Novel SNPs and haplotypes identified in the CD4 gene and their influence on deregressed MACE EBV indexes of milk-related traits in Simmental breed

Francesco Napolitano1, Francesco Grandoni1, Giovanna De Matteis1, Lorenzo Degano2, Daniele Vicario2 and Luca Buttaozzi1

1Consiglio per la ricerca in agricoltura e l’analisi dell’economia agraria (CREA) – Centro di ricerca Zootecnia e Acquacoltura, 00015 Monterotondo, Rome, Italy and 2Associazione Nazionale Allevatori Bovini di razza Pezzata Rossa Italiana, 33100 Udine, Italy

Abstract

Cluster of differentiation 4 (CD4) is the accessory protein non-covalently bound to the T cell receptor that recognizes an invariant region of MHC class II on antigen presenting cells. Its cytoplasmic tail, physically associated with a protein tyrosine kinase, is important in the activation of helper/inducer T lymphocytes. In Bos taurus, CD4 gene is located on chromosome 5 from which two isoforms are transcribed, with a different number of amino acids due to splicing of exon 7 and variation in the reading frame. The aim of this study was to investigate the sequence of the entire CD4 gene in Simmental sires to evaluate the effects of genomic variants on the indexes of the bulls for milk, fat and protein yields, as well as somatic cell score. The associations among genomic variants and indexes were analysed using the Allele and GLM procedures of SAS 9.4. The analysis indicated that only four of the thirty-one identified SNPs influenced the considered traits. Identified variants insist on coding zones and intronic sequences, where we revealed the presence of sites for transcription factors. To evaluate the existence of haplotypic effects, combinations among the four genomic variants (SNP 3, SNP 8, SNP 11 and SNP 19) were investigated. Six different haplotypic alleles were identified, but only four of them were frequent enough to allow for an evaluation of any haplotypic effect (at least six copies in the examined sample): Hap1, Hap2, Hap3 and Hap6. The analysis of associations between the selected haplotypes in the CD4 gene with milk related indexes showed that bulls with Hap2 (T-A-C-C) had better indexes for milk and protein yields (P < 0.05), whereas the presence of the Hap1 haplotype (A-G-A-T) caused a significant decrease of the index for protein yield (P< 0.05). Frequencies of the two alleles Hap1 and Hap2 (9 and 36% respectively) make them of interest for their possible inclusion in breeding schemes and support the hypothesis of considering this gene as a candidate for the improvement of milk-related traits in the Simmental breed.
Bioinformatic analysis was performed using the MATCH™ pro-
gramme (Kel et al., 2003) in the TRANSFAC® Professional 10.2
(http://www.biobaseinternational.com/) (Matys et al., 2006). A
profile search was performed for transcription factor binding sites
in the sequenced DNA fragments, focusing on the areas where SNPs
had been detected.

The ‘vertebrate binding’ and the ‘only high quality’ matrix
options were used in the profile selection, with cut-offs for core
and matrix similarity set to 0.75 and 0.7 respectively, without
any changes in other options.

Diversity analysis of CD4 SNPs
Thirty-one SNPs were identified in the described genomic
sequence (Table 1) sixteen of them causing sequence variations
in coding or regulatory regions of the CD4 gene (SNPs: 1, 2,
11 ÷ 17, 21 ÷ 24, 27 and 30 ÷ 31), and the others located in regions
not expected to affect gene functionality. Among those positioned
in coding regions, SNPs 11 ÷ 17 (exon 2), 22 (exon 4), 23 and 24
(exon 5), produce changes in the amino acid sequence, while the
other two are synonymous (21 of exon 3 and 27 of exon 6). Three
SNPs were newly found in the CD4 gene, namely: SNP 19 and 20,
identified in intron 2 and SNP 30 present in the 3’-UTR regula-
tory region.
Most of the SNPs are informative markers (online Supplementary Table S1), with a PIC value over 0.20. Only genotypes at SNPs 7 and 19 were not in Hardy-Weinberg equilibrium ($\chi^2 = 21.19$ and 14.74, respectively; $P < 0.0001$), possibly due to some sort of selection.

Correlations among SNPs are reported in online Supplementary Table S2. Sixteen SNPs (1, 9 ÷ 18, 21, 22, 25, 28 and 29), one half of the total number, were 100% in linkage disequilibrium (LD = 1). Out of these, one was chosen for the association analysis based on its location in the coding region of the gene and because it had already been used by other authors (Usman et al., 2016; Zeb et al., 2020) in association tests in dairy cattle: SNP 11 (rs110955838) in exon 2. Ten more SNPs with LD < 0.9: SNP 2 (rs109536830), SNP 3 (rs211555032), SNP 5 (rs110147215), SNP 7 (rs110894017), SNP 8 (rs110774456), SNP 20 (novel), SNP 27 (rs136156940), SNP 30 (novel), and SNP 31 (rs136340758) were processed for the estimation of their associations with phenotypes (deregressed MACE EBV indexes of milk, fat and protein yields, and SCS) as function of their genotype.

Association analysis of each SNP on milk indexes

Table 2 reports the association of the genotypes at the 11 SNP loci with the indexes on milk, fat and proteins yields, and SCS. The association analysis indicated that only four of the eleven chosen markers had an effect on some of the considered traits. In particular, SNPs: 3, 8, 11, and 19, exhibited a significant effect ($P < 0.05$)
on protein yield index; SNP 19 exhibited a significant effect ($P < 0.05$) also on milk yield, while SNP 11 showed a numerical but non-significant difference ($P > 0.05$) on the same trait.

**Bioinformatic analysis on the detected SNPs**

The identification of SNPs affecting the considered indexes even if located in non-coding or regulatory regions (SNPs: 3, 8 and 19) led us to evaluate the existence of transcription factor sites near them.

Transcription factors identified in TRANFAC database were present in the regions of the CD4 gene modified by our markers. SNP 3 T > A was located into the ‘core’ of the consensus sequence (tttGATCT/Cct, g.103643038...029 on NC_037332) for GATA-3 (GATA-binding protein 3). No transcription factors were detected in the sequence affected by the marker SNP 8.

**Association analysis of haplotypes on milk indexes**

As reported in Table 2, 4 out of 11 analysed SNPs (3, 8, 11 and 19) produced significant effects on milk related indexes. To evaluate the existence of haplotypic effects, combinations among the four genomic variants were investigated.

Haplotype make-up and frequencies are listed in Table 3. Six different haplotypic alleles were identified, but only four of them were frequent enough to allow for an evaluation of any haplotypic effect (at least six copies in the sample examined): Hap1, Hap2, Hap3, and Hap6.

### Table 2. Association of SNPs of the CD4 gene with the deregressed multiple across country evaluation of estimated breeding values indexes of some milk-related traits in Simmental breed.*

| Name** | Genotype | N  | Milk (kg) Mean ± SE | Fat (kg) Mean ± SE | Protein (kg) Mean ± SE | SCS Mean ± SE |
|--------|----------|----|---------------------|-------------------|------------------------|---------------|
| SNP 2  | CC       | 19 | −629 ± 192          | −38 ± 15          | −20 ± 8                | 92 ± 3        |
|        | CT       | 13 | −19 ± 242           | 1 ± 19            | −6 ± 10                | 91 ± 4        |
|        | TT       |  2 | −150 ± 593          | −22 ± 46          | −9 ± 24                | 91 ± 9        |
| SNP 3  | TT       | 28 | −277 ± 163          | −24 ± 13          | −7 ± 6b                | 91 ± 3        |
|        | TA       |  6 | −837 ± 347          | −21 ± 27          | −46 ± 13b              | 93 ± 6        |
| SNP 5  | TT       | 28 | −378 ± 169          | −25 ± 13          | 11 ± 7                 | 92 ± 3        |
|        | CT       |  6 | −381 ± 359          | −14 ± 27          | −29 ± 14               | 91 ± 6        |
| SNP 7  | CC       |  4 | −516 ± 438          | −23 ± 34          | −12 ± 17               | 99 ± 8        |
|        | CT       | 30 | −359 ± 163          | −23 ± 13          | −14 ± 6                | 91 ± 3        |
| SNP 8  | AA       | 20 | −208 ± 193          | −24 ± 16          | −3 ± 7a                | 91 ± 3        |
|        | GA       | 13 | −695 ± 234          | −24 ± 19          | −33 ± 9b               | 93 ± 4        |
| SNP 11 | CC       | 26 | −237 ± 168a         | −23 ± 13          | −5 ± 6b                | 91 ± 3        |
|        | AC       |  8 | −820 ± 297b         | −24 ± 24          | −44 ± 11b              | 93 ± 5        |
| SNP 19 | TT       |  7 | −1102 ± 298b        | −37 ± 25          | −42 ± 6b               | 88 ± 6        |
|        | TC       | 27 | −183 ± 154a         | −19 ± 13          | −7 ± 6a                | 92 ± 3        |
| SNP 20 | TT       | 22 | −465 ± 190          | −31 ± 14          | −18 ± 7                | 91 ± 3        |
|        | TC       | 12 | −227 ± 251          | −8 ± 19           | −8 ± 10                | 92 ± 4        |
| SNP 27 | AA       |  9 | −577 ± 294          | −20 ± 23          | −14 ± 12               | 94 ± 5        |
|        | GA       | 18 | −290 ± 214          | −30 ± 16          | −10 ± 8                | 92 ± 3        |
|        | GG       |  7 | −338 ± 334          | −9 ± 26           | −24 ± 13               | 89 ± 6        |
| SNP 30 | CC       | 16 | −644 ± 212          | −23 ± 17          | −22 ± 9                | 94 ± 4        |
|        | CG       | 15 | −91 ± 227           | −25 ± 18          | −6 ± 9                 | 90 ± 4        |
|        | GG       |  3 | −301 ± 490          | −13 ± 39          | −11 ± 20               | 88 ± 8        |
| SNP 31 | CC       | 11 | −499 ± 268          | −19 ± 20          | −13 ± 10               | 93 ± 5        |
|        | CG       | 17 | −298 ± 222          | −31 ± 17          | −11 ± 9                | 92 ± 4        |
|        | GG       |  6 | −371 ± 362          | −9 ± 28           | −26 ± 14               | 88 ± 6        |

*Within each group, values with differing superscript letters mean a significant difference (a, b = $P < 0.05$) or numerical but non-significant difference (A, B = $P < 0.10$).

**Novel SNPs are marked in bold.
The analysis of the associations between the selected haplotypes in the CD4 gene with milk related indexes (Table 4) showed that bulls with haplotype Hap2 (T-A-C-C) had better indexes for milk and protein yields ($P < 0.05$), whereas the presence of the Hap1 haplotype (A-G-A-T) caused a significant decrease of the index for protein yield ($P < 0.05$). Frequencies of the two alleles Hap1 and Hap2 (9 and 36% respectively) make them of particular interest.

### Discussion

This study documents the occurrence of several polymorphisms in the genomic sequence of the entire CD4 gene in a sample of Simmental genomic sires. Many of these polymorphisms, as already reported by other authors (Usman et al., 2016; Zeb et al., 2020), were in perfect linkage disequilibrium in the surveyed 34 genomes. In the present work we investigated the possible effect of genomic variants in CD4 on the deregressed MACE EBV indexes for SCS, an indicator of udder health status and for milk quality and quantity.

The identified variants insist on coding regions (variants of exons 2, 3, 4, 5 and 6), in regulatory regions (5’ and 3’ UTR) and also within intronic sequences. Attention to these last two areas of the CD4 gene revealed the presence of sites for transcription factors. SNP 3 (rs211555032) located into the ‘core’ of the consensus sequence of binding site for CREB, determined a significant and negative effect on protein yield index. CREB is a transcription factor that regulates diverse cellular responses, including proliferation, survival, and differentiation. Its activation is required for the generation and maintenance of regulatory T cells (Wen et al., 2010). Furthermore, as reported by Pegolo et al. (2018) the transcription factor CREB is known to regulate milk protein synthesis.

In our sample of Simmental sires the SNP 8 (rs110774456) highlights a significant and negative effect on the protein yield index. No effect was detected on the other traits, in particular for the somatic cell score.

The analysis of the genetic variants of the CD4 gene in intron 6, which corresponds to the SNP g.13598 C > T of the CD4 gene in Simmental breed. This study documents the occurrence of several polymorphisms in the genomic sequence of the entire CD4 gene in a sample of Simmental genomic sires. Many of these polymorphisms, as already reported by other authors (Usman et al., 2016; Zeb et al., 2020), were in perfect linkage disequilibrium in the surveyed 34 genomes. In the present work we investigated the possible effect of genomic variants in CD4 on the deregressed MACE EBV indexes for SCS, an indicator of udder health status and for milk quality and quantity.

The identified variants insist on coding regions (variants of exons 2, 3, 4, 5 and 6), in regulatory regions (5’ and 3’ UTR) and also within intronic sequences. Attention to these last two areas of the CD4 gene revealed the presence of sites for transcription factors. SNP 3 (rs211555032) located into the ‘core’ of the consensus sequence of binding site for CREB, determined a significant and negative effect on protein yield index. CREB is a transcription factor that regulates diverse cellular responses, including proliferation, survival, and differentiation. Its activation is required for the generation and maintenance of regulatory T cells (Wen et al., 2010). Furthermore, as reported by Pegolo et al. (2018) the transcription factor CREB is known to regulate milk protein synthesis.

In our sample of Simmental sires the SNP 8 (rs110774456) highlights a significant and negative effect on the protein yield index. No effect was detected on the other traits, in particular for the somatic cell score.

The analysis of the genetic variants of the CD4 gene in intron 6, which corresponds to the SNP g.13598 C > T of the CD4 gene in Simmental breed.

### Table 3. Information on haplotypic alleles of the CD4 gene in Simmental breed.

| Haplotype    | N      | SNP 3 rs211555032 | SNP 8 rs110774456 | SNP 11 rs110955838 | SNP-19 novel | Frequency |
|--------------|--------|------------------|------------------|------------------|--------------|-----------|
| Hap1         | 6      | A                | G                | A                | T            | 0.09      |
| Hap2         | 26     | T                | A                | C                | C            | 0.36      |
| Hap3         | 27     | T                | A                | C                | T            | 0.42      |
| Hap4         | 2      | T                | G                | A                | T            | 0.03      |
| Hap5         | 1      | T                | G                | C                | C            | 0.01      |
| Hap6         | 6      | T                | G                | C                | T            | 0.09      |

### Table 4. Association of the haplotypic alleles of the CD4 gene with the deregressed multiple across country evaluation of estimated breeding values indexes of some milk traits in Simmental breed*

| Trait | Haplotype/Frequency | Copy | N    | Milk, kg Mean ± SE | Fat, kg Mean ± SE | Protein, kg Mean ± SE | SCS Mean ± SE |
|-------|---------------------|------|------|-------------------|------------------|----------------------|---------------|
|       | Hap1 (A-G-A-T)/9%   | 0    | 28   | −277 ± 163        | −24 ± 13         | 7 ± 6a               | 91 ± 3        |
|       |                     | 1    | 6    | −837 ± 347        | −21 ± 28         | 46 ± 13b              | 93 ± 6        |
|       | Hap2 (T-A-C-C)/36%  | 0    | 8    | −901 ± 291b       | −31 ± 24         | 36 ± 11b              | 88 ± 5        |
|       |                     | 1    | 26   | −211 ± 165a       | −20 ± 13         | −7 ± 6a               | 93 ± 3        |
|       | Hap3 (T-A-C-T)/42%  | 0    | 9    | −179 ± 294        | −7 ± 22          | −15 ± 12              | 93 ± 5        |
|       |                     | 1    | 23   | −459 ± 188        | −31 ± 14         | −14 ± 8               | 92 ± 3        |
|       | Hap6 (T-G-C-T)/9%   | 0    | 28   | −390 ± 169        | −24 ± 13         | −15 ± 7               | 92 ± 3        |
|       |                     | 1    | 6    | 328 ± 358         | −19 ± 27         | −11 ± 14              | 91 ± 6        |

*Within each group, values with differing superscript differ significantly, $P<0.05$ or greater.
SCS. The same authors identify possible reasons for these discrepancies in the differences in the breeds of cattle, the environment conditions, and population structure.

The novel SNP 19 was located in the ‘core’ of the consensus sequence of binding site for the transcription factor GATA-3 and determined a significant and positive effect on milk and protein yield indexes. The zinc finger transcription factors GATA-3 was defined as a master regulator of the mammary gland (Tong and Hotamisligil, 2007): its expression is required for normal embryonic mammary development, while in the postnatal mammary gland, GATA-3 is essential for the formation of the ductal tree and for differentiation of the epithelial cells (Kourous-Mehr et al., 2006; Tong and Hotamisligil, 2007; Kinoshita et al., 2014). GATA-3 expression is also essential for early T cell development in the thymus and for differentiation of naïve CD4+ T cells into committed T helper type 2 cells (Skapenko et al., 2004; Tindemans et al., 2014).

The haplotypic approach allows us to gain a systemic view of the probable role of the CD4 gene in the determination of livestock traits. This study, within the limits of our sample, highlighted the association of some of the newly found haplotypes with milk-related deregressed MACE EBV indexes in Simmental breed. Therefore, they could be used as genetic markers in future breeding programs. For example, it should be possible to increase the frequency of haplotype Hap2 (T-A-C-C) in the population, if the selection goals included milk quality and yield improvement. Protein yield appear to be significantly depressed by Hap1 (A-G-A-T). As it can be easily seen, the two haplotypes differ in all four SNPs contributing to their identification. This is important as an indirect confirmation of the results from SNP association analysis.

Although the role of CD4 gene during inflammation is well documented (Oyugi et al., 2009; Usman et al., 2016), little is known about its possible association with production traits. This work implements knowledge on CD4 genomic variability through a ‘candidate gene’ approach, highlighting three novel variants. These new SNPs, together with the many other ones previously identified by other authors, allowed us to identify new haplotypes considering the entire gene sequence. The significant association of CD4 haplotypes with milk and protein indexes observed in this study supports the hypothesis of considering this gene as a candidate for the improvement of milk-related traits in Simmental breed.

In conclusion, we investigated the sequence of the entire CD4 gene in Simmental sires to evaluate the effects of genomic variants on the indexes of the bulls for milk, fat and protein yields, as well as somatic cell score. Only four of thirty-one identified SNPs influenced these traits. This work implements knowledge on CD4 genomic variability through a ‘candidate gene’ approach, highlighting three novel variants. These new SNPs, together with the many other ones previously identified by other authors, allowed us to identify new haplotypes considering the entire gene sequence. The significant association of CD4 haplotypes with milk and protein indexes observed in this study supports the hypothesis of considering this gene as a candidate for the improvement of milk-related traits in Simmental breed.

**Supplementary material.** The supplementary material for this article can be found at https://doi.org/10.1017/S0022029921000832

**References**

Grandoni F, Napolitano F, Scata MC and De Matteis G (2020) A new polymorphic epitope of bovine CD4 antigen evidenced by flow cytometry. Veterinary Immunology and Immunopathology 219, 109957.

He Y, Chu Q, Ma P, Wang Y, Zhang Q, Sun D, Zhang Y, Yu Y and Zhang Y (2011) Association of bovine CD4 and STAT5b single nucleotide polymorphisms with somatic cell scores and milk production traits in Chinese Holsteins. Journal of Dairy Research 78, 242–249.

Hussein YM, El Tahoury SA, Mohamed RH, El-Shal AS, Abdo SA and Abdo M (2012) Association of CD4 enhancer gene polymorphisms with rheumatoid arthritis in Egyptian female patients. Rheumatology International 32, 2325–2330.

Kel AE, Gossling E, Reuter I, Cheremushkin E, Kel-Margoulis OV and Wingender E (2003) MATCHTM, a tool for searching transcription factor binding sites in DNA sequences. Nucleic Acids Research 31, 3576–3579.

Kinoshita Y, Yoshizawa K, Emoto Y, Yuki M, Yuri T, Shikata N and Tsubura A (2014) Similarity of GATA-3 expression between Rat and human mammary glands. Journal of Toxicologic Pathology 27, 159–162.

König R and Zhou W (2004) Signal transduction in T helper cells: CD4 coreceptors exert complex regulatory effects on T cell activation and function. Current Issues in Molecular Biology 6, 1–15.

Kouros-Mehr H, Slorach EM, Sternlicht MD and Werb Z (2006) GATA-3 maintains the differentiation of the luminal cell fate in the mammary gland. Cell 127, 1041–1055.

Kwong J, McCallum N, Sintchenko V and Howden BP (2015) Whole genome sequencing in clinical and public health microbiology. Pathology 47, 199–210.

Matys V, Kel-Margoulis OV, Fricke E, Liebich I, Land S, Barre-Dirrie A, Reuter I, Chekmenev D, Krull M, Hornischer K, Voss N, Stegmaier P, Lewicki-Potapov B, Saxel H, Kel AE and Wingender E (2006) TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. Nucleic Acids Research 34, D108–D110.

Oyugi JO, Vouriot FCM, Alimonti J, Wayne S, Luo M, Land AM, Ao Z, Yao X, Sekaly RP, Elliott LJ, Simonsen JN, Ball TB, Jaako W, Kimani J, Plummer FA and Foweke KR (2009) A common CD4 gene variant is associated with an increased risk of HIV-1 infection in Kenyan female commercial sex workers. The Journal of Infectious Diseases 199, 1327–1334.

Pegolo S, Mach N, Ramayo-Caldas Y, Schiavon S, Bittante G and Cecchinato A (2018) Integration of GWAS, pathway and network analyses reveals novel mechanistic insights into the synthesis of milk proteins in dairy cows. Scientific Reports 8, 566.

Skapenko A, Leipe J, Niesner U, Devriendt K, Beetz R, Radbruch A, Kalden JR, Lipsky PE and Schulze-Koops H (2004) CD4 and STAT5 single nucleotide polymorphisms influence the CD4+ T cell helper type 2 differentiation. Journal of Experimental Medicine 199, 423–428.

Tindemans I, Serafini N, Di Santo JP and Hendriks RW (2014) GATA-3 Function in innate and adaptive immunity. Immunity 41, 191–206. https://doi.org/10.1016/j.immuni.2014.06.006.

Tong Q and Hotamisligil GS (2007) Cell fate in the mammary gland. Nature 444, 724–726.

Usman T, Yu Y, Zhai L, Liu C, Wang X and Wang Y (2016) Association of CD4 SNPs with fat percentage of Holstein cattle. Genetics and Molecular Research 15, gmr.15038697.

Wang XS, Zhang Y, He YH, Ma PP, Fan LJ, Wang YC, Zhang YI, Sun DX, Zhang SL, Wang CD, Song JZ and Yu Y (2013) Aberrant promoter methylation of the CD4 gene in peripheral blood cells of mastitic dairy cows. Genetics and Molecular Research 12, 6228–6239.

Wen Y, Sakamoto KM and Miller LS (2010) The role of the transcription factor CREB in immune function. The Journal of Immunology 185, 6413–6419.

Zeb S, Ali N, Niaz S, Rasheed A, Khattak I, Khan NU, Wang Y and Usman T (2020) Association of SNPs in the coding regions of CD4 gene with mastitis susceptibility and production traits in dairy cattle. Thai Journal of Veterinary Medicine 50, 75–80.