Livestock Science 202 (2017) 7–12

Contents lists available at ScienceDirect

Livestock Science

journal homepage: www.elsevier.com/locate/livsci

A genome wide association study between CNVs and quantitative traits in Brown Swiss cattle

R.T.M.M. Prinsen[a], A. Rossoni[b], B. Gredler[c], A. Bieber[d], A. Bagnato[a], M.G. Strillacci[a,*]

[a] Università degli Studi di Milano, Department of Veterinary Medicine, Via Celoria 10, MI 20133, Italy
[b] Associazione Nazionale Allevatori Razza Bruna, Località Ferlina, Bussolengo, 204, VR 37012, Italy
[c] Qualitas AG, Chamerstrasse 56, 6300 Zug, Switzerland
[d] Research Institute of Organic Agriculture, Ackerstrasse 113, 5070 Frick, Switzerland

ARTICLE INFO

Keywords:
CNV
GWAS
Brown Swiss

ABSTRACT

Copy Number Variations (CNVs) are DNA sequences of 50 bp up to several Mb long, which can vary in number of copies in comparison with a reference genome. CNVs can be used in association studies to disclose genetic basis of quantitative traits phenotypic variation. Up to date, no genome-wide association study (GWAS) with CNVs and quantitative traits in such a large Brown Swiss population (i.e. with 1116 samples) has been described. The purpose of this study was to perform a GWAS using CNVs with functional, health and productive traits and to asses the impact on farming and breeding practices. The CNV – association studies were performed with the Golden Helix SVS 8.4.4 software using a correlation-trend test model. Genes within significant associated CNVs for each trait were annotated with a GO analysis using the DAVID Bioinformatics Resources 6.7. A total of 56 CNVs were significantly associated with one or more of the eight evaluated traits. The greatest association signals were given by three CNVs on chromosome 12 for the fat yield trait and on BTA23 for udder traits. The associated CNVs overlap with 23 different genes annotated on the *Bos taurus* genome assembly (UMD3.1).

1. Introduction

Different genomic features can regulate quantitative traits through the variation of gene expression level. Copy Number Variation (CNV), a type of genomic structural variation, is one of those genomic features (Redon et al., 2006). In CNVs a sequence of nucleotides can vary in copy number (i.e. gain or loss) when an individual's genome is compared with a reference one. The length of this structural variation according to Mills et al. (2011) can range from 50 bp up to several Mb.

CNVs can affect gene expression because of different number of copies of a specific sequence in the genome, which could consecutively have an impact on the phenotypic expression variability (Redon et al., 2006; Zhang et al., 2009). CNVs coverage represents approximately 4,6% of the entire bovine genome as recently shown by several studies (Hou et al., 2011; Bagnato et al., 2015; da Silva et al., 2016; Keel et al., 2016; Prinsen et al., 2016).

Genome-wide association studies (GWAS) using CNVs and phenotypes have recently been developed: de Almeida Santana et al. (2016) performed a GWAS using CNVs and feed conversion ratio in beef cattle; Durán Aguilar et al. (2016) performed a GWAS for milk somatic cell score in Holstein cattle using CNVs as markers; Sassi et al. (2016) developed a GWAS on Spanish Holstein with production, milk somatic cell count and type traits. Another study performed in cattle by Xu et al. (2014) revealed some CNVs associated with milk production traits in Holsteins. These types of studies have also been performed in other species such as in swine (Long et al., 2016) and in avian specie (Völker et al., 2010).

So far, no studies on the relationship between CNVs and economically important traits in the Brown Swiss breed using HD SNP chip data are available.

This study describes a GWAS between CNVs and eight production, udder and health related traits in the Italian and Swiss Brown Swiss populations in order to identify CNVs involved in genetic basis of traits used in selection programs.

2. Material and methods

2.1. Sampling and genotyping

A total of 1116 Brown Swiss dairy cattle breed samples (313 males and 803 females) was considered for this study. Among these, 160 samples were genotyped by Braunvieh Schweiz, 761 samples were

* Corresponding author.
E-mail addresses: raphaelle.prinsen@unimi.it (R.T.M.M. Prinsen), attilio.rossoni@anarb.it (A. Rossoni), Birgit.Gredler@qualitasag.ch (B. Gredler), anna.bieber@fibl.org (A. Bieber), alessandro.bagnato@unimi.it (A. Bagnato), maria.strillacci@unimi.it (M.G. Strillacci).

http://dx.doi.org/10.1016/j.livsci.2017.05.011
Received 19 December 2016; Received in revised form 15 May 2017; Accepted 17 May 2017

genotyped within the EC funded LowInputBreeds project, and 195 samples were genotyped within the EC funded Quantomics project. All samples were genotyped on an Illumina BovineHD Genotyping BeadChip featuring 777,962 loci in the bovine genome assembly. The reference genome used in this study, was the UMD3.1 bovine genome assembly.

The CNVs used in this study are those obtained on the same population and data set of a previous study (Prinsen et al., 2016) by using the SVS 8.4.4 software (Golden Helix Inc., Bozeman, MT, USA), where the CNAM univariate segmentation detected 25,030 CNVs, of which 837 were gains and 24,193 were losses.

2.2. Quantitative traits used in the analysis

The Estimated Breeding Values (EBVs) for the production traits (milk yield - MY, fat yield – FY and protein yield – PY), udder traits (fore udder attachment – FUA, rear udder height - RUH, rear udder width - RUW and overall udder score – UDD), and a health related trait (somatic cell score – SCS) were provided by Qualitas AG from their official genetic evaluation records on the Swiss EBV scale.

Deregressed proofs (DEBV) were calculated as:
$DEBV = PA + (EBV–PA)/REL_{dau}$

where:

$DEBV$ = Deregressed EBV; PA = EBV Parent Average; EBV = Estimated Breeding Value; REL_{dau} = Reliability from daughters

according to Van Raden et al. (2009) and used as dependent variable in the association analysis.

2.3. Genome wide association study and gene annotation

The SVS 8.4.4 Correlation-trend test was here used for the CNV-association with production, udder and health traits DEBVs. Significant CNVs were detected whenever their false discovery rate (FDR) adjusted p-values did not exceed a cutoff of 0.05. Only CNVs significantly associated with at least one trait were considered for the CNV annotation step.

Gene content of significant CNVs was determined against the Ensembl v.86 autosomal UMD3.1 gene set (http://www.ensembl.org/biomart/martview/32471e73613a45753e5689f2626f9add) using the intersectBed command of BEDTools software (Quinlan and Hall, 2010) with at least one bp overlapping. A gene ontology (GO) and pathway analysis was performed employing the DAVID online Bioinformatics Resources 6.7 (https://david.ncifcrf.gov/tools.jsp). The options selected were: a high classification stringency and a FDR correction.

3. Results and discussion

3.1. CNV-association and annotation results

Not all phenotypes were available for every sample (Table 1). Therefore, the association analysis for every trait was performed with a different group and number of samples.

Table 2 reports the 56 CNVs associated with the traits (with a FDR corrected p-value < 0.05) and the list of genes mapping within each significant CNV. Since there were no associated CNVs for the SCS trait, Table 2 does not indicate any significant corrected p-value for this trait. Figs. 1 and 2 represent the Manhattan plots of the association analyses of the CNVs with the production and the SCS traits, and the with the udder traits, respectively. In every Manhattan plot the 0.05 FDR significance threshold is shown (fuchsia colored line). There is no threshold line in the SCS plot as there were no significant CNVs associated with this trait.

Table 1
Descriptive statistics for the DEBVs of the productive, the functional and the health traits and the number of available samples per trait (N).

Trait	N	Mean	Standard Deviation	Min	Max
MY	755	−40.74	660.79	−2113.58	2113.36
FY	750	−4.06	29.66	−112.21	88.49
PY	757	−2.40	21.28	−76.90	63.49
FUA	449	−1.42	16.55	−66.45	44.99
RUH	538	−3.09	16.11	−55.59	39.88
RUW	294	−3.44	14.18	−52.63	34.58
UDD	571	−5.74	18.54	−82.70	42.44
SCS	938	2.10	14.59	−53.13	45.28

For the production traits (Fig. 1), 11, 22 and 12 CNVs were significantly associated with MY, FY and PY, respectively. In Fig. 2 the CNVs significantly associated with FUA (22), RUH (1), RUW (17) and UDD (4) are shown.

The chromosome with the highest number of associated CNVs was chromosome 12 which had 12 significantly associated CNVs. Chromosomes 3, 7, 8, 13, 14, 15, 16, 21, 22, 25 and 27 do not have any significant associated CNV for any of the traits. Among the 56 associated CNVs, 38 CNVs are significantly associated with only one of the traits considered. Some of the associated CNVs detected were contemporaneously associated with more than one trait. In particular the CNV on chromosome 9 (CNV_10 in Table 2) was associated at the same time with the MY, FY, PY, FUA and UDD traits.

The annotation analysis allowed the identification of 23 different genes overlapping the associated 16 CNVs as reported in Table 2. On the contrary there are 40 CNVs that do not contain annotated genes (Table 2). This may be due to the partial annotation, especially in non-human species.

The GO and pathway analysis performed with the DAVID online Bioinformatics Resources 6.7, classified the genes contained in these CNVs, in 5 cellular components GO terms: i) GO:0005886~plasma membrane, genes *NOSTRIN, CLDN10, ADGRG3*; ii) GO:0016021~integral component of membrane, genes *HS6ST3, CERS6, BOLA-DRA, BOLA-DQB, CLDN10, OXGR1, BOLA-DRB3, ADGRG3*; iii) GO:0005634~nucleus, genes *NOSTRIN, CDYL2 and CERS6*; iv) GO: 0042613~MHC class II protein complex, genes *BOLA-DQB, BOLA-DRA* and *BOLA-DRB3*; v) GO: 000250~antigen processing and presentation of peptide or polysaccharide antigen via MHC class II.

The genes BOLA-DQB, BOLA-DRA and BOLA-DRB3 were also classified in different KEGG pathways: the autoimmune thyroid disease (bta05320); the Inflammatory bowel disease (bta03521); the Antigen processing and presentation (bta04612); the toxoplasmosis (bta05145); the rheumatoid arthritis (bta05323); the type I diabetes mellitus (bta04940); the Asthma (bta05310), the Graft-versus-host disease (bta05332); Viral myocarditis (bta05416).

Chromosome 10 has the highest number of genes mapped within the associated CNVs. These genes are: *TRAV29DV5* (*T cell receptor alpha variable 29/delta variable 5*), *TRDC* (*T cell receptor delta constant*), *TRAC* (*T-cell receptor alpha constant*), *TRAV14DV4* (*T cell receptor alpha variable 14/delta variable 4*), *TRAV17* (*T cell receptor alpha variable 17*), *TRAV18* (*T cell receptor alpha variable 18*) and *TRAV16* (*T cell receptor alpha variable 16*). Some of these genes are present in more significant associated CNVs of chromosome 10.

All these genes are protein-coding involved in the organization of the T cell receptor molecular structure necessary to recognize foreign processed antigens and bound to major histocompatibility complex molecules at the surface of antigen presenting cells. As reported by Connelley et al. (2014) the variable domains of the T cell receptor chains are formed by recombination of single discontinuous variable (V), diversity (D) and joining (J) genes selected from multiple genomic copies of these genes. The genes in the associated CNVs on chromosome 10 are part of the TRA/TRD locus which contains the genes for the V(D)