Identification and validation of copy number variants in Italian Brown Swiss dairy cattle using Illumina Bovine SNP50 Beadchip®

Alessandro Bagnato,1,2 Maria G. Strillacci,1 Laura Pellegrino,1 Fausta Schiavini,1 Erika Frigo,1 Attilio Rossoni,3 Luca Fontanesi,2 Christian Maltecca,4 Raphaëlle T.M.M. Prinsen,1 Marlies A. Dolezal5

1 Dipartimento di Scienze Veterinarie per la Salute, la Produzione Animale e la Sicurezza Alimentare, University of Milan, Italy
2 Dipartimento di Scienze e Tecnologie Agro-Alimentari, University of Bologna, Italy
3 Associazione Nazionale Allevatori Razza Bruna, Bussolengo (VR), Italy
4 Department of Animal Science, North Carolina State University, Raleigh, NC, USA
5 Institut für Populationsgenetik Veterinärmedizinische, University of Wien, Austria

Introduction

The understanding of the genetic variation in livestock species, such as cattle, is crucial to associate genomic regions to the traits of interest. Copy Number Variations (CNV) is defined as a variable copy number of DNA segments ranging from 50bp to several megabases (Mb) compared with a reference genome (Mills et al., 2011). The CNVs are important sources of genetic diversity and provide structural genomic information comparable to single nucleotide polymorphism (SNP) data; they influence gene expression, phenotypic variation, environmental adaptability and disease susceptibility (Wang et al., 2007).

The development of SNP arrays allowed the identification of CNVs by high-throughput genotyping on different cattle breeds. CNV loci were identified in several indicine and taurine (Hou et al., 2012; Bickhart et al., 2012). In livestock, recent studies underlined the effects of the CNVs in intron 1 of the SOX5 gene causing the pea-comb phenotype in chickens (Wright et al., 2009), in the STX17 gene responsible for premature hair greying and susceptibility to melanoma in horses (Rosenberg et al., 2008). Also, the CNVs in the ASIP gene are responsible in the leading of different coat colours in goats (Fontanesi et al., 2009). In cattle, Meyers et al. (2010), identified the association between CNVs in a deletion state in the SLC4A2 gene and osteoporosis in Red Angus cows. Additionally, it has been reported that a Copy Number Variation Region (CNVR) located on BTA18 is associated with the index of total merit and protein production, fat production and herd life in Holstein cattle (Seroussi et al., 2010). Several CNV detection algorithms based on SNP array are available (Xu et al., 2013). Winchester et al. (2009), Pinto et al. (2011) and Tsuang et al. (2010) recommended the use of a minimum of two algorithms for the identification of CNVs in order to reduce the false discovery rates as the algorithms differ in performance and impact in CNV calling (Xu et al., 2013).

The Italian Brown Swiss breed represents the Italian strain of the Swiss Brown Alpine Breed, originally native of central Switzerland. The typical rusticity of the breed, together with its good production attitude, has led its spread all over many European and American countries, with the differentiation of different genetic groups in relation to various environmental conditions. The milk of the Italian Brown Swiss breed has a good cheese-making attitude due to the low frequency of the allele A of the K-casein, in respect to other breeds (http://www.anarb.it/).

Nowadays in literature, there is not a whole-genome CNV map for the Italian Brown Swiss in a large population dataset. The aim of this study was to obtain a consensus CNV genome map in the Italian Brown Swiss cattle based on the Illumina Bovine SNP50 BeadChip® and two SNP based CNV calling algorithms.

Materials and methods

Sampling and genotyping

The National Association of Italian Brown Swiss breeder (ANARB) provided commercial semen samples for 1342 bulls. Genomic DNA was extracted from semen using the ZR Genomic DNA TM Tissue MiniPrep (Zymo, Irvine, CA, USA). Sample DNA was quantified using NanoQuant Infinite®m200 (Tecan, Männedorf, Switzerland) and diluted to 50 ng/L as required to apply the Illumina Infinium protocol. DNA samples were genotyped using Illumina Bovine SNP50 BeadChip® (Illumina Inc., San Diego, CA, USA) containing 54,001 polymorphic SNPs with an average probe spacing of 51.5 kb and a median spacing of 37.3 kb. In this study, the UMD3.1 assembly was used as the reference genome.

Editing data

All SNPs were clustered and genotyped using the Illumina BeadStudio software V2.0 (Illumina Inc.). Samples that showed a call rate below 98% were excluded for the CNV detection on autosomal chromosomes. The signal intensity data of Log R Ratio (LRR) and B allele frequency (BAF) were exported from the Illumina BeadStudio software and the overall distribution of derivative log ratio spread (DLRS) values was used in the SVS7 software (Golden Helix Inc.) to identify and filter outlier samples, as described by Pinto et al., (2011). Principal component analysis (PCA) for LRR was performed using the SVS7 software to detect the presence of batch effects and correct the signal intensity values accordingly. Samples with extreme wave factors were excluded from the analysis through the SVS7 software wave correction algorithm.

Genomic waves occur when even after normalization the log ratio data still have a long-range wave outline when charted in a genomic
log ratio graph. Waviness is hypothesized to be correlated with the GC content of the probes themselves in addition to the GC content of the region around the probes (Diskin et al., 2008).

**Copy number variations detection**

Two software were chosen for the detection of CNVs: PennCNV (http://www.openbioinformatics.org/penncnv/) and Copy Number Analysis Module (CNAM) of SVS7 software. The use of two software based on different algorithms has the final aim to reduce the false discovery calls resulting from the limitations of the identification of CNVs based on the Illumina Bovine SNP50 BeadChip.

**PennCNV detection**

The open access PennCNV online software is nowadays one of the most utilized CNV calling software in bovine studies; it considers multiple sources of information as the LRR and BAF for every SNP. Furthermore, the software performs quality control measurements for each single CNV analysis. Individual-based CNV calling was performed by PennCNV for all autosomes, using the default parameters of the Hidden Markov Model (HMM) that integrates multiple sources of information to infer CNV calls for individual genotyped samples. To reduce the false discovery rate in CNVs calling we used high quality samples with a standard deviation (SD) of LRR <0.30 and with default set of BAF drift as 0.01. In addition, we deleted the CNVs which overlapped at least 10% of telomere length (the first and last 500 kb of each autosome were considered representing the telomeres).

**SVS7 detection**

SVS7 software has a user-friendly graphical interface, efficient pipelines for analysis and workflow, optimized computational speed as well as a technical support. The univariate analysis was used for the CNVs identification. The univariate approach segments each sample independently. An extra covariate is generated every time a sample has a change-point. The covariates value is the mean intensity of the original segment for that sample. This results in a spreadsheet displaying all change-points found within the samples. The following options in SVS7, were utilised as suggested in the software manual:

1. Univariate outlier removal;
2. Maximum number of segments: search for up to 10 per 10,000 markers;
3. Minimum markers per segment: 1;
4. Maximum pairwise permuted P value: 0.005 (2000 permutations per segment pair).

**Copy number variation regions definition**

CNVRs were defined as in Redon et al. (2006) with the BedTools software (Quinlan and Hall, 2010) within software. In addition, consensus regions were created among those identified within the two software, using the Wain et al. (2009)’s approach, which identified only CNVRs that fully overlapped each other.

**Copy number variation regions validation by quantitative polymerase chain reaction**

Quantitative PCR (qPCR) experiments were performed to validate the CNVRs among those identified. The BTF3 gene was selected as a reference location for all qPCR experiments (Bae et al., 2010). Primers for the selected target regions and for the reference gene were designed with the Primer Express® Software v3.0.1 (Life Technologies™, Carlsbad, CA, USA) using the minor groove binder (MGB) quantification parameters. All the qPCR experiments were run in quadruple using the qPCR protocol described by TaqMan® Copy Number Assays Kit (Life Technologies™) on 7500 Fast Real-time PCR System instrument (Applied Biosystems by Life Technologies™). The samples for each qPCR experiment were randomly selected with or without CNVs for each CNVR. The analysis of the crossing thresholds (CT) for each samples tested was carried out using CopyCaller™ software (Applied Biosystems). The validated CNVR positions were converted from Bos taurus UMD3.1 to Btau_4.6.1 assembly using the Batch Coordinate Conversion option in the UCSC database (https://genome.ucsc.edu/) in order to identify potential candidate CNV genes for complex traits.

**Copy number variation regions annotation**

The full Ensembl v76 gene set for the autosomal chromosomes was downloaded (http://www.ensembl.org/biomart/martview?76
dl1cab099658c68bde77f7da55117e). Gene ontology (GO) and pathways analyses, using the DAVID Bioinformatics Resources 6.7 (http://david.abcc.ncifcrf.gov/), were performed (with the high classification stringency and the false discovery rate (FDR) correction). The analyses allowed the identification of molecular functions, biological processes, cellular components and pathways for the genes included in the consensus CNVRs.

**Results and discussion**

The application of stringent quality filters described above reduced the number of bull samples to be analysed to 651. In this way we filtered most of the potentially problematic data that would have reduced the reliability of the called CNVs. Anyway, it is clear that the remaining bulls constitute a good representation of the Italian Brown Swiss sire population that was used for artificial insemination during the last decades, including several half sib-families.

### Table 1. Descriptive statistics for copy number variations identified with PennCNV and SVS7 software.

| Copy number | Number of events | Mean       | Median     | Total length | Min length | Max length |
|-------------|------------------|------------|------------|--------------|------------|------------|
| **PennCNV** |                  |            |            |              |            |            |
| 0           | 97               | 311,345.2  | 245,646.2  | 30,200,500   | 46,665     | 1,053,143  |
| 1           | 2086             | 159,066.4  | 134,534.5  | 331,711,379  | 40,374     | 1,688,267  |
| 3           | 2915             | 488,559.5  | 385,138.7  | 1,423,739,019| 41,449     | 4,457,756  |
| 4           | 1                | 511,301.5  | 511,301.5  | 511,301      | 511,301    | 511,301    |
| **Total**   | 5099             |            |            |              |            |            |
| **SVS7**    |                  |            |            |              |            |            |
| Loss        | 762              | 94,830.3   | 57,612.5   | 72,260,727   | 11,315     | 1,440,751  |
| Gain        | 527              | 80,324.2   | 37,591.4   | 42,330,968   | 20,342     | 770,044    |
| **Total**   | 1289             |            |            |              |            |            |

0, homozygous deletion; 1, heterozygous deletion; 3, heterozygous duplication; 4, homozygous duplication; loss, homozygous or heterozygous deletion; gain, homozygous or heterozygous duplication.
Copy number variations and copy number variation regions detection

Table 1 shows the descriptive statistics of the identified CNVs length using PennCNV and SVS7 software.

Using PennCNV, a total of 5099 CNVs were detected, located in all 29 autosomes with a mean size of 350 kb (±165,259) ranging from 40.4 kb to 4.46 Mb (median=230 kb). The highest number of CNVs was detected on BTA7 (8.4%). In detail, the homozygous deletion, heterozygous deletion and the heterozygous duplication CNVs with the highest frequency were observed on BTA5 (12.4%), BTA7 (13.4%), BTA2 (7.9%), respectively. Only one homozygous duplication CNV was identified on BTA25. A total of 1289 CNVs were identified by SVS7 in all the 29 autosomes. The length of the CNVs ranged from 11.3 kb to 1.4 Mb with median and average values equal to 45 kb and 88.9 kb, respectively. The highest frequency of gain (23.9%) and loss (21%) CNVs were detected on BTA28, which also showed the highest number of CNVs in total (22.2%).

The discrepancies among the number of CNVs detected from the two software packages is ascribed to the lack of identification of shorter CNVs of the SVS7 univariate approach (here used) (http://doc.goldenhelix.com/). A graphical representation of CNVs obtained by PennCNV and SVS7 software for each chromosome was visualized by HD-CNV software (http://daleylab.org/lab/?page_id=125) and reported in Figure 1. The graph files allow the visualization of the regions where CNVs were observed across samples overlapping consistently.

A total of 1101 CNVRs were mapped with PennCNV software (Table 2). The total length of the sequence covered by the CNVRs was 682 Mb, which corresponded to 27.14% of the bovine autosome genome in the Brown Swiss breed. The percentage of sequence covered by CNVRs by chromosome ranged from 16.59 (BTA 12) to 50.14 (BTA 19). The CNVs identified with SVS7 software were summarized at the population level according to Redon et al. (2006)’s approach, resulting into 277 CNVRs (Table 2). The total length of the sequence covered by the CNVRs was 33.71 Mb (1.35%) of the bovine autosome. The percentage by chromosome of sequence covered by CNVRs ranged from 0.12 (BTA 10) to 3.5 (BTA 12). The highest frequency of CNVRs were identified on BTAs 8 and 4 for PennCNV and SVS7 software, respectively. The consensus performed between the two software generated 150 consensus CNVRs with a total length of 17.1 Mb (0.68 % of the autosomes), as shown in Supplementary File 1.

Table 2. Descriptive statistics for copy number variation regions identified with PennCNV and SVS7 software.

| CNVRs  | Number of events | Mean      | Median      | Total length | Min length | Max length |
|--------|------------------|-----------|-------------|--------------|------------|------------|
| **PennCNV** |                  |           |             |              |            |            |
| Loss   | 220              | 210,454.3 | 148,427.5   | 46,299,963   | 40,754     | 977,685    |
| Gain   | 774              | 596,255.2 | 403,827.3   | 461,501,583  | 45,465     | 3,873,856  |
| Complex| 107              | 1,625,208.1| 1,068,260.7 | 173,897,266  | 179,707    | 6,703,707  |
| Total  | 1101             | 210,454.4 | 148,427.5   | 46,299,963   | 40,754     | 977,685    |
| **SVS7** |                  |           |             |              |            |            |
| Loss   | 185              | 116,378.6 | 61,523.6    | 21,530,044   | 11,314     | 1,440,750  |
| Gain   | 56               | 115,358.1 | 83,498.5    | 6,460,049    | 20,341     | 460,833    |
| Complex| 36               | 158,863.1 | 127,525.5   | 5,719,073    | 21,916     | 770,043    |
| Total  | 277              |           |             |              |            |            |

CNVRs, copy number variation regions; Loss, homozygous or heterozygous deletion; gain, homozygous or heterozygous duplication; complex, CNVRs defined both as deletion and duplication across samples.

Table 3. Comparison between results of this study and results from literature.

| Reference         | Methods of detection | Total CNVRs | Number of samples | Breeds | Length (Mb) | Count | Percentage overlap | Total length (Mb) |
|-------------------|----------------------|-------------|------------------|--------|-------------|-------|--------------------|-------------------|
| Bae et al. (2010) | SNP-based Studies (54k) | 388         | 265              | 1      | 63.1        | 13    | 8.7%               | 4.1               |
| Hou et al. (2011) | SNP-based Studies (54k) | 682         | 521              | 21     | 158.0       | 57    | 38.0%              | 22.4              |
| Jiang et al. (2013) | SNP-based Studies (HD) | 367         | 96               | 1      | 42.7        | 15    | 10.0%              | 0.92              |
| Liu et al. (2010) | CGH-based Studies    | 177         | 90               | 17     | 28.1        | 3     | 2.0%               | 1.3               |
| Fadista et al. (2010) | CGH-based Studies | 304         | 20               | 4      | 22.0        | 4     | 2.7%               | 1.3               |
| Bickhart et al. (2012) | Resequencing-based Studies | 1,265   | 5                 | 3      | 55.6        | 12    | 8.0%               | 2.3               |
| This study        | SNP-based Studies (54k) | 1101        | 210,454.4        | 148,427.5 | 46,299,963 | 40,754 | 977,685            |                   |
identified in our study (405028-4476378 Mb). The comparison between CNVRs here identified with PennCNV software and those detected in the study of Hou et al. (2011) in Brown Swiss cattle, using the same software, provided five common CNVRs on BTAs 2, 9, 12, 14, 18. Table 4 shows a list of the QTL included in these regions (http://www.animalgenome.org/QTLdb/doc/genameversion#UMD_3.1). Some QTL regions enclosed the CNVR but only those that were not further than 1Mb from the beginning and/or end of the CNVR were taken into account. Additionally, also the QTL regions that were included within the CNVR were considered.

Copy number variation regions validation by quantitative polymerase chain reaction

Eleven CNVRs were selected for the validation taking into account the possibility to test with a molecular approach (quantitative PCR) the results obtained by the different in silico approaches we used to call CNVRs; three of which were in common between PennCNV and SVS7 software (a molecular validation of in silico consensus), six and two of which were randomly chosen among the CNVRs identified with the two software, respectively. The validation of the selected regions for each software was the best that could be done taking into account the combination of both the primer design and an adequate number of samples.

Supplementary File 2 reports the primer list for the eleven regions. Ten CNVRs (91%) were confirmed by qPCR experiments. Additionally, the proportions of confirmed positive CNVs in each sample varied from 50% to 100% in each of the confirmed CNVRs; however, the average of false negative rate was equal to 25%. Jiang et al. (2013) reported similar values rates in the Holstein breed.

Copy number variation regions annotation

Supplementary File 3 shows the 252 Ensembl annotated genes, which correspond to 218 gene symbols in the consensus CNVRs. Supplementary File 4 (spreadsheet genes_clustered by DAVID and spreadsheet genes not clustered by DAVID) report the GO and the pathways analyses for 158 genes among those included in Supplementary File 3. The KEGG pathway analysis revealed that these genes are mainly represented in the pathway of immune system. This aspect could be interesting considering that disease resistance, that is a complex trait, might be affected, at least in part, by many polymorphisms related to CNVs in addition to many other genetic factors. However, GO and pathway enrichment should be considered with caution in this context. The number of CNVRs could be biased by the detection strategy that was based to define consensus derived CNVRs. That means that we probably obtained a low false positive calling rate but we might have a higher false negative rate (as also mentioned above). For these reasons we focused our discussion on a few specific genes that might be interesting, according to their potential functional roles and not on the general picture that might be derive from the preliminary GO and pathway enrichment analysis. Among the identified genes, in Table 5 we highlighted those showing differential expression of association with various traits in cattle reported in literature.

In detail, Lewandowska-Sabat et al. (2013) in their in vitro study, highlighted the role of the TREM1 (triggering receptor expressed on myeloid cells 1) signaling pathway in which the FCGR2B [(Fc fragment of IgG, low affinity...
Illb, receptor (CD32)] gene is included. The TREM1, in synergy with the TLR2 (Toll-like receptor 2) pathway, are involved in phagocytosis and production of proinflammatory cytokines, determining optimal host defense during bovine mastitis. The PPAR gene, encoding for the peroxisome proliferator-activated receptor-α, maps in the CNVRs found in this study on BTA5. The role of this gene in the fatty acid metabolism is widely described in literature (Bionaz et al., 2013). The CNVRs detected on BTA12 in our study enclose the KATNAL1 (Katanin p60 subunit A-like 1) and DNAJC15 (DnaJ (Hsp40) homolog, subfamily C, member 15) genes. Zhang et al. (2014) characterized the KATNAL1 gene and found the promoter polymorphism associated with semen traits in Chinese bulls. Moreover, SNPs in the DNAJC15 gene were found to be associated with bovine blastocyst rate by Zhang et al. (2011). The polymorphism in the PTK2 (Protein tyrosine kinase 2) gene, annotated in one of the CNVR on BTA14 was found to be associated with milk production traits in Chinese Holstein (Wang et al., 2013). On the same BTA, the gene encoding for the thyroglobulin (TG) was associated with fat distribution, carcass and meat traits in beef cattle (Bennett et al., 2013). Fernández et al. (2014) also found the association of SNPs in the TG gene with age of puberty in bulls. The STAT3 ([signal transducer and activator of transcription 3 (acute-phase response factor)], the STAT5B (signal transducer and activator of transcription 5B) and the STAT5A (signal transducer and activator of transcription 5A) genes are annotated in the CNVR on BTA19 detected in our study. The main bovine STAT family members STAT3 and STAT5 are involved in prolactin receptor (PRLR) signaling by JAK/STAT pathway (Janus kinase (JAK) and Signal Transducer and Activator of Transcription) that activates the expression of milk protein genes (Zhang et al., 2010). The NPM1 gene (nucleophosmin 1) is located in

### Table 4. Common copy number variation regions in the Brown Swiss breed between Hou et al. (2011) and this study.

| BTA | Start CNVR | End CNVR | Length CNVR | Start CNVR | End CNVR | Length CNVR | Start QTL | End QTL | Trait (ID) |
|-----|------------|----------|-------------|------------|----------|-------------|-----------|---------|------------|
| 2   | 8,788,219  | 9,113,368 | 325,150     | 8,788,219  | 9,040,720 | 252,501     | 9,005,563 | 9,687,863| trans-15: C18:1 fatty acid content (2650) |
| 9   | 4,366,338  | 4,386,831 | 81,494      | 4,086,328  | 4,476,378 | 410,555     | 2,148,415 | 9,159,784| Interval from first to last insemination (5006) |
| 12  | 31,968,562 | 31,795,957| 311,396     | 30,099,199 | 31,555,734| 1,456,535   | 30,751,141| 30,912,533| Longissimus muscle area (11733) |
| 14  | 17,328,950 | 17,457,836| 7,178,987   | 17,322,658 | 17,457,836| 135,178     | 9,884,020 | 19,204,282| Calving ease (maternal) (10585) |
| 18  | 57,396,406 | 57,639,303| 83,938      | 56,369,575 | 58,090,176| 7,725,430   | 16,752,141| 18,440,442| Shear force (20791) |
| 16  | 55,977,394 | 55,928,978| 68,417      | 55,181,080 | 56,364,657| 1,174,577   | 55,181,080 | 60,030,732| Birth index (15200) |
| 17  | 55,777,394 | 56,364,657| 588,263     | 55,777,394 | 56,364,657| 588,263     | 55,181,080 | 60,030,732| Birth index (15200) |

CNVR, copy number variation region; QTL, quantitative trait locus.

### Table 5. List of genes located in identified Brown Swiss copy number variation regions reported in literature.

| BTA | Start | End | Start | End | Ensembl code | Gene symbol |
|-----|-------|-----|-------|-----|--------------|-------------|
| 3   | 795,7960 | 798,3149 | 792,8113 | 794,4607 | ENSBTAG000000021842 | FCGR2B Lewandowska-Sabat et al. (2013) |
| 5   | 116,895,329 | 117,247,824 | 117,151,549 | 117,233,112 | ENSBTAG00000008063 | PPAR α Bionaz et al. (2013) |
| 12  | 304,161,11 | 306,460,942 | 305,198,52 | 305,822,10 | ENSBTAG0000009340 | KATNAL1 Zhang et al. (2014) |
| 12  | 131,796,96 | 132,041,37 | 131,837,34 | 132,663,10 | ENSBTAG00000034785 | DNAJC15 Zhang et al. (2011) |
| 14  | 388,5798 | 401,720 | 387,0893 | 406,5010 | ENSBTAG0000009578 | PTK2 Wang et al. (2013) |
| 14  | 930,0228 | 934,5140 | 926,2251 | 950,8938 | ENSBTAG0000007823 | TG Fernández et al. (2014) |
| 19  | 429,768,59 | 431,025,6 | 430,5660 | 431,3262 | ENSBTAG0000002152 | STAT3 Bennett et al. (2013) |
| 19  | 429,768,59 | 431,025,6 | 429,602,26 | 429,966,71 | ENSBTAG0000001012 | STAT5B Zhang et al. (2010) |
| 19  | 429,768,59 | 431,025,6 | 430,335,97 | 430,407,5 | ENSBTAG0000009496 | STAT5A Zhang et al. (2010) |
| 20  | 288,0532 | 318,9118 | 311,1198 | 312,3880 | ENSBTAG0000001531 | NPM1 Huang et al. (2010) |
| 22  | 590,1940 | 602,439,16 | 600,198,95 | 602,458,6 | ENSBTAG0000001570 | GATA2 Bai et al. (2011) |
| 25  | 602,241 | 938,359 | 72,444 | 775,899 | ENSBTAG0000001974 | LMFI Ren et al. (2011) |
| 26  | 258,298,73 | 259,822,93 | 258,564,75 | 258,655,94 | ENSBTAG0000001771 | ECHS1 Strillacci et al. (2014) |

CNVR, copy number variation region.
the CNVR found on BTA20. This gene encoding a multifunctional nuclear phosphoprotein that plays a crucial role in the control of various aspects of cell growth and homeostasis, is a candidate gene for growth traits in cattle (Huang et al., 2010). The GATA2 (GATA binding protein 2) gene, included in the CNVR on BTA22, is involved in the regulation of trophoblast-specific gene transcription in bovine trophoblast CT-1 cells, as described by Bai et al. (2011). The LMFI (lipase maturation factor 1) gene maps in the CNVR found on BTA25. The mutations of this gene are involved in glyceridemia and hypertriglyceridemia in human and animals, playing an important role in the lipase maturation (Ren et al., 2011). The ECHS1 (enoyl coenzyme A hydratase, short chain, 1) gene, included in the CNVR on BTA26, was associated with the conjugated linoleic and vaccenic acids in milk in a quantitative trait loci mapping study of Strillacci et al. (2014) in the Italian Brown Swiss cattle breed.

Conclusions

In this study, the first on this breed on a such a large number of individuals, we detected CNVs in the Italian Brown Swiss cattle population based on whole genome SNP genotyping data, using two software packages (PennCNV and SVST), with the aim to reduce the high error rate commonly recognised in copy number discovery.

The detection from different software for the same CNV call increases confidence in the data giving clearer indications of the dimensions of the CNV identified. CNVRs identified by PennCNV software overlapped in part with the SVST data, which emphasised the diversities and the shared features of the two detection methods. The GO and pathway analyses here conducted identified genes that have shown differential expression of association with production traits, carcass and meat traits, reproduction traits and growth traits in cattle. The results enrich the bovine CNVs map providing new information for association studies with economic and health-related traits of interest.

References

Bae, J.S., Cheong, H.S., Kim, L.H., Gung, S.N., Park, T.J., Chun, J.Y., Kim, J.Y., Pasaje, C.F., Lee, J.S., Shin, H.D., 2010. Identification of copy number variations and common deletion polymorphisms in cattle. BMC Genomics 11:232.

Bai, H., Sakurai, T., Someya, Y., Konno, T., Ieda, A., Ayagi, Y., Imakawa, K., 2011. Regulation of trophoblast-specific factors by GATA2 and GATA3 in bovine trophoblast CT-1 cells. J. Reprod. Dev. 57:518-525.

Bennett, G.L., Shackelford, S.D., Wheeler, T.L., King, D.A., Casas, E., Smith, T.P.L., 2013. Selection for genetic markers in beef cattle reveals complex associations of thyroglobulin and casein-s1 with carcass and meat traits. J. Anim. Sci. 91:565-571.

Bickhart, D.M., Hou, Y., Schroeder, S.G., Alkan, C., Cardone, M.F., Matukumalli, L.K., Song, J., Schnabel, R.D., Ventura, M., Taylor, J.F., Garcia, J.F., Van Tassell, C.P., Sonstegard, T.S., Eichler, E.E., Liu, G.E., 2012. Copy number variation of individual cattle genomes using next-generation sequencing. Genome Res. 22:778-790.

Bionaz, M., Chen, S., Khan, M.J., Loor, J.J., 2013. Functional role of PPARs in ruminants: potential targets for fine-tuning metabolism during growth and lactation. PPAR Res. 2013:684159.

Diskin, S.I., Li M., Hou, C., Yang, S., Glessner, J., Hakonarson, H., Bucan, M., Maris, J.M., Wang, K., 2008. Adjustment of genomic waves in signal intensities from whole-genome SNP genotyping platforms. Nucl. Acid. Res. 36:e126.

Fadista, J., Thomesen, B., Holm, L.E., Bendixen, B.M.C., 2010. Copy number variation in the bovine genome. Genomics 11:284.

Fernández, M., Goszczyński, D., Prando, A., Peral-García, P., Baldo, A., Giovambattista, G., Liron, J., 2014. Assessing the association of single nucleotide polymorphisms in thyroglobulin gene with age of puberty in bulls. J. Anim. Sci. Tech. 56:17.

Fontanesi, L., Beretti, F., Riggio, V., Gómez-Escañ, E., Dall’Olio, S., Davoli, R., Russo, V., Portolano, B., 2008. Copy number variation and missense mutations of the agouti signaling protein (ASIP) gene in goat breeds with different coat colors. Cytogenet. Genome Res. 126:333-347.

Huang, Y.Z., Zhang, E.P., Hou, Y., Bickhart, D.M., Alkan, C., DeNise, S., Wiggans, G.R., Sonstegard, T.S., Van Tassell, C.P., Liu, G.E., 2012. Fine mapping of copy number variations on two cattle genome assemblies using high density SNP array. BMC Genomics 13:376.

Hou, Y., Liu, G.E., Bickhart, D.M., Cardone, M.F., Wang, K., Kim, E., Matukumalli, L.K., Ventura, M., Song, J., VanRaden, P.M., Sonstegard, T.S., Van Tassell, C.P., 2011. Genomic characteristics of cattle copy number variations. BMC Genomics 12:127.

Jiang, L., Jiang, J., Yang, J., Liu, X., Wang, J., Wang, H., Ding, X., Liu, J., Zhang, Q., 2013. Genome-wide detection of copy number variations using high-density SNP genotyping platforms in Holsteins. BMC Genomics 14:131.

Lewandowska-Sabar, A.M., Boman, G.M., Downing, A., Talbot, R., Storset, A.K., Olsaker, I., 2013. The early phase transcriptome of bovine monocyte-derived macrophages infected with Staphylococcus aureus in vitro. BMC Genomics 14:891.

Liu, G.E., Hou, Y., Zhu, B., Cardone, M.F., Jiang, L., Cellamare, A., Mitra, A., Alexander, L.J., Coutinho, L.L., Dell’Aquila, M.E., Gasbarre, L.C., Lacalandra, G., Li, R.W., Matukumalli, L.K., Nonneman, D., Regitano, L.C., Smith, T.P., Song, J., Sonstegard, T.S., Van Tassell, C.P., Ventura M., Eichler, E.E., McDaniel, T.G., Keefe, J.W., 2010. Analysis of copy number variations among diverse cattle breeds. Genome Res. 20:693-703.

Matukumalli, L.K., Lawley, C.T., Schnabel, R.D., Taylor, J.F., Allan, M.F., Heaton, M.P., Connell, J., Moore, S.S., Smith, T.P., Sonstegard, T.S., Van Tassell, C.P., 2009. Development and characterization of a high density SNP genotyping assay for cattle. PLoS One 4:e5350.

Mills, R.E., Walker, R., Stewart, C., Handsaker, R.E., Chen, K., Alkan, C., Abyzov, A., Yoon, S.C., Ye, K., Cheetham, R.K., Chinnwalla, A., Conrad, D.F., Fu, Y., Grubert, F., Hajirasouliha, I., Hormozdiari, E., Iakoucheva, L.M., Iqbal, Z., Kang, S., Kidd, J.M., Konkel, M.K., Korn, J., Kur throw, E., Liao, Y., Li, Y., Lin C.Y., Luo, R., Mu, X.J., Nemesh, J., Peckham, H.E., Rausch, T., Scally, A., Shi, X., Stromberg, M.P., Stütz, A.M., Urban, A.E., Walker, J.A., Wu, J., Zhang, Y., Zhang, Z.D., Batzer, M.A., Ding, L., Marth, G.T., McVean, G., Sebat, J., Snyder, M., Wang, J., Ye, K., Eichler, E.E., Gerstein, M.B., Hurles, M.E., Lee, C., McCarthy, S.A., Kerbel, J.O., 1000 Genomes Project, 2011. Mapping copy number variation by population-scale genome sequencing. Nature 470:59-65.

Meyers, S.N., McDaniel, T.G., Swist, S.L., Marron, B.M., Steffen, D.J., O’Toole, D., O’Connell, J. R., Beever, J.E., Sonstegard, T.S., Smith, T.P., 2010. A deletion mutation study of common deletion polymorphisms in cattle. BMC Genomics 11:232.

Sonstegard, T.S., Van Tassell, C.P., Eichler, E.E., McDaniel, T.G., Keefe, J.W., 2010. Analysis of copy number variations among diverse cattle breeds. Genome Res. 20:693-703.

Wang, K., 2008. Adjustment of genomic waves in signal intensities from whole-genome SNP genotyping platforms. Nucl. Acid. Res. 36:e126.

Xi, Z.J., Hua, Y., Ma, L., Yan, X.Y., Ren, G., Lei, C.Z., Fang, X.T., Wang, J.Q., 2010. Novel 12-bp deletion in the coding region of the bovine PMI gene affects growth traits. J. Appl. Genet. 51:199-202.

Yamamoto, A., DeNise, S., Wiggans, G.R., Sonstegard, T.S., Van Tassell, C.P., Liu, G.E., 2012. Fine mapping of copy number variations on two cattle genome assemblies using high density SNP array. BMC Genomics 13:376.
in bovine SLC4A2 is associated with osteopetrosis in Red Angus cattle. BMC Genomics 11:337.

Pinto, D., Darvishi, K., Shi, X., Rajan, D., Rigler, D., Fitzgerald, T., Lionel, A.C., Thiruvahindrapuram, B., Macdonald, J.R., Mills, R., Prasad, A., Noonan, K., Gribble, S., Prigmore, E., Donahoe, P.K., Smith, R.S., Park, J.H., Hurles, M.E., Carter N.P., Lee C, Scherer S.W., Feuk L., 2011. Comprehensive assessment of array-based platforms and calling algorithms for detection of copy number variants. Nat. Biotechnol. 29:512-520.

Quinlan, A.R., Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841-842.

Redon, R., Ishikawa, S., Fitch, K.R., Feuk, L., Perry, G.H., Andrews, T.D., Fiegler, H., Shapero, M. H., Carson, A.R., Chen, W., 2006. Global variation in copy number in the human genome. Nature 444:444-454.

Ren, G., Liu, J.X., Li, F., Lan, X.Y., Li, M.J., Zhang, Z.Y., Chen, H., 2011. A novel missense mutation of bovine lipase maturation factor 1 (LMF1) gene and its association with growth traits. African J. Biotech. 10:7562-7566.

Rosengren Pielberg, G., Golovko, A., Sundstrom, E., Curik, I., Lennartsson, J., Seltenhammer, M.H., Druml, T., Binns, M., Fitzsimmons, C., Lindgren, G., Sandberg, K., Baumung, R., Vetterlein, M., Strömberg, S., Grabherr, M., Wade, C., Lindblad-Toh, K., Pontén, F., Hedén, C.H., Sölöker, J., Andersson, L., 2008. A cis-acting regulatory mutation causes premature hair graying and susceptibility to melanoma in the horse. Nat. Genet. 40:1004-1009.

Seroussi, E., Glick, G., Shirak, A., Yakobson, E., Weller, J.I., Ezra, E., Zeron, Y., 2010. Analysis of copy loss and gain variations in Holstein cattle autosomes using BeadChip SNPs. BMC Genomics 11:673.

Strillacci, M.G., Frigo, E., Canavesi, F., Ungar, Y., Schiavini, F., Zaniboni, L., Reghizzi, L., Cozzi, M.C., Samorò, A.B., Kashi, Y., Shimoni, E., Tal-Stein, R., Soller, M., Lipkin, E., Bagnato, A., 2014. Quantitative trait loci mapping for conjugated linoleic acid, vaccenic acid and Δ(9) -desaturase in Italian Brown Swiss dairy cattle using selective DNA pooling. Anim. Genet. 45:485-499.

Tsuang, D.W., Millard, S.P., Ely, B., Chi, P., Wang, K., Raskind, W.H., Kim, S., Brkanac, Z., Yu, C.E., 2010. The effect of algorithms on copy number variant detection. PLoS One 5:e14456.

Xu, L., Hou, Y., Bickhart, D.M., Song, J., Liu, G.E., 2013. Comparative Analysis of CNV calling algorithms: literature survey and a case study using bovine high-density SNP Data. Microarrays 2:171-185.

Wain, L.V., Armour, J.A.L., Tobin, M.D., 2009. Genomic copy number variation, human health, and disease. Lancet 374:340-350.

Wang, H., Jiang, L., Liu, X., Yang, J., Wei, J., Xu, J., Zhang, Q., Liu, J.F., 2013. A post-GWAS replication study confirming the PTK2 gene associated with milk production traits in Chinese Holstein. PLoS One 8:e83625.

Winchester, L., Yau, C., Ragoussis, J., 2009. CNVdetection methods for SNP arrays. Brief Funct. Genomic Proteomic 8:353-366.

Wright, D., Boije, H., Meadows, J.R., Bedt, B., Gourichon, D., Vieaud, A., Tixier-Boichard, M., Rubin, C.J, Imsland, F., Hallböök, F., Andersson, L., 2009. Copy number variation in intron 1 of SOX5 causes the Pea-comb phenotype in chickens. PLoS Genet. 5:e1000512.

Zhang, B., Peñagaricano, F., Driver, A., Chen, H., Khatib, H., 2011. Differential expression of heat shock protein genes and their splice variants in bovine preimplantation embryos. J. Dairy Sci. 94:4174-4182.

Zhang, F., Huang, J., Li, Q., Ju, Z., Li, J., Shi, F., Zhong, J., Wang, C., 2010. Novel single nucleotide polymorphisms (SNPs) of the bovine STAT4 gene and their associations with production traits in Chinese Holstein cattle. African J. Biotech. 9:4003-4006.

Zhang, X., Wang, C., Zhang, Y., Ju, Z., Qi, C., Wang, X., Huang, J., Zhang, S., Li, J., Zhong, J., Shi, F., 2014. Association between an alternative promoter polymorphism and sperm deformity rate is due to modulation of the expression of KATNAL1 transcripts in Chinese Holstein bulls. Anim. Genet. 45:641-651.