Missense mutations of NCPAG gene affect calving ease in Piedmontese cattle: preliminary evidences

Giovanni Chillemi, Silvia Bongiorni, Silvia Gioiosa, Tiziano Flati, Tiziana Castrignanò, Marco Milanesi, Paolo Ajmone Marsan & Alessio Valentini

To cite this article: Giovanni Chillemi, Silvia Bongiorni, Silvia Gioiosa, Tiziano Flati, Tiziana Castrignanò, Marco Milanesi, Paolo Ajmone Marsan & Alessio Valentini (2017): Missense mutations of NCPAG gene affect calving ease in Piedmontese cattle: preliminary evidences, Italian Journal of Animal Science, DOI: 10.1080/1828051X.2017.1370362

To link to this article: http://dx.doi.org/10.1080/1828051X.2017.1370362

© 2017 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

Published online: 30 Aug 2017.

Article views: 35

View Crossmark data
Missense mutations of NCPAG gene affect calving ease in Piedmontese cattle: preliminary evidences

Giovanni Chillemi, Silvia Bongiorni, Silvia Gioiosa, Tiziano Flati, Tiziana Castrignan, Marco Milanesi, Paolo Ajmone Marsan and Alessio Valentinib

ABSTRACT
A previous genome scan on 323 Piedmontese individuals identified a cluster of 13 SNPs significantly associated with direct calving ease and centred on the three genes LAP3, LCORL and NCPAG in chromosome 6. We investigated missense mutations affecting calving ease in Piedmontese cattle in the identified region using sequences from the whole exome in eight Piedmontese individuals chosen from the extremes of the direct calving ease estimated breeding values distribution for this trait. The present study has not found missense variants in LAP3 and LCORL, while two were identified on NCPAG by three different variant calling methods. Other gene candidates in the same region harbour missense mutations, such as PPM1K, PKD2, SPP1 and MEPE, but both SIFT analysis and chi-square test on frequency of alleles make us hypothesise that NCPAG is the single gene responsible for the trait variation. The two SNPs on NCPAG are in complete linkage disequilibrium in our samples; therefore, further investigations are needed in order to discriminate their role.

ARTICLE HISTORY
Received 17 March 2017
Revised 22 May 2017
Accepted 28 June 2017

KEYWORDS
NCPAG; Piedmontese; calving ease; whole-exome sequencing; NGS

Introduction
In the past years, genetic improvement programmes generated an increase in livestock production. However, several traits different from production received less attention, particularly those related to reproduction. Indeed, they are difficult to record and the low heritability made them less appealing as objectives of selection using traditional methods. In recent years, molecular genetics tools helped dissecting difficult traits and started elucidating also the biological role of genes and regions. Molecular markers, notably SNP panel genotyping, have become the standard procedure for exploring chromosomes to find regions associated with particular traits. The literature has amply demonstrated the potentiality of this approach, a few recent examples in cattle are focussed on different traits: lactation persistency in Holstein (Do et al. 2017); temperament in Guzerat (Santos et al. 2017) and beef production traits in Marchigiana (Sorbolini et al. 2017). The finding of candidate positions in a chromosome, although valuable, the ultimate goal is to find the causal mutation behind a trait variation. In this respect, the diacylglycerol O-acyltransferase 1 gene (DGAT1), is exemplary. The K232A mutation was found affecting milk yield in Holstein–Friesian (Grisart 2004), while the R251L and R396W mutations are responsible for reducing milk fat content in goat (Martin et al. 2017).

Currently, genome sequencing has become feasible and relatively cheap, allowing to gather much more information than SNP panels. Particularly new is the possibility to sequence only the exome, thus capturing mostly the translated regions. In this way, it is possible to obtain a much higher sequencing depth than possible with a full-genome sequencing. Arguments against are that several regions, particularly those controlling the gene regulation, are lost; arguments for are that it is possible to call base variation in an accurate way due to the high multiplicity of reads.

Whole-exome sequencing has successfully been applied to the discovery of putative causative mutations in regions already identified by SNP-genotyping analysis, for example, in relation to bovine embryo...
development (McClure 2014) or high-altitude pulmonary hypertension in cattle (Newman et al. 2015). Here, following the same approach, we carried out a whole-exome analysis on eight bulls of Piedmontese cattle breed, chosen at the extreme of direct calving ease estimated breeding values (EBV), with the aim to investigate for causative mutations in the region of BTA6 already found significantly associated with calving ease (Bongiorni 2012). Direct calving ease is the ability for a cow to deliver a calf without problems. While we obtained the sequence of whole exome, in this study, we focussed in a region in BTA6 that has been previously found implied in calving ease (Bongiorni 2012). The analogous region in horses has been deemed as a candidate locus for withers height (Tetens 2013) and for skeletal variation (Staiger 2016); as well as for body weight in sheep (Al-Mamun 2015).

We evaluated the potential effect on protein functionality of missense SNPs, thus narrowing the search for the causative variation.

Materials and methods

We have chosen two groups of four Piedmontese bulls from the extremes of the distribution of the EBV for direct calving ease. Positive values are indicative of a better calving ease ability and vice versa. Although the Breeders Associations computes also the indirect calving ease, we used only the direct one since the former was available for too few individuals. The groups were formed using only the EBV and not the SNP genotype as this would have biased the choice towards ‘similar’ genotypes by descent or by population structure. In the following the two groups are named ebv+ and ebv−, respectively. Ebv+ and ebv− groups had means that were 1.998 and −2.655 standard deviations from the population mean respectively.

Genomic DNA was extracted from semen using the NucleoSpin® Tissue kit (MACHEREY-NAGEL GmbH & Co, Düren, Germany) according to manufacturer’s instruction. DNA was checked for quality on agarose gel and quantified using a DTX® microplate reader (Beckman Instruments, Inc, Fullerton, CA) after staining with PicoGreen® (Invitrogen; Thermo Fisher Scientific Inc, Waltham, MA). Libraries were sequenced on the HiSeq® instrument (Illumina Inc, San Diego, CA) with a 100-bp paired-end protocol.

The adopted bioinformatics NGS data analysis is close to the one described in Chiara et al. (2016), which takes advantage of three different variant callers algorithms and builds a robust set of highly reliable results by merging all the predictions found by at least two variant calling methods. In our case, in order to maximise the reliability, we considered only the variants confirmed by the three methods. In particular, a read quality trimming was executed with the software v.0.33 (Aachen, Germany) (Bolger 2014), a tool that discards low-quality reads after quality check and removes primer/adaptor sequences. A quality check with the fastQC (Andrews 2010) code was executed before and after the trimming phase, to verify the read trimming efficacy.

The resulting set of reads were then aligned to the Bos taurus reference genome (UCSC BosTa8 version/Ensembl UMD3.1.1 version) by means of the Bowtie2 software (Langmead and Salzberg 2012) a quick and accurate aligner that carries out a fine sensitivity tuning of the alignments.

SAMtools (Li 2009) was then used to convert the aligned reads to the binary Alignment/Map format (BAM) and to sort them by chromosomal coordinates. Moreover, the PicardTools v. 1.119 (Cambridge, MA) (Available online at: https://broadinstitute.github.io/picard/) markDuplicates utility was used to remove possible PCR artefacts introduced during the library preparation. After this step, the BAM files from each individual were indexed through the samtools index command and passed as joint input to FreeBayes v. 0.9.2.1 (Garrison and Marth 2012) and VarScan2 v. 2.3.7 (Koboldt et al. 2012), two of the three variant calling software used in our bioinformatics pipeline.

Independently, on the same BAM files, the GATK BaseRecalibrator (McKenna 2010; DePristo 2011) utility was carried out to re-calibrated quality score after indel realignment.

Detection of single-nucleotide and indel polymorphisms (SNPs and DIPs) has been also carried out by means of the main algorithm of GATK, HaplotypeCaller.

In order to perform a joint variant calling, the GenotypeGVCFs application was used to merge individual gVCF records independently produced by the HaplotypeCaller and hard filters were applied, as suggested by GATK best practices (Van der Auwera 2013).

In the end, custom Perl scripts were used to compare the filtered vcf files obtained by each of the three variant callers to produce the final consensus call set.

To confirm the accuracy of each call, the Phred-like quality score was associated with variant results. Furthermore, each detected variant has been compared to a set of known B. taurus variants (dbsnp v.148 obtained from http://hgdownload.cse.ucsc.edu/ goldenPath/bosTa8/database/) and annotated with further information such as gene name, genome position, amino acid substitution and functional effect, by using the ANNOVAR tool (2016 Feb. 01 Release) (Wang
et al. 2010; Yang and Wang 2015). SIFT algorithm was used to evaluate the effect of missense variations (Kumar 2009).

The chi-square test has been applied to evaluate differences between frequency of alleles with a significant threshold of $10^{-3}$ for the $p$ value.

### Results and discussion

In a previous work (Bongiorni 2012), we found that Piedmontese cattle shows a sharp and unique probability peak in chromosome 6 for direct calving ease EBV. Here, we report a more detailed search for all variants within the above region in coding regions. We have examined them for their frequency in the tails of the EBV distribution and their effect on coded proteins in order to get closer to the most probable causative one. These results are preliminary and need further studies, but anyway they bring our knowledge a step further in comparison to the current status.

The number of reads per sample are reported in Supplemental Table S1. Due to technical reasons, the mapped reads vary greatly from a minimum of 19.4 M–84.7 M. The variant calling procedure applied by us; however, being based on the validation performed by three different algorithms, allowed us to obtain very reliable results.

In the region 37.8–39.0 Mb, previously identified as a candidate region of hosting a variation affecting calving ease (Bongiorni 2012), our bioinformatics analysis highlighted 30 variants respect the Btau8 sequence on nine different genes.

All SNPs segregate in our samples but rs110493242. Due to our limited number of individuals, we cannot assert with high probability if this genotype is private to Piedmontese, but it will be worth to further explore its frequency in a larger data set, both for traceability and for studying the specific effect on the makeup of this breed.

Only 10 in five genes are coding for missense variants (Table 1). Moreover, according to SIFT score only variants in Polycystin 2, Transient Receptor Potential Cation Channel (PKD2), Matrix Extracellular Phosphoglycoprotein (MEPE) and Non-SMC Condensin I Complex Subunit G (NCAPG) can be considered ‘deleterious’.

As shown in Table 1, most of the missense variants are ‘non-deleterious’ and a few of them are in non-coding regions, probably resulting from a dragging effect in the exome capturing.

No missense variations were found in two of the genes deemed involved in Bongiorni et al. (2012), that is, Ligand-Dependent Nuclear Receptor Corepressor Like (LCORL) and Leucine Aminopeptidase 3 (LAP3). Since the region is quite short and gene dense, we can hypothesise that LCORL and LAP3 are identified by the genome scan of SNP in the chip due to a hitch-hiking effect.

The only gene suggested in Bongiorni et al. (2012) and confirmed here as carrying missense variants is NCAPG. Interestingly, both variations in this gene show the lowest SIFT score, that is, they are the most ‘deleterious’ of the set. Our analysis highlighted two variants in NCAPG. The first variation observed occurs on exon 9 at position 38,777,311 and it is a T $\Rightarrow$ G transversion. The second one appears on exon 17 at position 38,808,241 and it is a C $\Rightarrow$ A transversion. Interestingly, in all eight samples, these two genomic loci are completely correlated, thus indicating a linkage disequilibrium relationship.

In addition to NCAPG, our analysis reveals two other genes lying on the genomic region of interest and affected by a deleterious SIFT score: MEPE (rs134577412, SIFT score = 0.02) and PKD2 (rs383697460, SIFT score = 0.04).

Dividing the individuals in two sets of equal size according to the EBV, the Chi square test reveals highly significant differences between frequency of alleles ($p = 9.0 \times 10^{-6}$) only for the SNPs in the NCAPG gene (Supplemental Table S2). Note that a closer look to our data, reveals that rs109570900 is supported by a slightly greater GT (genotype quality) than

Table 1. Missense variations in the region 37.8–39.0 Mb of BTA6 identified in sequenced exomes of Piedmontese individuals.

| Gene      | Position (bp) | AA Change | dbSNP         | SIFT  |
|-----------|---------------|-----------|---------------|-------|
| PPM1K     | 37,882,970    | c.G13A    | p.A5T         | rs384876247  |
| PPM1K     | 37,883,034    | c.G77A    | p.R26H        | rs134464339  |
| PKD2      | 38,058,254    | c.G1567A  | p.A523T       | rs383697460  |
| PKD2      | 38,068,274    | c.A967G   | p.I323V       | rs378159798  |
| SPP1      | 38,121,192    | c.G937T   | p.T218M       | rs133929040  |
| MEPE      | 38,280,292    | c.C1336T  | p.P446S       | rs208465095  |
| MEPE      | 38,280,348    | c.A1280C  | p.Q427P       | rs134577412  |
| MEPE      | 38,280,426    | c.T1202C  | p.L401P       | rs110493242  |
| MEPE      | 38,286,780    | c.C113T   | p.T38M        | rs211478936  |
| NCA5G     | 38,777,311    | c.T1326G  | p.J442M       | rs109570900  |
| NCA5G     | 38,808,241    | c.C2629A  | p.L877M       | rs110251642  |

Underlined values for SIFT indicate a ‘deleterious’ substitution.
rs110251642, thus suggesting that the first variant is detected in our samples with a higher confidence rate.

The NCAPG gene is a regulatory subunit of the mammalian condensin I complex and it plays a role in during mitotic cell division (Dej 2004). NCAPG was associated with foetal growth by microsatellite markers in cattle (Setoguchi et al. 2009, 2011) and with human height (Gudbjartsson 2008). Transcriptome analyses in placentomes from bovine cloned foetuses revealed that NCAPG is one of the genes most differentially expressed in the placentomes of embryos originating from somatic nuclear cloning versus placentomes originating from embryos generated by conventional artificial insemination (Eberlein 2009). Therefore, somatic cloning of bovine embryos is known to frequently result in extremely large offspring and significant abnormalities of the placenta (Hill 1999).

Our results confirm that NCAPG is a fundamental gene in determining calving ease and points to two SNPs that are unequivocally implied because their effects on the translated protein and the segregation of wild/mutated alleles in ebv+ and ebv− individuals.

Further work is needed to disentangle the effect of the two SNPs. One way could be to collect a sample population large enough to permit the detection of recombination between the SNPs and their consequences. Another way could consist on modelling the effects of the mutations by molecular dynamics (Bongiorni 2016). The latter method could also reveal if the effect on the trait is due indeed to the combination of the two SNPs as a haplotype.

**Conclusions**

By using a whole-exome sequencing we were able to narrow the search for a causative variation affecting calving ease in Piedmontese cattle. Our previous genome scan using the Illumina 54k beadchip pointed to three genes responsible for this phenotype (Bongiorni 2012). The exome-sequencing experiment presented here excluded two of them but revealed that other genes carry allelic variations that could influence the EBV. However, SIFT analysis and genotypes of ebv+ vs ebv− individuals clearly highlighted a single gene (NCAPG) as the most probable one. Even though our data suggest that rs109570900 is the SNP detected with greater accuracy, further work is still needed to demonstrate which of the two identified SNPs in complete linkage disequilibrium is the causative one. Alternatively, it can be hypothesised that both of them constitute the causative haplotype of the EBV variation.

**Acknowledgements**

We kindly thank the Piedmontese Breeders Association for providing genetic indices and biological material. Part of the work was carried out with funds from the Innovagen project (MIPAAF) and part from PRIN-MIUR GEN2PHEN. We would like to thank ELIXIR-IIB (Italian ELIXIR Node) and Cineca for providing computational resources.

**Disclosure statement**

No potential conflict of interest was reported by the authors.

**Funding**

This study was supported by the Innovagen (MIPAAF) and PRIN (MIUR) projects.

**ORCID**

Giovanni Chillemi http://orcid.org/0000-0003-3901-6926
Marco Milanesi http://orcid.org/0000-0001-6244-7455
Paolo Ajmone Marsan http://orcid.org/0000-0003-3165-4579
Alessio Valentini http://orcid.org/0000-0003-4733-7519

**References**

Al-Mamun HA, Kwan P, Clark SA, Ferdosi MH, Tellam R, Gondro C. 2015. Genome-wide association study of body weight in Australian Merino sheep reveals an orthologous region on OAR6 to human and bovine genomic regions affecting height and weight. Genet Sel Evol. 47:66.
Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc
Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 30:2114–2120.
Bongiorni S, Mancini G, Chillemi G, Pariet L, Valentini A. 2012. Identification of a short region on chromosome 6 affecting direct calving ease in Piedmontese cattle breed. PLoS One. 7:e50137.
Bongiorni S, Valentini A, Chillemi G. 2016. Structural and dynamic characterization of the C313Y mutation in myostatin dimeric protein, responsible for the “double muscle” phenotype in Piedmontese cattle. Front Genet. 7:14.
Chiara M, Chillemi G, D’antonio M, D’Onorio D, Meo P, Flati T, Gioiosa S, Picardi E, Pesole G, Castrignanò T. 2016. IVaCS: an integrated variant calling system. PeerJ Preprints. 4:e2213v1.
Dej KJ, Ahn C, Orr-Weaver TL. 2004. Mutations in the Drosophila condensin subunit dCAP-G: defining the role of condensin for chromosome condensation in mitosis and gene expression in interphase. Genetics. 168:895–906.
McClure MC, Bickhart D, Null D, Vanraden P, Xu L, Wiggans G, Liu G, Schroeder S, Glasscock J, Armstrong J, et al. 2014. Bovine exome sequence analysis and targeted SNP genotyping of recessive fertility defects BH1, HH2, and HH3 reveal a putative causative mutation in SMC2 for HH3. PLoS One. 9:e92769.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. 2010. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20:1297–1303.

Newman JH, Holt TN, Cogan JD, Womack B, Phillips JA III, Li C, Kendall Z, Stenmark KR, Thomas MG, Brown RD, et al. 2015. Increased prevalence of EPAS1 variant in cattle with high-altitude pulmonary hypertension. Nat Commun. 6:6863.

Santos FCD, Peixoto MG, Fonseca PA, Pires MF, Ventura RV, Rosse ID, Brunali FA, Machado MA, Carvalho MR. 2017. Identification of candidate genes for reactivity in Guzerat (Bos indicus) cattle: a Genome-wide Association Study. PLoS One. 12:e0169163.

Setoguchi K, Furuta M, Hirano T, Nagao T, Watanabe T, Sugimoto Y, Takasuga A. 2009. Cross-breed comparisons identified a critical 591-kb region for bovine carcass weight QTL (CW-2) on chromosome 6 and the Ile-442-Met substitution in NCAPG as a positional candidate. BMC Genet. 10:43.

Setoguchi K, Watanabe T, Weikard R, Albrecht E, Köhn C, Kinoshita A, Sugimoto Y, Takasuga A. 2011. The SNP c.1326T > G in the non-SMC condensin I complex, subunit G (NCAPG) gene encoding a p.Ile442Met variant is associated with an increase in body frame size at puberty in cattle. Anim Genet. 42:650–655.

Sorbolini S, Bongiorni S, Cellesi M, Gaspa G, Dimarco A, Valentini A, Macciotta NP. 2017. Genome-wide association study on beef production traits in Marchigiana cattle breed. J Anim Breed Genet. 134:43–48.

Staiger EA, Al Abri MA, Pflug KM, Kalla SE, Ainsworth DM, Miller D, Raudsepp T, Sutter NB, Brooks SA. 2016. Skeletal variation in Tennessee walking horses maps to the LCORL/NCAPG gene region. Physiol Genomics. 48:325–335.

Tetens J, Widmann P, Köhn C, Thaller G. 2013. A genome-wide association study indicates LCORL/NCAPG as a candidate locus for withers height in German Warmblood horses. Anim Genet. 44:467–471.

Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Angel GD, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, et al. 2013. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics. 43:1–33.

Wang K, Li M, Hakonarson H. 2010. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 38:e164.

Yang H, Wang K. 2015. Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. Nat Protoc. 10:1556–1566.