. **b** Dot plot showing pairwise Spearman's rank correlation coefficients between each cell subpopulations and serum immunoglobulin and cytokine levels. Two-sided p -values were adjusted using the Benjamini–Hochberg false discovery rate (FDR) correction. **c** Unsupervised hierarchical clustering of the cytokine responses in PBMCs and WWBCs induced by various stimulations using Spearman correlations as the measure of similarity. Red depicts positive correlation and blue indicates negative correlation. **d** Non-metric MultiDimensional Scaling (NMDS) analysis for production of cytokines after ex vivo

stimulation of PBMCs and WWBCs. Clustering was performed using pairwise Spearman's correlation as the measure of similarity. Differences between two unpaired groups were evaluated using a two-sided Wilcoxon rank-sum test. **e** Dot plot showing pairwise Spearman's rank correlation coefficients between cytokine production and percentages of immune cells in blood. Clustering was performed using pairwise Spearman's correlation as the measure of similarity. Two-sided p -values were adjusted using the Benjamini–Hochberg FDR correction. This approach ranks and adjusts all p -values to control the estimated number of false positives among all statistically significant associations. NMDS, Nonmetric multidimensional scaling; FDR, false discovery rate. The correlations among immunophenotypes were calculated using the datasets from T1 ($n = 246$ animals).

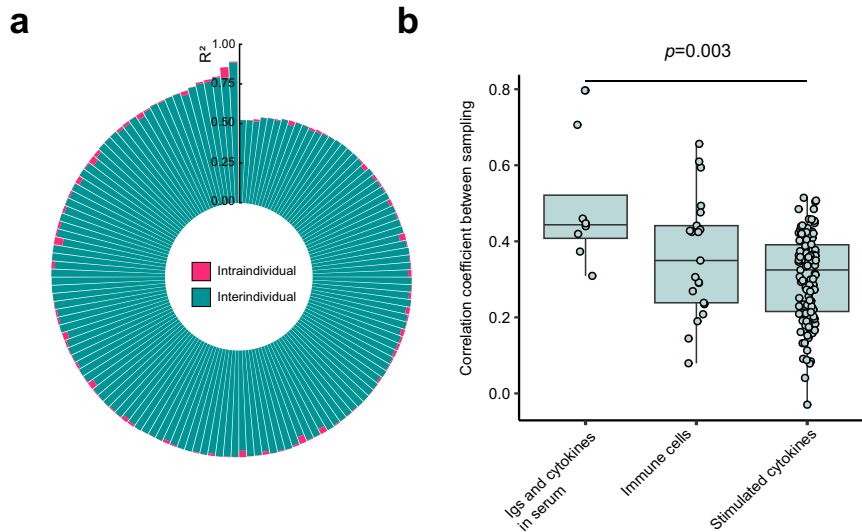


Fig. 3 | The effect of sampling point on the immune variation. **a** The R^2 values from all individuals attributable to either interindividual or intraindividual differences were assessed based on T1 and T2 data. Each division of the circle represents a distinct immunophenotype. For each immunophenotypes, R^2 value was calculated using T1 and T2 data from 211 animals. **b** Spearman correlation coefficient of each immunophenotypes between two time points. Each dot represents a distinct immunophenotype ($n = 8$ for Igs and cytokines in serum; $n = 21$ for immune cells;

$n = 116$ for stimulated cytokines). The center line in the boxplot represents the median value; box boundaries represent the 25th and 75th percentiles; whiskers extend to 1.5 times the interquartile range of the 25th and 75th percentiles; all the values are shown on the graph. The pairwise test (two-sided) for multiple comparisons of mean rank sums (Dunn's test) was used to check the differences and p -values were adjusted by Benjamini–Hochberg FDR correction.

(Supplementary Fig. 9c) (Bonferroni-corrected $p = 1.37e-08$). We therefore assessed the underlying factors contributing to LF6 to better understand the biology behind these sampling time-point differences. It revealed that IL-17 production and percentages of $\gamma\delta^+$ and $\gamma\delta^-$ T cells were the main constituents of LF6 reflecting the fact that T2 samples displayed lower percentages of $\gamma\delta^+$ T cells at steady state and an increased capacity of producing IL-17 upon stimulation (Supplementary Fig. 9e). Collectively, these findings suggest that, as seen in human, individual specificities are the main source of variation of immunophenotypes and the repeatability of the measures between sampling points is high. Anyway, some immunophenotypes display variation explained at least partially by sampling point and could be used in the future to monitor longitudinal variation.

Effect of age and season on the interindividual variation of immune parameters

Age significantly influences immune profiles^{10,16,43,44}. We therefore investigated age effect on the data collected at T1 using a Spearman correlation test after correcting for the effect of season. Our analysis revealed both negative and positive correlations between immune parameters and age (Fig. 4a). As the variation of ages is limited in the cohort (Fig. 4b), we conducted permutation tests to underscore the robustness of the age-related impact on immune traits. 95.8% of immunophenotypes yielded results consistent with the actual estimates (Supplementary Data 1). Specifically, age was linked to an

increase in CD21+IgM- B cells, accounting for 8.9% of the variance, and elevated serum IgM levels (FDR < 0.05) (Fig. 4a, b). Age was also related to increased serum TNF- α levels (FDR < 0.05). In terms of cytokine production induced by stimulations, the levels of IL-10 induced by several stimulations in PBMCs decreased with age (FDR < 0.05) (Fig. 4a, b and Supplementary Data 1). Despite the presence of some correlations, age exerted only a moderate influence (average adjusted $R^2 = 0.013$, range from $2.35e-06$ to 0.089) on immune profile variation within our cohort of healthy, young calves, which showed limited age variability.

Seasonality is another important factor that has previously been shown to influence immune cell populations and gene expression^{10,16,47}. Similarly, we observed a seasonal effect on immune cell subpopulations and cytokine productions (Fig. 4c, d). Importantly, this was not related to some batch effect due to sampling point as these differences were observed from data collected on a weekly basis over two years and shown by month. During spring, B and T cell subsets were more abundant. Conversely, intermediate (CD14+ CD16+ in CD172+) and non-classical monocytes (CD14- CD16+ in CD172+) were found at higher levels during winter and classical monocytes (CD14+ CD16- in CD172+) during summer. Neutrophils levels displayed a peak during autumn (Fig. 4c, d and Supplementary Data 1). In line with immune cell patterns, cytokine production was similarly influenced by season. Interestingly, similar to classical monocytes, the levels of IL-6 (monocyte-derived cytokines) induced by stimulation of PBMCs were higher

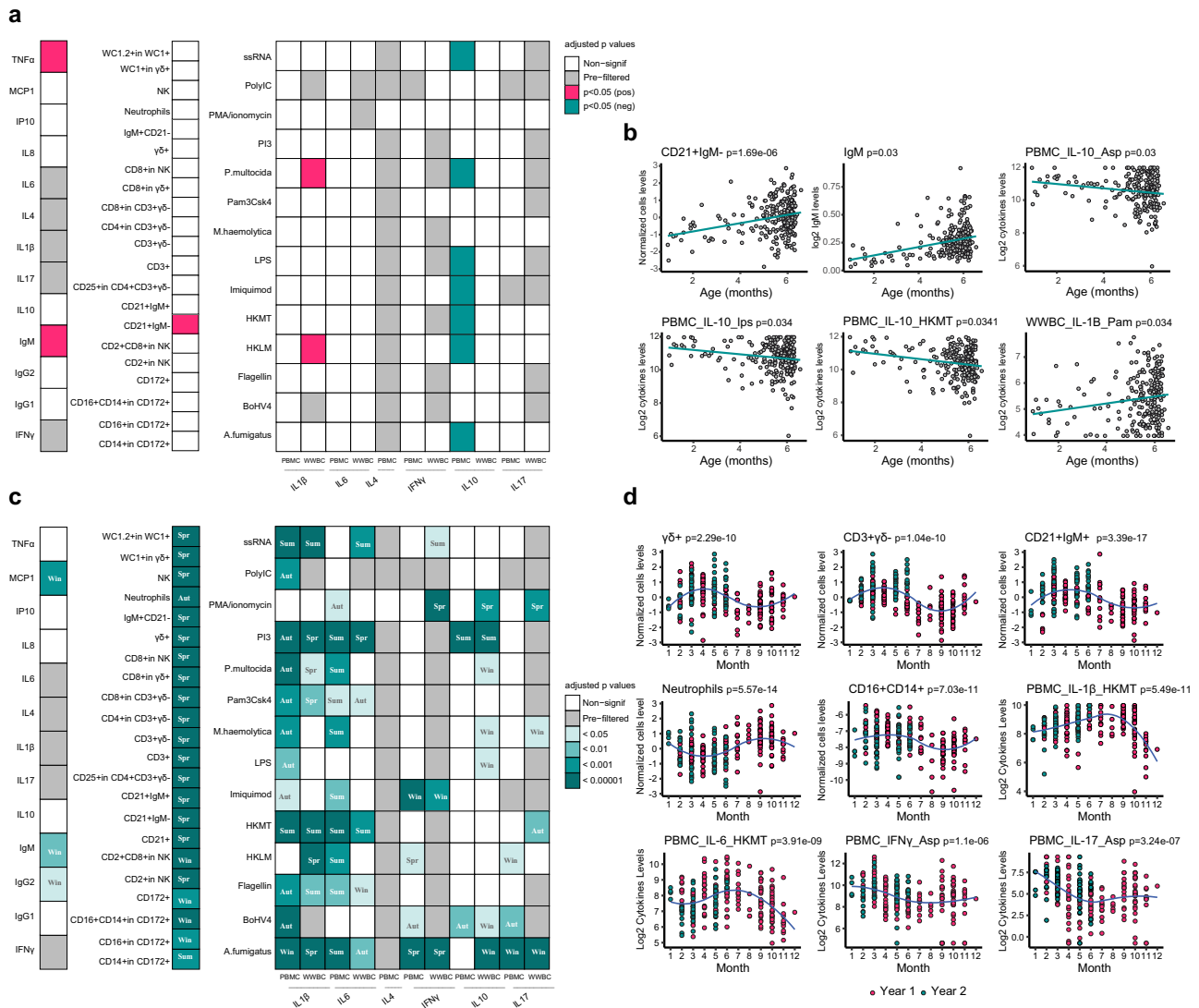


Fig. 4 | The effect of age and season on immune variation. **a** Heatmap depicting the statistical significance of the correlation between age and immunophenotypes. Pink and teal color correspond to positive and negative correlations, respectively. A gray color means that no analysis could be performed. *p*-values were derived from linear models adjusted for the effect of season and corrected for multiple testing using the Benjamini-Hochberg procedure to control FDR. **b** Examples of immunophenotypes that were significantly associated with age. *p*-values were derived from linear models adjusted for the effect of season and corrected for multiple testing using the Benjamini-Hochberg procedure to control FDR. **c** Heatmap depicting the statistical significance of the association between season and immunophenotypes. Darker teal colors indicate higher levels of significance. The month which is statistically associated with significant difference compared to

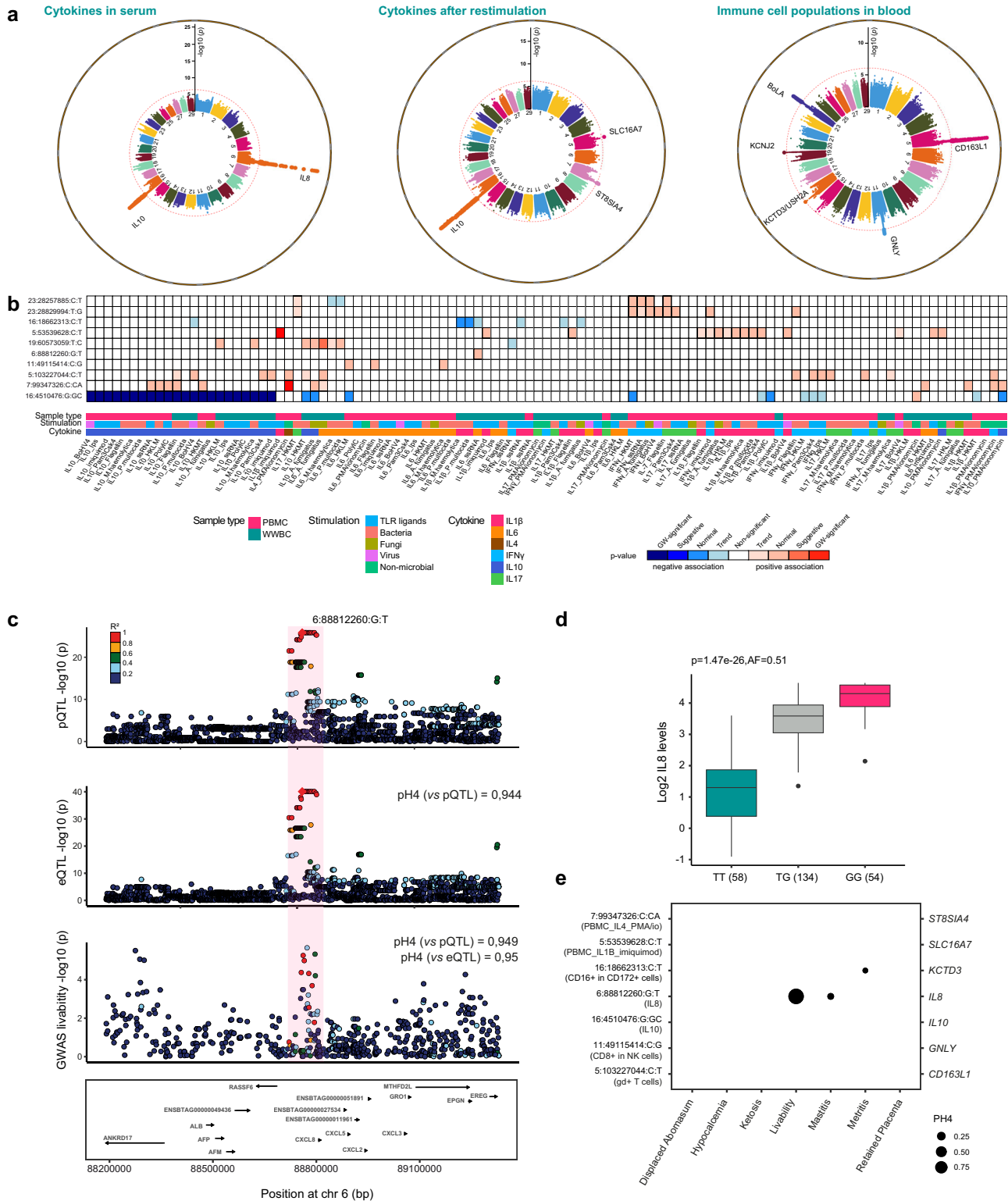
others is indicated in the corresponding box of the table. A gray color means that no analysis could be performed. Statistical significance was determined using a variance test comparing the full model with an alternative regression model (see Methods). *p*-values from all season-immunophenotype associations were adjusted using the Benjamini-Hochberg procedure to control FDR. **d** Examples of immunophenotypes that were significantly associated with season. Statistical significance was determined using a variance test comparing the full model with an alternative regression model (see Methods). *p*-values from all season-immunophenotype associations were adjusted using the Benjamini-Hochberg procedure to control FDR. The associations of age and season on immune variation were calculated using the datasets from T1.

in summer (Fig. 4c, d and Supplementary Data 1). Through a resampling approach, 96% of immunophenotypes showed consistent results when compared to the original full dataset in more than 70% of iterations (Supplementary Fig. 10). Overall, these findings underscore the significant role of seasonal environmental factors in modulating variations in immune cell populations and cytokine production.

Identification of genetic variations that impact immunophenotypes

Based on these results, we conducted genome-wide association studies (GWAS) for each immunophenotype using a mixed linear model correcting for the effect of age and season to identify common genetic variants (minor allele frequency > 0.05) influencing

immunophenotypes. This analysis unveiled ten independent genome-wide significant quantitative trait loci (QTLs) ($p < 1.3e-07$) from the immunophenotypes collected at T1 (Fig. 5a and Supplementary Data 2-4). When we repeated those analyses for immunophenotypes collected at T2, 6 regions could again be identified as significant at T2 through the identification of significant loci ($p < 1.3e-07$) in high linkage disequilibrium (LD) with loci identified at T1 (Supplementary Fig. 11a and Supplementary Data 2). Finally, we re-sampled an independent cohort of 239 calves to test the replication of the signals obtained from the cytokine measurements performed directly on serum (IL-8 and IL-10). In both cases, we again identified significant loci ($p < 1.3e-07$), each of which was in strong LD with the loci identified in the first cohort at both T1 and T2 (Supplementary Fig. 11b and



Supplementary Data 2). We then conducted conditional GWAS by adjusting those immunophenotypes on the 10 leading associated variants and found no independent locus that reached genome-wide significance. Finally, we looked at the effect of these independent QTLs on all the immunophenotypes in order to reveal some patterns of effects (Fig. 5b). Most of these QTLs displayed trend, nominal or suggestive signals with other immunophenotypes suggesting their general importance in bovine immune responses. We then specifically investigated these 10 different loci.

Genetic variants associated with cytokines production. We first identified a QTL on chromosome 6 that governs serum IL-8 levels and accounts for 31.6% of the phenotypic variation. This QTL was associated with genetic variation in the vicinity of the *IL-8* gene (*CXCL8*, Fig. 5c). These variants were located either in an intronic region of the gene or in 5' or 3' untranslated regions (Supplementary Data 3). The lead SNP identified from T1 dataset (6:88812260:G:T) was associated with significant differences in IL-8 concentrations in serum with the reference TT allele producing less detectable IL-8 (Fig. 5d). Then, based

Fig. 5 | Genome-wide significant associations with immunophenotypes.

a Manhattan plots showing the genome-wide association results for the cytokines in serum, the stimulated cytokines and immune cells, respectively. Each dot represents a SNP. p -values were obtained from mixed linear models including age and season as fixed effects and kinship as a random effect (see Methods). To account for multiple comparisons, significance was assessed using a permutation-based approach. p -values $< 1.3 \times 10^{-7}$ (the threshold is highlighted in red) were considered genome-wide significant. **b** An overview of the association of ten independent lead variants reaching genome-wide significance for at least one immune phenotype with all other immune phenotypes. p -values were obtained from mixed linear models including age and season as fixed effects and kinship as a random effect (see Methods). Darkest colors indicate genome-wide significant associations, whereas red and blue colors distinguish a positive or negative direction of effect, respectively. GW significant, suggestive, nominal, and trend correspond to p -values $< 1.3 \times 10^{-7}$, $< 2.6 \times 10^{-6}$, < 0.01 , and < 0.05 , respectively.

on transcriptome-wide gene expression profiling of resting whole blood cells from the same animals (Dr C. Charlier, ENA accession number: PRJEB94322), we analyzed expression quantitative trait loci (eQTL) governing the expression of *IL-8* gene (Fig. 5c) and colocalized them with genetic variants associated with variation of IL-8 detection in serum. A strong colocalization evidence was revealed between the *cis*-pQTL and *cis*-eQTL of *IL-8* ($PH4 = 0.944$) suggesting that *cis*-acting variants regulate *IL-8* expression at steady state. Finally, in order to identify the possible consequences of this genetic variation on animal health, we tested the colocalization of those SNPs with loci recently associated with clinical traits in the Holstein breed⁴⁸. These pQTL and eQTL colocalized with a region associated with cow livability, a health-related herd management trait that refers to the ability of a cow to remain alive and stay in the herd during a specific period ($PH4 = 0.949$ and 0.95 , respectively), suggesting that it could have a significant impact on the health of those animals (Fig. 5c, e and Supplementary Data 4).

Similarly, the region associated with variation in IL-10 production was located on chromosome 16 and encompassed the *IL-10* gene itself (Supplementary Fig. 12). Alleles associated with lower IL-10 production bear 4 coding SNPs (LD with lead SNP > 0.8 , Supplementary Data 3). However, these coding mutations, potentially impeding protein recognition, do not appear to be the cause of the difference observed, as detection of the protein remained possible (albeit at lower levels) in animals homozygous for the mutant allele (Supplementary Fig. 12). Unexpectedly, this pQTL region is quite extensive, encompassing numerous SNPs that could account for the variability in IL-10 production. When we tried to colocalize it with *cis*-eQTL for *IL-10* expression identified as described above (transcriptomic data from resting whole blood cells from the same animals), we did not detect any colocalization, suggesting that this difference in protein detection was not associated with gene expression differences. Interestingly, one of the mutations identified is located at a site potentially involved in the splicing of the *IL-10* mRNA. In the future, additional analyses will be necessary to pinpoint the molecular mechanisms responsible for the observed defect in IL-10 production within specific genotypes and to decipher the potential functional consequences on the health of those animals.

Two additional loci functioned mainly as *trans*-pQTLs (> 1 Mb from the gene encoding the protein; Supplementary Fig. 12). The first, chr5:53539628:C:T, was associated with IL-1 β production in PBMCs following imiquimod stimulation, explaining 11% of phenotypic variance. This intergenic SNP is located near *SLC16A7* (*MCT2*), and individuals with the TT genotype showed higher cytokine levels than CC homozygotes. The second locus, chr7:99347326:C:CA, was linked to IL-4 production after PMA/Ionomycin stimulation. The lead SNP lies upstream of *FAM174A* and *ST8SIA4*, and the minor allele was associated with increased IL-4 concentrations ($p = 4.2 \times 10^{-8}$). These two signals were, however, not replicated with T2 data (Supplementary Data 2).

c Colocalization analyses of *cis*-pQTL for IL-8 with eQTL of IL-8 and GWAS of cattle livability in the 1-Mb region around 6:88812260:G:T. We evaluated whether the *cis*-pQTL for IL-8 shared the same causal variant with eQTL of IL-8 and GWAS of cattle livability using colocalization as described in the methods. Each dot is an SNP; color indicates linkage disequilibrium (r^2), with the lead SNP (6:88812260:G:T) in the square. **d** Box-plot of levels of IL-8 in serum with the genotypes of lead SNP (6:88812260:G:T) in (C). $n = 246$ (the numbers of animals for each genotype are indicated between brackets). The center line in the boxplot represents the median value of normalized IL-8 levels; box boundaries represent the 25th and 75th percentiles; whiskers extend to 1.5 times the interquartile range of the 25th and 75th percentiles; and outliers are shown by points outside the whiskers. The reported p -value has been obtained as in (a) and (b). **e** Overview of colocalization evidence of *cis*-pQTL of indicated proteins and 6 clinical diseases. The GWAS described here were calculated using the datasets from T1.

Genetic variants associated with proportions of immune cells. Six independent genetic loci were identified as being significantly associated with variations in the percentages of certain immune cell populations. Firstly, genetic polymorphism in the region of *WCI* genes, containing 13 genes in cattle⁴⁹, was associated with the percentage of *WCI2+* in $\gamma\delta + T$ cells (Supplementary Fig. 12). The results point towards a dominant allele negatively controlling the expression of *WCI2* antigens on $\gamma\delta + WCI + T$ cells. It will be important to analyze the functional consequences of this variation in the future, given that the differential expression of these genes determines the type of $\gamma\delta + T$ cell subpopulations and that each of these subpopulations has specific functional properties against pathogens frequently encountered by cattle⁴⁹.

A second signal linked to variation in the percentage of *WCI2+* cells among $\gamma\delta + WCI + T$ cells was identified near the *KCNJ2/KCNJ16* genes, which encode potassium channel components (Supplementary Fig. 12). Colocalization analysis showed moderate evidence ($PH4 = 0.58$) between the *KCNJ2 cis*-eQTL (identified as above from resting whole blood cell transcriptomes) and this *trans*-pQTL, suggesting that *KCNJ2* expression may regulate this T cell subset and play a role in cattle immunity. This signal was, however, not replicated with T2 dataset (Supplementary Data 2).

Thirdly, genetic variation on chromosome 11 upstream of the *GPLY* gene (distance = 143,003 bp) was found to be linked with the percentage of CD8+ in NK cells ($p = 2.2 \times 10^{-8}$) (Supplementary Fig. 12). *GPLY* is an essential effector molecule of killer lymphocytes involved in immune response defense against bacterial and viral infection⁵⁰.

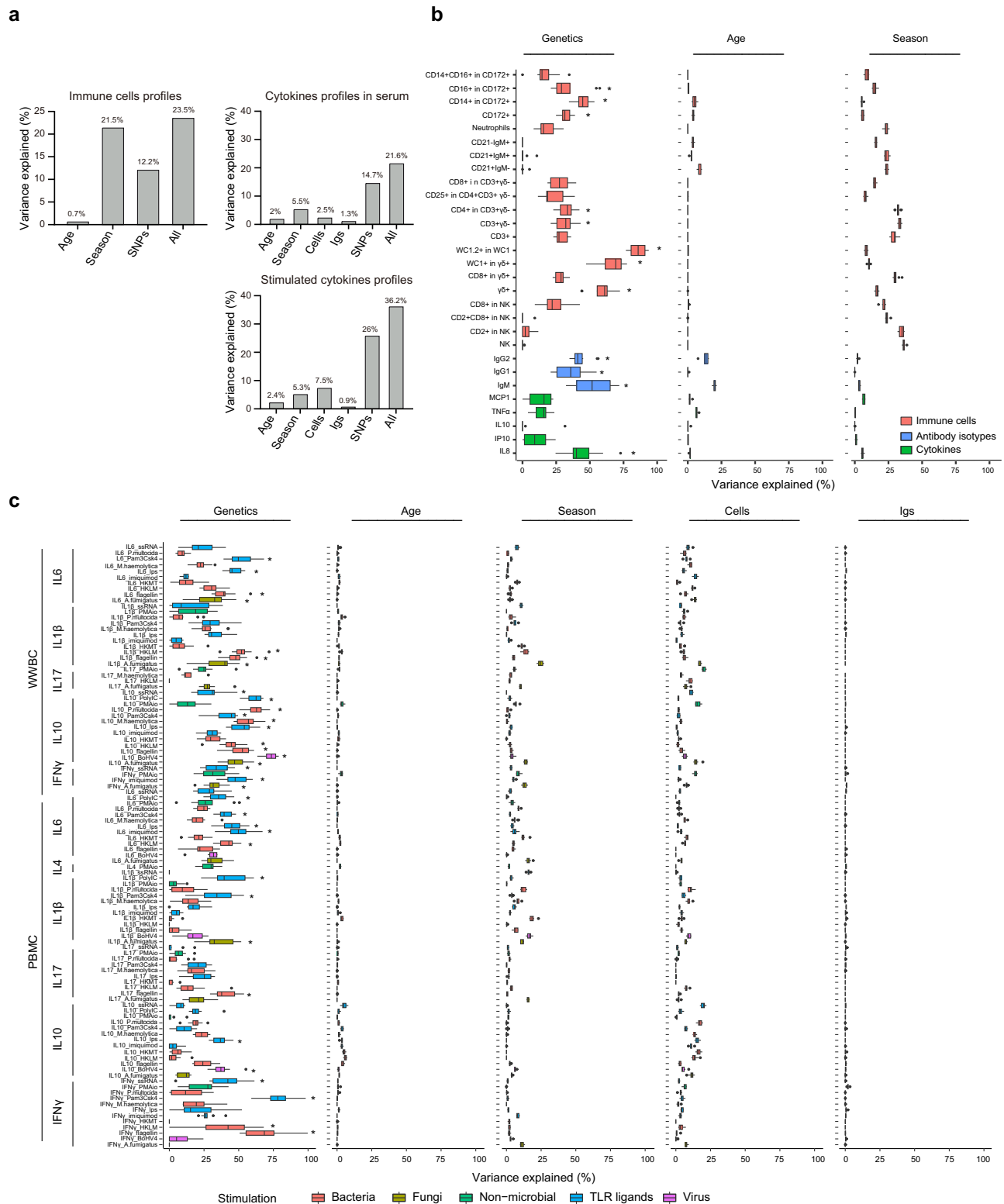
Fourthly, genetic variations around the *MHC* locus on chromosome 23 were associated with variations of the percentage of CD4+ in CD3+ cells (distance = 105,595 bp, $p = 1.16 \times 10^{-7}$) and of the percentage of CD8+ in CD3+ cells (distance = 421,160 bp, $p = 2.49 \times 10^{-9}$) (Supplementary Fig. 12). The LD between the lead SNPs linked with the percentage of CD4+ in CD3+ cells and the percentage of CD8+ in CD3+ cells is minimal ($r^2 = 0.556076$), suggesting that the percentages of CD4+ and CD8+ T cells were determined by distinct variations, even if the quality of genetic information at that location is poor.

Finally, genetic polymorphism on chromosome 16 (chr16:18662313:C:T) was associated with variation in the different monocyte populations among CD172+ cells (Supplementary Fig. 12). However, no SNP hit was highlighted at T2.

In addition, we identified 93 genome-wide suggestive signals ($p < 2.6 \times 10^{-6}$; Supplementary Data 5) warranting further investigation.

Determinants of interindividual variation of cytokine profiles at baseline or after stimulation ex vivo

We then systematically quantified the effects of genetic and non-genetic factors on the immunophenotypes. To assess the relative contributions of age, season, and genetics (based on significant and suggestive SNPs) to interindividual variation in immune cell



proportions, we used PERMANOVA to estimate the variance explained by each factor (Fig. 6a and Supplementary Data 6)^{51,52}. Notably, without controlling for covariates, season (21.5%) emerged as the primary contributor to this interindividual variation, overshadowing the effect of the GWAS loci (12.2%). Using a similar approach, we calculated the proportion of variance explained by age, season, immune cells, immunoglobulins, and the same genetic loci for the entire profile of cytokines measured in serum or induced by stimulations (Fig. 6a). In contrast to immune cell profiles, genetic factors accounted for the

largest proportion of variance in cytokine traits, consistent with human studies⁵³. By comparison, immune cell proportions and season had moderate effects (5.3% and 7.5%, respectively) across most cytokine-stimulus combinations, while circulating immunoglobulin levels and age had minimal impact. These factors cumulatively explained 21.6% and 36.2% of the variance in cytokine production at steady (serum) and activated states, respectively (Fig. 6a). Subsequently, we calculated the proportion of variance explained by genetic (using here all genetic markers from SNP array (38,828 SNPs across the

Fig. 6 | Interindividual variation of cytokine profiles at baseline or after ex vivo stimulation. **a** Interindividual variation in each dataset (eg., immune cells and cytokines production) explained by the indicated factors, estimated using the PERMANOVA method with 1,000 times permutation. All, all of the indicated factors combined; Igs, immunoglobulin levels. **b** Contribution of genetics, age and season to interindividual variation for immune cells and cytokine production in serum. The variance explained through the multivariate linear models (MVLMS) trained on ten random subsets of the full dataset (representing 90% of all samples) is represented. The center line in the boxplot represents the median value; box boundaries represent the 25th and 75th percentiles; whiskers extend to 1.5 times the interquartile range of the 25th and 75th percentiles; and outliers are shown by points outside the whiskers. The variance explained through the MVLMS using the entire

T1 dataset is shown in Supplementary Data 6. The heritability was considered significant if it was in the top 5% of the permuted distribution of immunophenotype estimates. * $p < 0.05$. **c** Each category's contribution to interindividual variation for the stimulated cytokines. The variance explained by multivariate linear models (MVLMS) trained on ten random subsets of the full dataset (representing 90% of all samples) is represented. The center line in the boxplot represents the median value; box boundaries represent the 25th and 75th percentiles; whiskers extend to 1.5 times the interquartile range of the 25th and 75th percentiles; and outliers are shown by points outside the whiskers. The variance explained through the MVLMS using the entire T1 dataset is shown in Supplementary Data 7. The heritability was considered significant if it was in the top 5% of the permuted distribution of immunophenotype estimates. * $p < 0.05$.

genome)) and non-genetic factors for each independent immunophenotype (Fig. 6b, c and Supplementary Data 7). Interestingly, and in line with previous study¹⁰, we found that genetic factors explained a significantly greater proportion of variance in T-cell immune traits compared to B-cell traits (Student's t test, $p = 0.01$) (Fig. 6b). In contrast, unlike previously reported in human⁷, we did not highlight differences in variance explained by genetic factors between innate and adaptive immune cells (Student's t test, $p = 0.582$). This result could be attributed to the small size of the cohort and to the fact that we used the percentage of immune cell populations as parameters instead of Mean Fluorescence Intensity (MFI) measurements of individual receptor expression.

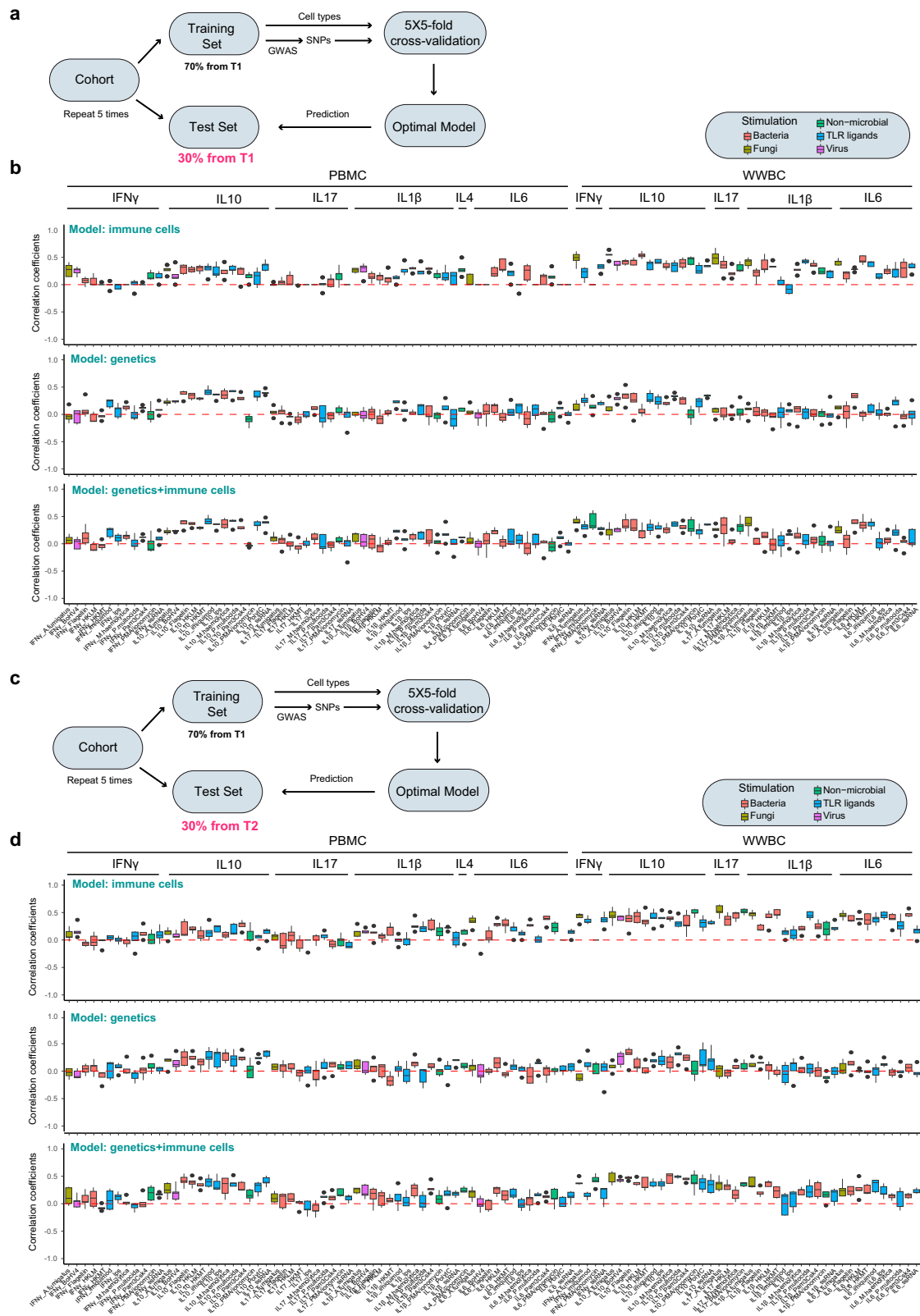
In addition, we determined which was the dominant factor (i.e. the factor accounting for more variance than the others) for each immunophenotype. For proportion of immune cells, season was the dominant factor in 57% of the cases and genetics was in 43% (Fig. 6b and Supplementary Data 6). In contrast, interindividual differences in stimulated cytokine production were found to be primarily explained by genetics in 77% of cases. In 10.5% of the cases, season was the dominant factor and for 11.5% of them, it was immune cells (Fig. 6c and Supplementary Data 7). To further gauge the significance of the estimated contribution of genetics (h^2_{SNP}) to the observed total variance, we performed 1000 permutations of the animal IDs in our dataset and then applied the analysis pipeline to the permuted data to obtain the empirical distribution of the estimates of h^2_{SNP} (null distribution). The estimate of h^2_{SNP} from the original data was then compared to the estimate of h^2_{SNP} from the permuted data. Among the 21 immune cell phenotypes and the 104 stimulated cytokines immunophenotypes (Figs. 6c), 7 and 35 exhibited significant h^2_{SNP} estimates, respectively ($p < 0.05$, Fig. 6b, c). In particular, 12 of these belonged to IL-10 production induced by stimuli in WWBCs and PBMCs.

Finally, we compared data on stimulus-dependent cytokine production from the two different types of cells (PBMCs and WWBCs) from the same individuals used in our study. This highlighted that genetic factors, proportions of immune cell populations, and the concentration of immunoglobulin levels had a greater effect on stimulus-dependent cytokine production in WWBCs than in PBMCs (Mann-Whitney U test, $p < 0.05$) (Supplementary Fig. 13a). When we compared the effect of these different factors in function of the stimulations used, we did not identify any major effect of any of these factors in function of a particular type of stimulus (Supplementary Fig. 13b). In contrast, while genetics was still the major factor explaining variation for the production of each cytokine, it appeared that season explains more variance for IL-1 β production (and to a lesser extent for IL-6) than for the other cytokines (IFN γ , IL-10, IL-17 and IL-4) (Supplementary Fig. 13c). For example, season explained 18.7% and 16.3% of the variance of IL-1 β production by PBMCs following *P.multocida* and *M.haemolytica*-stimulation, respectively. These results also show that, although genetics is the most important factor contributing to interindividual variation in cytokine production (mean heritability = 0.241, ranging from $1e-06$ to 0.756) (Fig. 6c), the majority of the

variation in our population could still be explained by nongenetic factors, which is consistent with human studies^{8,53}.

Cytokine production in response to stimulations can be partly predicted on the basis of the immune cell composition and genetic factors

Previous studies have proposed that cytokine responses to ex vivo stimulation could be predicted from steady state measurements (i.e. genetics and blood cell composition) through machine learning approaches with various degrees of success⁵³⁻⁵⁶. We tested therefore the possibility of predicting cytokine production using a cross-validation strategy that involved randomly splitting of the data into training (70% of the animals) and validation (30% of the animals) sets multiple times to predict cytokine productions. We initially explored whether immune cells proportions could be used to predict the cytokine response to ex vivo stimulations as previously shown in human (Fig. 7a, b and Supplementary Data 8)⁵⁵. The average correlation coefficient between predicted and measured stimulus-induced cytokine productions was 0.214 (range -0.083 to 0.542). Interestingly, in comparison to PBMCs, cytokines induced by stimulation in WWBCs are more likely to be predicted by immune cell composition (Wilcoxon signed-rank test, $p = 1.574e-09$). Notably, five cytokine-stimulus pairs exhibited prediction accuracy exceeding 0.45 in WWBCs, including *A.fumigatus*-induced IL10 production (mean 0.542), HKMT-induced IL-10 production (mean 0.537), *A.fumigatus*-induced IL-17 production (mean 0.495), *A.fumigatus*-induced IFN γ production (mean 0.478), and HKMT-induced IL-6 production in WWBCs (mean 0.454) (Fig. 7b). Secondly, genetic factors (corresponding here to SNPs with a p -value $< 5e-05$ in the GWAS analysis) were tested to predict cytokine production in response to stimulation (see Methods)⁵³. When we used genetic information from GWAS analyses performed on the entire cohort, correlations of approximately 0.75 were observed between predicted and measured values (Supplementary Fig. 14). However, in this setup, the test animals contribute to the identification of GWAS signals and thus to model construction. Therefore, in a second approach, to ensure the models were completely independent from the test cohorts, we repeated the GWAS analyses for each model using only the data from the corresponding training cohort (Fig. 7a, b). In this setting, the average correlation between predicted and measured stimulus-induced cytokine production was 0.096 among the cytokine-stimulus pairs (range -0.102 to 0.421). However, some, such as the prediction of IL-10 production, consistently showed a correlation of the order of 0.3. Finally, the inclusion of both immune cell proportions and genetic factors in the model yielded to notable enhancement in predictive accuracy compared to genetics only (FDR < 0.05 , Student's t test) with an average correlation of 0.151 between predictions and measures (range -0.066 to 0.422) (Fig. 7b). To further validate the accuracy of prediction, we trained the model on 70% of the cohort with cell proportions measurements from T1 sampling and predicted cytokine production for the other 30% of the cohort at T2 (Fig. 7c, d and Supplementary Data 9). The predictions made



using immune cell proportions, genetics, or genetics plus immune cells resulted in comparable accuracies at T2 than the ones obtained for T1, with mean values of 0.204, 0.070, and 0.205, respectively. In general, models based on cell population percentages gave better results than those based on genetics alone (Fig. 7). However, a clear improvement was observed for IL10 production estimates, which could be expected given the prior identification of a highly

significant pQTL region for production of this cytokine (Fig. 5 and Supplementary Fig. 12).

Discussion

While immune variation has been extensively studied in humans, much less is known about these dynamics and their determinants in other species, particularly cattle. This study is the first to apply a systems

Fig. 7 | Prediction of ex vivo cytokine production upon ex vivo stimulation. **a** Design of the analytical pipeline for **(b)**. The cytokine levels were predicted using an Elastic Net model. To validate the prediction, a 5×5 cross-validation strategy was used. Data from T1 were randomly split into training and test sets. GWAS for each cytokine-stimulus pair in the training set were conducted, and SNPs with an association to a cytokine-stimulation pair with $p < 5 \times 10^{-5}$ were kept as genetic factors. The prediction accuracy was assessed by computing Spearman correlations between measured cytokine levels and Elastic Net model predictions for the test sets. **(b)** The prediction of cytokine production in response to stimulation using immune cells, SNPs, and SNPs plus immune cells in T1. The prediction accuracy was assessed by computing Spearman correlations between measured cytokine levels and Elastic Net model predictions for the test sets. $N = 5$. The center line represents the median; box boundaries represent the 25th and 75th percentiles; whiskers extend to 1.5 times the interquartile range of the 25th and 75th percentiles; and outliers are shown by points outside the whiskers.

shown by points outside the whiskers. **c** Design of analytical pipeline for **(d)**. The cohort was randomly divided into two sets: training (from T1) and test (from T2). GWAS for each cytokine-stimulus pair in the training set were conducted, and SNPs with an association to a cytokine-stimulation pair with $p < 5 \times 10^{-5}$ were kept as genetic factors. The prediction accuracy was assessed by computing Spearman correlations between measured cytokine levels and Elastic Net model predictions for the test sets. **d** The prediction of cytokine production in response to stimulation using immune cells, SNPs, and SNPs plus immune cells in T2 using the model build from T1 data. The prediction accuracy was assessed by computing Spearman correlations between measured cytokine levels and Elastic Net model predictions for the test sets. $N = 5$. The center line represents the median; box boundaries represent the 25th and 75th percentiles; whiskers extend to 1.5 times the interquartile range of the 25th and 75th percentiles; and outliers are shown by points outside the whiskers.

immunology approach combined with GWAS to broadly assess immune traits in a domestic animal population. We examined how immune variation is shaped by genetic and non-genetic factors, including age and seasonality. Notably, we identified genome-wide significant variants strongly associated with immunophenotypes. Unlike immune cell profiles, cytokine responses were mainly driven by genetic factors, echoing patterns seen in humans. Importantly, genetic variants explained more variation in stimulated cytokine production (26%) than in baseline serum concentrations (14.7%), suggesting that genetics has a greater impact on immune reactivity than on steady-state cytokine levels, in line with previous studies^{28,53}. Finally, we demonstrated that future immune responses in cattle could be partly predicted from blood-based immune profiles (immune cell populations) and genotypes using models trained on a reference cohort. This opens possibilities for genetic selection and health management in livestock. Improving model performance will require larger training datasets and further testing over time, possibly integrating additional factors such as hormones, metabolites, and microbiota, an approach already shown to enhance prediction in humans⁵³.

Regarding immune parameter correlations, our results showed that cytokine production was primarily grouped by the cytokines produced, rather than by the stimuli used. This was unexpected, as immune responses are generally thought to be stimulus-specific, with distinct pathways triggered by different microorganisms⁵⁷. However, for the cytokines studied (IL-1 β , IL-6, IL-10, IL-17, and IFN- γ), we observed no clear distinction between responses to bacterial (HKMT, HKLM, *M. haemolytica*, *P. multocida*, Flagellin) and fungal (*A. fumigatus*) stimuli. Notably, IL-10 production was highly correlated across all stimuli. Similarly, despite some exceptions, monocyte-derived cytokines (IL-1 β and IL-6) showed consistent positive correlations across stimuli (cluster 2, Fig. 2c), as did certain lymphocyte-derived cytokines (IL-17 and IFN- γ) (Fig. 2c). Intriguingly, in contrast to human studies reporting weak correlations between monocyte- and T-helper-derived cytokines^{15,58}, we observed strong positive correlations between IL-1 β and IFN- γ in PBMCs (cluster 3, Fig. 2c). While recent human functional genomics studies have also found cytokine-driven clustering^{59–61}, other reports describe clustering by stimulus^{15,23}. These divergent findings suggest that multiple factors shape the regulation of stimulation-specific immune responses. Future studies across diverse cattle breeds and species could clarify these patterns. Conversely, we also observed negative correlations between certain lymphocyte populations and IL-6 production (a monocyte-derived cytokine) by PBMCs (Fig. 2e), consistent with reports that higher baseline levels of adaptive immune cells associate with reduced monocyte-derived cytokine production following ex vivo stimulation^{58,62}.

Although our results indicate that genetics plays a major role in shaping immune responses in the animal tested, a substantial proportion of the variance is attributable to non-heritable factors. Among these, age has previously been shown to significantly influence

immune profiles^{10,16,43,44}. However, in our study, age was not a significant contributor to variation in immune cell composition or cytokine production. This may be due to the narrow age range of the cohort (mostly 5–6 months old), suggesting that age-related effects require dedicated experimental designs in cattle. Similarly, as our study included only male animals, we could not assess the impact of sex on immune variability. While this is a limitation, it also serves as a strength by minimizing confounding and enhancing the detection of variability attributable to genetics and other factors such as seasonality. Future studies should involve also female animals.

Accordingly, we found that season significantly influences variation in immune cell frequencies and cytokine production, consistent with previous studies^{10,47,63}. Proportions of T cells, B cells, and NK cells peaked in spring, contrasting with human studies where these populations peak in summer or winter^{10,47}. Monocyte proportions, except for CD14⁺CD16⁺ in CD172⁺ cells, were highest in winter, aligning more closely with human data⁴⁷. These discrepancies may reflect species-specific biology but also lifestyle differences, as seasonal environmental changes are typically more pronounced for farm animals than for humans. While housing at CBS is standardized (ventilation, temperature, feed), animals originate from farms where conditions likely vary seasonally. Notably, in spring, most animals, including calves, are housed indoors under higher population density and infection pressure. In contrast, during summer and autumn, animals are often kept outdoors in lower-density settings, potentially reducing immune activation. Such environmental differences may underlie the seasonal shifts observed in lymphocyte population dynamics.

Another notable finding was the identification of genome-wide significant QTLs influencing immune cell composition and cytokine production. In particular, genetic variation affecting IL-8 expression appears especially relevant to cattle health, as the associated loci colocalize with a region linked to livability in Holsteins⁴⁸. Similar variants influencing IL-8 expression have also been reported in other breeds^{64–66}. The locus associated with IL-10 production is also of interest, given IL-10's key immunoregulatory role. Reduced IL-10 production, as seen in IL-10-deficient mice⁶⁷, may predispose animals to immunopathologies. Variants at the IL-8 and IL-10 loci may contribute to exaggerated inflammatory responses^{68,69}. In this context, it is important to note that the BWB breed is known for its increased susceptibility to pronounced inflammatory responses, particularly in association with pneumonia⁷⁰, and psoroptic mange⁷¹.

Given the importance of these immune traits, it is reasonable to assume they may significantly impact animal health. However, clinical follow-up data for the animals studied are currently lacking, and implementing such follow-up remains challenging due to the variable clinical manifestations these differences may produce. Nonetheless, collecting robust and standardized clinical data will be essential to further explore the biological relevance of the observed signals. A promising strategy would be to longitudinally monitor animals

carrying specific genotypes, identified through *in vitro* profiling as in this study, alongside matched controls. This would help determine whether, and especially in what way, the observed immunophenotypic variation influences animal health. Only with such evidence can these immune traits be considered promising candidates for selective breeding or health monitoring frameworks. Although similar clinical databases are being developed in other breeds, achieving genetic colocalization across breeds is difficult given that the number of shared polymorphisms is often limited, and key mutations may be rare or absent in certain populations. Moreover, disease susceptibility and relevance vary by breed and production context, for example, mastitis is of greater concern in Holsteins. It is therefore essential to collect clinical phenotypes within the breed of interest to assess the true clinical impact of immune-related genetic variation identified in the laboratory. Simultaneously recording relevant production traits will also be critical to understanding how immune loci influence broader performance metrics.

The methodology employed in this study has some limitations. (1) The modest sample size reduces statistical power to detect additional loci associated with immunophenotypes. (2) While cytokine production was assessed following stimulation of WWBCs and PBMCs, recent evidence supports whole blood stimulation as a more physiologically relevant method for assessing immune function⁷². Unlike PBMCs or WWBCs, whole blood retains serum components and cellular diversity (e.g., red blood cells, platelets, granulocytes), better preserving the *in vivo* immune environment and enabling direct cell-cell communication during stimulation. In contrast, the isolation of PBMCs and WWBCs is labor-intensive and may cause pre-activation of immune cells, introducing bias, especially in PBMC-based assays. Whole blood stimulation protocols have recently been adapted for cattle^{73,74}. Furthermore, variable cytokine levels in FCS, such as IL-10, may represent a potential confounding factor influencing cytokine production following stimulation and should be taken into account in future analyses. (3) Seasonal variation in immunity was assessed by comparing different individuals, each sampled in a different month of the year. Longitudinal sampling of the same animals, ideally over multiple years, would better capture recurring seasonal effects^{41,47,63,75}. Longitudinal follow-up would also allow us to better assess the modifications associated with immune system development^{43,44}. For instance, the observed decline in $\gamma\delta$ + T cells and increased IL-17 production from T1 to T2 may reflect developmental changes. Our age-restricted cohort limits conclusions on immune development across the lifespan and on the onset of immunosenescence. Broader longitudinal studies incorporating larger cohorts and multi-omics approaches will also be essential to characterize the stability and context-dependence of genetic associations. (4) Our analyses focused on blood cells, yet immune responses are highly compartmentalized. Future work should examine immune variation in tissues^{24,76}, and at single-cell resolution^{77–79}, to identify both genetic and non-genetic contributors to this complexity. (5) Only limited clinical and metadata variables were included. Although non-genetic factors accounted for much inter-individual variation, variables such as BMI, vitamin D levels^{80,81}, sex, infection and vaccination history, metabolism, and epigenetics warrant further investigation. (6) The breed-specific context of Belgian Blue cattle offers insights into the relationship between muscle hypertrophy and immune function but may not reflect other breeds selected for different production traits. Validation in additional breeds with diverse genetic backgrounds and selected for other purposes is therefore necessary to assess the broader relevance of our findings³⁶. Conversely, integrating data across cohorts and breeds, as done for clinical data, has inherent limitations that have to be addressed by collecting all variables within a single, well-characterized cohort. This will also enable all genetic variability to be retained for colocalization studies, rather than restricting them to loci retained between breeds. (7) through the exclusion of outlier immune phenotypes, we have

potentially constrained our analysis to the investigation of common genetic variants^{82,83}. However, rare or private genetic variants could also be of main interest⁸⁴. In particular, selection for production traits and excessive reliance on specific animal specimens could quickly lead to the spread of rare genetic variants deleterious for immune responses that would remain hidden until animals face the emergence of some pathogens for example. (8) Finally, copy number variants at immune-related loci may also contribute to phenotypic diversity and should be explored in future studies. A systems immunology approach integrating these factors will be key to advancing our understanding of immune variability.

This study provides a comprehensive analysis of how non-genetic host factors, the environment, and genetics shape immune cell composition and cytokine production in cattle. These insights offer a foundation for understanding disease susceptibility and immune function in this species. While defining an optimal immune response remains challenging, identifying genetic determinants of immune traits may in the long-term support genomic selection strategies to improve resilience in cattle breeds. In the shorter term, this approach holds promise for detecting and managing genetic defects affecting immune function. Moreover, the historical focus on human and mouse models, largely due to technical constraints, has somewhat limited the scope of a broader understanding of mammalian immune responses, leaving minimal room for insights from alternative animal species. However, with the democratization of immunological and genetic methodologies, the avenue is now open for addressing more fundamental questions using information from distant species like bovines. This study therefore opens the way to the development of translational immunology studies from diverse animal species which will undoubtedly allow a better understanding of the functioning of the immune system within the species studied but which will also make it possible to highlight in a transversal way the similarities and the differences of functioning among mammals in a holistic manner.

Methods

Ethics

All procedures were conducted in accordance with the European Convention for the Protection of Vertebrate Animals used for Experimental and Other Scientific Purposes (CETS 123) and were approved by the University of Liège Animal Ethics Committee (protocol number 2509).

Population cohorts

246 BWB male calves housed at the center for bovine selection (CBS) in Ciney have been enrolled in the first cohort of this study. All calves were sampled on the first week after entry (time point T1) and 211 individuals were sampled 4 weeks later (time point T2). The 35 animals not included in the second sampling (T2) were excluded primarily due to logistical constraints. Importantly, no mortality occurred during the study period. On the week after T1, all calves were vaccinated and treated with anti-parasite drugs (Supplementary Table 1). To validate certain GWAS hits in an independent cohort of animals, an additional 239 male calves were sampled one month after entry at the CSB.

Cell Processing

Blood was sampled at the jugular vein in K2 EDTA tubes with a Vacutainer system. For WWBCs purification, blood was incubated in red blood cell (RBC) lysis buffer according to the manufacturer's instructions (ThermoFisher), with some minor modifications. Briefly, blood (5 mL) was first incubated for 7 minutes with lysis buffer (10 mL) before the addition of 30 mL PBS containing 2% fetal calf serum (FCS). Cells were spun down (335 g for 5 minutes at 4 °C) and once again treated with 500 μ L of RBC lysis buffer for 3 min and spun down at 335 g for 5 min at 4 °C after addition of 30 mL PBS containing 2% FCS. Cells were resuspended in 10 mL PBS 2% FCS and counted.

For PBMCs isolation, 5 mL of blood was mixed with the same volume of PBS 2% FCS and added on top of 5 mL of Ficoll (Ficoll-Paque PLUS, Fischer Scientific) before centrifugation at 697g for 10 min at RT, no brake. Interphase containing purified PBMCs was transferred to a new tube and washed with 10 mL PBS 2% FCS and centrifuged at 335g for 5 min at 4 °C. Cell pellets were treated with 500 µL of RBC lysis buffer for 3 minutes, washed with 45 mL PBS 2% FCS and spun down at 335g for 5 min at 4 °C. Cells were resuspended in 10 mL PBS 2% FCS and counted.

Flow cytometry analysis

Before staining, most unlabeled antibodies were labeled with lightning-link kits (Abcam) following the manufacturer's instructions. 10⁶ WWBCs were stained with standard panels of immunophenotyping antibodies (Supplementary Table 2 for a list of antibodies, clones, fluorochromes, and manufacturers) and incubated for 15 minutes at 4 °C with a volume of 50 µL per well. Cells were washed with 200 µL fluorescence-activated cell sorting (FACS) buffer (PBS containing 2 mM EDTA, 0.5% bovine serum albumin (BSA), and 0.1% NaN₃) and centrifuged at 335g for 5 min at 4 °C. Next, cells were incubated with second mix in FACS buffer for 15 min at 4 °C and washed with PBS with a volume of 200 µL per well. Finally, cells were stained with Zombie aqua (BioLegend) according to the manufacturer's instructions and suspended in 200 µL FACS buffer. Data were acquired with a BD LSR Fortessa X-20 flow cytometer equipped with 50-mW violet 405-nm, 50-mW blue 488-nm, 50-mW yellow-green 561-nm and 40-mW red 633-nm lasers and a ND1.0 filter in front of the forward scatter (FSC) photodiode, using BD FACSDiva software (BD Bioscience). Data was analyzed using FlowJo software (TreeStar).

Each sample was analyzed with 5 multicolor antibody panels: (i) NK cells, (ii) γδ T cells, (iii) CD8 + CD4 + T cells, (iv) Monocytes, and (v) B-neutrophils illustrated in Supplementary Fig. 1. For each panel, the live leukocytes subpopulation (P1) was discriminated using side-scattered area (SSC-A) and forward scatter area (FSC-A) to select cells, and forward scatter height (FSC-H) and FSC-A to select single-cell population, and finally Zombie aqua staining to select live cells. In panel 1, within the lymphocytes selected by FCS-A/SSC-A, γδ T cells populations (γδ +, CD8+ in γδ +, WC1+ in γδ +, WC1.2+ in WC1+ γδ +) were identified. In panel 2, T cells populations (CD3+γδ-, CD8+ in CD3+γδ-, CD4+ in CD3+γδ-, CD25+ in CD4 + CD3+γδ-) were identified. In panel 3, NK cells (CD335 +, CD2+ in CD335 +, and CD8+ in CD335 +) were characterized. In panel 4, granulocytes and lymphocytes were discriminated by FSC-A and SSC-A and further characterized by staining with the CH138A antibody (identifying granulocyte and therefore mostly neutrophils in blood) and for IgM and CD21 expression to identify B cell subsets. In panel 5, subpopulations within monocytes were separated as CD172 +, CD14+ in CD172 +, and CD16+ in CD172 +.

Stimuli

Bacteria. HKMT (Heat-killed *Mycobacterium tuberculosis*), HKLM (Heat-killed *Listeria monocytogenes*) and Flagellin (Recombinant flagellin from *S. typhimurium*) were purchased from InvivoGen and used at a final concentration of 100 µg/mL (HKMT), 10⁸ cell/mL (HKLM) or 1 µg/mL (Flagellin). *Mannheimia haemolytica* and *Pasteurella multocida* were produced and heat inactivated by the Bacteriology Laboratory from the veterinary faculty (ULiège) and used at 4 × 10⁵ cfu/mL (*Mannheimia haemolytica*), and 2 × 10³ cfu/mL (*Pasteurella multocida*).

Fungi. *Aspergillus fumigatus* extract was purchased at the UK National Institute for Biological Standards and Control (NIBSC) and used at 1% dilution.

Virus. Bovine herpesvirus 1 (BoHV-1) and BoHV-4 were amplified on Madin-Darby bovine kidney cells (ATCC CCL-22) and purified on

sucrose gradient. The purified productions were heat inactivated at 56 °C for 60 min and used at 10⁵ pfu/mL (BoHV-1) and 5.10⁴ pfu/mL (BoHV-4). Parainfluenza 3 (PI3) was amplified on bovine turbinate (BT) cells and concentrated by ultracentrifugation (100,000g, 2 h at 4 °C) after removal of the cell debris by low-speed centrifugation (1000g, 10 min at 4 °C).

TLR ligands and non-microbial stimuli. Pam3CSK4 (Pam3Cys-SerLys4), PolyI:C (High molecular weight), LPS (E. coli serotype O55:B5), Imiquimod (R837), and ssRNA were purchased from InvivoGen and used at a final concentration of 1 µg/mL (Pam3CSK4, LPS, Imiquimod, and ssRNA), 10 µg/mL (PolyI:C). Phorbol 12-myristate 13-acetate (PMA) and Ionomycin were purchased from Sigma and used at 50 ng/ml and 1 µg/mL, respectively.

Ex vivo stimulations

2.5 × 10⁵ WWBCs or PBMCs were added to a 96-well plate and cultured with stimuli in 250 µL of IMDM+glutamax (Gibco) 10% FCS, 2% penicillin-streptomycin, 1% amino acid, 50 µM β-mercaptoethanol for 48 h at 37 °C and 5% CO₂ except for PMA and Ionomycin which were incubated at 37 °C for 5 h. Supernatants were collected and stored at -20 °C until used for ELISA.

ELISA analysis of cytokines and immunoglobulin levels

Concentrations of bovine IL-1β, IL-6, IL-4, IFN-γ, IL-17, TNF-α, IP-10 (CXCL-10), IL-8, and IL-10 as well as serum levels of IgG1, IgG2, and IgM were determined by ELISA following manufacturer's recommendations (see Supplementary Table 3 for a list of antibodies, clones, and manufacturers). Briefly, the 96-well microtiter plates were coated with capture antibody in PBS overnight at 4 °C and then were blocked for 2 h at 20 °C with 200 µL of 0.1% PBS-Tween 20 for anti-Ig ELISA or 5% bovine serum albumin (BSA) for cytokines ELISA. 50 µL of serum dilution or two times diluted supernatant were added into the wells and incubated for 2 h at 20 °C, then the detection antibody was added into wells after washing four times with wash buffer (PBS containing 0.1% tween 20) and incubated for 1 hour at room temperature (RT). Finally, streptavidin-HRP was incubated for 30 minutes at RT. After each incubation, the plates were washed four times with wash buffer. The plates were developed using 3,3',5,5'-tetramethylbenzidine. The reaction was stopped after suitable color development by the addition of 50 µL of 0.2 M sulfuric acid and the absorbance at 450 nm was read using an ELISA plate reader. Detection ranges were 31.25 pg/mL-2000 pg/mL (IL-6), 15.625 pg/mL-1000 pg/mL (IL-1β and IL-4), 25 pg/mL-6250 pg/mL (IFN-γ), 62.5 pg/mL-4000 pg/mL (IL-10), 0.6 pg/mL-632 pg/mL (IL-17), 125 pg/mL-8000 pg/mL (TNF-α), 0.123 ng/mL-30 ng/mL (IP-10), 0.39 ng/mL-25 ng/mL (IL-8), 0.048 ng/mL-200 ng/mL (IgG1 and IgG2) or 0.024 ng/mL-100 ng/mL (IgM).

Genomic DNA extraction, SNP array genotyping, and 57,979 SNPs

Genomic DNA was extracted from whole blood (EDTA) on a King-Fisher™ instrument following standard procedures. SNP genotyping for all the animals was conducted at the GIGA Genomics platform (ULiège) using two custom bovine medium density (MD) BeadChip arrays (Illumina) following the manufacturer's instructions. The two arrays, namely EuroGenomics MDv1 and MDv2, interrogate, respectively, 50,809 and 57,979 genomic positions. For each animal, final genome-wide genotypes were obtained with the GenomeStudio Genotyping (GT) Module from Illumina. After intersecting the two array positions and excluding markers on sex chromosomes, a total of 43,173 genetic markers from SNP array genotyping was kept for subsequent analysis. In parallel, a total of 156 animals were selected for whole-genome sequence by random selection. The sequencing was performed by Illumina HiSeq 2000 instrument in the GIGA Genomics platform. The bioinformatic analysis to call variants from the sequence

data is fully described in Gualdrón Duarte et al.⁸⁵ and is summarized here. Sequence reads were mapped to the bovine reference genome ARS-UCD1.2 using BWA-MEM software (version 0.7.5)⁸⁶. After sorting and marking duplicates, variant calling was performed on individual recalibrated BAM files using HaplotypeCaller (GATK4 v4.1.7.0)⁸⁷ to obtain individual genomic variant call format (GVCF) files. Joint genotyping of all these GVCF files was then done using GenotypeGVCFs (GATK4 v4.1.7.0)⁸⁷. After applying the variant quality recalibration procedure with (GATK4 v4.1.7.0)⁸⁷, variants with a quality score below the 97.5 threshold, minor allele frequency (MAF) < 0.01 and multi-allelic sites were filtered out. Genotype imputation to the sequence level was then performed for the individuals genotyped on an MD array using ShapeIT4.2 in combination with Minimac4⁸⁵, using sequenced Belgian Blue individuals as a reference. Finally, a total of 11,486,822 SNPs were included in the genome-wide association studies, with 246 calves from T1 sampling. As a validation cohort, 211 animals from T2 sampling were used.

Quantification and statistical analysis

Data pretreatment. The Shapiro-Wilk normality test was used to test the normality of both raw (Supplementary Fig. 3) and log-transformed data (Supplementary Fig. 4) data. A threshold of $p > 0.05$ was employed to define a normal distribution. Percentages of the various immune cell populations, were normalized using an inverse rank transformation (IRT) algorithm¹⁰. A log₂ transformation was used to transform immunoglobulin levels as well as all cytokine levels (Supplementary Fig. 4). In cases where the response fell outside of the ELISA measurement range, the corresponding lower or upper value of the ELISA plate detection range was used as the measured value. Raw cytokine production followed a non-Gaussian distribution (Supplementary Fig. 3) while some log₂ transformed cytokine concentrations followed a normal distribution (Supplementary Fig. 4). The Wilcoxon signed rank test was used to compare the significance of cytokine responses after different types of stimulation to the unstimulated condition (IMDM medium). Considering the fact that some cytokines were not normally distributed, the data were further normalized using the IRT algorithm⁸⁸, as described above.

Correlations among immune parameters. To determine the correlation between immune parameters including cytokines concentrations, immunoglobulin isotypes concentrations, and immune cells percentages, we first corrected the normalized immunophenotypes for age and seasonal effects using a linear model. Pairwise Spearman correlation analysis was used to calculate associations using the normalized and corrected immunophenotypes, and an unsupervised hierarchical clustering approach (hclust) was used using these correlations as distance. The same methodology was used to calculate the relationship between immune cells percentages and immunoglobulin concentrations, as well as the relationship between immune cell percentages and cytokine production. After correction for multiple testing (for each category of dataset), the significance of the correlations was considered (FDR < 0.05)⁸⁹. As part of the preprocessing for multidimensional scaling analysis (MDS) using monoMDS from the vegan package or hierarchical clustering with heatmap in stats package, Euclidean distances were calculated from correlation matrices (using Spearman).

Genome-wide association analyses. Prior to the GWAS, values of immunophenotype data that were more than 1.5 times the interquartile range above the upper quartile and less than 1.5 times the lower quartile were removed (removal of outliers). The association analysis was used to identify common genetic variants that influence interindividual variation in immunophenotypes using leave-one-chromosome-out (LOCO) mixed linear model, as implemented in GCTA (v1.26.0) software⁹⁰, with age and season as covariables. The

approach used by Druet et al.⁹¹ was applied to estimate genome-wide significance at $p < 0.05$ using Bonferroni corrections for multiple-testing. By comparing uncorrected and corrected p -values, the number of independent tests was set as 382,000. As a result, the p -value threshold for genome-wide significance association was $1.3e-07$ ($0.05/382,000$). Suggestive association threshold was set to $p < 2.6 \times 10^{-6}$ ($1/382,000$). Inflation or deflation in p -values was assessed by genomic inflation factor (λ) and also visually inspected by quantile-quantile (Q-Q) plot. λ is calculated as the median of the χ^2 test statistics (1 degree of freedom) divided by its theoretical median under the null distribution. To further evaluate whether the genetic loci of immunophenotype contain several independent signals, we conducted conditional and joint analysis utilizing the stepwise model selection approach in GCTA-COJO. The proportion of variance in phenotype explained by a given SNP (PVE) was calculated⁹². Subsequently, Variant Effect Predictor (VEP) (<http://www.ensembl.org/Tools/VEP>) was used for the identification of the candidate genes located in significant loci and predicted functional consequences of identified mutations⁹³.

Colocalization analysis. Immune-related QTLs identified in this study were colocalized with eQTLs derived from transcriptomic data from resting whole blood cells from the same animals (ENA accession number: PRJEB94322) or with genomic regions associated with health-related phenotypes in cattle using the R package ‘coloc’ (cran.r-project.org/web/packages/coloc). For genomic regions associated with health-related phenotypes, summary statistics from large-scale single-marker GWAS for seven cow health-related variables in Holstein cattle, including cow livability, mastitis, ketosis, hypocalcemia, displaced abomasum, metritis, and retained placenta were extracted from a recent study⁴⁸. The chromosomal locations were initially transformed from UMD3.1 to ARS-UCD1.2 using liftOver. Bayesian test was then used to assess the likelihood of QTL and clinical traits sharing the same underlying causative variants⁹⁴. The Bayesian co-localization produces posterior probabilities for five hypotheses related to a shared variant between the two traits: (H0) no association with either trait; (H1) association with expression only; (H2) association with disease only; (H3) association with both traits but independent causal variants; and (H4) association with both traits driven by a shared causal variant. Strong evidence of colocalization was defined as a posterior probability of ≥ 0.8 for H4.

Estimation of explained variance. Feature selection based on permutational multivariate analysis of variance using distance matrices (PERMANOVA) was firstly used to estimate the contributions of different factors to interindividual variations of each dataset (e.g., cytokines production and immune cells)⁵² (Fig. 6a). In brief, interindividual variations for each dataset was estimated using vegan’s (version 2.5.5) adonis function with 1000 times permutation⁹⁵. Only response variables and candidate explanatory variables with a permutational FDR < 0.05 were kept. Only suggestive and significant SNPs from SNP-genotyping array were retained as genetic variants in this analysis. To address collinearity of the selected factors, the set of candidate explanatory variables was correlated to itself. When candidate explanatory factors displayed Spearman correlation greater than 0.4, the factor displaying the lowest correlation with the immunophenotype was removed from the model⁵³. All independent explanatory factors were further included in PERMANOVA to estimate the combined contribution to interindividual variation. In order to estimate the variance explained by the different factors on individual immunophenotypes (Fig. 6b, c), a correlation-based factor-selection approach was used⁵³. In brief, if a factor (age, season, antibody concentration and frequency of immune cell population) had a significant association (Spearman FDR < 0.05) with the immunophenotype, it was added to the set of candidate explanatory variables. To address collinearity of the selected factors, the set of candidate explanatory variables was

selected using the same procedure stated above. The adjusted R^2 was used as the measure of explained variance, accounting for the inflation caused by adding predictors. The heritability was estimated using genome-based restricted maximum likelihood (GREML) implemented in GCTA software, with age and season as covariates. Meanwhile, we permuted the immunophenotypes and covariate datasets 1000 times randomly before running the GWAS with these datasets to generate 1000 random profiles for each immunophenotype. Using the pipeline described above, the heritability for the permuted immunophenotypes dataset for each run was calculated. This analysis produced a distribution of 1000 heritability estimates for each immunophenotype. A measured estimate was considered significant if it was in the top 5% of the permuted distribution of immunophenotype estimates.

Cytokine level prediction. For machine learning, Elastic Net was used to predict cytokine levels⁹⁶. 206 individuals sampled at both T1 and T2 were selected (5 animals out of 211 were excluded based on missing values). These animals were firstly randomly divided into training (70% of the dataset) and test sets (30% of the dataset) to validate the prediction. Predictions (always based on measured made at T1) were either made for cytokine production measured at T1 (Fig. 7a, b) or at T2 (Fig. 7c, d). The training dataset consisted of either percentages of immune cells alone, or genetic information alone, or immune cell percentages + genetic information. Importantly, in order to define the training genetic information, we used for prediction either SNPs with a p -value $< 5 \times 10^{-5}$ obtained from the whole cohort (Supplementary Fig. 14) or SNPs with a p -value $< 5 \times 10^{-5}$ obtained from GWAS analyses performed on the training animals only (Fig. 7) in order to avoid that the same animals would contribute to the GWAS results while being part of the test set. Briefly, for the training dataset, a 5 × 5-fold cross-validation approach was applied and the trained model with the lowest MSE (Mean Squared Error) for each run was chosen for the prediction of test set. All the procedure was repeated 5 times and the prediction accuracy was assessed by calculating Spearman correlations between the measured cytokine levels and the predictions. A t test was then used to investigate whether there was a significant improvement in power for cytokine production prediction when immune cell percentages and genetics were used together in the model.

Other statistical analyses. To investigate the effect of sampling time on immune variation, linear regression modeling was performed using the base R function (lm), with R^2 proportions calculated using the relaimpo package⁴³. The majority of the modeling shown uses a categorical variable (such as subject sampling point or calf number) to model a continuous variable (immune parameter). lm provides an ANOVA model in this case (for example, immune parameter-sampling point + calf's ID). The Bonferroni correction was used for multiple testing, as implemented in base R. Wilcoxon signed rank test or dependent t -test were used to compare the different immunophenotypes from the same individual at different time points (Supplementary Fig. 7). To test for an association of immunophenotypes with age, we normalized the immune traits as described above, then used a linear model to correct for the effects of season. The significance and direction of the correlations were determined using Spearman correlation analysis. A cosinor model was fitted with the month of the year, in a one-year cycle period, using age as covariates to determine the effect of season on the immune traits measured. The fitted model was compared to an alternative regression model that did not include these terms, and significance (multiple correction testing; FDR < 0.05) was determined using a variance test between the full model and the alternative regression model^{16,47}. After controlling for multiple testing (FDR < 0.05), significance was declared. MOFA was performed in the R statistical environment using the MOFA+ package (version 1.0)⁹⁷, and latent factors were determined using MOFA's TopWeights function.

Data availability

The RNA-seq data generated in this study have been deposited in the ENA under accession code [PRJEB94322](https://doi.org/10.1038/s41467-025-68234-x). The genotype and GWAS summary statistics are deposited in zenodo (<https://doi.org/10.5281/zenodo.15967035>) under restricted access for consent as data are partially owned by the BWB herdbook association. The access to genotype data, full GWAS summary statistics, and raw phenotype data can be requested to Prof. Laurent Gillet (l.gillet@uliege.be). The time-frame for response to requests is within one month. Data on clinical phenotypes from Holstein cattle are publicly available (<https://doi.org/10.1186/s12864-020-6461-z>). Source data are provided with this paper.

Code availability

Codes used for the analyses in the study are available at GitHub (https://github.com/frucelee/System-immunology_BWB) and Zenodo (<https://doi.org/10.5281/zenodo.17514186>)⁹⁸.

References

- De Jager, P. L. et al. ImmVar project: Insights and design considerations for future studies of 'healthy' immune variation. *Semin. Immunol.* **27**, 51–57 (2015).
- Kumar, V., Wijmenga, C. & Xavier, R. J. Genetics of immune-mediated disorders: From genome-wide association to molecular mechanism. *Curr. Opin. Immunol.* **31**, 51–57 (2014).
- Liston, A., Carr, E. J. & Linterman, M. A. Shaping Variation in the Human Immune System. *Trends Immunol.* **37**, 637–646 (2016).
- Ye, C. J. et al. Intersection of population variation and autoimmunity genetics in human T cell activation. *Science* **345** (2014).
- Duffy, D. Understanding immune variation for improved translational medicine. *Curr. Opin. Immunol.* **65**, 83–88 (2020).
- Sanz, J., Randolph, H. E. & Barreiro, L. B. Genetic and evolutionary determinants of human population variation in immune responses. *Curr. Opin. Genet. Dev.* **53**, 28–35 (2018).
- Patin, E. et al. Natural variation in the parameters of innate immune cells is preferentially driven by genetic factors. *Nat. Immunol.* **19**, 302–314 (2018).
- Brodin, P. et al. Variation in the human immune system is largely driven by non-heritable influences. *Cell* **160**, 37–47 (2015).
- Notarangelo, L. D., Bacchetta, R., Casanova, J. L. & Su, H. C. Human inborn errors of immunity: an expanding universe. *Sci. Immunol.* **5** (2020).
- Aguirre-Gamboa, R. et al. Differential effects of environmental and genetic factors on T and B cell immune traits. *Cell Rep* **17**, 2474–2487 (2016).
- Davis, M. M. A prescription for human immunology. *Immunity* **29**, 835–838 (2008).
- Davis, M. M., Tato, C. M. & Furman, D. Systems immunology: Just getting started. *Nat. Immunol.* **18**, 725–732 (2017).
- Davis, M. M. & Brodin, P. Rebooting human immunology. *Annu. Rev. Immunol.* **36**, 843–864 (2018).
- Hagan, T., Nakaya, H. I., Subramaniam, S. & Pulendran, B. Systems vaccinology: enabling rational vaccine design with systems biological approaches. *Vaccine* **33**, 5294–5301 (2015).
- Li, Y. et al. A functional genomics approach to understand variation in cytokine production in humans. *Cell* **167**, 1099–1110.e14 (2016).
- Ter Horst, R. et al. Host and environmental factors influencing individual human cytokine responses. *Cell* **167**, 1111–1124.e13 (2016).
- Brodin, P. New approaches to the study of immune responses in humans. *Hum. Genet.* **139**, 795–799 (2020).
- Eckhardt, M., Hultquist, J. F., Kaake, R. M., Hüttenhain, R. & Krogan, N. J. A systems approach to infectious disease. *Nat. Rev. Genet.* **21**, 339–354 (2020).
- Pulendran, B. & Davis, M. M. The science and medicine of human immunology. *Science* **369** (2020).

20. Wang, L. et al. An Atlas of genetic variation linking pathogen-induced cellular traits to human disease. *Cell Host Microbe* **24**, 308–323.e6 (2018).
21. Astle, W. J. et al. The allelic landscape of human blood cell trait variation and links to common complex disease. *Cell* **167**, 1415–1429.e19 (2016).
22. Lagou, V. et al. Genetic architecture of adaptive immune system identifies key immune regulators. *Cell Rep* **25**, 798–810.e6 (2018).
23. Li, Y. et al. Inter-individual variability and genetic influences on cytokine responses to bacteria and fungi. *Nat. Med.* **22**, 952–960 (2016).
24. Marderstein, A. R. et al. Demographic and genetic factors influence the abundance of infiltrating immune cells in human tissues. *Nat. Commun.* **11**, 1–14 (2020).
25. Orrù, V. et al. Genetic variants regulating immune cell levels in health and disease. *Cell* **155**, 242–256 (2013).
26. Oosting, M. et al. Functional and genomic architecture of *Borrelia burgdorferi*-induced cytokine responses in humans. *Cell Host Microbe* **20**, 822–833 (2016).
27. Roederer, M. et al. The genetic architecture of the human immune system: a bioresource for autoimmunity and disease pathogenesis. *Cell* **161**, 387–403 (2015).
28. Flori, L. et al. Immunity traits in pigs: Substantial genetic variation and limited covariation. *PLoS ONE* **6** (2011).
29. Ballester, M. et al. Genetic parameters and associated genomic regions for global immunocompetence and other health-related traits in pigs. *Sci. Rep.* **10**, 1–15 (2020).
30. Braun, R. O. et al. System immunology-based identification of blood transcriptional modules correlating to antibody responses in sheep. *npj Vaccines* **3** (2018).
31. Maroilley, T. et al. Deciphering the genetic regulation of peripheral blood transcriptome in pigs through expression genome-wide association study and allele-specific expression analysis. *BMC Genomics* **18**, 1–19 (2017).
32. Emam, M. et al. The effect of host genetics on in vitro performance of bovine monocyte-derived macrophages. *J. Dairy Sci.* **102**, 9107–9116 (2019).
33. Thompson-Crispi, K. A., Sewalem, A., Miglior, F. & Mallard, B. A. Genetic parameters of adaptive immune response traits in Canadian Holsteins. *J. Dairy Sci.* **95**, 401–409 (2012).
34. Mach, N. et al. The peripheral blood transcriptome reflects variations in immunity traits in swine: Towards the identification of biomarkers. *BMC Genomics* **14** (2013).
35. Mallard, B. A., Wilkie, B. N., Kennedy, B. W. & Quinton, M. Use of estimated breeding values in a selection index to breed Yorkshire Pigs for high and low immune and innate resistance factors. *Anim. Biotechnol.* **3**, 257–280 (1992).
36. Martin, P. et al. Identification of genome regions and promising candidate genes linked to innate immune capacity on young Holstein calves. In *Proceedings of 12th World Congress on Genetics Applied to Livestock Production (WCGALP)* (ed. Veerkamp, R. F. and de Haas, Y.) 426–429 (Wageningen Academic, 2022).
37. Espinosa, R., Tago, D. & Treich, N. Infectious diseases and meat production. *Environ. Resour. Econ* **76**, 1019–1044 (2020).
38. Ikhimiukor, O. O., Odih, E. E., Donado-Godoy, P. & Okeke, I. N. A bottom-up view of antimicrobial resistance transmission in developing countries. *Nat. Microbiol.* **7**, 757–765 (2022).
39. Georges, M., Charlier, C. & Hayes, B. Harnessing genomic information for livestock improvement. *Nat. Rev. Genet.* **20**, 135–156 (2019).
40. Wells, S. B. et al. Multimodal profiling reveals tissue-directed signatures of human immune cells altered with age. *bioRxiv* <https://doi.org/10.1101/2024.01.03.573877> (2024).
41. ter Horst, R. et al. Seasonal and nonseasonal longitudinal variation of immune function. *J. Immunol.* **207**, 696–708 (2021).
42. Kapellos, T. S. et al. Human monocyte subsets and phenotypes in major chronic inflammatory diseases. *Front. Immunol.* **10**, 1–13 (2019).
43. Carr, E. J. et al. The cellular composition of the human immune system is shaped by age and cohabitation. *Nat. Immunol.* **17**, 461–468 (2016).
44. Hill, D. L. et al. Immune system development varies according to age, location, and anemia in African children. *Sci. Transl. Med.* **12** (2020).
45. Cotugno, N. et al. Virological and immunological features of SARS-CoV-2-infected children who develop neutralizing antibodies. *Cell Rep.* **34** (2021).
46. Argelaguet, R. et al. Multi-omics profiling of mouse gastrulation at single-cell resolution. *Nature* **576**, 487–491 (2019).
47. Dopico, X. C. et al. Widespread seasonal gene expression reveals annual differences in human immunity and physiology. *Nat. Commun.* **6** (2015).
48. Freebern, E. et al. GWAS and fine-mapping of livability and six disease traits in holstein cattle. *BMC Genomics* **21**, 41 (2020).
49. Baldwin, C. L. et al. Special features of $\gamma\delta$ T cells in ruminants. *Mol. Immunol.* **134**, 161–169 (2021).
50. Dotiwala, F. & Lieberman, J. Granulysin: killer lymphocyte safeguard against microbes. *Curr. Opin. Immunol.* **60**, 19–29 (2019).
51. Anderson, M. J. Permutational multivariate analysis of variance (PERMANOVA). *Wiley StatsRef Stat. Ref. Online* 1–15 <https://doi.org/10.1002/9781118445112.stat07841> (2017).
52. Chen, L. et al. Influence of the microbiome, diet and genetics on inter-individual variation in the human plasma metabolome. *Nat. Med.* <https://doi.org/10.1038/s41591-022-02014-8> (2022).
53. Bakker, O. B. et al. Integration of multi-omics data and deep phenotyping enables prediction of cytokine responses. *Nat. Immunol.* **19**, 776–786 (2018).
54. Chu, X. et al. Integration of metabolomics, genomics, and immune phenotypes reveals the causal roles of metabolites in disease. *Genome Biol.* **22**, 1–22 (2021).
55. Kaczorowski, K. J. et al. Continuous immunotypes describe human immune variation and predict diverse responses. *Proc. Natl. Acad. Sci. USA.* **114**, E6097–E6106 (2017).
56. Lin Da, J. et al. Rewilding Nod2 and Atg16l1 mutant mice uncovers genetic and environmental contributions to microbial responses and immune cell composition. *Cell Host Microbe* **27**, 830–840.e4 (2020).
57. Netea, M. G., Wijmenga, C. & O’Neill, L. A. J. Genetic variation in Toll-like receptors and disease susceptibility. *Nat. Immunol.* **13**, 535–542 (2012).
58. Chu, X. et al. A genome-wide functional genomics approach uncovers genetic determinants of immune phenotypes in type 1 diabetes. *Elife* **11**, 1–17 (2022).
59. Boahen, C. K. et al. A functional genomics approach in Tanzanian population identifies distinct genetic regulators of cytokine production compared to European population. *Am. J. Hum. Genet.* **109**, 471–485 (2022).
60. Kullberg, B. et al. A comprehensive genetic map of cytokine responses in Lyme borreliosis. 1–15 <https://doi.org/10.1038/s41467-024-47505-z> (2024).
61. Saint-André, V. et al. Smoking changes adaptive immunity with persistent effects. *Nature* **626**, 827–835 (2024).
62. Dong Kim, K. et al. Adaptive immune cells temper initial innate responses. *Nat. Med.* **13**, 1248–1252 (2007).
63. Lakshmikanth, T. et al. Human immune system variation during 1 year. *Cell Rep.* **32** (2020).
64. Stojkovic, B., McLoughlin, R. M. & Meade, K. G. In vivo relevance of polymorphic Interleukin 8 promoter haplotype for the systemic immune response to LPS in Holstein-Friesian calves. *Vet. Immunol. Immunopathol.* **182**, 1–10 (2016).

65. Meade, K. G., O’Gorman, G. M., Narciandi, F., MacHugh, D. E. & O’Farrelly, C. Functional characterisation of bovine interleukin 8 promoter haplotypes in vitro. *Mol. Immunol.* **50**, 108–116 (2012).
66. Stojkovic, B., Mullen, M. P., Donofrio, G., McLoughlin, R. M. & Meade, K. G. Interleukin 8 haplotypes drive divergent responses in uterine endometrial cells and are associated with somatic cell score in Holstein-Friesian cattle. *Vet. Immunol. Immunopathol.* **184**, 18–28 (2017).
67. Berg, D. J. et al. Enterocolitis and colon cancer in interleukin-10-deficient mice are associated with aberrant cytokine production and CD4⁺ Th1-like responses. *J. Clin. Invest.* **98**, 1010–1020 (1996).
68. Cesta, M. C. et al. The Role of Interleukin-8 in Lung Inflammation and Injury: Implications for the Management of COVID-19 and Hyperinflammatory Acute Respiratory Distress Syndrome. *Front. Pharmacol.* **12**, 1–7 (2022).
69. Carlini, V. et al. The multifaceted nature of IL-10: regulation, role in immunological homeostasis and its relevance to cancer, COVID-19 and post-COVID conditions. *Front. Immunol.* **14**, 1–19 (2023).
70. Pardon, B. et al. Longitudinal study on morbidity and mortality in white veal calves in Belgium. *BMC Vet. Res.* **8** (2012).
71. Meyermans, R. et al. Genetic and genomic analysis of Belgian Blue’s susceptibility for psoroptic mange. *Genet. Sel. Evol.* **56**, 1–12 (2024).
72. Duffy, D. et al. Functional analysis via standardized whole-blood stimulation systems defines the boundaries of a healthy immune response to complex stimuli. *Immunity* **40**, 436–450 (2014).
73. Reid, C., Beynon, C., Kennedy, E., O’Farrelly, C. & Meade, K. G. Bovine innate immune phenotyping via a standardized whole blood stimulation assay. *Sci. Rep.* **11**, 1–13 (2021).
74. Lesueur, J. et al. Standardized whole blood assay and bead-based cytokine profiling reveal commonalities and diversity of the response to bacteria and TLR ligands in cattle. *Front. Immunol.* **13**, 1–15 (2022).
75. Sailani, M. R. et al. Deep longitudinal multiomics profiling reveals two biological seasonal patterns in California. *Nat. Commun.* **11**, 1–12 (2020).
76. de Goede, O. M. et al. Population-scale tissue transcriptomics maps long non-coding RNAs to complex disease. *Cell* **184**, 2633–2648.e19 (2021).
77. Nathan, A. et al. Single-cell eQTL models reveal dynamic T cell state dependence of disease loci. *Nature* **606**, 120–128 (2022).
78. Perez, R. K. et al. Single-cell RNA-seq reveals cell type-specific molecular and genetic associations to lupus. *Science* **376** (2022).
79. Yazar, S. et al. Single-cell eQTL mapping identifies cell type-specific genetic control of autoimmune disease. *Science* **376** (2022).
80. Reid, C. et al. Long-term in vivo vitamin D3 supplementation modulates bovine IL-1 and chemokine responses. *Sci. Rep.* **13**, 1–13 (2023).
81. Flores-Villalva, S. et al. Low serum vitamin D concentrations in Spring-born dairy calves are associated with elevated peripheral leukocytes. *Sci. Rep.* **11**, 1–11 (2021).
82. Barnett, I. J., Lee, S. & Lin, X. Detecting rare variant effects using extreme phenotype sampling in sequencing association studies. *Genet. Epidemiol.* **37**, 142–151 (2013).
83. Bjørnland, T., Bye, A., Ryeng, E., Wisløff, U. & Langaas, M. Powerful extreme phenotype sampling designs and score tests for genetic association studies. *Stat. Med.* **37**, 4234–4251 (2018).
84. van Deuren, R. C. et al. Impact of rare and common genetic variation in the interleukin-1 pathway on human cytokine responses. *Genome Med.* **13**, 1–17 (2021).
85. Gualdrón Duarte, J. L. et al. Sequenced-based GWAS for linear classification traits in Belgian Blue beef cattle reveals new coding variants in genes regulating body size in mammals. *Genet. Sel. Evol.* **55**, 1–17 (2023).
86. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
87. Depristo, M. A. et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–501 (2011).
88. Nath, A. P. et al. Multivariate genome-wide association analysis of a cytokine network reveals variants with widespread immune, haematological, and cardiometabolic pleiotropy. *Am. J. Hum. Genet.* **105**, 1076–1090 (2019).
89. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing author (s): Yoav Benjamini and Yosef Hochberg Source. *J. R. Stat. Soc.* **57**, 289–300 (1995).
90. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: A tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
91. Druet, T. et al. Selection in action: Dissecting the molecular underpinnings of the increasing muscle mass of Belgian Blue Cattle. *BMC Genomics* **15**, 1–12 (2014).
92. Shim, H. et al. A multivariate genome-wide association analysis of 10 LDL subfractions, and their response to statin treatment, in 1868 Caucasians. *PLoS ONE* **10**, 1–20 (2015).
93. McLaren, W. et al. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics* **26**, 2069–2070 (2010).
94. Giambartolomei, C. et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* **10** (2014).
95. Dixon, P. Computer program review VEGAN, a package of R functions for community ecology. *J. Veg. Sci.* **14**, 927–930 (2003).
96. Friedman, J., Hastie, T. & Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33**, 1–22 (2010).
97. Argelaguet, R. et al. MOFA+: A statistical framework for comprehensive integration of multi-modal single-cell data. *Genome Biol.* **21**, 1–17 (2020).
98. Li, S. & Gillet, L. Genetic and non-genetic factors distinctly shape the variation of the immune response in cattle. *GitHub Repos* <https://doi.org/10.5281/zenodo.17514186> (2025).

Acknowledgements

C.D. is a research fellow of the FRIA. We thank the bacteriology lab of FARAH for providing the bacteria used in our study. We are grateful to the Walloon Region for providing funding through the Resistomics Project (L.G., C.C., P.M.). This research was supported by the European Regional Development Fund (FEDER-SYSTIMM; L.G.). We thank the Consortium des Equipements de Calcul Intensif en Fédération Wallonie Bruxelles (CECI), funded by the F.R.S.-FNRS for providing the super-computing facilities for the genome-wide association studies. We also thank all the people in CBS (Ciney, Belgium) for their assistance in sample collection.

Author contributions

S.L., F.M. and C.D. measured all immunophenotypes; J.J. and R.S. provided support to perform analyses in the lab; P.C., M.D. and P.M. provided support for the collection of data on animals; L.T., G.C.M., J.L.G.D., T.D., and C.C. performed genomic imputations and transcriptomic analyses and provided support for GWAS and colocalization analyses; S.L., F.M., C.D., and L.G. analyzed the data; T.D., M.N., M.G., and C.C. critically reviewed the analyses; S.L., F.M., C.C., and L.G. conceived the approach; S.L., F.M., and L.G. drafted the paper; all authors contributed to the final paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-68234-x>.

Correspondence and requests for materials should be addressed to Laurent Gillet.

Peer review information *Nature Communications* thanks Cliona O'Farrelly and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

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

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2026