Identification of genome regions and promising candidate genes linked to innate immune capacity on young Holstein calves

P Martin, S Fresco, L Jouneau, S Taussat, A Vinet, R Lefebvre, D Descamps, C Ferret, C Chottin, V Pietralunga, et al.

Identification of genome regions and promising candidate genes linked to innate immune capacity on young Holstein calves

P. Martin[1], S. Fresco[1,2*], L. Jouneau[3], S. Taussat[2], A. Vinet[1], R. Lefebvre[1], D. Descamps[3], C. Ferret[3], C. Chottin[3], V. Pietralunga[3], T. Pezier[4], S. Lacroix-Lamandé[4], M. Boussaha[1], S. Barbey[5], V. Denizot[6], D. Boichard[1], S. Riffault[3], F. Laurent[4]

[1] Université Paris-Saclay, INRAE, AgroParisTech, GABI, 78350, Jouy-en-Josas, France; [2] Allice, 149 rue de Bercy, 75012, Paris, France; [3] Université Paris-Saclay, INRAE, UVSQ, VIM, 78350, Jouy-en-Josas, France; [4] INRAE, UMR1282 Infectiologie et Santé Publique, 37380 Nouzilly, France; [5] INRAE, UE326 Domaine Expérimental du Pin, 61310 Gouffern en Auge, France; [6] Innoval, GDS Bretagne, 35538 Noyal-sur-Vilaine, France; *solene.fresco@inrae.fr

Abstract

In order to investigate the genomic control of innate immune response, the concentration levels of 14 cytokines produced naturally or in response to whole blood stimulation with three pattern recognition receptors agonists were determined in 623 Holstein calves enrolled in the Healthycalf project. The effect of the immunostimulants and the relationship among the cytokines produced were observed through a principal components analysis, and 50K GWAS were performed on all cytokine x condition traits. Results showed that most cytokines are oriented together following the first component that explained 70% of the variability and partially discriminated the agonistic conditions. Additionally, 40 genome regions were found in association with at least one trait with some regions being largely shared between conditions and cytokines. Seven positional candidate genes were considered as especially interesting due to their functional link to innate immune response reported from the literature.

Introduction

In most national breeding programs, the included health traits relate to adult animal health. However, calf health is also a major economical and welfare burden. In France, depending on the breed, between 10 and 20% of the dairy calves die before six months of age (Leclerc *et al.*, 2016). If a large part of deaths occurs in the first 48h and can be related to calving, a considerable and variable part happens between 48h and 2 months of age with a detrimental influence of diarrheas (Jegou *et al.*, 2006). A better knowledge of the mechanisms and genetic determinism of calves' innate immune response could help reducing the prevalence of diarrheas and respiratory symptoms and possible subsequent deaths. In this study, we aim to identify genome regions and candidate genes associated with innate immune response through the concentration level of various cytokines.

Materials and methods

Between 2018 and 2021, a large immune monitoring study was performed on Holstein calves from various commercial farms and one experimental farm (https://doi.org/10.15454/1.5483257052131956E12) all located in the Brittany and Normandy regions of France. A total of 623 calves were sampled at 1-2 weeks of age by veterinarians or qualified technicians. The authors would like to especially acknowledge H. Chapuy, C. Cazals, K. Toukmidine and C. Mechin for this work. Blood samples from the jugular vein were collected in four monovette syringes (Sarstedt, Marnay, France) containing heparin and pre-filled with pattern recognition receptors-agonists (LPS, NOD2L, R848) or a non-stimulated

condition with vehicle only. Samples were incubated at 38.5°C for 24h and plasma recovered stored at -20°c for further cytokine determination. An additional sample of blood or cartilage (when twin births) was taken for genotyping. This study is part of the Healthycalf project that was funded by Apis-Gene (https://www.apis-gene.com/).

Determination of cytokine concentrations. Concentrations of 15 cytokines/chemokines (IFN-γ, IL-1α et IL-1β, IL-1RA, IL-2, IL-4, IL-6, IL-8, IL-10, IL-17a, IP-10, CCL2, CCL3, CCL4 and TNFα) were measured in plasma samples with a custom bovine MILLIPLEX® MAP assay (SPRCUS706, Merck Millipore). Data were recorded on a MAGPIX® using xPONENT® software (Luminex). This system allows the quantification of a large number of samples and of 15 cytokines concomitantly. The wide assay range generally allows for accurate measurement of the 15 cytokines at a given plasma dilution (1/10). When a concentration was out of range (below or above), we assigned the minimum/maximum threshold for this cytokine. For IL-8, more than half the samples were out of range and therefore this cytokine was discarded from the analyses. For the other cytokines, the proportion of assigned samples varied from 1 to 22%.

Genotyping and quality controls. All animals were genotyped using the EuroGMD SNPchip designed for EuroGenomics (Illumina Inc., San Diego, CA, USA). DNA extraction and genotyping were performed at LABOGENA (www.labogena.fr, Jouy-en-Josas, France). The following quality controls were applied: an individual call rate higher than 95%, a SNP call rate higher than 90%, a minor allele frequency higher than 2% and genotype frequencies in Hardy-Weinberg equilibrium with $P>10^{-4}$. A total of 573 animals and 46, 910 SNP from the 29 autosomes passed the quality controls.

Statistical analysis. In order to visualize the effect of the condition (LPS, NOD2L, R848 or non-stimulated) and the relationships among the 14 cytokines, a principal component analysis (PCA) was performed using the miXOmics package of the R Software (Rohart *et al*., 2017).

Genome-wide association study. Single-trait association analyses were performed on 64 traits, corresponding to the 14 cytokines x 4 conditions, plus 8 traits corresponding to the coordinates of each sample under each condition on the two first components of the PCA analysis. The following mixed linear model was used with the mlma option of GCTA software:

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{x}b + \mathbf{u} + \mathbf{e} \quad (1)$$

where $\mathbf{y}$ is the vector of phenotypes; μ is the overall mean; b is the additive fixed effect of the SNP to be tested for association; $\mathbf{x}$ is the vector of predicted allele assays ranging from 0 to 2; $\mathbf{u} \sim N(\mathbf{0},\mathbf{G}\sigma_u^2)$ is the vector of random polygenic effects (with $\mathbf{G}$ being the genomic relationship matrix and σ_u^2 being the estimated polygenic variance based on the null model) and $\mathbf{e} \sim N(\mathbf{0},\mathbf{I}\sigma_e^2)$ is the vector of random residual effects (with $\mathbf{I}$ being the identity matrix and σ_e^2 the residual variance). In a first intend, SNP with –logP > 5 were selected, this threshold being on the range of the chromosome-wide threshold of significance under a Bonferroni correction of α=5%.

Results

The results for the PCA are reported in Figure 1. The first axis accounts for more than 70% of the variability while the second axis explains only 10%. Most of the cytokine concentrations are correlated following the first component, which is also able to discriminate the different conditions, reflecting the overall intensity of immune reaction depending on the agonist. The

second axis discriminates some specific cytokines (CCL2 and CCL4 versus IP10 and IFNγ) but to a lower extent. No structure is observed depending on the farm, the year or the calf sex (results not shown).

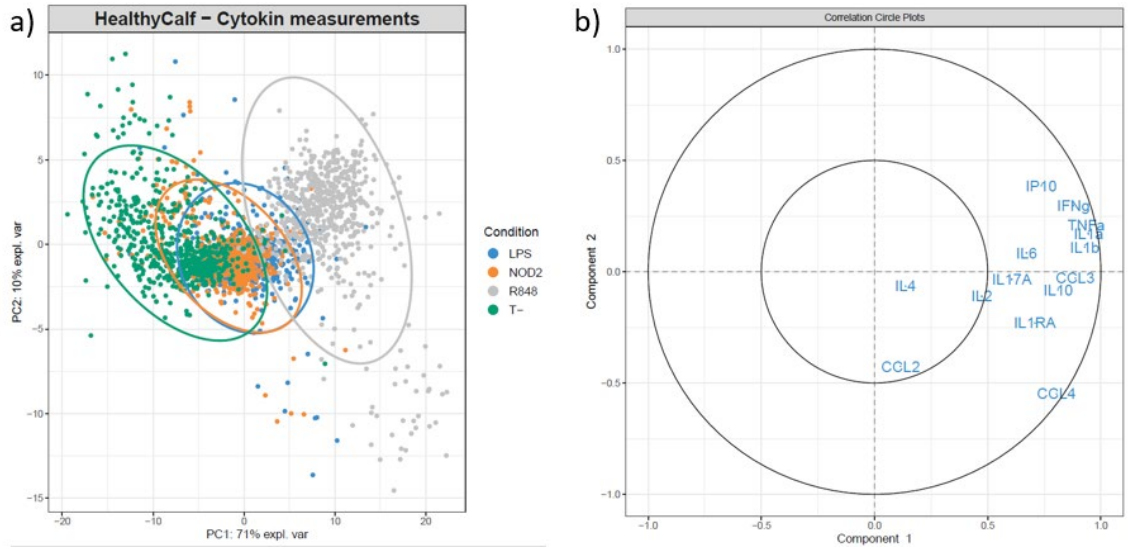

Figure 1. Principal component analysis results with a) the plot of all samples colorized by condition and b) the correlation circle plot among the cytokines.

Table 1. Details of the 10 regions identified by the GWAS associated with at least 6 different traits.

Chromosome	SNPmax Position (mb)	Minor allele frequency	Associated cytokines (with number of associated conditions)
2	36.0	0.13	IL1RA (4), IL6 (2)
2	72.7	0.12	IFNγ (1), IL1α(1), IL1RA (3), IL6 (3), TNFα (1)
3	70.4	0.09	IL1RA (4), IL6 (3), TNFα (1)
7	85.4	0.16	IL1α (1), IL1RA (4), IL6 (2), TNFα (1)
8	82.2	0.11	IL1RA (4), IL6 (2)
12	51.9	0.14	IL1RA (3), IL6 (3)
17	66.1	0.08	IFNγ (1), IL10 (1), IL1RA (4), IL6 (3), TNFα (1)
19	31.8	0.03	IL1RA (2), IL2 (2), IL6 (2), TNFα (1)
19	54.5	0.32	CCL4 (2), IFNγ (2), IL10 (2), IL1α (3), IL1β (3), IL1RA (2), IL2 (1), IL4 (4), PC1_LPS, TNFα (3)
24	12.5	0.24	IL17A (1), IL1α (1), IL1b (2), IL4 (1), PC1_LPS, TNFα (1)

Considering the GWAS results for the 64 traits, a total of 145 SNP reached the threshold, corresponding to 40 different chromosomic regions. From these regions, 18 are detected associated with one trait only while six regions are shared between two traits, four between

three traits, two between four traits and 10 regions, presented in Table 1, are shared between at least six traits. The same region was frequently associated with more than one condition of the same cytokine but different cytokines also shared some regions. The cytokines sharing the highest number of regions were the ones highly correlated with the first axis of the PCA, while cytokines like CCL2 were mostly associated with regions specific to them. It is however interesting to note that the number of regions associated with the PC traits was small (only five regions overall the eight traits).

Discussion

These results show that the mechanisms of innate immune response are complex with a genomic control that is partially shared between some cytokines but mostly specific for others and also with possible variations depending on the agonist. Among the numerous regions found associated with the traits here, some contain genes that have already been linked with immunity processes and are especially interesting candidate genes: LYAR (Yang *et al.*, 2019) and OTOP1 (Wang *et al.*, 2014), both located in a region of chromosome 6 associated with CCL4; SLC46A2 (Paik *et al.*, 2017), located in a region associated with IL6 on chromosome 8; SEC14L (Li *et al.*, 2013) found in the largely shared region of chromosome 19; CXCL17 (Lee *et al.*, 2013) on chromosome 18 in a region associated with IL6; OTULIN (Fiil *et al*., 2013) on chromosome 20 associated with IL4 and finally TREX1, on chromosome 22 associated with IL1β (Yan *et al.*, 2010). The next step will be to confront these cytokine profiles with the clinical status of the animals to determine if there is a link between the cytokine profile observed and the health status of the animals allowing the identification of predictive markers of sensitivity/resistance to diseases in young calves.

References

Fiil B.K., Damgaard R.B., Wagner S.A., Keusekotten K., Fritsch M., *et al.* (2013). Mol Cell 50:818–830. doi:10.1016/j.molcel.2013.06.004.

Jegou V., Porhiel J.Y., Brunschwig P., and Jouanne D. (2006). Proc. of 13eme Renc. Rech. Ruminants, Paris, France.

Leclerc H., Lefebvre R., Douguet M., Phocas F., and Mattalia S. (2016). Proc. of 23eme Renc. Rech. Ruminants, Paris, France.

Lee W.-Y., Wang C.-J., Lin T.-Y., Hsiao C.-L., and Luo C.-W. (2013). Am J Physiol Endocrinol Metab 304:E32-40. doi:10.1152/ajpendo.00083.2012.

Li M.-T., Di W., Xu H., Yang Y.-K., Chen H.-W., *et al*. (2013). J Virol 87:10037–10046. doi:10.1128/JVI.01073-13.

Paik D., Monahan A., Caffrey D.R., Elling R., Goldman W.E., *et al*. (2017). J Immunol 199:263–270. doi:10.4049/jimmunol.1600409.

Rohart F., Gautier B., Singh A., and Le Cao K.-A. (2017). PLOS Computational Biology 13:e1005752. doi:10.1371/journal.pcbi.1005752.

Wang G.-X., Cho K.W., Uhm M., Hu C.-R., Li S., *et al.* (2014). Diabetes 63:1340–1352. doi:10.2337/db13-1139.

Yan N., Regalado-Magdos A.D., Stiggelbout B., Lee-Kirsch M.A., and Lieberman J. (2010). Nat Immunol 11:1005–1013. doi:10.1038/ni.1941.

Yang C., Liu X., Cheng T., Xiao R., Gao Q., *et al*. (2019). J Virol 93:e00769-19. doi:10.1128/JVI.00769-19.