Genome-wide associations of immune-associated traits in dairy cows

S.J. Denholm[1], T.N. McNeilly[2], M.P Coffey[1], G.C. Russell[2], G. Banos[1,3], A. Tolkamp[1], J.E. Coe[1], A. Bagnall[1], M.C. Mitchell[2] & E. Wall[1]

[1] SRUC, Kings Buildings, West Mains Road, Edinburgh, EH9 3JG, UK
scott.denholm@sruc.ac.uk (Corresponding Author)
[2] Moredun Research Institute, Pentlands Science Park, Penicuik, Midlothian, EH26 OPZ, UK
[3]The Roslin Institute, The University of Edinburgh, Easter Bush, Midlothian, EH25 9RG, UK

Summary

Previous work by our group has consistently identified relationships between immune-associated (IA) traits and dairy cow health, fertility and productivity. Furthermore, we observed that these associations are partly under genetic control. SNP genotype data relating to 548 Holstein-Friesian dairy cows were used to carry out a genome-wide association study of IA traits previously linked to production, health and fertility. In this study, we discovered 96 SNPs that were significantly associated with defined IA traits, most notably SNP on chromosome 17 and natural antibodies in a region previously linked with ptosis, intellectual disability, retarded growth and mortality (PIRM) syndrome in cattle. Moreover, 22 SNP were found to be common across 2 to 3 IA traits. A functional cluster analysis was carried out on significant SNPs and showed 9 clusters of genes linked to various immune response and inflammation genes as well as significant associations with biological pathways such as systemic lupus erythematosus and metabolic processes. Results from the present study have highlighted potentially useful SNPs for inclusion in future animal breeding programs.

Keywords: dairy cow, GWAS, immune-associated trait

Introduction

Immune-associated (IA) traits are measurable phenotypes obtainable from blood and milk and can be used to predict susceptibility and general immunity of individual animals (Clapperton et al., 2008; Thompson-Crispi et al., 2012) and have been shown to be heritable, repeatable and associated with dairy cow health, fertility and production (Banos et al., 2013; Denholm et al., 2017). Specifically, natural antibodies and T-cell subsets have been shown to have heritabilities from 0.17-0.37 and 0.41-0.46 respectively. These results support that genetic selection for IA traits may offer a pathway to improving dairy cow health, fitness, and fertility (Mallard et al., 2011; Denholm et al., 2017).

Since the advent of low-cost high-throughput genotyping technology, genome-wide association studies (GWAS) have been used to dissect the genetics of complex traits in many species. In cattle, GWAS has been used to increase the accuracy of breeding values, via the identification of markers, as well as increasing our understanding of the genetic control of economically important and hard-to-record phenotypes (Pryce et al., 2010). They have also provided a successful and insightful tool in understanding and identifying genomic regions

associated with a variety of traits from production and fertility (*e.g.,* Pryce et al., 2010; Berry et al., 2012) to health (*e.g.,* Wijga et al., 2012) and more recently, immune responsiveness (Thompson-Crispi et al., 2014).

The aim of the present study was to determine genomic regions significantly associated with cellular and serological immune-associated phenotypes that have been previously linked to health, production and fertility traits in dairy cows.

Material and methods

Animals

All animals were Holstein-Friesians (n=548) from the Langhill lines of dairy cattle. As part of a long running selection experiment the study population was equally divided between two genetic lines (average vs. high merit), with each line assigned one of two distinct diets (high vs. low energy). All cows were routinely monitored for productivity, health and fertility.

Sampling protocol and measurement of immune-associated traits

Repeated blood samples were collected at 20 bi-monthly intervals between July 2010 and March 2015. Milk samples were also collected at 15 bi-monthly intervals between April 2013 and March 2015. Whole blood was collected into plain vacutainers (BD) and allowed to coagulate before centrifugation at 2,000×*g* for 10 min. Milk was centrifuged at 3,000×*g* for 30 minutes and the skimmed milk fraction was retained. Samples were stored at -20^{o}C.

Serological IA traits were derived by an enzyme-linked immunosorbent assay (ELISA) for natural antibodies (NAb) binding keyhole limpet hemocyanin (NAb_{KLH}) and lipopolysaccharide (NAb_{LPS}), acute phase proteins (haptoglobin, Hp) and pro-inflammatory cytokines (Tumour Necrosis Factor Alpha, TNFα). Additionally, blood leukocytes were analysed by flow cytometry to derive cellular IA traits, including granulocytes (eosinophils, neutrophils) B cells (lymphocytes), natural killer (NK) cells ($NKp46^{+}$) and %T cell subsets ($CD4^{+}$, $CD8^{+}$).

Genotypes and quality control

DNA was extracted from frozen whole blood (collected in EDTA vacutainers), via a Qiasymphony automated DNA extraction platform (Qiagen, Hilden, Germany). Of the 548 cows with IA trait information, 436 cows (~80%) had already been genotyped as part of a previous genotyping exercise using the Illumina BovineSNP50 BeadChip (Illumina Inc., San Diago CA, USA) containing data for 54,001 SNPs. For the present study a further 112 cows required genotyping. To ensure an overlap with as many of the same SNPs as the previously genotyped cows as possible, cows were genotyped using the GeneSeek GGP Bovine 150k BeadChip (GeneSeek, Neogen Corporation, Lansing, MI, USA) containing over 134,000 SNPs.

SNP and sample level filters were applied to the raw data (34,143 common SNPs) to remove SNPs with a minor allele frequency (MAF) < 0.01 and/or a call rate < 95%; and individuals with a sample call rate < 85%.

Genome-wide association analysis

The association of all QC-filtered SNPs with IA traits was carried out via the following method. Estimated breeding values (EBV) for cows were calculated via (1):

$$\boldsymbol{y} = \boldsymbol{Xa} + \boldsymbol{Z_1 b} + \boldsymbol{Z_2 c} + \boldsymbol{e} \quad (1)$$

Where, $\boldsymbol{y}$ is a vector of IA trait observations; $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$, and $\boldsymbol{e}$ are vectors of fixed, random permanent environmental, random additive genetic and random residual effects, respectively; $\boldsymbol{X}$, $\boldsymbol{Z_1}$, $\boldsymbol{Z_2}$ are incidence matrices linking phenotypic records to fixed, permanent environmental and additive effects, respectively. Fixed effects included diet group, genetic group, lactation week, calving age, calving year and record year by month interaction. Cow was fitted as a random effect and a permanent environmental effect of cow was also fitted to account for the use of repeated records. A pedigree relationship matrix was fitted to account for genetic relationships between cows and included 2,793 animals within 7 generations.

The EBV obtained above were then deregressed (dEBV), and parent average effects removed, according to the methods described in Garrick *et al.* (2009). These dEBV were then used in a single marker regression via the model in (2):

$$\boldsymbol{y} = \mu + \boldsymbol{X}g + \boldsymbol{e} \quad (2)$$

Where, $\boldsymbol{y}$ is a vector of dEBV; μ is the overall mean; $\boldsymbol{X}$ is a design matrix allocating records to SNP effects; g is the effect of the SNP (coded 0, 1 or 2); and, $\boldsymbol{e}$ represents the residual error. Multiple testing was accounted for by applying a Bonferroni correction (*0.05 / number of SNPs*). This was further validated by calculating the false discovery rate (*FDR = number of tests × [P-value / number significant SNPs]*) for each IA trait.

Functional cluster analysis

A list of genes within ±500kbp of SNP significantly associated with IA traits was compiled using the *Bos taurus* UMD3.1 assembly (Ensembl 90, Yates et al., 2016). Genes were subsequently analysed using the online tool DAVID (https://david.ncifcrf.gov) to investigate functional annotation clusters, *i.e.*, genes grouped by functionality (Huang et al., 2009).

Results and discussion

Quality control and genetic analysis

The common SNP data contained 34,143 SNPs for 548 individuals with an average of 4.5 missing genotypes per SNP. Applying QA filters on the SNP and sample data removed 1,637 SNP and 13 sample records respectively. The resultant SNP data for analyses was comprised of 535 samples × 32,506 SNPs. Genetic analyses were carried out on IA-trait data to obtain variance components and genetic parameters. Heritability and repeatability of the IA traits ranged from 0.02 to 0.44 and 0.09 to 0.80 respectively. All estimates of repeatability were significant, as were the heritabilities of all cellular IA traits and NAb. Summary statistics and genetic parameters aligned with literature values (Table 1).

Genome-wide association analysis

Analyses identified 96 SNP across the genome significantly associated with defined IA traits (Figure 1); 22 SNP were found to be common across 2 to 3 IA traits giving a total of 118 SNP-IA trait associations. Significant SNPs ($P<1.5\times10^{-6}$, post Bonferroni correction) were found for all IA traits except for % monocytes and ranged from 1 to 15 SNP per IA trait. The false discovery rare for all traits was within the range $5\times10^{-7}<P<1.5\times10^{-6}$.

The most significant association observed (Figure 1B) was between Hp and SNP on chr18 ($P=3.4\times10^{-15}$, $b=0.57$). Multiple significant SNPs were discovered on chr17 (~20% of all significant SNPs, Figure 1A) and a clear association between SNP and NAb was observed (Figure 2). Interestingly, this region on chr17 has previously been linked to bovine PIRM-syndrome, a combination of severe symptoms including ptosis, intellectual disability, retarded growth and mortality, (Venhoranta et al., 2014). A significant association between SNP on chr17 and Hp was also observed in the same region.

The largest SNP effect (defined as the coefficient of the regression, *b*) observed was a SNP on chr19 associated with TNF-α (ARS-BFGL-NGS-52992, $b=19.54$, $P=5.3x10^{-7}$). Of the 5 largest SNP effects (Figure 1C), a further 3 were associated with TNF-α (ARS-BFGL-NGS-33785 on chr24, $b=-18.24$, $P=1.3x10^{-7}$; ARS-BFGL-NGS-19906 on chr20, $b=12.84$, $P=3.5x10^{-10}$; BTB-01524979 on chr10, $b=-8.37$, $P=1.1x10^{-6}$) and 1 with % lymphocytes (ARS-BFGL-NGS-10954 on chr26, $b=6.28$, $P=3.3x10^{-7}$).

Functional cluster analysis

Functional cluster analysis was carried out to determine associations between significant SNP and known genes. In total, 9 clusters containing 129 genes were observed (Table 2). As expected a strong association with immune and inflammatory response was observed with gene cluster 1 linked with phospholipase A2 and gene cluster 2 linked with interleukin (IL1F10, IL36A/B/G/RN and IL37). The present study also found 301, 100 and 6 genes associated with response to stimulus, immune system processes and detoxification respectively, including genes belonging to the major histocompatibility complex (MHC) class I family. Significant associations with biological pathways were also found such as systemic lupus erythematosus ($P=2.5\times10^{-4}$), an autoimmune disease, and an association previously highlighted (Thompson-Crispi et al., 2014). Other pathways observed included fat digestion and absorption ($P=2.4\times10^{-6}$), pancreatic secretion ($P=8.8\times10^{-4}$), and tuberculosis ($P=5\times10^{-2}$) as well as various metabolic related pathways.

Conclusion

Even though the detection power of GWAS in the present study was limited by the number of cows available, interesting and potentially novel genomic regions were observed. The present study highlighted several SNP significantly associated with various IA traits, most notably SNP associated with natural antibodies on chr17. Moreover, the study has enabled the identification of potentially useful SNPs for inclusion in future animal breeding programs.

Acknowledgments

The Langhill experiment at Crichton Dairy Research Centre and authors are supported by the Scottish Government Rural Affairs, Food and the Environment Strategic Research Portfolio

(SRP) 2016-2021. The collection of immune associated trait and initial analyses was supported by the Scottish Government (SRP2011-2016) and BBSRC (grant no., BB/K002260/1). The authors gratefully acknowledge all staff at Crichton farm for collecting samples and managing animals and Ian Archibald for managing the Langhill database and assisting with data extraction.

List of References

Banos, G., E. Wall, M.P. Coffey, A. Bagnall, S. Gillespie, et al. 2013. Identification of Immune Traits Correlated with Dairy Cow Health, Reproduction and Productivity. PLoS One 8:e65766.

Berry, D.P., J.W.M. Bastiaansen, R.F. Veerkamp, S. Wijga, E. Wall, et al. 2012. Genome-wide associations for fertility traits in Holstein–Friesian dairy cows using data from experimental research herds in four European countries. Animal 6:1206–1215.

Clapperton, M., E.J. Glass, and S.C. Bishop. 2008. Pig peripheral blood mononuclear leucocyte subsets are heritable and genetically correlated with performance.. Animal 2:1575–84.

Denholm, S.J., T.N. McNeilly, G. Banos, M.P. Coffey, G.C. Russell, et al. 2017. Estimating genetic and phenotypic parameters of cellular immune-associated traits in dairy cows. J. Dairy Sci. 100:2850–2862.

Garrick, D.J., J.F. Taylor, and R.L. Fernando. 2009. Deregressing estimated breeding values and weighting information for genomic regression analyses. Genet. Sel. Evol. 41:55.

Huang, D.W., B.T. Sherman, and R.A. Lempicki. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 4:44–57.

Mallard, B.A., H. Atalla, S. Cartwright, B.C. Hine, B. Hussey, et al. 2011. Genetic and epigenetic regulation of the bovine immune system: Practical implications of the high immune response technology. Pages 53–61 in Proc. National Mastitis Council 50th Annual Meeting. National Mastitis Council, Madison, WI.

Pryce, J.E., S. Bolormaa, a J. Chamberlain, P.J. Bowman, K. Savin, et al. 2010. A validated genome-wide association study in 2 dairy cattle breeds for milk production and fertility traits using variable length haplotypes. J. Dairy Sci. 93:3331–3345.

Thompson-Crispi, K.A., M. Sargolzaei, R. Ventura, M. Abo-Ismail, F. Miglior, et al. 2014. A genome-wide association study of immune response traits in Canadian Holstein cattle. BMC Genomics 15:559.

Thompson-Crispi, K.A., A. Sewalem, F. Miglior, and B.A. Mallard. 2012. Genetic parameters of adaptive immune response traits in Canadian Holsteins. J. Dairy Sci. 95:401–9.

Venhoranta, H., H. Pausch, K. Flisikowski, C. Wurmser, J. Taponen, et al. 2014. In frame exon skipping in UBE3B is associated with developmental disorders and increased mortality in cattle. BMC Genomics 15:890.

Wijga, S., J.W.M. Bastiaansen, E. Wall, E. Strandberg, Y. de Haas, et al. 2012. Genomic associations with somatic cell score in first-lactation Holstein cows. J. Dairy Sci. 95:899–908.

Yates, A., W. Akanni, M.R. Amode, D. Barrell, K. Billis, et al. 2016. Ensembl 2016. Nucleic Acids Res. 44:D710–D716.

*Table 1. Summary statistics and genetic parameters of IA trait data. Standard errors are given in parenthesis and significant values (P<0.05) are in **bold**.*

IA trait	*n*	Mean	SD	CV[1]	h^2 *(se)*	*Rep (se)*
PBMC[2, 3]	2,262	58.38	10.20	17.47	**0.29** (0.067)	**0.33** (0.030)
Eosinophils[3]	2,262	3.62	3.44	95.03	**0.16** (0.064)	**0.38** (0.028)
Lymphocytes[3]	2,177	45.39	11.03	24.30	**0.34** (0.029)	**0.34** (0.029)
Monocytes[3]	2,177	12.95	6.35	49.03	**0.15** (0.049)	**0.18** (0.025)
Neutrophils[3]	2,262	37.77	10.06	26.63	**0.27** (0.063)	**0.31** (0.030)
$CD4^+$ [4]	2,229	25.53	6.29	24.64	**0.44** (0.102)	**0.69** (0.022)
$CD8^+$ [4]	2,256	11.30	3.41	30.18	**0.41** (0.095)	**0.75** (0.018)
$CD4^+$:$CD8^+$ ratio	2,229	2.38	0.73	30.67	**0.41** (0.111)	**0.80** (0.015)
$NKp46^+$ [4]	2,257	2.31	1.53	66.23	**0.41** (0.083)	**0.57** (0.027)
Hp[5]	3,561	83.87	369.79	440.89	0.02 (0.020)	**0.09** (0.014)
NAb_{KLH}[5]	2,687	0.94	0.30	31.79	**0.21** (0.055)	**0.41** (0.024)
NAb_{LPS}[5]	3,570	1.15	0.52	44.90	**0.25** (0.086)	**0.47** (0.028)
TNF-α[5]	3,568	1841.74	6435.26	349.41	0.04 (0.029)	**0.21** (0.019)
Hp[6]	2,667	0.97	6.39	659.36	0.09 (0.061)	**0.32** (0.027)
NAb_{KLH}[6]	2,667	0.81	0.35	43.84	**0.17** (0.068)	**0.36** (0.028)
NAb_{LPS}[6]	2,667	0.34	0.24	71.61	**0.37** (0.083)	**0.41** (0.031)
TNF-α[6]	2,667	103.29	375.23	363.29	0.03 (0.046)	**0.30** (0.026)

[1] Coefficient of variation (%)
[2] Peripheral Blood Mononuclear Cells
[3] % of total leukocytes that were PBMC, eosinophils, lymphocytes, monocytes or neutrophils
[4] % of PBMC that were CD4, CD8 and NKp46 positive
[5] Serum serological immune-associated trait
[6] Milk serological immune-associated trait

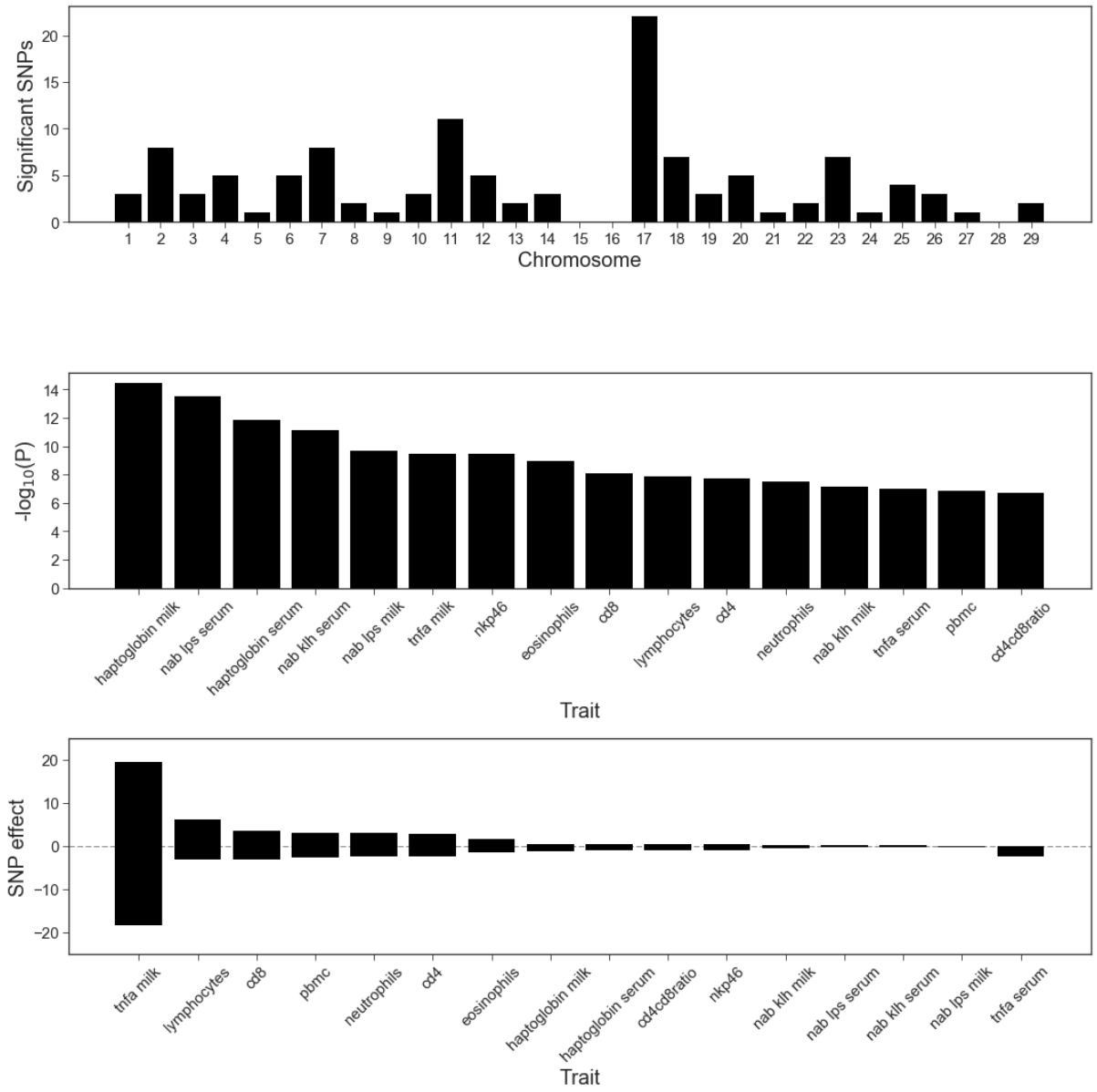

Figure 1. Summary of significant SNP ($P<1.5\times10^{-6}$) for 17 immune-associated traits across chr1 to chr29. Distribution of SNP across chromosomes (A); level of significance of association by trait (B); and, range of SNP effect size per trait (C).

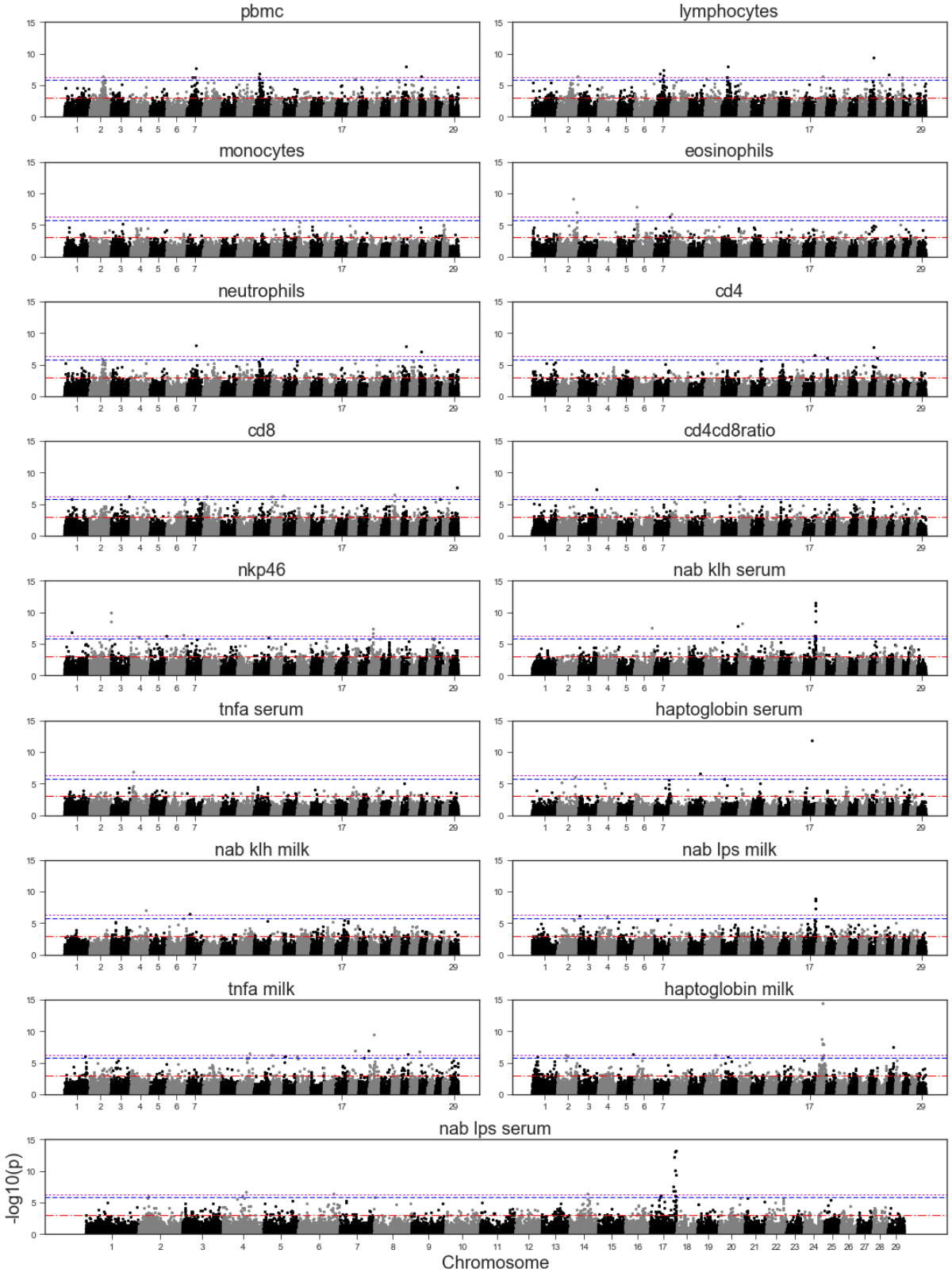

Figure 2. Manhattan plots resulting from a genome-wide association analysis of $-log_{10}(P)$ of SNP for 17 immune-associated traits. Dotted (· · ·), dashed (– – –) and dash-dot (— · —) lines represent significance levels of $P<5\times10^{-7}$, $P<1.5\times10^{-6}$ and $P<0.001$ respectively.

Table 2. Summary of the 9 functional annotated gene clusters associated with significant SNP found via a genome-wide association analysis.

Cluster	*Number of genes*	Enrichment Score[‡]	Cluster names
1	10	3.6	phospholipases
2	6	3.1	Interleukins
3	9	2.4	Histone cluster proteins
4	54	1.1	olfactory receptor
5	6	1.0	homeobox
6	20	0.4	zinc finger protein
7	6	0.3	myelin associated glycoprotein
8	6	0.3	transmembrane proteins
9	5	0.3	transmembrane proteins

[‡] defined as the geometric mean (-log scale) of member's p-values in a corresponding annotation cluster and used to rank biological significance. The top ranked annotation groups most likely have consistent lower p-values for their annotation members (Huang et al., 2009)