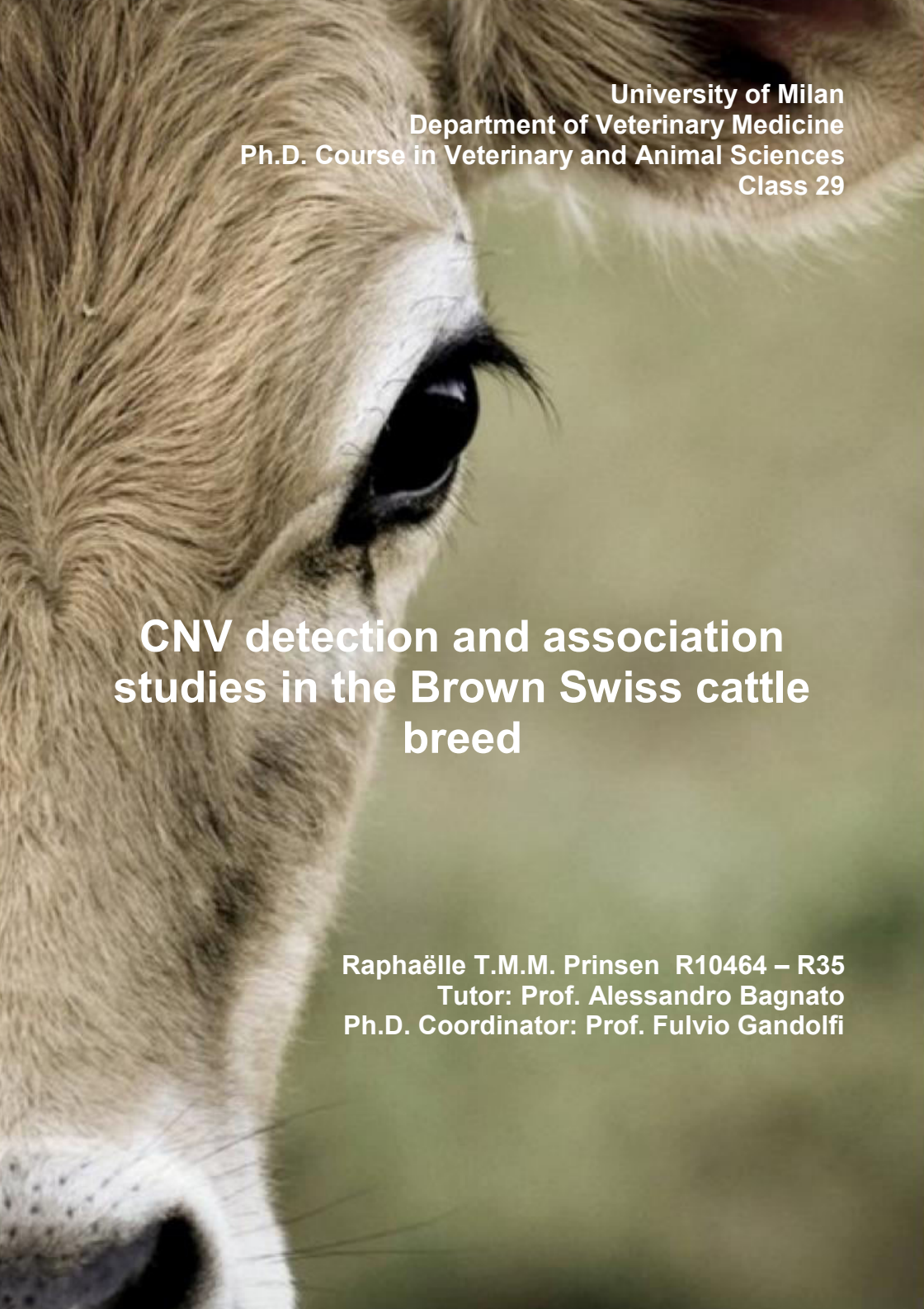
University of Milan
Department of Veterinary Medicine
Ph.D. Course in Veterinary and Animal Sciences
Class 29



CNV detection and association studies in the Brown Swiss cattle breed

Raphaëlle T.M.M. Prinsen R10464 – R35
Tutor: Prof. Alessandro Bagnato
Ph.D. Coordinator: Prof. Fulvio Gandolfi

CNV detection and association studies in the Brown Swiss cattle breed

Raphaëlle T.M.M. Prinsen

To Andrea, my beloved husband, who gave me endless love, support, encouragement and endurance to complete this hard work.

… and to our little baby: you are the most beautiful gift I could ever imagine.

Contents

Sintesi – Italiano

Gli scopi, i materiali, i metodi usati, i risultati e le conclusioni dei tre studi sono organizzati in tre capitoli. La sintesi generale dei tre studi e quindi divisa in base a questi tre capitoli.

Capitolo 1

La determinazione dei "copy number variants" (CNV) è fondamentale per la valutazione dei tratti genomici in diverse specie in quanto rappresentano una fonte principale della variabilità genetica, influenzando l'espressione genica, la variabilità fenotipica, la adattabilità e la predisposizione all'insorgenza di malattie. Lo scopo di questo studio è stato quello di ottenere una mappa genomica di CNV utilizzando i dati ottenuti dall'Illumina Bovine SNP50 BeadChip di 651 tori di razza Bruna Italiana. Per l'identificazione dei CNV e delle regioni CNV (CNVR) sono stati usati i software PennCNV e SVS7 (Golden Helix).

Sono stati identificati un totale di 5,099 e 1,289 CNVs con i software PennCNV ed SVS7 rispettivamente. Questi CNV sono stati raggruppati a livello di popolazione in 1,101 (220 delezioni, 774 duplicazioni e 107 complex) e 277 (185 delezioni, 56 duplicazioni e 36 complex) CNVR. Dieci dei CNVR selezionati sono stati validati sperimentalmente attraverso qPCR. La GO e la pathway analysis effettuate hanno identificato i geni (corretti per la false discovery rate) localizzati nelle CNVR e correlati a diversi processi biologici, componenti cellulari, funzioni metaboliche e vie metaboliche. Tra questi, sono stati identificati i geni *FCGR2B*, *PPARalpha*, *KATNAL1*, *DNAJC15*, *PTK2*, *TG*, *STAT family*, *NPM1*, *GATA2*, *LMF1* e *ECHS1*, già noti in letteratura, per la loro associazione con diversi caratteri quantitativi nei bovinia. Sebbene ci sia una variabilità nell'identificazione dei CNVR attraverso l'utilizzo di diversi metodi e piattaforme, questo studio ha permesso l'identificazione dei CNVR nella Bruna Italiana, sovrapponendo quelli già identificati in altre razze e identificandone

dei nuovi, producendo quindi nuove conoscenze per gli studi di associazione con caratteri quantitativi di interesse nei bovini.

Capitolo 2

Scoprire variazioni genetiche come i Copy Number Variants (CNVs) nei bovini, fornisce l'opportinità di studiare la loro associazione con caratteri quantitativi. I CNVs sono sequenze di DNA di lunghezza 50 bp fino a diverse Mb, che possono variare in numero di copie rispetto ad un genoma di riferimento. Lo scopo di questo studio è stato quello di identificare i CNVs in 1,410 campioni di razza Bruna Svizzera usando informazioni derivanti dall' Illumina Bovine HD SNP chip, che include 777,962 SNPs. Dopo uno stringente controllo di qualità, i CNVs sono stati identificati con i software Golden Helix SVS 8.3.1 (SVS) e PennCNV e sono stati raggruppati in regioni CNV (CNVRs) a livello di popolazione (i.e. CNVs sovrapposti) utilizzando il software BEDTools. I CNVR comuni ai due software sono stati definiti come regioni consensus. I geni all'interno delle CNVR consensus sono stati annotati con un'analisi GO utilizzando DAVID Bioinformatics Resources 6.7. Per poter validare i risultati, sono state eseguite PCR quantitative su 15 CNVR selezionate.

Con il software SVS sono stati identificati 25,030 CNVs successivamente raggruppati in 398 CNVR, che comprendevano 30 duplicazioni, 344 delezioni e 24 complex CNVR (che contenevano sia duplicazioni che delezioni) coprendo il 3.92% del genoma bovino. Il software PennCNV ha identificato 62,341 CNV, corrispondenti a 5,578 CNVRs che comprendevano 2,638 duplicazioni, 2,404 delezioni e 537 complex CNVR, coprendo il 7.68% del genoma bovino. La lunghezza di queste CNVR variava da 1,244 bp a 1,381,355 bp. Sono state trovate 563 CNVR consensus che coprivano il 2.29% del UMD 3.1 bovine genome assembly. Di queste, 24 erano duplicazioni, 300 erano delezioni e 239 erano CNVR complex. Un totale di 775 official gene IDs sono stati annotati nelle CNVR consensus. Tra i 537 geni con informazioni funzionali, la GO e la pathway analysis è stata riportata per quelli

che clusterizzavano con un p-value < 0.05. Le PCR quantitative hanno validato con successo 14 delle 15 CNVR selezionate.

Il risultato di questo studio è una prima analisi genomica integrale della razza Bruna Svizzera basata sull'Illumina Bovine HD SNP chip su un numero cosi grande di animali che arricchisce la mappa CNV nel genoma bovino. I risultati forniscono inoltre informazioni preziose per successivi studi sui CNV. Infine, i risultati della mappa CNVR sono informativi per i caratteri funzionali, produttivi e sanitari considerati nei programmi di selezione nella razza Bruna Svizzera.

Capitolo 3

I Copy Number Variations (CNV) possono essere usati negli studi di associazione per rivelare la base genetica della variazione fenotipica di caratteri quantitativi. I CNV sono sequenze di DNA di 50 bp fino a qualche Mb, che possono variare in numero di copie rispetto ad un genoma di riferimento. Fino ad oggi, nessuno studio di associazione genome-wide (GWAS) con i CNV e caratteri quantitavivi è stato descritto in una popolazione cosi ampia (cioè di 1,116 campioni) della razza bovina Bruna Svizzera. Lo scopo di questo studio era quello di eseguire delle GWAS utilizzando i CNV precedentemente mappati, con caratteri funzionali, produttivi e sanitari al fine di valutare il loro impatto sull'allevamento e sulla selezione. Gli studi di associazione con i CNV sono stati effettuati con il software Golden Helix SVS 8.4.4 utilizzando un correlation-trend test model. I geni all'interno dei CNV significativamente associati per ogni carattere sono stati annotati con un'analisi GO usando DAVID Bioinformatics Resources 6.7. Sono stati identificati 56 CNV significativamente associati con uno o più degli otto caratteri valutati. I segnali di associazione più forti erano dati da tre CNV sul cromosoma 12 per il carattere grasso. I CNV associati si sovrappongono con 23 geni diversi, annotati sul *Bos taurus* genome assembly (UMD3.1).

Abstract – English

The aims, material and methods, results and conclusions of the three studies are organized in three different chapters. The general abstract is therefore divided according to these chapters.

Chapter 1

The determination of copy number variation (CNV) is very important for the evaluation of genomic traits in several species because they are a major source for the genetic variation, influencing gene expression, phenotypic variation, adaptation and the development of diseases. The aim of this study was to obtain a CNV genome map using the Illumina Bovine SNP50 BeadChip data of 651 bulls of the Italian Brown Swiss breed. PennCNV and SVS7 (Golden Helix) software were used for the detection of the CNVs and Copy Number Variation Regions (CNVRs). A total of 5,099 and 1,289 CNVs were identified with PennCNV and SVS7 software, respectively. These were grouped at the population level into 1101 (220losses, 774 gains, 107 complex) and 277 (185losses, 56 gains and 36 complex) CNVRs. Ten of the selected CNVRs were experimentally validated with a qPCR experiment. The GO and pathway analyses were conducted and they identified genes (false discovery rate corrected) in the CNVR related to biological processes, cellular component, molecular function and metabolic pathways. Among those, we found the *FCGR2B*, *PPARalpha*, *KATNAL1*, *DNAJC15*, *PTK2*, *TG*, *STAT family*, *NPM1*, *GATA2*, *LMF1*, *ECHS1* genes, already known in literature because of their association with various traits in cattle. Although there is variability in the CNVRs detection across methods and platforms, this study allowed the identification of CNVRs in Italian Brown Swiss, overlapping those already detected in other breeds and finding additional ones, thus producing new knowledge for association studies with traits of interest in cattle.

Chapter 2

Detecting genetic variation such as Copy Number Variants (CNVs) in cattle provides the opportunity to study their association with quantitative traits. CNVs are DNA sequences of 50 bp up to several Mb long, which can vary in copy number in comparison with a reference genome. The aim of this study was to investigate CNVs in 1,410 samples of the Brown Swiss cattle breed using Illumina Bovine HD SNP chip information, which includes 777,962 SNPs. After stringent quality control, CNVs were called with the Golden Helix SVS 8.3.1 (SVS) and PennCNV software and were summarized to CNV regions (CNVRs) at a population level (i.e. overlapping CNVs), using BEDTools. Additionally, common CNVRs between the two software were set as consensus regions. Genes within consensus CNVRs were annotated with a GO analysis using the DAVID Bioinformatics Resources 6.7. In order to validate these results, quantitative PCRs were executed on 15 selected CNVRs.

The SVS software identified 25,030 CNVs summarized to 398 CNVRs, which comprised 30 gains, 344 losses and 24 complex CNVRs (i.e. containing both losses and gains), covering 3.92% of the bovine genome. The PennCNV software identified 6,2341 CNVs summarized to 5,578 CNVRs, which comprised 2,638 gains, 2,404 losses and 537 complex CNVRs, covering 7.68% of the bovine genome. The length of these CNVRs ranged from 1,244 bp to 1,381,355 bp. A total of 563 consensus CNVRs were found covering 2.29 % of the UMD 3.1 bovine genome assembly. Of these, 24 were gains, 300 were losses and 239 were complex CNVRs. A total of 775 official gene IDs were annotated in the consensus CNVRs. Among the 537 genes with functional information, the GO and pathway analysis was reported for those who clustered with a p-value < 0.05. The quantitative PCRs successfully validated 14 (93.33%) of the selected CNVRs.

The result of this study is the first comprehensive genomic analysis of the Brown Swiss breed based on the Illumina Bovine HD SNP chip on such a large number of animals that enriches the CNV map in the bovine genome. These findings also provide valuable information for further CNV studies. Finally, the results of the CNVR map delivers new information for functional, health and productive traits considered in selection programs of the Brown Swiss breed.

Chapter 3

Copy Number Variation (CNV) can be used in association studies to disclose genetic basis of quantitative traits phenotypic variation. 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. Up to date, no genome-wide association study (GWAS) with CNVs and quantitative traits in such a large Brown Swiss population (i.e. with 1,116 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).

Abbreviations

- aCGH (Comparative genomic hybridization array)
- ANARB (National Association of Italian Brown Swiss breeders)
- BAF (B allele frequency)
- CNAM (Copy Number Analysis Module)
- CNV (Copy number variant)
- CNVR (Copy number variant region)
- DEBV (deregressed estimated breeding value)
- DLRS (derivative log ratio spread)
- FDR (false discovery rate)
- FUA (fore udder attachment)
- FY (fat yield)
- GO (Gene ontology)
- GWAS (genome-wide association study)
- HMM (hidden markov model)
- Kb (kilobase)
- KEGG (Kyoto Encyclopedia of Genes and Genomes)
- LRR (log R ratio)
- Mb (megabase)
- MGB (minor groove binder)
- MY (milk yield)
- NCBI (National Center for Biotechnology Information)
- NGS (next generation sequencing)
- PCA (Principal component analysis)
- PY (protein yield)
- qPCR (quantitative polymerase chain reaction)
- QTL (quantitative trait loci)
- RUH (rear udder hight)
- RUW (rear udder width)
- SCS (somatic cell score)
- SNP (single nucleotide polymorphism)
- SV (single variation)
- UDD (overall udder score)

General Introduction

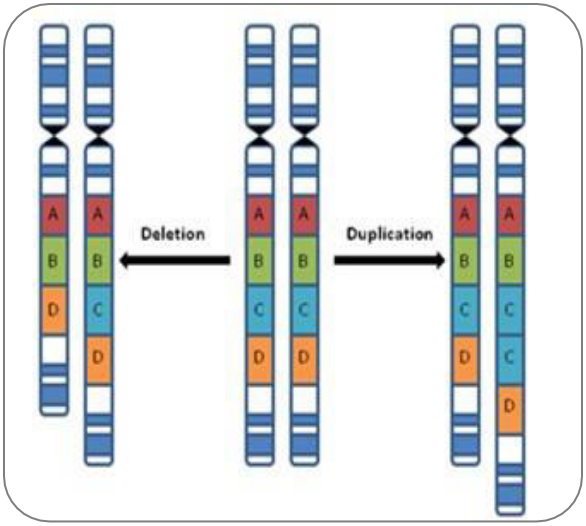


Figure 1 Graphical representation of a CNV. Here a "C/C" genotype is represented in light blue. A duplication "C/CC" represent a gain, while a deletion "C/-" represents a loss.

Copy Number Variants

Genomic sequences were assumed to be practically always present in two copies in a genome. However, certain segments of DNA, can vary in copy number. Copy number variants (CNVs) are DNA segments of 0.5 kilobases (Kb) up to several megabases (Mb) long, which vary in copy number in comparison to a reference genome (Jiang et al., 2013). According to the number of copies of the segment, CNVs can be classified in deletions (or losses), duplications (or gains) or both (complex CNVs).

Even if CNVs are much less abundant compared to SNPs, as they are much longer, inter-individual variability based on CNVs is much higher (Redon et al., 2006).

Iafrate et al. (2004), and Redon et al. (2006) published the first comprehensive human CNV map. This type of genetic structural variant covers approximately 4.8% of the human genome (Zarrei et al., 2015) and 4.6% of the bovine genome (Hou et al., 2011). CNVs therefore contribute to explain the genetic diversity and offer similar structural genomic knowledge as single nucleotide polymorphisms (SNPs), since they can impact gene expression and consequently the phenotypes deriving from them, as well as environmental adaptability and disease susceptibility (Wang et al., 2012). Basically, CNVs contribute to individuals' uniqueness (Varki et al., 2008).

CNVs identification and CNV studies

There are several techniques to identify CNVs in the genome. In particular, SNP genotyping microarrays, next-generation sequencing (NGS) technologies and comparative genomic hybridization arrays (aCGH) can be employed. Each of these techniques has its advantages and disadvantages, which makes the comparison of the results not easy. Neither of the array-based assays provides information on the localization of additional copies and orientation of a single variation (SV). Whenever the genotyping process is used, specific software tools are needed to identify the CNVs after the genotyping process. Winchester et al. (2009), Pinto et al. (2011) and Tsuang et al. (2010) suggested the use of a minimum of two algorithms for the CNV detection since this reduces eventual false discovery rates.

There are different CNV detection algorithms that can be used depending on the genotyping platform that has been employed in the previous step. What regards Illumina data, three Hidden Markov Model (HMM) based algorithms are nowadays available: the QuantiSNP (Colella et al., 2007), the PennCNV (Wang et al., 2007) and the R package genoCN (Sun et al., 2009) algorithms. Besides, also other algorithms such as the Copy Number Analysis Module (CNAM) of the SVS8 software (Golden Helix) can be used.

The identification of CNVs by genotyping arrays in different cattle breeds could have been accomplished thanks to the progress of the SNP genotyping arrays. Together with this progress, some studies identified CNVs in different cattle breeds based on medium and high-density SNP arrays and with comparative genomic hybridization array experiments. Resolution and accurate breakpoint definition of the identified CNV regions (CNVRs) is dependent on the probe spacing on the array (i.e. SNP spacing on the reference genome). CNVs have been identified in different cattle breeds such as the African, the Indicine, and the Taurine breeds (Matukumalli et al., 2009). Furthermore, Bae et al. (2010) and Fadista et al. (2010) created two CNV maps of the bovine genome using SNP genotyping and CGH arrays. Genome wide studies of CNVs in other primates such as the chimpanzee (Perry et al., 2008) and rhesus macaque (Gibbs et al. 2007), as well as in model organisms such as Drosophila melanogaster (Dopman and Hartl, 2007), mouse (Adams et al., 2005), rat (Guryev et al., 2008) and livestock e.g. cattle (Liu et al., 2010; Hou et al., 2011), goat (Fontanesi et al., 2010), sheep (Fontanesi et al., 2011) and horse

(Dupuis et al., 2012) followed. Clop et al. (2012) published a review on CNV in the genomes of domestic animals including mammals and avian species. Since then CNV studies have advanced from mere discovery to studying population genetics of structural variants and assessing the influence of CNVs on phenotypic variation.

CNVs and Phenotypes

There are different ways by which CNVs can alter phenotypes including gene dosage effects. For instance, a gene in a hemizygous state causes haplo-insufficiency, or a duplication produces too much gene product, whereas a loss of the gene function can occur due to disrupted open reading frames. There can also be a loss or disruption of regulatory elements, or an unmasking of recessive mutants by deletions, or else, duplications that can rescue phenotypes and cause incomplete penetrance, or even new regulatory environments can be created due to translocations and inversions (Beckmann et al., 2007).

Phenotypic consequences of CNVs depend on their size and the kind of genetic elements they contain (Lee and Scherer, 2010). Consequently CNVs can also play a crucial role in an individual's susceptibility to several complex diseases and common disorders (Estivil and Armengol, 2007; Wain et al. 2009; Lee and Scherer, 2010; and Almal and Padh, 2011). Some famous examples include Crohn's disease, susceptibility to HIV/AIDS and Kawasaki disease, autism, bipolar disorder, schizophrenia, age-related macular degeneration and cancer. Other studies demonstrated that also the global CNV load can have phenotypic effects. For example, Dauber et al. (2011) showed that individuals with short stature have increased odds to carry a higher

overall burden of low-frequency deletions. A finding the authors observed in a clinical cohort but were able to confirm in a population based design, suggests that CNVs might contribute to genetic variation in stature in the general population and thus might account for part of the missing heritability of human height.

CNVs and Genome-Wide Association Studies

With the availavility of single nucleotide polymorphism (SNP) genotyping arrays, the genome-wide association studies (GWAS) have been used to establish the genetic influence on quantitative traits. CNVs are an alternative genomic marker to SNPs, that can be used in GWASs and that can be identified from SNP chips themselves as described previously.

Genome-wide association studies (GWASs) are commonly used to uncover genetic variants, like CNVs, in livestock species, such as cattle, to verify associations between genetic variants and phenotypes of interest. Several GWASs underlined the role of CNVs in determining certain phenotypes in different species: for example CNVs in intron 1 of the SOX5 gene triggers the pea-comb phenotype in chickens (Wright et al., 2009), CNVs in the STX17 gene cause premature hair greying and vulnerability to melanoma in horses (Rosengren et al., 2008) and CNVs in the ASIP gene produce different coat colors in goats (Fontanesi et al., 2009). Meyers et al. (2010) identified an association among a CNV in a loss state in the SLC4A2 gene and osteoporosis in Red Angus cattle. In 2011, a GWAS study in Israeli Holstein revealed a CNV significantly associated with their fertility (Glick et al., 2011). CNVs

have also been associated with resistance and susceptibility to gastrointestinal nematodes in Angus cattle (Hou et al., 2011) and residual feed intake in the Holstein cattle breed (Hou et al., 2012).

Aims

A whole-genome CNV map for the Italian Brown Swiss cattle breed in a large population dataset was still not available. Therefore, the first aim 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.

The second aim was be to investigate genomic structural variation in cattle using dense SNP information in more than 1000 samples of the Italian and Swiss Brown Swiss breed genotyped on Illumina HD Bovine BeadChips.

CNVs' potential influence on economically relevant traits in the Brown Swiss dairy cattle breed had to be fully investigated. Therefore, a third objective was to evaluate and perform a GWAS with CNVs' and deregressed estimated breeding values of productive, functional and health traits.

The different aims are all focused on the Brown Swiss cattle breed. This breed is bred both for its milk of high quality (with high cheese yields due to the low frequency of the "A" allele of the K-casein, in comparison to other cattle breeds (http://www.anarb.it/ and http://www.anarb.it/dossier_qualita_latte/Testimonianze_Scientifiche/023_Summer_Mende2016.pdf)) and for its meat production. Furthermore, this breed has a good environmental adaptability to the most varying

agricultural and climatic conditions - both in the mountains and on the plain. The typical rusticity of the breed, together with its good production attitude and its resistance to thermal stress, have lead its spread all over many European and American countries, with the differentiation of different genetic groups in relation to various environmental conditions.

References

Adams, D.J., Dermitzakis, E.T., Cox, T., Smith, J., Davies, R., Banerjee, R., Bonfield, J., Mullikin, J.C., Chung, Y.J., Rogers, J., Bradley, A., 2005. Complex haplotypes, copy number polymorphisms and coding variation in two recently divergent mouse strains. Nat Genet. 37(5):532-6.

Almal, S.H., Padh, H., 2011. Implications of gene copy-number variation in health and diseases. J Hum Genet. 57(1):6-13.

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.

Beckmann, J.S., Estivill, X., Antonarakis, S.E., 2007. Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability. Nature Reviews Genetics. 8: 639-646.

Colella, S., Yau, C., Taylor, J.M., Mirza, G., Butler, H., Clouston, P., Bassett, A.S., Seller, A., Holmes, C.C., Ragoussis, J., 2007. QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. Nucleic Acids Res. 35(6): 2013-25.

Clop, A.,Vidal, O., Amills, M., 2012. Copy number variation in the genomes of domestic animals. Anim Genet. 43, 503-17.

Dauber, A., Yu, Y., Turchin, M.C., Chiang, C.W., Meng, Y.A., Demerath, E.W., Patel, S.R., Rich, S.S., Rotter, J.I., Schreiner, P.J., Wilson, J.G., Shen, Y., Wu, B.L., Hirschhorn, J.N., 2011. Genome-wide association of copy-number variation reveals an association between short stature and the presence of low-frequency genomic deletions. Am J Hum Genet. 89(6):751-9.

Dopman, E.B., Hartl, D.L., 2007. A portrait of copy-number polymorphism in Drosophila melanogaster. Proc Natl Acad Sci USA. 104(50): 19920-5.

Dupuis, M.C., Zhang, Z., Durkin, K., Charlier, C., Lekeux, P., Georges, M., 2013. Detection of copy number variants in the horse genome and

examination of their association with recurrent laryngeal neuropathy. Anim Genet. 44(2): 206-8.

Estivill, X., Armengol, L., 2007. Copy number variants and common disorders: filling the gaps and exploring complexity in genome-wide association studies. PLoS Genet. 3(10):1787-99.

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

Fontanesi, L., Beretti, F., Martelli, PL., Colombo, M., Dall'Olio, S., Occidente, M., Portolano, B., Casadio, R., Matassino, D., Russo, V., 2011. A first comparative map of copy number variations in the sheep genome. Genomics. 97(3): 158 – 165.

Fontanesi, L., Beretti, F., Riggio, V., Gómez González, E., Dall'Olio, S., Davoli, R., Russo, V., Portolano, B., 2009. 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.

Fontanesi, L., Martelli, P.L., Beretti, F., Riggio, V., Dall'Olio, S., Colombo, M., Casadio, R., Russo, V., Portolano, B., 2010. An initial comparative map of copy number variations in the goat (Capra hircus) genome. BMC Genomics. 11:639.

Gibbs, R.A., Rogers, J., Katze, M.G., Bumgarner, R., Weinstock, G.M., Mardis, E.R., Remington, K.A., Strausberg, R.L., Venter, J.C., Wilson, R.K., Batzer, M.A., Bustamante, C.D., Eichler, E.E., Hahn, M.W., Hardison, R.C., Makova, K.D., Miller, W., Milosavljevic, A., Palermo, R.E., Siepel, A., Sikela, J.M., Attaway, T., Bell, S., Bernard, K.E., Buhay, C.J., Chandrabose, M.N., Dao, M., Davis, C., Delehaunty, K.D., Ding, Y., Dinh, H.H., Dugan-Rocha, S., Fulton, L.A., Gabisi, R.A., Garner, T.T., Godfrey, J., Hawes, A.C., Hernandez, J., Hines, S., Holder, M., Hume, J., Jhangiani, S.N., Joshi, V., Khan, Z.M., Kirkness, E.F., Cree, A., Fowler, R.G., Lee, S., Lewis, L.R., Li, Z., Liu, Y.S., Moore, S.M., Muzny, D., Nazareth, L.V., Ngo, D.N., Okwuonu, G.O., Pai, G., Parker, D., Paul, H.A., Pfannkoch, C., Pohl, C.S., Rogers, Y.H., Ruiz, S.J., Sabo, A., Santibanez, J., Schneider, B.W., Smith, S.M., Sodergren, E., Svatek, A.F., Utterback, T.R., Vattathil, S., Warren, W., White, C.S., Chinwalla, A.T., Feng, Y., Halpern, A.L., Hillier, L.W., Huang, X., Minx, P., Nelson, J.O., Pepin, K.H., Qin, X., Sutton, G.G., Venter, E., Walenz, B.P., Wallis, J.W., Worley, K.C., Yang, S.P., Jones, S.M., Marra, M.A., Rocchi, M., Schein, J.E., Baertsch, R., Clarke, L., Csürös, M., Glasscock, J., Harris, R.A., Havlak, P., Jackson, A.R., Jiang, H., Liu, Y., Messina, D.N., Shen, Y., Song, H.X., Wylie, T., Zhang, L., Birney, E., Han, K., Konkel, M.K., Lee, J., Smit, A.F., Ullmer, B., Wang, H., Xing, J., Burhans, R., Cheng, Z., Karro, J.E., Ma, J., Raney, B., She, X., Cox, M.J., Demuth, J.P., Dumas, L.J., Han, S.G., Hopkins, J., Karimpour-Fard, A., Kim, Y.H., Pollack, J.R., Vinar, T., Addo-Quaye, C., Degenhardt, J., Denby, A., Hubisz, M.J., Indap, A., Kosiol, C., Lahn, B.T., Lawson, H.A., Marklein, A., Nielsen, R., Vallender, E.J., Clark, A.G., Ferguson, B., Hernandez, R.D., Hirani, K.,

Kehrer-Sawatzki, H., Kolb, J., Patil, S., Pu, L.L., Ren, Y., Smith, D.G., Wheeler, D.A., Schenck, I., Ball, E.V., Chen, R., Cooper, D.N., Giardine, B., Hsu, F., Kent, W.J., Lesk, A., Nelson, D.L., O'brien, W.E., Prüfer, K., Stenson, P.D., Wallace, J.C., Ke, H., Liu, X.M., Wang, P., Xiang, A.P., Yang, F., Barber, G.P., Haussler, D., Karolchik, D., Kern, A.D., Kuhn, R.M., Smith, K.E., Zwieg, A.S., 2007. Evolutionary and biomedical insights from the rhesus macaque genome. Science. 316(5822): 222-34.

Glick, G., Shirak, A., Seroussi, E., Zeron, Y., Ezra, E., Weller, J.I., Ron, M., 2011. Fine Mapping of a QTL for Fertility on BTA7 and Its Association With a CNV in the Israeli Holsteins. G3 (Bethesda). 1(1):65-74.

Guryev, V., Saar, K., Adamovic, T., Verheul, M., van Heesch, S.A., Cook, S., Pravenec, M., Aitman, T., Jacob, H., Shull, J.D., Hubner, N., Cuppen, E., 2008. Distribution and functional impact of DNA copy number variation in the rat. Nat Genet. 40(5): 538-45.

Hou, Y., Bickhart, D.M., Hvinden, M.L., Li, C., Song, J., Boichard, D.A., Fritz, S., Eggen, 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.

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.

Iafrate, A.J., Feuk, L., Rivera, M.N., Listewnik, M.L., Donahoe, P.K., Qi, Y., Scherer, S.W., Lee, C., 2004. Detection of large-scale variation in the human genome. Nat Genet. 36(9): 949-51.

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.

Lee, C., Scherer, S.W., 2010. The clinical context of copy number variation in the human genome. Expert Rev Mol Med. 12:e8.

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., McDaneld, T.G., Keele, 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.

Meyers, S.N., McDaneld, 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 in bovine SLC4A2 is associated with osteopetrosis in Red Angus cattle. BMC Genomics 11:337.

Perry, G.H., Yang, F., Marques-Bonet, T., Murphy, C., Fitzgerald, T., Lee, A.S., Hyland, C., Stone, A.C., Hurles, M.E., Tyler-Smith, C., Eichler, E.E., Carter, N.P., Lee, C., Redon, R., 2008. Copy number variation and evolution in humans and chimpanzees. Genome Res. 18(11): 1698–1710.

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.

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.

Rosengren Pielberg, G., Golovko, A., Sundström, 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., Heldin, C.H., Sölkner, J., Andersson, L., 2008. A cis-acting regulatory mutation causes premature hair graying and susceptibility to melanoma in the horse. Nat Genet. 40(8): 1004-9.

Sun, W., Wright, F.A., Tang, Z., Nordgard, S.H., Van Loo, P., Yu, T., Kristensen, V.N., Perou, C.M., 2009. Integrated study of copy number states and genotype calls using high-density SNP arrays. Nucleic Acids Res. 37(16):5365-77.

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. PLoSOne 5:e14456.

Varki, A., Geschwind, D.H., Eichler, E.E., 2008. Explaining human uniqueness: genome interactions with environment, behaviour and culture. Nat Rev Genet. 9(10): 749–763.

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

Wang, J., Jiang, J., Fu, W., Jiang, L., Ding, X., Liu, J., Zhang, Q., 2012. A genome-wide detection of copy number variations using SNP genotyping arrays in swine. BMC Genomics. 13:273.

Wang, K., Li, M., Hadley, D., Liu, R., Glessner,J., Grant, S.F., Hakonarson, H., Bucan, M., 2007. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Res. 17:1665-1674.

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

Wright, D., Boije, H., Meadows, J.R., Bed'hom, 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(6): e1000512.

Zarrei, M., MacDonald, J.R., Merico, D., Scherer, S.W., 2015. A copy number variation map of the human genome. Nat Rev Genet. 16(3):172-83.

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

Published in: Italian Journal of Animal Science 14:3900 · September 2015. DOI: 10.4081/ijas.2015.3900.

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]Raphaelle T.M.M. Prinsen,[1]Marlies A. Dolezal[5]

[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 RazzaBruna, Bussolengo (VR), Italy
[4]Department of Animal Science, NorthCarolina State University, Raleigh, NC,USA
[5]Institut für Populationsgenetik Veterinärmedizinische, University of Wien, Austria

Abstract

The determination of copy number variation (CNV) is very important for the evaluation of genomic traits in several species because they are a major source for the genetic variation, influencing gene expression,

phenotypic variation, adaptation and the development of diseases. The aim of this study was to obtain a CNV genome map using the Illumina Bovine SNP50 BeadChip data of 651 bulls of the Italian Brown Swiss breed. PennCNV and SVS7 (Golden Helix) software were used for the detection of the CNVs and Copy Number Variation Regions (CNVRs). A total of 5,099 and 1,289 CNVs were identified with PennCNV and SVS7 software, respectively. These were grouped at the population level into 1101 (220losses, 774 gains, 107 complex) and 277 (185losses, 56 gains and 36 complex) CNVR. Ten of the selected CNVR were experimentally validated with a qPCR experiment. The GO and pathway analyses were conducted and they identified genes (false discovery rate corrected) in the CNVR related to biological processes, cellular component, molecular function and metabolic pathways. Among those, we found the *FCGR2B*, *PPARalpha*, *KATNAL1*, *DNAJC15*, *PTK2*, *TG*, *STAT family*, *NPM1*, *GATA2*, *LMF1*, *ECHS1* genes, already known in literature because of their association with various traits in cattle. Although there is variability in the CNVRs detection across methods and platforms, this study allowed the identification of CNVRs in Italian Brown Swiss, overlapping those already detected in other breeds and finding additional ones, thus producing new knowledge for association studies with traits of interest in cattle.

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 segment ranging from 50bp to several megabases (Mb) in comparison with a reference genome (Mills et al., 2011). CNVs are an 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 breeds, and CNV maps of the bovine genome, using SNPs, Next Generation Sequencing (NGS) and Comparative genome hybridization (aCGH) arrays, were reported (Matukumalli et al., 2009; Bae et al., 2010; Fadista et al., 2010; 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 (Rosengren et al., 2008). Also, the CNVs in the *ASIP* gene are responsible in the leading of different coat colors 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, have 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.

Material and Methods

Sampling and genotyping

The National Association of Italian Brown Swiss breeders (ANARB) provided commercial semen samples for 1,342 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 50ng/32L 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 Bead Studio software V.2.0 (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). A principal component analysis (PCA) for LRRs 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.openbioinfor-matics.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 such 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 with at least 10% of telomeres' length (the first and last 500 kb of each autosome were considered representing the telomeres).

SVS7 detection

The 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 CNV 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 utilized as suggested in the software manual:
i) univariate outlier removal; ii) maximum number of segments: search for up to 10 per10,000 markers; iii) minimum markers per segment: 1; iv) max pairwise permuted P value: 0.005 (2000 permutations per segment pair).

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 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® Softwarev3.0.1 (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 Tecnologies™) 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 the Bos_taurus_UMD3.1 to the 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/76d1cab099658c68bde77f7daf55117e). 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 option 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 analyzed 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.

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 from40.4 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 682Mb, which corresponded to 27.14% of the bovine autosomal 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's et al. (2006) approach, resulting into 277 CNVRs (Table 2). The total length of the sequence covered by the CNVRs was 33.71Mb (1.35%) of the bovine autosomes. 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 was 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 17.1Mb (0.68% of the autosomes), as shown in Supplementary File 1.

Table 3 shows the comparison between the CNVRs detected and those reported in literature, confirming both the existence of high variability in CNVRs detection across platforms, methods, population size, cattle breeds and species. Only a small number of CNVR of those detected in our study (150 regions) have already been detected in other studies: the range of overlapping varies between 2 and 38%. An additional factor that could explain this feature, except for the reasons listed above, is the fact that the CNVR set of this study is the result of a consensus between the two software, which work through different algorithms. This strengthens the solidness of the CNVR discovered in our study, but weakens on the other hand the possibility of these lasts to overlap with CNVR of other studies.

The highest overlapping coverage (38%) was found with the study of Hou et al. (2011), in which CNVs detection was performed using Bovine SNP50 assay including animals from Taurine dairy and beef breeds, breeds of predominantly Indicine background, Taurine -

Indicine composite and African groups. The previous mentioned dataset included 24 Brown Swiss individuals in which 22 CNVRs were identified on 13 BTAs. Only one CNVR on BTA9 from Hou et al. (2011) (4305338 -4386831Mbp) resulted in common with the region identified in our study (4050528-4476378Mbp). 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/genomever-sion#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 an 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 [page 556] [Ital J Anim Sci vol.14: 2015] 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 IIb*, receptor (*CD32*)] gene is included. The TREM1, in synergy with the TLR2 (Toll-like receptor 2) pathway, are involved in phagocytosis and production of pro inflammatory cytokines, determining optimal host defense during bovine mastitis. The *PPAR41* gene, encoding for the peroxisome proliferator-activated receptor-41, 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 DNAJC15gene 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). Fernàndez 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 STATs 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 the CNVR found on BTA20. This gene encoding a multifunctional nucleolar phospho-protein 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 *LMF1* (*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 SVS7), with the aim to reduce the high error rate commonly recognized 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 SVS7 data, which emphasized 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., Ideta, A., Aoyagi, Y., Imakawa, K., 2011. Regulation of trophoblast-specific factorsby GATA2 and GATA3 in bovine trophoblastCT-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 casein1-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 rumi-nants: potential targets for fine-tuning metabolism during growth and lactation.PPAR Res. 2013:684159.

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

Fernández, M., Goszczynski, 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 González, E., Dall'Olio, S., Davoli, R., Russo, V., Portolano, B., 2009. 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., Chen, H., Wang, J., Li,Z.J., Huai, Y.T., Ma, L., Lan, 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 NPM1 gene affects growth traits. J. Appl. Genet. 51:199-202.

Hou, Y., Bickhart, D.M., Hvinden, M.L., Li, C., Song, J., Boichard, D.A., Fritz, S., Eggen, 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.

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-Sabat, 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., McDaneld, T.G., Keele, 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., Walter, K., Stewart, C., Handsaker, R.E., Chen, K., Alkan, C., Abyzov, A., Yoon, S.C., Ye, K., Cheetham, R.K., Chinwalla, A., Conrad, D.F., Fu, Y., Grubert, F., Hajirasouliha, I., Hormozdiari, F., Iakoucheva, L.M., Iqbal, Z., Kang, S., Kidd, J.M., Konkel, M.K., Korn, J., Khurana, E., Kural, D., Lam, H.Y., Leng, J., Li, R., 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., McCarroll, S.A., Korbel, J.O.,1000 Genomes Project, 2011. Mapping copy number variation by population-scale genome sequencing. Nature 470:59-65.

Meyers, S.N., McDaneld, 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 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., Heldin, C.H., Sölkner, 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., Reghenzani, L., Cozzi, M.C., Samoré, 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 D(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. PLoSOne 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, K., Li, M., Hadley, D., Liu, R., Glessner,J., Grant, S.F., Hakonarson, H., Bucan, M., 2007. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Res. 17:1665-1674.

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 One8:e83625.

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

Wright, D., Boije, H., Meadows, J.R., Bed'hom,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 chick-ens. 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 pre-implantation 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. 9:4003-4008.

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.

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.

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.

		Results from different studies				Overlapped CNVRs of this study		
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)	368	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)	150			17.1			

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

Hou *et al.* (2011) (UMD3.1 assembly) PennCNV				This study (UMD3.1 assembly) PennCNV				QTL		
bta	start CNVR	end CNVR	lenght CNVR	bta	start CNVR	end CNVR	lenght CNVR	start QTL	end QTL	trait (ID)
2	8,788,219	9,113,368	325,150	2	8,788,219	9,040,720	252,501	9,003,563	9,867,063	trans-15-C18:1 fatty acid content (20510)
9	4,305,338	4,386,831	81,494	9	4,050,528	4,476,378	425,850	2,148,415	9,159,784	Interval from first to last insemination (5006)
12	31,368,562	31,679,957	311,396	12	30,099,199	31,555,734	1,456,535	30,751,141	30,912,583	*Longissimus* muscle area (11733)
14	17,378,950	17,457,836	78,887	14	17,322,658	17,457,836	135,178	9,884,020	19,204,282	Calving ease (maternal) (10958)
								12,169,925	20,562,022	*Longissimus* muscle area (4550)
								16,752,147	18,440,442	Shear force (20791)
18	57,565,406	57,659,303	93,898	18	56,364,657	58,090,087	1,725,430	55,181,080	60,030,732	Stillbirth (direct) (15198), (15199)
								55,181,080	60,030,732	Birth index (15200)
								55,181,080	60,030,732	Calving ease (direct) (15201) (15202)
								55,181,080	60,030,732	Calf size (direct) (15203), (15204)
								55,777,394	55,928,978	Marbling score (10014)
								55,860,765	63,144,054	Bilateral convergent strabismus with exophthalmus (10051)
								56,701,305	56,852,890	Dystocia (direct) (11363)

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.

Consensus CNVRs			Genes in Consensus CNVRs				References
Bta	Start	End	Start	End	Ensembl code	Gene symbol	
3	7957960	7983149	7928113	7944607	ENSBTAG00000021842	*FCGR2B*	Lewandowska-Sabat *et al.* (2013)
5	116895329	117247824	117151549	117233112	ENSBTAG00000008063	*PPARα*	Bionaz *et al.* (2013)
12	30418611	30646042	30519852	30558210	ENSBTAG00000009340	*KATNAL1*	Zhang *et al.* (2014)
12	13179696	13204137	13183734	13266310	ENSBTAG00000034785	*DNAJC15*	Zhang *et al.* (2011)
14	3885798	4017201	3870893	4065010	ENSBTAG00000009578	*PTK2*	Wang *et al.* (2013)
14	9300228	9345140	9262251	9508938	ENSBTAG00000007823	*TG*	Fernàndez *et al.* (2014) Bennett *et al.* (2013)
19	42976859	43170256	43056660	43132624	ENSBTAG00000021523	*STAT3*	Zhang *et al.* (2010)
19	42976859	43170256	42960226	42996671	ENSBTAG00000010125	*STAT5B*	
19	42976859	43170256	43033597	43054075	ENSBTAG00000009496	*STAT5A*	
20	2880532	3189118	3111198	3123860	ENSBTAG00000015316	*NPM1*	Huang *et al.* (2010)
22	59951940	60243916	60016985	60024586	ENSBTAG00000019707	*GATA2*	Bai *et al.* (2011)
25	609241	983759	724446	775899	ENSBTAG00000019745	*LMF1*	Ren *et al.* (2011)
26	25828973	25982293	25856475	25865594	ENSBTAG00000017710	*ECHS1*	Strillacci *et al.* (2014)

CNVR, copy number variation region.

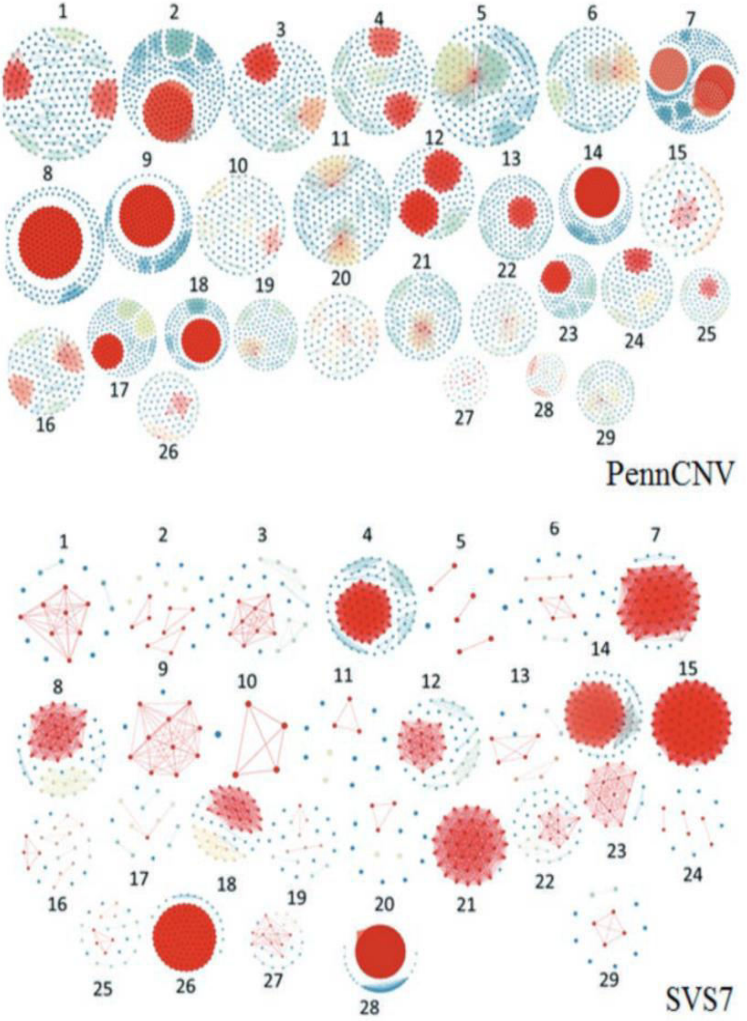


Figure 1. Graphical representation of copy number variations (CNVs) visualized by HD-CNV software. Each graph represents CNVs identified by PennCNV and SVS7 software for each of the 29 bovine autosomes. The size of the graphical representation of each chromosome depends on their total length. Each note represents a CNV, and edges are added between nodes that share 40% (default) overlap. The darker the red spots the more overlapping of CNVs across samples. The blue spots identify the unique events.

Supporting information

Table 1. List of the copy number variant regions (CNVRs) identified using pennCNV and SVS7 softwares and the CNVRs consensus generated according to Wain et al. (2009) method.

pennCNV (CNVRs)			**lenght**
chr1	1359951	1603944	243993
chr1	1625471	2013659	388188
chr1	4727118	5010974	283856
chr1	5319965	5741816	421851
chr1	12889396	13019983	130587
chr1	17021138	17160556	139418
chr1	18170722	18980052	809330
chr1	19829143	20192420	363277
chr1	27357510	27761036	403526
chr1	28034525	28216247	181722
chr1	41315418	41465449	150031
chr1	43317853	43861492	543639
chr1	46321775	46648008	326233
chr1	47389301	49230292	1840991
chr1	52191701	52316632	124931
chr1	56505444	56901848	396404
chr1	60753573	61329180	575607
chr1	66188527	66630647	442120
chr1	68066718	69250962	1184244
chr1	71229187	71623173	393986
chr1	72573950	73640200	1066250
chr1	80875014	80974985	99971
chr1	81018906	81160609	141703
chr1	82059960	82468782	408822
chr1	83030921	83672872	641951
chr1	88086502	88262684	176182
chr1	91031267	91358383	327116
chr1	92958471	93279488	321017
chr1	95793312	95910972	117660
chr1	97276664	97484342	207678
chr1	101103098	101366448	263350
chr1	101403755	101664632	260877
chr1	101773849	102056710	282861
chr1	103728420	103926075	197655
chr1	104842361	105264358	421997
chr1	106228836	106295981	67145
chr1	106512965	106824629	311664
chr1	107094743	107172243	77500
chr1	108154057	108782509	628452
chr1	109649036	109869888	220852
chr1	113631332	113780811	149479
chr1	114945541	115185215	239674
chr1	115389278	115456741	67463

chr1	117619160	118226740	607580
chr1	118532290	118882592	350302
chr1	120915343	121003901	88558
chr1	123921894	124196948	275054
chr1	126070556	126846004	775448
chr1	127491286	127749327	258041
chr1	128320901	129505485	1184584
chr1	131423039	132312674	889635
chr1	132497074	132647596	150522
chr1	135194234	136476195	1281961
chr1	138583183	138832098	248915
chr1	139409122	139590378	181256
chr1	141903958	145280015	3376057
chr1	145470565	146632014	1161449
chr1	146862570	147076679	214109
chr1	147121098	147195759	74661
chr1	147255491	148388848	1133357
chr1	148684318	148841496	157178
chr1	149101763	149335782	234019
chr1	149969397	150574361	604964
chr1	150900517	151929994	1029477
chr1	152189521	155224721	3035200
chr1	155654021	156710174	1056153
chr1	157039952	157328169	288217
chr2	904065	1247436	343371
chr2	1795004	1982197	187193
chr2	4087123	4165880	78757
chr2	4587203	5601419	1014216
chr2	6645590	6831955	186365
chr2	8788219	9040720	252501
chr2	12533969	12632490	98521
chr2	12789975	12954788	164813
chr2	14496313	14565726	69413
chr2	15972368	16087638	115270
chr2	17627839	17830735	202896
chr2	23334606	23405887	71281
chr2	25359234	25826013	466779
chr2	26781358	28215944	1434586
chr2	28376835	28573158	196323
chr2	32056065	32376219	320154
chr2	39976359	40021512	45153
chr2	52966242	53016506	50264
chr2	55118828	55566541	447713
chr2	56123710	56267092	143382
chr2	56504541	56636575	132034
chr2	57260012	57449775	189763
chr2	59569007	59927024	358017
chr2	63411488	64300614	889126
chr2	65044427	65249007	204580
chr2	65469391	65700940	231549
chr2	70665916	70895063	229147
chr2	70978502	71382630	404128
chr2	71848377	73436684	1588307
chr2	76112733	76563133	450400
chr2	79017908	79414331	396423

chr2	82707674	82808175	100501
chr2	83008944	83050544	41600
chr2	90155450	90355514	200064
chr2	93065483	93172942	107459
chr2	103864593	104026376	161783
chr2	105666691	105958636	291945
chr2	106526130	108549488	2023358
chr2	109740844	110393192	652348
chr2	112515240	112756530	241290
chr2	114112197	114317194	204997
chr2	115475433	116087072	611639
chr2	119450394	120251367	800973
chr2	122409628	124922135	2512507
chr2	125653839	125716884	63045
chr2	126957475	127860492	903017
chr2	128251331	128695085	443754
chr2	128947842	132821698	3873856
chr2	132921343	133258339	336996
chr2	133504852	136531159	3026307
chr3	970386	1223441	253055
chr3	3050190	3320855	270665
chr3	5505466	5552259	46793
chr3	6900453	7052779	152326
chr3	7866803	8021336	154533
chr3	9579325	9751086	171761
chr3	13635591	13842528	206937
chr3	13974787	14321565	346778
chr3	14791163	17319185	2528022
chr3	21371931	22326368	954437
chr3	23409559	23534653	125094
chr3	26394698	26438136	43438
chr3	26738133	28144691	1406558
chr3	29187683	29381184	193501
chr3	29852694	30194236	341542
chr3	30963937	31131027	167090
chr3	31331920	31750523	418603
chr3	33276539	34344799	1068260
chr3	35046094	35317787	271693
chr3	39097420	39143931	46511
chr3	39456939	39599273	142334
chr3	40263643	40364628	100985
chr3	40903270	41432313	529043
chr3	42791404	42893462	102058
chr3	49544712	49766904	222192
chr3	51126976	52402934	1275958
chr3	69716313	69968011	251698
chr3	75364755	75413638	48883
chr3	76367708	76821191	453483
chr3	77923168	78108702	185534
chr3	78971559	79333053	361494
chr3	79353705	79581142	227437
chr3	79652820	79777779	124959
chr3	80889108	81630254	741146
chr3	82623974	82677195	53221
chr3	85100892	85450050	349158

chr3	85629524	85860004	230480
chr3	86762502	87293789	531287
chr3	87904203	88170287	266084
chr3	91703200	93847303	2144103
chr3	94758229	94980177	221948
chr3	97132442	97923006	790564
chr3	98123394	99655499	1532105
chr3	100442675	100982336	539661
chr3	101253585	101377479	123894
chr3	101785952	101942771	156819
chr3	103021431	103358134	336703
chr3	104093652	104249570	155918
chr3	104716951	105536719	819768
chr3	107362756	108033653	670897
chr3	108257297	110293079	2035782
chr3	111214429	111547601	333172
chr3	112203165	112286655	83490
chr3	113271662	114220425	948763
chr3	114451860	115236331	784471
chr3	115630359	118715700	3085341
chr3	118934229	119113936	179707
chr3	120191150	120860510	669360
chr4	4258236	5312990	1054754
chr4	6248795	6337362	88567
chr4	7779209	8466191	686982
chr4	10528978	10933419	404441
chr4	10967124	11215149	248025
chr4	15645122	15947178	302056
chr4	18447308	18689580	242272
chr4	25373005	25646689	273684
chr4	28413603	29123987	710384
chr4	29449548	29602230	152682
chr4	34794353	34856797	62444
chr4	36408463	36928225	519762
chr4	37145925	37631099	485174
chr4	38166835	38548833	381998
chr4	41556810	41657868	101058
chr4	44792807	45202155	409348
chr4	45835326	45904319	68993
chr4	47137606	47396270	258664
chr4	47821591	47949379	127788
chr4	48615363	48726071	110708
chr4	50300237	50547931	247694
chr4	55594630	55687238	92608
chr4	58512174	58998553	486379
chr4	60521073	60917347	396274
chr4	63420711	63674872	254161
chr4	66583105	66978951	395846
chr4	68942061	69220543	278482
chr4	69955754	70729858	774104
chr4	70839756	70931951	92195
chr4	72920828	73085109	164281
chr4	73699663	73791282	91619
chr4	73861517	74266771	405254
chr4	76868216	77839843	971627

chr4	78440044	78523846	83802
chr4	80154498	80407501	253003
chr4	83897939	84274238	376299
chr4	85983005	86286457	303452
chr4	86324014	86480802	156788
chr4	92235197	92436001	200804
chr4	93211337	93739544	528207
chr4	94760042	94973039	212997
chr4	95336471	95465677	129206
chr4	96034920	97128968	1094048
chr4	99542142	99635737	93595
chr4	100289958	101072503	782545
chr4	103231866	103687583	455717
chr4	106276208	106404042	127834
chr4	106885171	107060437	175266
chr4	107416505	107732347	315842
chr4	108867683	108974924	107241
chr4	111788712	113614764	1826052
chr4	114396713	114688732	292019
chr4	115713661	116485311	771650
chr4	117170573	120412745	3242172
chr5	3180610	3434356	253746
chr5	10969920	11526654	556734
chr5	11647027	11800467	153440
chr5	19128124	19594448	466324
chr5	22943453	23171537	228084
chr5	27973786	28024945	51159
chr5	33953799	34325053	371254
chr5	34371932	34446901	74969
chr5	34896625	35300909	404284
chr5	42166749	42436216	269467
chr5	46132243	46306149	173906
chr5	51141012	51604559	463547
chr5	60556520	60711494	154974
chr5	65922088	66577741	655653
chr5	69351812	69571158	219346
chr5	69856333	70258225	401892
chr5	72938296	73334655	396359
chr5	73976636	74593114	616478
chr5	74778313	74921819	143506
chr5	75002613	75627333	624720
chr5	76599679	76659850	60171
chr5	76724602	77702287	977685
chr5	85160180	85672503	512323
chr5	95459836	95666368	206532
chr5	96623157	97593586	970429
chr5	97828652	98753414	924762
chr5	103539042	104230386	691344
chr5	105275043	105387546	112503
chr5	106713645	107704991	991346
chr5	107840177	107984945	144768
chr5	108811739	111930616	3118877
chr5	113250833	113874131	623298
chr5	114543256	114859696	316440
chr5	115010779	117247824	2237045

chr5	117738204	120783915	3045711
chr6	496589	643965	147376
chr6	1125213	1641898	516685
chr6	2245446	2642053	396607
chr6	7363787	7902806	539019
chr6	9914189	10295237	381048
chr6	10424905	11363207	938302
chr6	12558447	12846571	288124
chr6	13211945	13417489	205544
chr6	13475512	13610388	134876
chr6	14851230	15252879	401649
chr6	15580216	15808841	228625
chr6	17282916	17563402	280486
chr6	19769454	19996175	226721
chr6	20924952	21126480	201528
chr6	23148412	23217855	69443
chr6	24184712	24764626	579914
chr6	25033485	25356965	323480
chr6	25474941	25925916	450975
chr6	30694723	31067604	372881
chr6	39069719	39753141	683422
chr6	40893067	42446118	1553051
chr6	42493184	42866573	373389
chr6	43037439	43614199	576760
chr6	44305092	44894050	588958
chr6	45479538	46178717	699179
chr6	46970861	47188068	217207
chr6	56952870	57139728	186858
chr6	60654151	61132085	477934
chr6	63996365	64542255	545890
chr6	66307093	66594604	287511
chr6	66747934	66802712	54778
chr6	69979797	70123449	143652
chr6	76696893	76754229	57336
chr6	77228984	77520815	291831
chr6	79766485	79817258	50773
chr6	79942163	79982917	40754
chr6	81467492	81604925	137433
chr6	82128469	82480772	352303
chr6	84819700	84996399	176699
chr6	91984771	92242383	257612
chr6	94562464	95080255	517791
chr6	98198721	98363600	164879
chr6	99187864	99408048	220184
chr6	100428424	100827998	399574
chr6	100903198	101297489	394291
chr6	101357896	101661011	303115
chr6	101749615	101983370	233755
chr6	104453601	105309108	855507
chr6	106197672	110591345	4393673
chr6	110697360	111110390	413030
chr6	113157379	113606872	449493
chr6	114747565	115490762	743197
chr6	117013231	117148273	135042
chr6	118004635	118986727	982092

chr7	1127439	2008771	881332
chr7	2565484	6669157	4103673
chr7	7822111	7910128	88017
chr7	7986664	8315512	328848
chr7	13279972	13823591	543619
chr7	14330447	14434623	104176
chr7	17490481	18476120	985639
chr7	18707985	18845261	137276
chr7	18910887	19084659	173772
chr7	19352717	21347602	1994885
chr7	21462645	23074262	1611617
chr7	24233634	24778430	544796
chr7	24946051	25232158	286107
chr7	26984291	27258897	274606
chr7	27973088	28567404	594316
chr7	33658873	34011987	353114
chr7	34534432	34709657	175225
chr7	35003374	35581341	577967
chr7	37537822	38075277	537455
chr7	38613662	39679687	1066025
chr7	39868831	41875877	2007046
chr7	42788788	43479777	690989
chr7	44490836	44901489	410653
chr7	45097887	46109256	1011369
chr7	46293900	47213804	919904
chr7	47954688	48218360	263672
chr7	49999470	50541246	541776
chr7	53736537	54081460	344923
chr7	55434939	55655916	220977
chr7	57357439	57641183	283744
chr7	59766632	60012102	245470
chr7	60628965	61330104	701139
chr7	61879960	61992242	112282
chr7	62400698	62978076	577378
chr7	63255889	63778905	523016
chr7	64766846	65245810	478964
chr7	66061402	66450039	388637
chr7	66951622	67774508	822886
chr7	68382283	68861661	479378
chr7	72264142	72424026	159884
chr7	74320957	74632687	311730
chr7	78762342	78987067	224725
chr7	81863002	82943896	1080894
chr7	84971705	85071041	99336
chr7	85387851	85579651	191800
chr7	86669785	86879535	209750
chr7	87700528	87790787	90259
chr7	88192479	88268249	75770
chr7	91670145	91974966	304821
chr7	95531504	95600027	68523
chr7	97691853	98013150	321297
chr7	103159234	103375625	216391
chr7	104730721	105465008	734287
chr7	108096309	108307729	211420
chr7	108685133	109268080	582947

chr7	109991711	110422232	430521
chr7	111259941	111891998	632057
chr8	494866	1047065	552199
chr8	1208626	1745249	536623
chr8	8340612	8725932	385320
chr8	10320524	11500414	1179890
chr8	21220037	21525831	305794
chr8	23493443	24103669	610226
chr8	27557552	27831038	273486
chr8	27908523	28559343	650820
chr8	28726561	29068846	342285
chr8	32611423	32749041	137618
chr8	33668705	33771490	102785
chr8	34584476	35464882	880406
chr8	40470452	41173061	702609
chr8	41930432	42545530	615098
chr8	47092657	47224522	131865
chr8	47540190	47898948	358758
chr8	48914474	49550836	636362
chr8	50626573	50938776	312203
chr8	52557297	53980316	1423019
chr8	54402516	54894631	492115
chr8	55084725	55239532	154807
chr8	56512655	56625850	113195
chr8	58505244	58729476	224232
chr8	58926246	59487308	561062
chr8	59541510	60071322	529812
chr8	60589007	61042106	453099
chr8	61877938	62108589	230651
chr8	62469487	62575787	106300
chr8	63435892	63956656	520764
chr8	65228923	65659020	430097
chr8	65960156	66134324	174168
chr8	66721299	67180774	459475
chr8	68168243	68442169	273926
chr8	68689710	68850097	160387
chr8	69285264	69622989	337725
chr8	71263375	71627282	363907
chr8	73238931	73435613	196682
chr8	76005495	76281603	276108
chr8	76530816	76620508	89692
chr8	79314073	79817203	503130
chr8	80944503	81334383	389880
chr8	82099108	82223990	124882
chr8	84780381	84896726	116345
chr8	85645138	86810388	1165250
chr8	88545459	89313850	768391
chr8	89619126	90249304	630178
chr8	90412852	91278924	866072
chr8	94115663	94386951	271288
chr8	94579362	95277410	698048
chr8	96266647	96712245	445598
chr8	97543154	98056575	513421
chr8	100873051	101206057	333006
chr8	101446930	101529840	82910

chr8	102608032	103188091	580059
chr8	104220117	105033919	813802
chr8	105486423	106196262	709839
chr8	107107709	107324529	216820
chr8	108492195	108868929	376734
chr8	110455623	111180295	724672
chr8	111668506	111713971	45465
chr8	112020602	112966306	945704
chr9	1926314	2284075	357761
chr9	2434402	2728202	293800
chr9	2779925	3115242	335317
chr9	4050528	4476378	425850
chr9	8858542	9018195	159653
chr9	10343145	10765698	422553
chr9	16768069	17298279	530210
chr9	18790503	18862613	72110
chr9	22281170	22396640	115470
chr9	25126494	25185694	59200
chr9	26179360	26847057	667697
chr9	28887462	29029183	141721
chr9	31510537	32049723	539186
chr9	32343804	32549297	205493
chr9	34681064	34802818	121754
chr9	50378425	50694137	315712
chr9	55532332	55630797	98465
chr9	56966851	57069740	102889
chr9	60280329	61209187	928858
chr9	65284951	65672050	387099
chr9	70248005	70491698	243693
chr9	73259605	74512290	1252685
chr9	76776811	77070363	293552
chr9	81195326	81277154	81828
chr9	83692030	83837068	145038
chr9	86288986	86971466	682480
chr9	87008260	87254861	246601
chr9	88342363	88714313	371950
chr9	89847787	90002616	154829
chr9	90807695	90924836	117141
chr9	90996268	91158204	161936
chr9	95127819	95952795	824976
chr9	96131728	97191669	1059941
chr9	97260495	99334002	2073507
chr9	100022759	101470808	1448049
chr9	102546533	105462864	2916331
chr10	2244249	2361595	117346
chr10	3229515	3343140	113625
chr10	4415581	4534148	118567
chr10	4742149	5598788	856639
chr10	5767015	6724993	957978
chr10	8132805	8394690	261885
chr10	8675119	8898592	223473
chr10	9558767	10846884	1288117
chr10	11654004	11707725	53721
chr10	11998905	12386269	387364
chr10	12486527	13040145	553618

chr10	15051277	16162053	1110776
chr10	16460889	17625573	1164684
chr10	18041832	18772359	730527
chr10	20322987	20721372	398385
chr10	21571180	22339193	768013
chr10	26406685	26767851	361166
chr10	28070182	28397659	327477
chr10	28508237	28703125	194888
chr10	35828126	36419341	591215
chr10	36946284	37133985	187701
chr10	41389231	41514241	125010
chr10	44694452	45506957	812505
chr10	46486647	46802159	315512
chr10	47309122	48110941	801819
chr10	48837583	49981642	1144059
chr10	50213424	51383075	1169651
chr10	51502722	51803725	301003
chr10	52933550	53318692	385142
chr10	56285758	57193699	907941
chr10	58028735	58318595	289860
chr10	67290043	67694539	404496
chr10	68324855	68570791	245936
chr10	68963110	69107598	144488
chr10	69311079	69658256	347177
chr10	70894537	71359437	464900
chr10	72658520	72846830	188310
chr10	73798107	73979984	181877
chr10	76604316	76779458	175142
chr10	78044879	78549784	504905
chr10	80910121	81113308	203187
chr10	81788224	82290325	502101
chr10	83241282	83389165	147883
chr10	84096660	84899748	803088
chr10	84951492	85085819	134327
chr10	86036359	86214669	178310
chr10	86868858	87139659	270801
chr10	87684721	88133750	449029
chr10	88673513	89365394	691881
chr10	89717052	91982227	2265175
chr10	92695820	93019985	324165
chr10	98420963	98817467	396504
chr10	99103087	99542820	439733
chr10	100804186	101130681	326495
chr10	101521757	102210987	689230
chr10	102544951	103699875	1154924
chr11	740136	1920092	1179956
chr11	2603799	3192673	588874
chr11	5539221	6243988	704767
chr11	6701679	6777332	75653
chr11	7190271	7585747	395476
chr11	8149768	8713134	563366
chr11	9153560	9473158	319598
chr11	10271653	10800426	528773
chr11	15788936	15919622	130686
chr11	20755386	20960589	205203

chr11	22558808	23085190	526382
chr11	24354910	26162970	1808060
chr11	27989531	28207767	218236
chr11	28992259	29367145	374886
chr11	30651364	31130270	478906
chr11	31173613	31259588	85975
chr11	36713471	37261133	547662
chr11	38733905	39104584	370679
chr11	46307696	47430509	1122813
chr11	51963535	52677772	714237
chr11	54193489	54246625	53136
chr11	60128321	60572203	443882
chr11	63686208	64313748	627540
chr11	67498111	68298497	800386
chr11	73934314	74197538	263224
chr11	75295616	76060924	765308
chr11	77235847	79314962	2079115
chr11	80450475	80839258	388783
chr11	82833655	82959206	125551
chr11	87124625	89371911	2247286
chr11	89552045	89680768	128723
chr11	90738123	90911719	173596
chr11	91961667	92500678	539011
chr11	92963716	93050541	86825
chr11	93546324	93612322	65998
chr11	94640611	94837245	196634
chr11	95150490	99180813	4030323
chr11	99497380	99738023	240643
chr11	99774304	104856815	5082511
chr11	105593624	106825407	1231783
chr12	2058297	2273370	215073
chr12	2814820	3705325	890505
chr12	9079069	9165632	86563
chr12	13140896	13273888	132992
chr12	15726574	16098884	372310
chr12	18472407	18584913	112506
chr12	20079121	20402284	323163
chr12	20510073	20800774	290701
chr12	21073393	21441958	368565
chr12	22664255	22863845	199590
chr12	25141005	25255436	114431
chr12	27006686	27133552	126866
chr12	29520358	29778424	258066
chr12	30099199	31555734	1456535
chr12	31639399	31848902	209503
chr12	34480373	34635671	155298
chr12	35189028	35703502	514474
chr12	41384223	41479027	94804
chr12	45002070	45573176	571106
chr12	47300401	47528422	228021
chr12	48758328	49421215	662887
chr12	50125707	50324576	198869
chr12	52546816	52663318	116502
chr12	55097796	55674695	576899
chr12	55867003	56204081	337078

chr12	56367819	56695055	327236
chr12	57077252	57211328	134076
chr12	57931622	58461348	529726
chr12	63969950	64152261	182311
chr12	64384089	64761912	377823
chr12	66068259	66877440	809181
chr12	69969069	70092297	123228
chr12	77109094	77347204	238110
chr12	78099072	78212571	113499
chr12	78276736	78750520	473784
chr12	81184160	81586853	402693
chr12	83477477	83777475	299998
chr12	87579030	87891233	312203
chr12	88638167	90826206	2188039
chr13	1247948	1912749	664801
chr13	2823073	5324891	2501818
chr13	9139641	9276071	136430
chr13	9424085	10663054	1238969
chr13	12587622	12759014	171392
chr13	13067288	13907892	840604
chr13	14356314	14509110	152796
chr13	14585745	15088689	502944
chr13	15923779	16424596	500817
chr13	18657767	19388240	730473
chr13	21363395	21676028	312633
chr13	25212012	25818921	606909
chr13	30108205	30337948	229743
chr13	31580781	31876709	295928
chr13	33141153	34047662	906509
chr13	34808642	35148331	339689
chr13	35563916	35875350	311434
chr13	36079542	36576754	497212
chr13	36730055	37334004	603949
chr13	38086947	38618369	531422
chr13	38943133	41366870	2423737
chr13	41414256	44168398	2754142
chr13	44847602	45158832	311230
chr13	45381514	47680091	2298577
chr13	51732402	52065761	333359
chr13	53890455	54829615	939160
chr13	55288364	56292896	1004532
chr13	58836000	59223837	387837
chr13	60468277	60781323	313046
chr13	61426148	61782627	356479
chr13	62534909	63500701	965792
chr13	66774859	68766380	1991521
chr13	69042143	69187742	145599
chr13	70860749	71713558	852809
chr13	71860088	73230677	1370589
chr13	73478012	73681829	203817
chr13	73746516	73866335	119819
chr13	74078322	74436097	357775
chr13	76330479	76932968	602489
chr13	77022083	77956325	934242
chr13	78148257	78501277	353020

chr13	78705310	79775537	1070227
chr13	79866776	82559505	2692729
chr13	83520405	83662585	142180
chr14	1435005	3664511	2229506
chr14	3885798	6371334	2485536
chr14	6850767	8264685	1413918
chr14	8385937	10549180	2163243
chr14	10586742	12151878	1565136
chr14	13146789	13680187	533398
chr14	17322658	17457836	135178
chr14	17674401	18296407	622006
chr14	20089569	20635979	546410
chr14	21037828	22226073	1188245
chr14	24482969	24787245	304276
chr14	25254540	25501417	246877
chr14	27669598	27875890	206292
chr14	34046332	34189618	143286
chr14	35554436	35744766	190330
chr14	38344250	38789482	445232
chr14	39201271	39907342	706071
chr14	40484309	40646356	162047
chr14	46959846	47141282	181436
chr14	48614875	48929525	314650
chr14	50323759	50680387	356628
chr14	51457342	51540635	83293
chr14	53415847	54164119	748272
chr14	56053388	57510609	1457221
chr14	58262807	58740723	477916
chr14	60333729	61100170	766441
chr14	61792567	61965465	172898
chr14	63431800	63695912	264112
chr14	65147770	65511994	364224
chr14	69028708	70003286	974578
chr14	70358208	70763413	405205
chr14	77123730	77222049	98319
chr14	77900043	78092917	192874
chr14	79178022	79486476	308454
chr14	80475984	81414515	938531
chr14	83306326	83504134	197808
chr14	83685686	83773615	87929
chr15	1492400	1676811	184411
chr15	5001246	5134162	132916
chr15	12027285	12520480	493195
chr15	13154168	13226449	72281
chr15	15595454	16042857	447403
chr15	16937958	17080803	142845
chr15	21325465	22026329	700864
chr15	23188415	23448971	260556
chr15	24673090	24930212	257122
chr15	25535567	25719311	183744
chr15	25925175	26790090	864915
chr15	29305818	29448587	142769
chr15	30573210	31873933	1300723
chr15	32044715	32395964	351249
chr15	33565804	33698965	133161

chr15	34171356	35252207	1080851
chr15	37027165	37107956	80791
chr15	37149431	38251520	1102089
chr15	41707747	41971248	263501
chr15	43308530	43652444	343914
chr15	44417808	44584661	166853
chr15	45634800	46058920	424120
chr15	46906029	47196374	290345
chr15	52590087	53166998	576911
chr15	53250782	53459520	208738
chr15	54617640	54939673	322033
chr15	57260972	57592823	331851
chr15	58628094	59434020	805926
chr15	61958209	62309986	351777
chr15	65590902	66483616	892714
chr15	67185489	68123068	937579
chr15	75482866	76166896	684030
chr15	76265233	76598267	333034
chr15	77046320	77451913	405593
chr15	78230054	78401390	171336
chr15	78544866	78681531	136665
chr15	81335502	82335513	1000011
chr16	900145	1104865	204720
chr16	2134973	3621602	1486629
chr16	3987821	4610955	623134
chr16	6146432	6494072	347640
chr16	8486467	8769236	282769
chr16	9883787	10805391	921604
chr16	12130589	12365230	234641
chr16	13695211	14500474	805263
chr16	24241296	24401108	159812
chr16	27248557	27432134	183577
chr16	29441057	29661958	220901
chr16	30466202	30885863	419661
chr16	32341971	32837610	495639
chr16	45083406	45425579	342173
chr16	46069918	46213318	143400
chr16	46509837	48879755	2369918
chr16	49355913	49491331	135418
chr16	50281187	52535473	2254286
chr16	66863251	67192692	329441
chr16	68014396	68455943	441547
chr16	69911268	71125864	1214596
chr16	71349209	74046418	2697209
chr16	75305757	76681045	1375288
chr16	77098989	78356180	1257191
chr16	78939515	79896422	956907
chr16	80351106	80924801	573695
chr17	1041832	2084655	1042823
chr17	2881443	3228062	346619
chr17	4217941	4675045	457104
chr17	5792606	6436523	643917
chr17	8776227	9027765	251538
chr17	9206816	9558819	352003
chr17	10140780	10557794	417014

chr17	11169306	11231535	62229
chr17	14357124	14596438	239314
chr17	16758551	16985065	226514
chr17	18186024	19008492	822468
chr17	19109434	19315294	205860
chr17	23322903	23431642	108739
chr17	24320366	24539327	218961
chr17	27416064	27516228	100164
chr17	30676454	30949325	272871
chr17	31310645	31790508	479863
chr17	32731007	32964041	233034
chr17	39963957	40071626	107669
chr17	42425636	42688661	263025
chr17	44413031	44616245	203214
chr17	45356254	46814289	1458035
chr17	47170772	48276271	1105499
chr17	48479069	48788398	309329
chr17	49531815	49828442	296627
chr17	53180044	54052291	872247
chr17	54312738	54945167	632429
chr17	55366744	55658355	291611
chr17	55713369	55941040	227671
chr17	56963536	57203284	239748
chr17	59778912	60165864	386952
chr17	60418182	60746006	327824
chr17	62070942	63480469	1409527
chr17	63515768	63753770	238002
chr17	64912981	68253146	3340165
chr17	68468706	69360236	891530
chr17	70290307	72741339	2451032
chr17	72987826	74598498	1610672
chr18	1094150	1751859	657709
chr18	1875406	2793266	917860
chr18	4899308	6776532	1877224
chr18	34379769	35085008	705239
chr18	35971459	36040190	68731
chr18	39793889	42136113	2342224
chr18	42638878	42826428	187550
chr18	43327273	44064689	737416
chr18	45870264	46373028	502764
chr18	46552279	46972002	419723
chr18	47259533	47443804	184271
chr18	47627016	47983685	356669
chr18	48960277	49167271	206994
chr18	51321651	52024379	702728
chr18	52664641	52739093	74452
chr18	53132012	53224638	92626
chr18	53970861	54765715	794854
chr18	55309510	55892476	582966
chr18	56121920	56319394	197474
chr18	56364657	58090087	1725430
chr18	61095214	61597742	502528
chr18	61920892	62751093	830201
chr18	63029071	64901743	1872672
chr19	1664089	1903390	239301

chr19	2484709	3247474	762765
chr19	5057128	5718616	661488
chr19	6822436	7273862	451426
chr19	8052163	9417473	1365310
chr19	11049355	12032389	983034
chr19	13492031	13634621	142590
chr19	14274648	17923922	3649274
chr19	20439349	20635700	196351
chr19	21388415	21733030	344615
chr19	22038801	23650335	1611534
chr19	24917540	26039760	1122220
chr19	27529791	27602402	72611
chr19	30220186	30572014	351828
chr19	30834957	31226718	391761
chr19	33533491	33811153	277662
chr19	34371541	36710214	2338673
chr19	37732974	38018129	285155
chr19	42089455	42352691	263236
chr19	42393606	43170256	776650
chr19	43295532	43441993	146461
chr19	43610147	46202442	2592295
chr19	46396064	46770465	374401
chr19	47270773	47725481	454708
chr19	48368103	48449016	80913
chr19	48630588	48801884	171296
chr19	49727374	49826620	99246
chr19	50087148	50567992	480844
chr19	51148913	52911677	1762764
chr19	53038373	59742080	6703707
chr19	59817808	61550668	1732860
chr19	61753819	62986532	1232713
chr20	808587	861460	52873
chr20	1627053	2585844	958791
chr20	2880532	5096097	2215565
chr20	5478758	6029026	550268
chr20	6360647	6416157	55510
chr20	7327625	7409378	81753
chr20	8164878	8539206	374328
chr20	8636338	9028131	391793
chr20	9919863	10295382	375519
chr20	10738152	10871544	133392
chr20	17710059	17946080	236021
chr20	18099389	19040448	941059
chr20	23843046	23886196	43150
chr20	28186842	28463308	276466
chr20	29684981	30161777	476796
chr20	37066726	37374576	307850
chr20	37777328	38606353	829025
chr20	40728408	41777888	1049480
chr20	42696358	43004103	307745
chr20	48715588	49251437	535849
chr20	50128719	50347483	218764
chr20	51044020	51205535	161515
chr20	52816141	53333822	517681
chr20	53798275	54060327	262052

chr20	55865295	56196291	330996
chr20	56662612	57263704	601092
chr20	58071204	59109201	1037997
chr20	60902173	60954990	52817
chr20	62247834	62736738	488904
chr20	63356077	63771728	415651
chr20	64754802	65099050	344248
chr20	65715521	67038772	1323251
chr20	67462559	68355681	893122
chr20	69020321	69504928	484607
chr20	69554351	71535445	1981094
chr21	2749974	3027827	277853
chr21	7022598	8519331	1496733
chr21	10788801	11360403	571602
chr21	13052680	13110087	57407
chr21	14362108	14654478	292370
chr21	14723518	15659059	935541
chr21	19051950	22397530	3345580
chr21	22520337	23220048	699711
chr21	26116828	26715346	598518
chr21	26936862	28283039	1346177
chr21	28387366	28612871	225505
chr21	28675178	29923689	1248511
chr21	30608016	30908358	300342
chr21	33327048	35267234	1940186
chr21	35777432	36083384	305952
chr21	37293354	37705700	412346
chr21	42389217	42835553	446336
chr21	46018333	46063657	45324
chr21	47963417	48099076	135659
chr21	49439452	49533074	93622
chr21	49594234	50244378	650144
chr21	52139183	52608666	469483
chr21	56118709	56785345	666636
chr21	57156944	58442333	1285389
chr21	58944572	59248804	304232
chr21	59787612	60553762	766150
chr21	61094140	63286443	2192303
chr21	65272069	67220188	1948119
chr21	67593887	69554017	1960130
chr21	70089833	71136925	1047092
chr22	632381	1428683	796302
chr22	2808095	3126680	318585
chr22	5022403	5353497	331094
chr22	6814022	7263366	449344
chr22	7623264	8838211	1214947
chr22	9918880	11246732	1327852
chr22	11756783	13223791	1467008
chr22	14086186	14300175	213989
chr22	14514444	15177059	662615
chr22	15589177	15974506	385329
chr22	16219978	16342830	122852
chr22	16437316	16727330	290014
chr22	17882779	18494857	612078
chr22	19409002	19588936	179934

chr22	20826586	20920119	93533
chr22	21963292	22075972	112680
chr22	23268971	23458863	189892
chr22	23834157	24538255	704098
chr22	29101885	29655985	554100
chr22	32990568	33583017	592449
chr22	35771697	36047120	275423
chr22	37590098	38303302	713204
chr22	39545402	39702951	157549
chr22	40167931	41573980	1406049
chr22	42177755	42616545	438790
chr22	42791065	44030499	1239434
chr22	46902036	48406016	1503980
chr22	50705290	52187729	1482439
chr22	52706213	53254546	548333
chr22	54440045	61040701	6600656
chr23	2570211	3664434	1094223
chr23	4083573	4455923	372350
chr23	4622146	4737838	115692
chr23	4781751	5233255	451504
chr23	6948746	8124702	1175956
chr23	9951185	10240332	289147
chr23	10526857	11251946	725089
chr23	12910040	13155828	245788
chr23	13770707	13866915	96208
chr23	14063300	14981935	918635
chr23	16601543	20402236	3800693
chr23	22716563	23507190	790627
chr23	24803563	25507676	704113
chr23	27174246	27583474	409228
chr23	27776075	28311070	534995
chr23	37292633	37692330	399697
chr23	40182681	42669248	2486567
chr23	45009074	45666148	657074
chr23	46578332	51938161	5359829
chr24	541784	1094942	553158
chr24	1450456	2859956	1409500
chr24	3237803	5054955	1817152
chr24	5828401	6047561	219160
chr24	7322657	7943787	621130
chr24	11753230	11888700	135470
chr24	12775555	12916861	141306
chr24	14841074	15259118	418044
chr24	17317994	17513135	195141
chr24	19267711	19370939	103228
chr24	21571435	22300490	729055
chr24	24302542	24582206	279664
chr24	28175885	28218879	42994
chr24	28400451	28546447	145996
chr24	28853585	29405619	552034
chr24	32346216	34201179	1854963
chr24	34621317	36201079	1579762
chr24	36301805	36566389	264584
chr24	39320770	39365195	44425
chr24	41289711	41540027	250316

chr24	42296195	43402173	1105978
chr24	45239241	47323079	2083838
chr24	47722933	47797164	74231
chr24	48140670	48226481	85811
chr24	48528178	48999628	471450
chr24	49755342	50568228	812886
chr24	52047342	52165674	118332
chr24	53328928	53576672	247744
chr24	55570429	55980406	409977
chr24	56379984	56564480	184496
chr24	56834044	56889750	55706
chr24	57257505	59703255	2445750
chr24	60394194	62191687	1797493
chr25	472458	5156189	4683731
chr25	5662943	7515881	1852938
chr25	9252279	10305794	1053515
chr25	12965823	13824693	858870
chr25	14216892	14569665	352773
chr25	15006757	15120011	113254
chr25	15365566	16017661	652095
chr25	17363060	17742193	379133
chr25	17891876	18206998	315122
chr25	18372002	18448555	76553
chr25	22566179	23040867	474688
chr25	25405137	26380285	975148
chr25	30183716	30269233	85517
chr25	32823036	33108530	285494
chr25	33412423	34112565	700142
chr25	34565592	36514994	1949402
chr25	36980815	38698430	1717615
chr25	38856905	39921068	1064163
chr25	40827689	42364359	1536670
chr26	2452597	2687667	235070
chr26	3524161	3618349	94188
chr26	4204767	4731392	526625
chr26	5258082	5526925	268843
chr26	8695518	8887209	191691
chr26	10689379	10995721	306342
chr26	16862899	16973274	110375
chr26	17815724	18967997	1152273
chr26	19323674	19657939	334265
chr26	20012464	20560149	547685
chr26	22826271	23603307	777036
chr26	24401843	24507253	105410
chr26	25501890	26124236	622346
chr26	27213271	27343629	130358
chr26	28723721	28784693	60972
chr26	32892811	33337772	444961
chr26	33716256	34460340	744084
chr26	36631883	37018466	386583
chr26	37203584	37579568	375984
chr26	38980475	39121205	140730
chr26	40903566	41183634	280068
chr26	41271740	41762178	490438
chr26	41956121	45302252	3346131

chr26	45642121	45954628	312507
chr26	46187669	46366100	178431
chr26	46673561	47837750	1164189
chr26	48007359	50273965	2266606
chr27	703314	967098	263784
chr27	2248130	2472812	224682
chr27	4357162	5000552	643390
chr27	6866166	6955584	89418
chr27	8781446	8848885	67439
chr27	12842240	15044426	2202186
chr27	15794249	16296679	502430
chr27	18710877	19195734	484857
chr27	20559148	21676934	1117786
chr27	22922395	23144459	222064
chr27	25295935	25728096	432161
chr27	27804403	28148660	344257
chr27	29009479	29300595	291116
chr27	31000749	32184354	1183605
chr27	32488843	32671451	182608
chr27	32724283	32889433	165150
chr27	33400664	33788321	387657
chr27	34036869	34166163	129294
chr27	35024154	35734689	710535
chr27	35900786	36285734	384948
chr27	36342691	36592652	249961
chr27	36935085	37479272	544187
chr27	38109791	39688231	1578440
chr27	40210986	41049981	838995
chr27	42055498	44148168	2092670
chr28	5123022	5965031	842009
chr28	6334557	6547497	212940
chr28	6626319	10044965	3418646
chr28	13424880	14091432	666552
chr28	16783056	17374797	591741
chr28	21762976	21872563	109587
chr28	23969284	24175313	206029
chr28	25899333	28205509	2306176
chr28	31130099	31197813	67714
chr28	33240420	35412301	2171881
chr28	35668756	36300699	631943
chr28	36504079	37052366	548287
chr28	37272033	37433520	161487
chr28	38961890	40191764	1229874
chr28	40957698	41239838	282140
chr28	42551127	42831281	280154
chr28	43088798	43877796	788998
chr28	44056044	44153620	97576
chr29	2522803	3074352	551549
chr29	5348843	5452376	103533
chr29	9020167	9454029	433862
chr29	9850630	10393660	543030
chr29	10872370	11753300	880930
chr29	11988163	12241827	253664
chr29	16131653	17796803	1665150
chr29	19200818	20146131	945313

chr29	25219612	26293574	1073962
chr29	28523337	29768788	1245451
chr29	30428085	30638566	210481
chr29	30827166	31693910	866744
chr29	31899837	32238809	338972
chr29	32516455	32944239	427784
chr29	33249348	34118132	868784
chr29	35051920	36669359	1617439
chr29	36761051	38015000	1253949
chr29	39930095	41264801	1334706
chr29	41552472	41664361	111889
chr29	42620218	45023665	2403447
chr29	45817015	50999092	5182077

SVS7 (CNVRs)			**lenght**
chr1	4626587	4648383	21796
chr1	52748123	52772080	23957
chr1	64501909	64525174	23265
chr1	91066621	91435586	368965
chr1	123825385	123921894	96509
chr1	144134237	144190645	56408
chr1	146587678	146632014	44336
chr1	155835816	155955828	120012
chr2	5757355	5802738	45383
chr2	14525350	14565726	40376
chr2	27489458	27724930	235472
chr2	28215944	28246133	30189
chr2	32108568	32154806	46238
chr2	39976359	39999947	23588
chr2	57373897	57416172	42275
chr2	58772041	59143632	371591
chr2	83029887	83050544	20657
chr2	110311653	110353189	41536
chr2	124137395	124204421	67026
chr3	7957960	7983149	25189
chr3	33514564	33741850	227286
chr3	39097420	39118317	20897
chr3	40977107	41079360	102253
chr3	68402961	68432145	29184
chr3	81379707	81411041	31334
chr3	91910014	92190368	280354
chr3	100468099	100493684	25585
chr3	105715727	105739637	23910
chr3	115888900	115937988	49088
chr3	116781408	116801749	20341
chr4	6248795	6337362	88567
chr4	10401625	10450547	48922
chr4	11472235	11500911	28676
chr4	17902915	17978432	75517
chr4	22093546	22180214	86668
chr4	24087424	24150445	63021
chr4	28384911	28413603	28692
chr4	41444423	41657868	213445

chr4	66781385	66830563	49178
chr4	69158293	69220543	62250
chr4	73699663	73791282	91619
chr4	78440044	78523846	83802
chr4	81497187	81554820	57633
chr4	86811243	86948025	136782
chr4	89850655	89921990	71335
chr4	99574406	99691481	117075
chr4	106980782	107007048	26266
chr4	108168742	108198485	29743
chr4	108867683	108904498	36815
chr4	111990062	112164314	174252
chr4	118608842	118655986	47144
chr5	30684691	30838906	154215
chr5	46279541	46306149	26608
chr5	58847022	58966295	119273
chr5	114543256	114698428	155172
chr5	116895329	118329917	1434588
chr5	119729902	119949553	219651
chr6	5025746	5086136	60390
chr6	9736332	9981135	244803
chr6	12648459	12703601	55142
chr6	22613578	22672648	59070
chr6	40107367	40208497	101130
chr6	50981312	51007189	25877
chr6	53428838	53449439	20601
chr6	68291351	68332579	41228
chr6	81551479	81604925	53446
chr6	90966250	90989420	23170
chr6	100620998	100709082	88084
chr6	104493834	104587477	93643
chr6	117013231	117148273	135042
chr7	1293067	1353317	60250
chr7	2597655	2680354	82699
chr7	22524899	22681472	156573
chr7	42788788	43132401	343613
chr7	43709405	43808593	99188
chr7	76886696	77011685	124989
chr7	78081511	78307528	226017
chr7	87158251	87359924	201673
chr8	9133270	9156221	22951
chr8	15417359	15441189	23830
chr8	15665796	15773921	108125
chr8	20018829	20067861	49032
chr8	29068846	29100768	31922
chr8	33356720	33747904	391184
chr8	34898163	34920926	22763
chr8	43628838	43779600	150762
chr8	73715997	73829090	113093
chr8	87038169	87141059	102890
chr8	94115663	94973599	857936
chr8	105683974	105695288	11314
chr9	4239500	4439872	200372
chr9	68973776	69002105	28329
chr9	102258435	102271805	13370

chr10	5437359	5540505	103146
chr10	39823420	39846476	23056
chr11	14979948	15029477	49529
chr11	16425876	16514661	88785
chr11	33642979	33682867	39888
chr11	46657176	46701073	43897
chr11	72555948	72598008	42060
chr11	84899274	84947879	48605
chr11	88028793	88377200	348407
chr11	93445185	93587894	142709
chr11	101750113	101802657	52544
chr11	105699664	105778702	79038
chr12	13179696	13204137	24441
chr12	20129895	20402284	272389
chr12	30418611	30646042	227431
chr12	41384223	41479027	94804
chr12	43601825	43638160	36335
chr12	50385487	50451289	65802
chr12	55867003	55931940	64937
chr12	57931622	58461348	529726
chr12	59437039	59609816	172777
chr12	67538730	67564989	26259
chr12	70649671	72090421	1440750
chr12	78212571	78235491	22920
chr12	82159124	82199690	40566
chr12	82450106	82661747	211641
chr13	5594384	5623697	29313
chr13	58043371	58070117	26746
chr13	65965727	66336246	370519
chr13	70496054	70523797	27743
chr13	80026050	80144645	118595
chr14	2721633	2803998	82365
chr14	3765019	4017201	252182
chr14	6778397	6850767	72370
chr14	8064004	8113083	49079
chr14	8499902	8551460	51558
chr14	9300228	9345140	44912
chr14	20119611	20157384	37773
chr14	30449596	30595032	145436
chr14	50955416	50996515	41099
chr14	51285167	51430094	144927
chr14	53415847	53436763	20916
chr14	54023420	54123146	99726
chr14	75571250	76043148	471898
chr14	76217573	76269650	52077
chr14	79178022	79322701	144679
chr14	80082712	80543545	460833
chr15	47780178	47819771	39593
chr15	76438547	76466667	28120
chr15	80369812	81139855	770043
chr16	4158997	4233985	74988
chr16	8317477	8582055	264578
chr16	9171622	9267151	95529
chr16	9670453	9729773	59320
chr16	15154212	15362423	208211

chr16	22272329	22302686	30357
chr16	29441057	29636822	195765
chr16	36768083	36817218	49135
chr16	41131268	41194530	63262
chr16	49355913	49455109	99196
chr16	68811897	68889225	77328
chr16	70906202	71125864	219662
chr17	8170089	8297151	127062
chr17	20275502	20484740	209238
chr17	27459029	27491589	32560
chr17	32762909	32964041	201132
chr17	39963957	40071626	107669
chr17	42425636	42661925	236289
chr17	55713369	55764236	50867
chr18	3080400	3114628	34228
chr18	35971459	36107915	136456
chr18	42659289	42826428	167139
chr18	50388296	50465387	77091
chr18	51571629	51592949	21320
chr18	53132012	53195763	63751
chr18	60978019	61054591	76572
chr18	61095214	61156737	61523
chr18	61438125	61920892	482767
chr18	63119361	63167945	48584
chr19	6709868	6768232	58364
chr19	11863651	11970132	106481
chr19	34836416	34905583	69167
chr19	35585081	35619269	34188
chr19	38519698	38546855	27157
chr19	42352691	42423488	70797
chr19	42976859	43170256	193397
chr19	46655940	46723662	67722
chr19	50336021	50395622	59601
chr19	51767413	51842198	74785
chr19	52175916	52264019	88103
chr19	54306610	54446207	139597
chr19	55379112	55527962	148850
chr19	56072306	56202223	129917
chr19	56754737	56837932	83195
chr20	1741145	1792368	51223
chr20	2880532	3189118	308586
chr20	6360647	6385223	24576
chr20	21018903	21048672	29769
chr20	33773531	33817557	44026
chr20	34241862	34264853	22991
chr20	34953795	34981347	27552
chr20	41239866	41289921	50055
chr20	45052283	45369517	317234
chr20	60902173	60928704	26531
chr20	64185456	64376028	190572
chr21	54186710	54208626	21916
chr21	70089833	71210609	1120776
chr22	12869969	12948282	78313
chr22	16219978	16407075	187097
chr22	19409002	19588936	179934

chr22	21431682	21455286	23604
chr22	25141851	25321072	179221
chr22	26527854	26604789	76935
chr22	26865100	26924506	59406
chr22	31649896	31675722	25826
chr22	36527685	36548339	20654
chr22	39545402	39657636	112234
chr22	44430993	44595995	165002
chr22	47510478	47537080	26602
chr22	48858472	49131324	272852
chr22	54028803	54183730	154927
chr22	57098389	57111693	13304
chr22	59951940	60243916	291976
chr22	60435042	60508872	73830
chr23	25250595	25339818	89223
chr23	28503248	28563533	60285
chr24	1027534	1137518	109984
chr24	3321961	3342966	21005
chr24	21071943	21129533	57590
chr24	24499452	24582206	82754
chr24	28060569	28083770	23201
chr24	28175885	28196203	20318
chr24	38640377	38694114	53737
chr24	39320770	39365195	44425
chr24	53328928	53434181	105253
chr25	609241	983759	374518
chr25	7952738	7992272	39534
chr25	12190414	12217941	27527
chr25	16017661	16048464	30803
chr25	18206998	18372002	165004
chr25	19082329	19145490	63161
chr25	22238007	22320002	81995
chr25	24125205	24228117	102912
chr25	37988321	38142895	154574
chr25	39286957	39424763	137806
chr25	39544407	39570754	26347
chr25	39785037	39844749	59712
chr26	5258082	5288263	30181
chr26	5472360	5504271	31911
chr26	12136498	12236803	100305
chr26	19686897	19942669	255772
chr26	21902497	21955137	52640
chr26	25828973	25982293	153320
chr26	28303383	28398156	94773
chr26	39655739	39681664	25925
chr26	48693316	48713332	20016
chr26	49027625	49090826	63201
chr27	4544917	4773381	228464
chr27	6922514	7188361	265847
chr27	8730441	8827679	97238
chr27	9096031	9191858	95827
chr27	12282518	12330184	47666
chr27	18036224	18164172	127948
chr27	33813284	33845584	32300
chr27	38025744	38233675	207931

chr27	43237090	43260976	23886
chr28	2313753	2638563	324810
chr28	6334557	6547497	212940
chr28	12717523	12973750	256227
chr28	13713042	13894573	181531
chr28	21982457	22148051	165594
chr28	26994978	27072121	77143
chr28	37514643	37624697	110054
chr28	38026506	38074472	47966
chr28	44030986	44056044	25058
chr29	7401774	7480356	78582
chr29	19618823	19701179	82356
chr29	21930571	21987120	56549
chr29	27184360	27465875	281515
chr29	28192104	28248785	56681
chr29	33329702	33353664	23962
chr29	35136093	35169599	33506
chr29	41212959	41264801	51842
chr29	48178151	48252404	74253

File 2. Summary of the results of the qPCR analysis of the eleven copy number regions (CNVRs) selected after the consensus analysis in the Italian Brown Swiss breed.

BTA	PRIMER (FORWARD)	PRIMER (REVERSE)	PROBE	START*	END*	LENGHT	PENN CNV STATE	CNAM CNV STATE	NUMBER OF SAMPLES	CONFIRMED SAMPLES	CONFIRMED RATE	VALIDATED	GENE (Btau4.6.1)
2	TGCATGCACACAGGAATGTTAC	TGCCCCTAAGAAGGAGTCGTT	ACTCTGTTCAGCCCTTC	63870167	64073748	203581	GAIN		11	8	0,73	YES	MGAT5
3	GACTAATGGCAAGAGCCGTGTA	AGGCAGGAACAGAAGGAGGAA	TGAGCATGTCACTTTAA	7957198	7966700	9502	GAIN	GAIN	3	2	0,67	YES	FCGR2B
4	GCCCGGCGGACACTAAG	CCAGCATTATGTCCTTCATCAACA	TCAGGAAGCTGTGGCCA	6228016	6450000	221984	GAIN		11	11	1,00	YES	-
4	TCCTGCCCAGATACCATATCCTT	CGAGGCAAGCTCTACAGGAAA	TGGCATTCAAATCAC	73686740	73785101	98361	LOSS	LOSS	10	9	0,90	YES	ZNF804B
5	GGAGATAGGATAGAAAGAAATGGAGAAC	ATGGGAGTGATGGAAAATTGAAG	CACTCTTAAATTCCC	58965609	59140571	174962	LOSS		10	7	0,70	YES	LOC787945
12	GGACAGTCACCTCAGGATGCA	TTGCCACAAGTGAGGCTTCTC	AACGGTCACCTAAGAGACA	58071208	58427000	355792	LOSS	LOSS	11	1	0,09	NO	-
12	GCCTGGTGTTGTCATGATGAA	CCCGTACACTGACACCAAAGTG	TTTTGCGCTTGAAGCAG	67534765	67579929	45164		LOSS	11	10	0,91	YES	-
13	TGCGAAATTCTGGAAGAGGAA	GGGTGCCTGGTGCAATTC	CCTGAGGACATGAAGTT	53931895	53983934	52039	GAIN		5	3	0,60	YES	SIRPB1
19	CAGTGAGCCAAAGCCAATCC	AATCCAACTTGCCGGCTAGTATT	CCTCCACAGGAATC	2585940	2607218	21278	LOSS		12	9	0,75	YES	-
28	ACATTCAGGCTGCCATTTTGT	GAGGCGGGATGTCACAGAAA	TCCAATATCGTCAACCATT	2553716	2635632	81916		LOSS	12	6	0,50	YES	OR5AS1
29	CACGGGCGCACCACTT	CCCCCGATGAATGGCTATC	AGCTCCCTGCTCGAC	5414600	5444000	29400	GAIN		10	9	0,90	YES	-
REF_BTF3	GCTGAGACAAAGCAACTGACAGA	TCGGCACCAAGCTGGTTTA	TGCTCCCCAGCATC										

File 3. Annotation of copy number variant regions (CNVRs).

CNVRs_pennCNV			CNVRs_SVS7		CNVRs_Consensus		CNVRs State	Genes included in Consensus CNVRs			
chr1	141903958	145280015	144134237	144190645	144134237	144190645	gain/loss	144176745	144180011	ENSBTAG00000030814	TFF2
chr1	155654021	156710174	155835816	155955828	155835816	155955828	gain/loss	155833805	156185921	ENSBTAG00000030581	TBC1D5
chr2	14496313	14565726	14525350	14565726	14525350	14565726	loss	14502890	14623643	ENSBTAG00000044009	PPP1R1C
chr2	26781358	28215944	27489458	27724930	27489458	27724930	gain/loss	27629813	27629887	ENSBTAG00000044462	bta-mir-2353
								27407917	27758923	ENSBTAG00000044179	CERS6
chr2	39976359	40021512	39976359	39999947	39976359	39999947	loss	39999717	40017015	ENSBTAG00000003650	NR4A2
chr2	109740844	110393192	110311653	110353189	110311653	110353189	gain/loss	110251546	110405363	ENSBTAG00000010030	EPHA4
chr3	7866803	8021336	7957960	7983149	7957960	7983149	gain	7928113	7944607	ENSBTAG00000021842	FCGR2B
chr3	33276539	34344799	33514564	33741850	33514564	33741850	gain/loss	33702816	33702894	ENSBTAG00000044953	bta-mir-2413
								33607139	33621030	ENSBTAG00000000283	CSF1
								33513768	33556281	ENSBTAG00000018893	AHCYL1
chr3	39097420	39143931	39097420	39118317	39097420	39118317	loss	39113552	39114954	ENSBTAG00000015180	none
chr3	91703200	93847303	91910014	92190368	91910014	92190368	gain/loss	91901853	91911965	ENSBTAG00000013241	BSND
								91919532	91925379	ENSBTAG00000046583	TMEM61
								91994928	91995058	ENSBTAG00000042369	SNORA8
								92098624	92100051	ENSBTAG00000040313	PARS2
								92023811	92054136	ENSBTAG00000017145	C1orf177
								92059023	92083335	ENSBTAG00000017132	TTC22
								92106927	92132082	ENSBTAG00000015931	TTC4
								91981619	92014282	ENSBTAG00000004688	DHCR24
								92136670	92190804	ENSBTAG00000030623	none
								92136706	92165969	ENSBTAG00000044141	HEATR8
chr3	100442675	100982336	100468099	100493684	100468099	100493684	gain/loss	100472063	100495789	ENSBTAG00000013322	POMGNT1
								100472499	100483275	ENSBTAG00000024144	LURAP1
chr3	115630359	118715700	115888900	115937988	115888900	115937988	gain/loss	115843770	116226449	ENSBTAG00000016504	none
			116781408	116801749							

chr4	66583105	66978951	66781385	66830563	66781385	66830563	gain	66785266	66851245	ENSBTAG00000016223	SCRN1
chr4	73699663	73791282	73699663	73791282	73699663	73791282	loss	73326980	73897041	ENSBTAG00000046430	ZNF804B
chr4	99542142	99635737	99574406	99691481	99574406	99635737	loss	99475015	99580189	ENSBTAG00000013953	CALD1
								99591042	99690829	ENSBTAG00000013976	AGBL3
chr4	106885171	107060437	106980782	107007048	106980782	107007048	gain/loss	106989816	106993917	ENSBTAG00000015510	none
								106996940	107013515	ENSBTAG00000024219	TRPV6
			111990062	112164314	111990062	112164314	gain/loss	112048065	112048217	ENSBTAG00000047873	none
chr4	117170573	120412745	118608842	118655986				112126680	112126805	ENSBTAG00000045871	none
chr5	114543256	114859696	114543256	114698428	114543256	114698428	gain/loss	114576427	114588012	ENSBTAG00000002413	MCAT
								114596006	114608640	ENSBTAG00000018073	TSPO
								114610664	114627819	ENSBTAG00000001708	TTLL12
								114644629	114767260	ENSBTAG00000011275	SCUBE1
chr5	115010779	117247824	116895329	118329917	116895329	117247824	gain/loss	117119385	117119458	ENSBTAG00000029772	bta-let-7a-3
								117119640	117119712	ENSBTAG00000045309	bta-mir-2443
								117120188	117120270	ENSBTAG00000036417	bta-mir-3596
								116959716	116966318	ENSBTAG00000009532	WNT7B
								117240033	117248391	ENSBTAG00000008065	CDPF1
								116756235	116897163	ENSBTAG00000009351	ATXN10
								117151549	117233112	ENSBTAG00000008063	PPARA
			116895329	118329917	117738204	118329917	gain/loss	117743264	117758670	ENSBTAG00000005595	TRMU
								117975282	118012354	ENSBTAG00000021803	GRAMD4
								118023267	118036020	ENSBTAG00000046654	CERK
								117677522	117738845	ENSBTAG00000007102	GTSE1
								117764821	117853214	ENSBTAG00000008036	CELSR1
								118086468	118343833	ENSBTAG00000012291	TBC1D22A
chr5	117738204	120783915	119729902	119949553	119729902	119949553	gain/loss	119821435	119829363	ENSBTAG00000019574	MAPK12
								119771356	119791795	ENSBTAG00000000647	SELO
								119814028	119819426	ENSBTAG00000011000	HDAC10
								119832293	119837101	ENSBTAG00000030182	MAPK11
								119840726	119854120	ENSBTAG00000014966	PLXNB2
								119874457	119885052	ENSBTAG00000024756	DENND6B
								119791795	119813398	ENSBTAG00000000650	TUBGCP6
								119926286	120029002	ENSBTAG00000018660	PPP6R2
chr6	81467492	81604925	81551479	81604925	81551479	81604925	gain/loss	81511554	81653990	ENSBTAG00000024826	TECRL
chr7	1127439	2008771	1293067	1353317	1293067	1353317	loss	1317088	1334619	ENSBTAG00000015602	C7H5orf45
								1334566	1346051	ENSBTAG00000015591	SQSTM1

								1273635	1313897	ENSBTAG00000015611	TBC1D9B
chr7	2565484	6669157	2597655	2680354	2597655	2680354	gain/loss	2586101	2598115	ENSBTAG00000001604	none
								2600732	2618482	ENSBTAG00000040028	MGC166429
chr7	21462645	23074262	22524899	22681472	22524899	22681472	gain/loss	22533096	22537819	ENSBTAG00000004521	LSM7
								22569683	22574912	ENSBTAG00000016477	C19orf35
								22577243	22580777	ENSBTAG00000018522	OAZ1
								22519686	22532977	ENSBTAG00000004524	SPPL2B
								22560008	22561849	ENSBTAG00000016478	LINGO3
								22629692	22680307	ENSBTAG00000009996	DOT1L
chr7	42788788	43479777	42788788	43132401	42788788	43132401	gain/loss	42787986	42788915	ENSBTAG00000007557	OR2AK2
								42811931	42813032	ENSBTAG00000047016	none
								42833645	42834607	ENSBTAG00000046417	none
								42868129	42869064	ENSBTAG00000027241	none
								42890345	42891283	ENSBTAG00000045733	none
								42913832	42914770	ENSBTAG00000046474	none
								42947455	42948392	ENSBTAG00000046042	none
								43044539	43045551	ENSBTAG00000040033	OR2AJ1
								43101693	43102628	ENSBTAG00000030725	none
								43119732	43120670	ENSBTAG00000047180	none
chr8	94115663	94386951	94115663	94973599	94115663	94386951	loss	94230962	94231065	ENSBTAG00000042843	U6
								94205191	94210057	ENSBTAG00000015608	CYLC2
chr11	46307696	47430509	46657176	46701073	46657176	46701073	gain/loss	46699166	46706152	ENSBTAG00000019665	IL1RN
chr11	87124625	89371911	88028793	88377200	88028793	88377200	gain/loss	88012967	88104077	ENSBTAG00000002329	ASAP2
chr11	93546324	93612322	93445185	93587894	93546324	93587894	gain/loss	93563425	93564411	ENSBTAG00000038726	none
								93584334	93585269	ENSBTAG00000037542	none
chr11	99774304	104856815	101750113	101802657	101750113	101802657	gain/loss	101728372	101793685	ENSBTAG00000020791	RAPGEF1
chr11	105593624	106825407	105699664	105778702	105699664	105778702	gain/loss	105698114	105702610	ENSBTAG00000030246	ENTPD8
								105702496	105711512	ENSBTAG00000012121	NOXA1
								105728961	105770612	ENSBTAG00000023788	EXD3
chr12	13140896	13273888	13179696	13204137	13179696	13204137	gain/loss	13183734	13266310	ENSBTAG00000034785	DNAJC15
chr12	30099199	31555734	30418611	30646042	30418611	30646042	gain	30587084	30587189	ENSBTAG00000045239	SNORA70
								30519852	30558210	ENSBTAG00000009340	KATNAL1
chr12	57931622	58461348	57931622	58461348	57931622	58461348	loss	58187519	58187639	ENSBTAG00000045992	none
chr13	79866776	82559505	80026050	80144645	80026050	80144645	gain/loss	80015601	80114072	ENSBTAG00000018270	NFATC2
chr14	1435005	3664511	2721633	2803998	2721633	2803998	gain/loss	2755206	2762197	ENSBTAG00000000158	LY6K
								2770551	2775678	ENSBTAG00000037824	none

								2715416	2742638	ENSBTAG00000004595	GML
								2801383	2803020	ENSBTAG00000034498	LY6D
chr 14	3885798	6371334	3765019	4017201	3885798	4017201	gain/loss	3870893	4065010	ENSBTAG00000009578	PTK2
chr 14	6850767	8264685	8064004	8113083	8064004	8113083	gain/loss	8080292	8080361	ENSBTAG00000029987	bta-mir-30d
								8084721	8084808	ENSBTAG00000029972	bta-mir-30b
chr 14	8385937	10549180	9300228	9345140	9300228	9345140	gain/loss	9334778	9371281	ENSBTAG00000007828	SLA
								9262251	9508938	ENSBTAG00000007823	TG
chr 14	53415847	54164119	54023420	54123146	54023420	54123146	loss	53901591	54429251	ENSBTAG00000038281	CSMD3
chr 14	79178022	79486476	79178022	79322701	79178022	79322701	gain/loss	79296713	79298474	ENSBTAG00000002851	none
chr16	3987821	4610955	4158997	4233985	4158997	4233985	gain/loss	4221210	4242621	ENSBTAG00000010432	EIF2D
								4144349	4218744	ENSBTAG00000010427	RASSF5
chr16	29441057	29661958	29441057	29636822	29441057	29636822	gain/loss	29552152	29561665	ENSBTAG00000033322	SRP9
								29624572	29655286	ENSBTAG00000002854	TMEM63A
								29585792	29624037	ENSBTAG00000000140	EPHX1
								29238992	29442791	ENSBTAG00000016185	ENAH
chr16	49355913	49491331	49355913	49455109	49355913	49455109	loss	49429155	49429874	ENSBTAG00000046062	none
								49447984	49564506	ENSBTAG00000021919	NAV1
chr16	69911268	71125864	70906202	71125864	70906202	71125864	gain/loss	71019535	71024141	ENSBTAG00000004790	UBE2T
								71077717	71077907	ENSBTAG00000033994	U2
								70925149	70928253	ENSBTAG00000047073	none
								70932238	71016271	ENSBTAG00000004789	LGR6
								71062132	71137463	ENSBTAG00000011772	PPP1R12B
								70902206	70917245	ENSBTAG00000003016	PTPN7
chr17	32731007	32964041	32762909	32964041	32762909	32964041	gain/loss	32712712	32889849	ENSBTAG00000003345	FAT4
chr17	55713369	55941040	55713369	55764236	55713369	55764236	loss	55707870	55719927	ENSBTAG00000004175	HPD
								55727012	55747669	ENSBTAG00000004172	SETD1B
								55759606	55773011	ENSBTAG00000032534	RHOF
18	35971459	36040190	35971459	36107915	35971459	36040190	gain/loss	36008030	36029234	ENSBTAG00000007488	ZFP90
chr18	42638878	42826428	42659289	42826428	42659289	42826428	gain/loss	42749252	42750804	ENSBTAG00000003856	none
chr18	51321651	52024379	51571629	51592949	51571629	51592949	gain/loss	51520760	51578983	ENSBTAG00000011723	GRIK5
								51587793	51603479	ENSBTAG00000018635	ATP1A3
chr18	53132012	53224638	53132012	53195763	53132012	53195763	loss	53154421	53158137	ENSBTAG00000013702	ZNF296
								53129172	53153714	ENSBTAG00000013697	CLASRP
								53160659	53167519	ENSBTAG00000010668	GEMIN7
								53169569	53207223	ENSBTAG00000018834	PPP1R37
chr18	61095214	61597742	61438125	61920892	61095214	61156737	gain/loss	61145844	61145922	ENSBTAG00000036392	bta-mir-371

								61091348	61143063	ENSBTAG00000038149	NLRP12
chr18	63029071	64901743	63119361	63167945	63119361	63167945	gain/loss	63119873	63124888	ENSBTAG00000045989	CDC42EP5
								63146729	63154412	ENSBTAG00000019547	none
chr19	11049355	12032389	11863651	11970132	11863651	11970132	gain/loss	11865527	11891936	ENSBTAG00000009968	TBX4
								11943185	11951411	ENSBTAG00000014278	TBX2
chr19	34371541	36710214	35585081	35619269	35585081	35619269	gain/loss	35557245	35646258	ENSBTAG00000010534	M-RIP
			34836416	34905583	34836416	34905583	gain/loss	34832961	34860869	ENSBTAG00000003705	FAM83G
								34878438	34899068	ENSBTAG00000014858	PRPSAP2
								34817325	34872403	ENSBTAG00000003700	SLC5A10
chr19	42393606	43170256	42352691	42423488	42393606	42423488	gain/loss	42413413	42418215	ENSBTAG00000047165	KRT9
chr19	46396064	46770465	46655940	46723662	42976859	43170256	gain/loss	43033597	43054075	ENSBTAG00000009496	STAT5A
								43148013	43162165	ENSBTAG00000039684	PTRF
								43056660	43132624	ENSBTAG00000021523	STAT3
								42960226	42996671	ENSBTAG00000010125	STAT5B
chr19	46396064	46770465	46655940	46723662	46655940	46723662	gain/loss	46650344	46775847	ENSBTAG00000012564	KANSL1
chr19	50087148	50567992	50336021	50395622	50336021	50395622	gain/loss	50388667	50536976	ENSBTAG00000015414	TBCD
	51148913	52911677	52175916	52264019	52175916	52264019	gain/loss	52189629	52189720	ENSBTAG00000029775	bta-mir-338
								52181458	52196943	ENSBTAG00000019049	AATK
								52198617	52263294	ENSBTAG00000019044	BAIAP2
chr19			51767413	51842198				51768184	51769908	ENSBTAG00000000354	PDE6G
								51781191	51829647	ENSBTAG00000019105	NPLOC4
								51833821	51842965	ENSBTAG00000019104	C17orf70
								51771254	51775213	ENSBTAG00000040573	TSPAN10
chr19	53038373	59742080	56754737	56837932	56754737	56837932	gain/loss	56754112	56818424	ENSBTAG00000004736	GRB2
			54306610	54446207	54306610	54446207	gain/loss	54343819	54343903	ENSBTAG00000047060	none
								54404340	54441590	ENSBTAG00000000675	PGS1
								54312243	54405245	ENSBTAG00000000920	DNAH17
			55379112	55527962	55379112	55527962	gain/loss	55419632	55419819	ENSBTAG00000044443	SCARNA16
			56072306	56202223	56072306	56202223	gain/loss	56072684	56078553	ENSBTAG00000013792	UBALD2
								56171932	56175103	ENSBTAG00000007916	FOXJ1
								56112050	56167541	ENSBTAG00000016240	RNF157

								56190846	56208380	ENSBTAG00000007910	EXOC7
chr20	1627053	2585844	1741145	1792368	1741145	1792368	gain/loss	1452376	1893741	ENSBTAG00000014612	DOCK2
chr20	2880532	5096097	2880532	3189118	2880532	3189118	gain	2985749	2986486	ENSBTAG00000034824	none
								3004314	3005537	ENSBTAG00000022684	none
								3064510	3066676	ENSBTAG00000010003	TLX3
								3111198	3123860	ENSBTAG00000015316	NPM1
								2680574	3054892	ENSBTAG00000024801	RANBP17
								3132655	3195021	ENSBTAG00000000128	FGF18
chr20	6360647	6416157	6360647	6385223	6360647	6385223	loss	6360600	6365489	ENSBTAG00000013873	MSX2
chr21	70089833	71136925	70089833	71210609	70089833	71136925	gain/loss	70113045	70118400	ENSBTAG00000026886	MP68
								70114200	70114755	ENSBTAG00000046186	none
								70133009	70134255	ENSBTAG00000046401	RD3L
								70717761	70727303	ENSBTAG00000006673	TMEM179
								70870027	70874415	ENSBTAG00000017622	SIVA1
								70903809	70905041	ENSBTAG00000005038	ZBTB42
								70999273	71004734	ENSBTAG00000015160	PLD4
								71011580	71016820	ENSBTAG00000046828	none
								71041900	71047128	ENSBTAG00000010370	C14orf79
								71052184	71052852	ENSBTAG00000031242	CDCA4
								71077800	71078924	ENSBTAG00000005647	GPR132
								70123917	70233306	ENSBTAG00000020402	TDRD9
								70269910	70292666	ENSBTAG00000017194	ASPG
								70313782	70351845	ENSBTAG00000021904	KIF26A
								70707408	70715785	ENSBTAG00000022775	C14orf180
								70827525	70839732	ENSBTAG00000007187	INF2
								70845971	70861972	ENSBTAG00000017616	ADSSL1
								70878138	70895537	ENSBTAG00000017636	AKT1
								70919570	70983506	ENSBTAG00000004802	CEP170B
chr22	16219978	16342830	16219978	16407075	16219978	16342830	gain	16310518	16348300	ENSBTAG00000016622	TCAIM
								16247929	16308333	ENSBTAG00000044167	TOPAZ1
chr22	19409002	19588936	19409002	19588936	19409002	19588936	gain/loss	18740484	19647747	ENSBTAG00000013047	GRM7
chr22	39545402	39702951	39545402	39657636	39545402	39657636	loss	39175038	40360572	ENSBTAG00000021911	PTPRG
chr22	54440045	61040701	60435042	60508872	60435042	60508872	gain/loss	60443564	60493810	ENSBTAG00000018248	MGLL
								60502719	60507367	ENSBTAG00000018238	ABTB1
								60507746	60556978	ENSBTAG00000018031	PODXL2
			59951940	60243916	59951940	60243916	gain/loss	60016985	60024586	ENSBTAG00000019707	GATA2

								60039910	60040605	ENSBTAG00000006179	DNAJB8
								59952594	59965985	ENSBTAG00000005191	RPN1
								60179708	60211026	ENSBTAG00000020998	RUVBL1
								60217808	60229341	ENSBTAG00000004937	SEC61A1
								60063340	60172924	ENSBTAG00000030962	EEFSEC
chr24	541784	1094942	1027534	1137518	1027534	1094942	gain/loss	1018253	1099817	ENSBTAG00000000656	NFATC1
chr24	39320770	39365195	39320770	39365195	39320770	39365195	gain/loss	39312037	39404644	ENSBTAG00000019251	EPB41L3
chr25	472458	5156189	609241	983759	609241	983759	gain/loss	613234	615345	ENSBTAG00000020198	METRN
								618418	619923	ENSBTAG00000002467	FAM173A
								624818	627214	ENSBTAG00000002481	HAGHL
								627456	633977	ENSBTAG00000002484	NARFL
								649253	652835	ENSBTAG00000000177	MSLN
								665058	668257	ENSBTAG00000000179	RPUSD1
								676665	677047	ENSBTAG00000033526	GNG13
								787493	790981	ENSBTAG00000020737	SOX8
								857167	858273	ENSBTAG00000039974	SSTR5
								865752	867111	ENSBTAG00000033481	C1QTNF8
								871131	877122	ENSBTAG00000012601	TEKT4
								620322	623768	ENSBTAG00000002470	CCDC78
								668529	676616	ENSBTAG00000019743	CHTF18
								724446	775899	ENSBTAG00000019745	LMF1
								959758	984950	ENSBTAG00000026461	CACNA1H
								653105	660655	ENSBTAG00000033580	MSLNL
chr25	36980815	38698430	37988321	38142895	37988321	38142895	gain	38041960	38053172	ENSBTAG00000045896	NPTX2
chr25	38856905	39921068	39286957	39424763	39544407	39570754	gain/loss	39565711	39589412	ENSBTAG00000012049	WIPI2
			39785037	39844749	39785037	39844749	gain/loss	39840013	39840088	ENSBTAG00000047050	bta-mir-2890
								39761774	39816244	ENSBTAG00000019310	FOXK1
			39286957	39424763	39286957	39424763	gain/loss	39292721	39302192	ENSBTAG00000003191	FSCN1
								39308651	39309366	ENSBTAG00000047781	none
								39343633	39347044	ENSBTAG00000026199	ACTB
								39359290	39376730	ENSBTAG00000010264	FBXL18
chr26	5258082	5526925	5472360	5504271	5258082	5288263	loss	5017714	5578654	ENSBTAG00000045905	PCDH15
					5472360	5504271	loss	5017714	5578654	ENSBTAG00000045905	PCDH15
chr26	25501890	26124236	25828973	25982293	25828973	25982293	gain/loss	25856475	25865594	ENSBTAG00000017710	ECHS1
								25872737	25875735	ENSBTAG00000005715	FUOM
								25881752	25885322	ENSBTAG00000013717	PRAP1

								25928928	25930145	ENSBTAG00000046499	none
								25938181	25941636	ENSBTAG00000012416	ZNF511
								25828813	25855778	ENSBTAG00000018321	PAOX
								25893499	25904421	ENSBTAG00000000791	CALY
								25960809	25972557	ENSBTAG00000023832	ADAM8
								25944197	25958640	ENSBTAG00000006395	TUBGCP2
chr27	4357162	5000552	4544917	4773381	4544917	4773381	gain	4677302	4677407	ENSBTAG00000043496	U6
								4766514	4783646	ENSBTAG00000007473	XKR5
								4313369	4554747	ENSBTAG00000011032	MCPH1
								4679600	4727246	ENSBTAG00000004922	AGPAT5
chr28	6334557	6547497	6334557	6547497	6334557	6547497	gain/loss	6492389	6559855	ENSBTAG00000004515	KCNK1
					26994978	27072121	gain/loss	26985668	27080093	ENSBTAG00000021177	ADAMTS14
chr29	35051920	36669359	35136093	35169599	35136093	35169599	gain	35154689	35575203	ENSBTAG00000010032	NTM
chr29	45817015	50999092	48178151	48252404	48178151	48252404	gain/loss	48167168	48194210	ENSBTAG00000006071	CTTN
								48217044	48378574	ENSBTAG00000003171	SHANK2

File 4. Sheet 1. Go and pathways analyses performed using DAVID on line database with high classification stringency option and the false discovery rate (FDR) correction.

Category	ID	Term	P-Value	FDR
Annotation Cluster 1	**Enrichment Score: 2.73**			
GOTERM_BP_FAT	GO:0043434	response to peptide hormone stimulus	1.22E+12	1.9E-1
	GO:0032870	cellular response to hormone stimulus	1.40E+12	2.1E-1
	GO:0009719	response to endogenous stimulus	2.67E+11	4.1E-1
	GO:0009725	response to hormone stimulus	1.46E-03	2.2E0
KEGG_PATHWAY	bta05221	Acute myeloid leukemia	3.29E-03	3.5E0
	bta05220	Chronic myeloid leukemia	5.08E-02	4.3E1
	bta04630	Jak-STAT signaling pathway	6.72E-02	5.3E1
AKT1, GRB2, STAT5A, STAT5B, NR4A2, STAT3				
Annotation Cluster 2	**Enrichment Score: 2.68**			
GOTERM_BP_FAT	GO:0060397	JAK-STAT cascade involved in growth hormone signaling pathway	2.57E+12	3.9E-1
	GO:0060396	growth hormone receptor signaling pathway	8.45E+11	1.3E0
	GO:0060416	response to growth hormone stimulus	8.45E+11	1.3E0
	GO:0007259	JAK-STAT cascade	3.69E-03	5.5E0
	GO:0040014	regulation of multicellular organism growth	5.23E-03	7.7E0
	GO:0019221	cytokine-mediated signaling pathway	2.25E-02	2.9E1
STAT5A, CSF1, STAT5B, STAT3				
Annotation Cluster 3	**Enrichment Score: 1.69**			
GOTERM_BP_FAT	GO:0045137	development of primary sexual characteristics	6.92E-03	1.0E1
	GO:0003006	reproductive developmental process	1.22E-02	1.7E1
	GO:0007548	sex differentiation	1.45E-02	2.0E1
	GO:0046661	male sex differentiation	2.08E-02	2.7E1
	GO:0008406	gonad development	4.87E-02	5.3E1
	GO:0048608	reproductive structure development	6.08E-02	6.2E1
FOXJ1, STAT5A, CSF1, STAT5B, DHCR24				
Annotation Cluster 4	**Enrichment Score: 1.60**			
GOTERM_BP_FAT	GO:0051056	regulation of small GTPase mediated signal transduction	4.08E-03	6.0E0

	GO:0030695	GTPase regulator activity	4.38E-02	4.4E1
	GO:0005083	small GTPase regulator activity	4.52E-02	4.5E1
	GO:0060589	nucleoside-triphosphatase regulator activity	4.92E-02	4.7E1
CSF1, TBC1D5, ASAP2, MGC166429, RAPGEF1, TBC1D22A, TBC1D9B				
Annotation Cluster 5	**Enrichment Score: 1.45**			
GOTERM_BP_FAT	GO:0002763	positive regulation of myeloid leukocyte differentiation	2.32E-03	3.5E0
	GO:0040014	regulation of multicellular organism growth	5.23E-03	7.7E0
	GO:0045639	positive regulation of myeloid cell differentiation	8.35E-03	1.2E1
	GO:0040018	positive regulation of multicellular organism growth	8.35E-03	1.2E1
	GO:0002761	regulation of myeloid leukocyte differentiation	1.33E-02	1.8E1
	GO:0045637	regulation of myeloid cell differentiation	3.56E-02	4.2E1
	GO:0070665	positive regulation of leukocyte proliferation	4.20E-02	4.8E1
	GO:0032946	positive regulation of mononuclear cell proliferation	4.20E-02	4.8E1
	GO:0045927	positive regulation of growth	4.42E-02	5.0E1
	GO:0045597	positive regulation of cell differentiation	5.02E-02	5.4E1
	GO:0032944	regulation of mononuclear cell proliferation	5.58E-02	5.8E1
	GO:0070663	regulation of leukocyte proliferation	5.58E-02	5.8E1
	GO:0051094	positive regulation of developmental process	8.86E-02	7.6E1
	GO:0048872	homeostasis of number of cells	1.02E-01	8.1E1
	GO:0030155	regulation of cell adhesion	1.17E-01	8.5E1
	GO:0008284	positive regulation of cell proliferation	1.71E-01	9.4E1
	GO:0051240	positive regulation of multicellular organismal process	2.53E-01	9.9E1
METRN, STAT5A, CSF1, STAT5B, STAT3				
Annotation Cluster 6	**Enrichment Score: 1.40**			
GOTERM_BP_FAT	GO:0032318	regulation of Ras GTPase activity	1.38E-02	1.9E1
	GO:0043087	regulation of GTPase activity	2.08E-02	2.7E1
	GO:0008047	enzyme activator activity	3.23E-02	3.4E1
	GO:0032313	regulation of Rab GTPase activity	3.56E-02	4.2E1
	GO:0032483	regulation of Rab protein signal transduction	3.56E-02	4.2E1
	GO:0005097	Rab GTPase activator activity	3.67E-02	3.8E1
	GO:0005083	small GTPase regulator activity	4.52E-02	4.5E1
	GO:0005096	GTPase activator activity	5.53E-02	5.2E1
	GO:0005099	Ras GTPase activator activity	6.16E-02	5.6E1
	GO:0051336	regulation of hydrolase activity	1.42E-01	9.0E1

TBC1D5, ASAP2, MGC166429, TBC1D22A, TBC1D9B, NOXA1				
Annotation Cluster 7	**Enrichment Score: 1.11**			
GOTERM_BP_FAT	GO:0006468	protein amino acid phosphorylation	5.68E-02	5.9E1
	GO:0006796	phosphate metabolic process	7.71E-02	7.1E1
	GO:0006793	phosphorus metabolic process	7.71E-02	7.1E1
	GO:0016310	phosphorylation	1.12E-01	8.4E1
PTPN7, AKT1, EPHA4, PTK2, MAPK12, PTPRG, STAT5A, STAT5B, MAPK11, PDE6G, AATK				
Annotation Cluster 8	**Enrichment Score: 1.03**			
GOTERM_BP_FAT	GO:0008344	adult locomotory behavior	2.25E-02	2.9E1
	GO:0030534	adult behavior	6.84E-02	6.6E1
	GO:0007610	behavior	2.04E-01	9.7E1
	GO:0007626	locomotory behavior	2.39E-01	9.8E1
EPHA4, ATP1A3, NR4A2, STAT3				
Annotation Cluster 9	**Enrichment Score: 1.03**			
GOTERM_MF_FAT	GO:0017076	purine nucleotide binding	6.80E-02	5.9E1
	GO:0001883	purine nucleoside binding	7.18E-02	6.1E1
	GO:0001882	nucleoside binding	7.46E-02	6.3E1
	GO:0032555	purine ribonucleotide binding	8.30E-02	6.7E1
	GO:0032553	ribonucleotide binding	8.30E-02	6.7E1
	GO:0000166	nucleotide binding	9.38E-02	7.2E1
	GO:0030554	adenyl nucleotide binding	1.17E-01	8.0E1
	GO:0005524	ATP binding	1.38E-01	8.5E1
	GO:0032559	adenyl ribonucleotide binding	1.45E-01	8.6E1
ACTB, PGS1, ADSSL1, SETD1B, TDRD9, ATP1A3, MAPK11, PDE6G, TPK1, AKT1, EPHA4, NAV1, MAPK12, PARS2, ENTPD8, RUVBL1, EEFSEC, RHOF, UBE2T, DHCR24, AATK				
Annotation Cluster 10	**Enrichment Score: 0.97**			
GOTERM_BP_FAT	GO:0032989	cellular component morphogenesis	2.99E-02	3.7E1
	GO:0000904	cell morphogenesis involved in differentiation	3.39E-02	4.1E1
	GO:0000902	cell morphogenesis	7.64E-02	7.0E1
	GO:0007409	axonogenesis	9.60E-02	7.9E1
	GO:0030182	neuron differentiation	1.03E-01	8.1E1
	GO:0048812	neuron projection morphogenesis	1.05E-01	8.1E1
	GO:0048667	cell morphogenesis involved in neuron differentiation	1.17E-01	8.5E1
	GO:0048858	cell projection morphogenesis	1.23E-01	8.6E1

	GO:0032990	cell part morphogenesis	1.42E-01	9.0E1
	GO:0031175	neuron projection development	1.48E-01	9.1E1
	GO:0048666	neuron development	2.39E-01	9.8E1
	GO:0006928	cell motion	4.03E-01	1.0E2
ACTB, EPHA4, PTK2, NR4A2, NFATC1, STAT3				
Annotation Cluster 11	**Enrichment Score: 0.83**			
GOTERM_BP_FAT	GO:0006417	regulation of translation	8.74E-02	7.5E1
	GO:0010608	posttranscriptional regulation of gene expression	1.85E-01	9.6E1
	GO:0032268	regulation of cellular protein metabolic process	2.00E-01	9.7E1
AKT1, CSF1, EEFSEC, SRP9				
Annotation Cluster 12	**Enrichment Score: 0.81**			
GOTERM_CC_FAT	GO:0005856	cytoskeleton	3.83E-02	3.5E1
	GO:0043228	non-membrane-bounded organelle	3.12E-01	9.8E1
	GO:0043232	intracellular non-membrane-bounded organelle	3.12E-01	9.8E1
AKT1, ACTB, FGF18, CYLC2, PTK2, TUBGCP6, EXOC7, CALD1, NPM1, TEKT4, TUBGCP2, RHOF				
Annotation Cluster 13	**Enrichment Score: 0.76**			
GOTERM_BP_FAT	GO:0009165	nucleotide biosynthetic process	1.65E-01	9.4E1
	GO:0034654	nucleobase, nucleoside, nucleotide and nucleic acid biosynthetic process	1.81E-01	9.5E1
	GO:0034404	nucleobase, nucleoside and nucleotide biosynthetic process	1.81E-01	9.5E1
ADSSL1, ENTPD8, ATP1A3, PRPSAP2				
Annotation Cluster 14	**Enrichment Score: 0.70**			
KEGG_PATHWAY	bta04664	Fc epsilon RI signaling pathway	4.92E-02	4.2E1
	bta04914	Progesterone-mediated oocyte maturation	2.34E-01	9.5E1
	bta04620	Toll-like receptor signaling pathway	2.89E-01	9.8E1
GOTERM_MF_FAT	GO:0004674	protein serine/threonine kinase activity	4.78E-01	1.0E2
AKT1, MAPK12, MAPK11, AATK, GRB2				
Annotation Cluster 15	**Enrichment Score: 0.66**			
GOTERM_BP_FAT	GO:0043066	negative regulation of apoptosis	1.56E-01	9.2E1
	GO:0043069	negative regulation of programmed cell death	1.60E-01	9.3E1
	GO:0060548	negative regulation of cell death	1.60E-01	9.3E1
	GO:0042981	regulation of apoptosis	2.96E-01	1.0E2
	GO:0043067	regulation of programmed cell death	3.03E-01	1.0E2
	GO:0010941	regulation of cell death	3.05E-01	1.0E2

MSX2, SIVA1, STAT5A, STAT5B, NR4A2				
Annotation Cluster 16	**Enrichment Score: 0.62**			
GOTERM_BP_FAT	GO:0019216	regulation of lipid metabolic process	4.87E-02	5.3E1
	GO:0010628	positive regulation of gene expression	1.41E-01	9.0E1
	GO:0045944	positive regulation of transcription from RNA polymerase II promoter	1.58E-01	9.3E1
	GO:0051254	positive regulation of RNA metabolic process	2.17E-01	9.8E1
	GO:0045893	positive regulation of transcription, DNA-dependent	2.17E-01	9.8E1
	GO:0045941	positive regulation of transcription	2.99E-01	1.0E2
	GO:0045935	positive regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	3.61E-01	1.0E2
	GO:0051173	positive regulation of nitrogen compound metabolic process	3.79E-01	1.0E2
	GO:0010557	positive regulation of macromolecule biosynthetic process	4.04E-01	1.0E2
	GO:0031328	positive regulation of cellular biosynthetic process	4.32E-01	1.0E2
	GO:0009891	positive regulation of biosynthetic process	4.39E-01	1.0E2
PPARA, STAT5A, CSF1, STAT5B, NR4A2				
Annotation Cluster 17	**Enrichment Score: 0.49**			
GOTERM_MF_FAT	GO:0003700	transcription factor activity	1.24E-01	8.2E1
	GO:0030528	transcription regulator activity	3.35E-01	9.9E1
GOTERM_BP_FAT	GO:0006355	regulation of transcription, DNA-dependent	3.91E-01	1.0E2
	GO:0051252	regulation of RNA metabolic process	4.11E-01	1.0E2
	GO:0045449	regulation of transcription	5.34E-01	1.0E2
MSX2, PPARA, TRMU, ZFP90, STAT5A, TBX4, STAT5B, NARFL, HDAC10, NR4A2, STAT3, NFATC1				
Annotation Cluster 18	**Enrichment Score: 0.30**			
GOTERM_CC_FAT	GO:0031981	nuclear lumen	4.50E-01	1.0E2
	GO:0070013	intracellular organelle lumen	5.03E-01	1.0E2
	GO:0043233	organelle lumen	5.04E-01	1.0E2
	GO:0031974	membrane-enclosed lumen	5.41E-01	1.0E2
ACTB, FGF18, NPM1, HDAC10, ECHS1, GEMIN7				
Annotation Cluster 19	**Enrichment Score: 0.25**			
GOTERM_BP_FAT	GO:0010558	negative regulation of macromolecule biosynthetic process	5.21E-01	1.0E2
	GO:0031327	negative regulation of cellular biosynthetic process	5.30E-01	1.0E2
	GO:0009890	negative regulation of biosynthetic process	5.47E-01	1.0E2
	GO:0010605	negative regulation of macromolecule metabolic process	6.40E-01	1.0E2

PPARA, HDAC10, SRP9				
Annotation Cluster 20	**Enrichment Score: 0.25**			
GOTERM_MF_FAT	GO:0046872	metal ion binding	5.35E-01	1.0E2
	GO:0043169	cation binding	5.61E-01	1.0E2
	GO:0043167	ion binding	5.82E-01	1.0E2
PPARA, NPLOC4, ADSSL1, TRMU, STAT5A, ZNF296, SCUBE1, STAT5B, ASAP2, NR4A2, CELSR1, KCNK1, STAT3, PRPSAP2, POMGNT1, SQSTM1, ZFP90, ENTPD8, HAGHL, ADAM8, SLC5A10, TBC1D9B, HPD				
Annotation Cluster 21	**Enrichment Score: 0.24**			
GOTERM_MF_FAT	GO:0032561	guanyl ribonucleotide binding	5.08E-01	1.0E2
	GO:0019001	guanyl nucleotide binding	5.15E-01	1.0E2
	GO:0005525	GTP binding	7.44E-01	1.0E2
ADSSL1, EEFSEC, PDE6G, RHOF				
Annotation Cluster 22	**Enrichment Score: 0.19**			
GOTERM_BP_FAT	GO:0006886	intracellular protein transport	5.61E-01	1.0E2
	GO:0034613	cellular protein localization	5.98E-01	1.0E2
	GO:0070727	cellular macromolecule localization	6.00E-01	1.0E2
	GO:0008104	protein localization	6.04E-01	1.0E2
	GO:0015031	protein transport	7.33E-01	1.0E2
	GO:0045184	establishment of protein localization	7.35E-01	1.0E2
	GO:0046907	intracellular transport	7.43E-01	1.0E2
RANBP17, LMF1, SEC61A1, SRP9, DHCR24				
Annotation Cluster 23	**Enrichment Score: 0.14**			
GOTERM_MF_FAT	GO:0005216	ion channel activity	7.11E-01	1.0E2
	GO:0022838	substrate specific channel activity	7.17E-01	1.0E2
	GO:0015267	channel activity	7.24E-01	1.0E2
	GO:0022803	passive transmembrane transporter activity	7.24E-01	1.0E2
GRIK5, CACNA1H, KCNK1				

File 4. Sheet2. Go and pathways analyses performed using DAVID on line database with high classification stringency option and the false discovery rate (FDR) correction.

Category	Term		Genes	FDR	P-Value
GOTERM_BP_FAT	GO:0009719	response to endogenous stimulus	AKT1, GRB2, STAT5A, STAT5B, NR4A2, STAT3	2.70E-04	2.67E+11
	GO:0010033	response to organic substance	MSX2, AKT1, GRB2, STAT5A, STAT5B, NR4A2, EPHX1, STAT3	1.90E-03	1.92E-03
	GO:0051056	regulation of small GTPase mediated signal transduction	CSF1, TBC1D5, ASAP2, MGC166429, RAPGEF1, TBC1D22A, TBC1D9B	4.10E-03	4.08E-03
	GO:0040014	regulation of multicellular organism growth	STAT5A, CSF1, STAT5B, STAT3	5.20E-03	5.23E-03
	GO:0046578	regulation of Ras protein signal transduction	CSF1, TBC1D5, ASAP2, MGC166429, TBC1D22A, TBC1D9B	6.30E-03	6.34E-03
	GO:0045137	development of primary sexual characteristics	FOXJ1, STAT5A, STAT5B, DHCR24	6.90E-03	6.92E-03
	GO:0007167	enzyme linked receptor protein signaling pathway	MSX2, EPHA4, GRB2, STAT5A, STAT5B, STAT3	9.20E-03	9.23E-03
	GO:0007169	transmembrane receptor protein tyrosine kinase signaling pathway	EPHA4, GRB2, STAT5A, STAT5B, STAT3	1.10E-02	1.06E-02
	GO:0003006	reproductive developmental process	FOXJ1, STAT5A, CSF1, STAT5B, DHCR24	1.20E-02	1.22E-02
	GO:0007243	protein kinase cascade	GRB2, STAT5A, STAT5B, PDE6G, STAT3	1.40E-02	1.40E-02
	GO:0007548	sex differentiation	FOXJ1, STAT5A, STAT5B, DHCR24	1.50E-02	1.45E-02
	GO:0000226	microtubule cytoskeleton organization	PTK2, TUBGCP6, TEKT4, TUBGCP2	1.80E-02	1.83E-02
	GO:0007010	cytoskeleton organization	ACTB, PTK2, TUBGCP6, TEKT4, TUBGCP2, RHOF	2.00E-02	1.97E-02
	GO:0046661	male sex differentiation	STAT5A, STAT5B, DHCR24	2.10E-02	2.08E-02
	GO:0030030	cell projection organization	EPHA4, PTK2, BAIAP2, NR4A2, TEKT4	2.20E-02	2.17E-02
	GO:0032989	cellular component morphogenesis	ACTB, EPHA4, PTK2, NR4A2, NFATC1	3.00E-02	2.99E-02

	GO:0000904	cell morphogenesis involved in differentiation	EPHA4, PTK2, NR4A2, NFATC1	3.40E-02	3.39E-02
	GO:0045596	negative regulation of cell differentiation	PPARA, PTK2, STAT5A, STAT5B	3.50E-02	3.50E-02
	GO:0040008	regulation of growth	PTK2, STAT5A, CSF1, STAT5B, STAT3	3.70E-02	3.72E-02
	GO:0006575	cellular amino acid derivative metabolic process	PAOX, STAT5A, STAT5B, NR4A2	4.40E-02	4.35E-02
	GO:0008406	gonad development	FOXJ1, STAT5A, STAT5B	4.90E-02	4.87E-02
	GO:0019216	regulation of lipid metabolic process	PPARA, STAT5A, STAT5B	4.90E-02	4.87E-02
	GO:0045597	positive regulation of cell differentiation	METRN, STAT5A, CSF1, STAT5B	5.00E-02	5.02E-02
	GO:0006468	protein amino acid phosphorylation	AKT1, EPHA4, PTK2, MAPK12, STAT5A, STAT5B, MAPK11, PDE6G, AATK	5.70E-02	5.68E-02
	GO:0048608	reproductive structure development	FOXJ1, STAT5A, STAT5B	6.10E-02	6.08E-02
	GO:0044271	nitrogen compound biosynthetic process	TPK1, ADSSL1, ENTPD8, ATP1A3, NR4A2, PRPSAP2	7.60E-02	7.64E-02
	GO:0000902	cell morphogenesis	EPHA4, PTK2, NR4A2, NFATC1	7.60E-02	7.64E-02
	GO:0008544	epidermis development	AKT1, PPARA, DHCR24	7.60E-02	7.64E-02
	GO:0006793	phosphorus metabolic process	PTPN7, AKT1, EPHA4, PTK2, MAPK12, PTPRG, STAT5A, STAT5B, MAPK11, PDE6G, AATK	7.70E-02	7.71E-02
	GO:0006796	phosphate metabolic process	PTPN7, AKT1, EPHA4, PTK2, MAPK12, PTPRG, STAT5A, STAT5B, MAPK11, PDE6G, AATK	7.70E-02	7.71E-02
	GO:0006357	regulation of transcription from RNA polymerase II promoter	PPARA, STAT5A, STAT5B, HDAC10, NR4A2, STAT3	8.10E-02	8.06E-02
	GO:0007398	ectoderm development	AKT1, PPARA, DHCR24	8.50E-02	8.46E-02
	GO:0006790	sulfur metabolic process	TPK1, STAT5A, STAT5B	8.50E-02	8.46E-02
	GO:0006417	regulation of translation	AKT1, EEFSEC, SRP9	8.70E-02	8.74E-02
	GO:0051094	positive regulation of developmental process	METRN, STAT5A, CSF1, STAT5B	8.90E-02	8.86E-02
	GO:0006631	fatty acid metabolic process	PPARA, STAT5A, STAT5B, ECHS1	9.00E-02	9.04E-02
	GO:0030182	neuron differentiation	EPHA4, PTK2, NR4A2, STAT3	1.00E-01	1.03E-01

	GO:0016310	phosphorylation	AKT1, EPHA4, PTK2, MAPK12, STAT5A, STAT5B, MAPK11, PDE6G, AATK	1.10E-01	1.12E-01
	GO:0007017	microtubule-based process	PTK2, TUBGCP6, TEKT4, TUBGCP2	1.30E-01	1.28E-01
	GO:0010628	positive regulation of gene expression	PPARA, STAT5A, CSF1, STAT5B, NR4A2	1.40E-01	1.41E-01
	GO:0010604	positive regulation of macromolecule metabolic process	AKT1, PPARA, STAT5A, CSF1, STAT5B, NR4A2	1.60E-01	1.56E-01
	GO:0008284	positive regulation of cell proliferation	FGF18, STAT5A, CSF1, STAT5B	1.70E-01	1.71E-01
	GO:0010608	posttranscriptional regulation of gene expression	AKT1, EEFSEC, SRP9	1.80E-01	1.85E-01
	GO:0032268	regulation of cellular protein metabolic process	AKT1, CSF1, EEFSEC, SRP9	2.00E-01	2.00E-01
	GO:0007610	behavior	EPHA4, ATP1A3, NR4A2, STAT3	2.00E-01	2.04E-01
	GO:0032940	secretion by cell	EXOC7, LMF1, SCRN1	2.20E-01	2.19E-01
	GO:0009100	glycoprotein metabolic process	POMGNT1, RPN1, DHCR24	2.30E-01	2.29E-01
	GO:0042127	regulation of cell proliferation	MSX2, FGF18, STAT5A, CSF1, STAT5B	2.50E-01	2.55E-01
	GO:0046903	secretion	EXOC7, LMF1, SCRN1	3.10E-01	3.05E-01
	GO:0009967	positive regulation of signal transduction	FGF18, CSF1, PDE6G	3.20E-01	3.19E-01
	GO:0010647	positive regulation of cell communication	FGF18, CSF1, PDE6G	3.40E-01	3.40E-01
	GO:0006355	regulation of transcription, DNA-dependent	MSX2, PPARA, ZFP90, STAT5A, TBX4, STAT5B, HDAC10, NR4A2, STAT3, NFATC1	3.90E-01	3.91E-01
	GO:0048609	reproductive process in a multicellular organism	CYLC2, STAT5A, STAT5B	4.00E-01	4.00E-01
	GO:0032504	multicellular organism reproduction	CYLC2, STAT5A, STAT5B	4.00E-01	4.00E-01
	GO:0043085	positive regulation of catalytic activity	CSF1, NR4A2, PDE6G	4.10E-01	4.07E-01
	GO:0051252	regulation of RNA metabolic process	MSX2, PPARA, ZFP90, STAT5A, TBX4, STAT5B, HDAC10, NR4A2, STAT3, NFATC1	4.10E-01	4.11E-01

	GO:0042592	homeostatic process	STAT5A, CSF1, STAT5B, NARFL, STAT3	4.20E-01	4.20E-01
	GO:0044093	positive regulation of molecular function	CSF1, NR4A2, PDE6G	4.90E-01	4.86E-01
	GO:0006350	transcription	PPARA, STAT5A, STAT5B, NR4A2, STAT3, NFATC1	4.90E-01	4.89E-01
	GO:0007242	intracellular signaling cascade	GRB2, STAT5A, STAT5B, PDE6G, RHOF, STAT3	5.00E-01	5.00E-01
	GO:0006811	ion transport	ATP1A3, GRIK5, CACNA1H, KCNK1, SLC5A10, NFATC1	5.20E-01	5.23E-01
	GO:0045449	regulation of transcription	MSX2, PPARA, TRMU, ZFP90, STAT5A, TBX4, STAT5B, NARFL, HDAC10, NR4A2, STAT3, NFATC1	5.30E-01	5.34E-01
	GO:0015672	monovalent inorganic cation transport	ATP1A3, KCNK1, SLC5A10	5.80E-01	5.82E-01
	GO:0008104	protein localization	RANBP17, LMF1, SEC61A1, SRP9, DHCR24	6.00E-01	6.04E-01
	GO:0006812	cation transport	ATP1A3, KCNK1, SLC5A10, NFATC1	6.50E-01	6.54E-01
	GO:0016192	vesicle-mediated transport	CALY, EXOC7, SCRN1	6.90E-01	6.91E-01
	GO:0050877	neurological system process	PTK2, ATP1A3, PDE6G	7.00E-01	6.99E-01
	GO:0007166	cell surface receptor linked signal transduction	CALY, GRB2, STAT5A, STAT5B, ATP1A3, GPR132, CELSR1, STAT3, MSX2, EPHA4, SSTR5, WNT7B, GRM7	7.20E-01	7.21E-01
	GO:0015031	protein transport	RANBP17, LMF1, SEC61A1, SRP9	7.30E-01	7.33E-01
	GO:0045184	establishment of protein localization	RANBP17, LMF1, SEC61A1, SRP9	7.30E-01	7.35E-01
	GO:0030001	metal ion transport	KCNK1, SLC5A10, NFATC1	7.50E-01	7.51E-01
	GO:0009057	macromolecule catabolic process	AKT1, UBE2T, DHCR24	8.20E-01	8.15E-01
	GO:0055085	transmembrane transport	CACNA1H, SLC5A10, SEC61A1	8.80E-01	8.84E-01
	GO:0006508	proteolysis	ADAM8, UBE2T, DHCR24	9.80E-01	9.82E-01
	GO:0007186	G-protein coupled receptor protein signaling pathway	SSTR5, CALY, GRM7, GPR132, CELSR1	1.00E+00	9.97E-01
GOTERM_CC_FAT	GO:0031252	cell leading edge	AKT1, CTTN, PTK2, BAIAP2	6.60E-03	6.59E-03
	GO:0005938	cell cortex	ACTB, CTTN, EXOC7, CALD1	2.10E-02	2.13E-02
	GO:0030027	lamellipodium	AKT1, CTTN, PTK2	2.40E-02	2.39E-02

	GO:0042995	cell projection	AKT1, CTTN, PTK2, BAIAP2, FSCN1, TEKT4	3.30E-02	3.27E-02
	GO:0005856	cytoskeleton	AKT1, ACTB, CYLC2, PTK2, TUBGCP6, EXOC7, CALD1, TEKT4, TUBGCP2, RHOF	3.80E-02	3.83E-02
	GO:0044448	cell cortex part	ACTB, EXOC7, CALD1	5.70E-02	5.66E-02
	GO:0005819	spindle	AKT1, TUBGCP6, TUBGCP2	8.80E-02	8.84E-02
	GO:0015630	microtubule cytoskeleton	AKT1, TUBGCP6, EXOC7, TEKT4, TUBGCP2	1.00E-01	9.98E-02
	GO:0044430	cytoskeletal part	AKT1, CYLC2, TUBGCP6, EXOC7, CALD1, TEKT4, TUBGCP2	1.00E-01	1.03E-01
	GO:0005815	microtubule organizing center	TUBGCP6, EXOC7, TUBGCP2	1.50E-01	1.54E-01
	GO:0048471	perinuclear region of cytoplasm	CYLC2, ATXN10, CSF1	2.10E-01	2.06E-01
	GO:0043228	non-membrane-bounded organelle	AKT1, ACTB, FGF18, CYLC2, PTK2, TUBGCP6, EXOC7, CALD1, NPM1, TEKT4, TUBGCP2, RHOF	3.10E-01	3.12E-01
	GO:0043232	intracellular non-membrane-bounded organelle	AKT1, ACTB, FGF18, CYLC2, PTK2, TUBGCP6, EXOC7, CALD1, NPM1, TEKT4, TUBGCP2, RHOF	3.10E-01	3.12E-01
	GO:0005654	nucleoplasm	ACTB, NPM1, HDAC10, GEMIN7	3.70E-01	3.66E-01
	GO:0031981	nuclear lumen	ACTB, FGF18, NPM1, HDAC10, GEMIN7	4.50E-01	4.50E-01
	GO:0012505	endomembrane system	POMGNT1, GRB2, SCRN1, SEC61A1	4.60E-01	4.62E-01
	GO:0005886	plasma membrane	AKT1, PTK2, CALY, CALD1, FSCN1, ENTPD8, GRIK5, CELSR1, RHOF, NTM, STAT3	4.90E-01	4.89E-01
	GO:0044451	nucleoplasm part	ACTB, HDAC10, GEMIN7	5.50E-01	5.49E-01
	GO:0000267	cell fraction	ACTB, CALD1, ENTPD8	5.70E-01	5.73E-01
	GO:0005783	endoplasmic reticulum	PGS1, LMF1, RPN1, SEC61A1	7.00E-01	6.98E-01
	GO:0005739	mitochondrion	PGS1, TRMU, TSPO, AGPAT5, ECHS1, MP68	7.10E-01	7.14E-01
	GO:0031090	organelle membrane	POMGNT1, GRB2, SCRN1, SEC61A1	8.20E-01	8.22E-01
	GO:0005576	extracellular region	TG, FGF18, WNT7B, METRN, FOXJ1, IL1RN	8.60E-01	8.61E-01
	GO:0031224	intrinsic to membrane	TSPO, CALY, CSF1, LMF1, SPPL2B, GRIK5, ATP1A3, GPR132, KCNK1, EPHA4, SSTR5, POMGNT1, TSPAN10, TECRL,	8.80E-01	8.85E-01

			GRM7, ENTPD8, RPN1, CACNA1H, SLC5A10, SEC61A1, NTM, AATK		
	GO:0016021	integral to membrane	TSPO, CALY, CSF1, LMF1, SPPL2B, ATP1A3, GRIK5, GPR132, KCNK1, EPHA4, SSTR5, POMGNT1, TSPAN10, TECRL, GRM7, ENTPD8, RPN1, CACNA1H, SLC5A10, SEC61A1, AATK	9.00E-01	8.98E-01
	GO:0044459	plasma membrane part	PTK2, CALY, GRIK5, RHOF	9.40E-01	9.37E-01
GOTERM_MF_FAT	GO:0008047	enzyme activator activity	NOXA1, TBC1D5, ASAP2, TBC1D22A, TBC1D9B	3.20E-02	3.23E-02
	GO:0030695	GTPase regulator activity	TBC1D5, ASAP2, MGC166429, RAPGEF1, TBC1D22A, TBC1D9B	4.40E-02	4.38E-02
	GO:0005083	small GTPase regulator activity	TBC1D5, ASAP2, MGC166429, TBC1D22A, TBC1D9B	4.50E-02	4.52E-02
	GO:0060589	nucleoside-triphosphatase regulator activity	TBC1D5, ASAP2, MGC166429, RAPGEF1, TBC1D22A, TBC1D9B	4.90E-02	4.92E-02
	GO:0019904	protein domain specific binding	PTK2, GRB2, SQSTM1, BAIAP2	6.30E-02	6.27E-02
	GO:0017076	purine nucleotide binding	ACTB, PGS1, ADSSL1, TDRD9, ATP1A3, MAPK11, PDE6G, TPK1, AKT1, EPHA4, MAPK12, PARS2, ENTPD8, RUVBL1, EEFSEC, RHOF, UBE2T, DHCR24, AATK	6.80E-02	6.80E-02
	GO:0001883	purine nucleoside binding	ACTB, PGS1, TDRD9, ATP1A3, MAPK11, PDE6G, TPK1, AKT1, EPHA4, MAPK12, PARS2, ENTPD8, RUVBL1, UBE2T, DHCR24, AATK	7.20E-02	7.18E-02
	GO:0001882	nucleoside binding	ACTB, PGS1, TDRD9, ATP1A3, MAPK11, PDE6G, TPK1, AKT1, EPHA4, MAPK12, PARS2, ENTPD8, RUVBL1, UBE2T, DHCR24, AATK	7.50E-02	7.46E-02
	GO:0032553	ribonucleotide binding	ACTB, PGS1, ADSSL1, TDRD9, ATP1A3, MAPK11, PDE6G, TPK1, AKT1, EPHA4, MAPK12, PARS2, ENTPD8, RUVBL1, EEFSEC, RHOF, UBE2T, AATK	8.30E-02	8.30E-02
	GO:0032555	purine ribonucleotide binding	ACTB, PGS1, ADSSL1, TDRD9, ATP1A3, MAPK11, PDE6G, TPK1, AKT1, EPHA4,	8.30E-02	8.30E-02

			MAPK12, PARS2, ENTPD8, RUVBL1, EEFSEC, RHOF, UBE2T, AATK		
	GO:0000166	nucleotide binding	ACTB, PGS1, ADSSL1, SETD1B, TDRD9, ATP1A3, MAPK11, PDE6G, TPK1, AKT1, EPHA4, NAV1, MAPK12, PARS2, ENTPD8, RUVBL1, EEFSEC, RHOF, UBE2T, DHCR24, AATK	9.40E-02	9.38E-02
	GO:0030554	adenyl nucleotide binding	ACTB, PGS1, TDRD9, ATP1A3, MAPK11, TPK1, AKT1, EPHA4, MAPK12, PARS2, ENTPD8, RUVBL1, UBE2T, DHCR24, AATK	1.20E-01	1.17E-01
	GO:0003700	transcription factor activity	MSX2, PPARA, STAT5A, TBX4, STAT5B, NR4A2, STAT3, NFATC1	1.20E-01	1.24E-01
	GO:0005524	ATP binding	ACTB, PGS1, TDRD9, ATP1A3, MAPK11, TPK1, AKT1, EPHA4, MAPK12, PARS2, ENTPD8, RUVBL1, UBE2T, AATK	1.40E-01	1.38E-01
	GO:0032559	adenyl ribonucleotide binding	ACTB, PGS1, TDRD9, ATP1A3, MAPK11, TPK1, AKT1, EPHA4, MAPK12, PARS2, ENTPD8, RUVBL1, UBE2T, AATK	1.40E-01	1.45E-01
	GO:0008083	growth factor activity	FGF18, FOXJ1, CSF1	2.10E-01	2.07E-01
	GO:0016563	transcription activator activity	PPARA, NR4A2, STAT3	2.60E-01	2.65E-01
	GO:0030528	transcription regulator activity	MSX2, PPARA, STAT5A, TBX4, STAT5B, HDAC10, NR4A2, STAT3, NFATC1	3.30E-01	3.35E-01
	GO:0005509	calcium ion binding	STAT5A, SCUBE1, ENTPD8, STAT5B, CELSR1, STAT3, TBC1D9B	3.60E-01	3.57E-01
	GO:0016879	ligase activity, forming carbon-nitrogen bonds	ADSSL1, UBE2T, TTLL12	3.90E-01	3.89E-01
	GO:0019899	enzyme binding	SQSTM1, HDAC10, STAT3	4.70E-01	4.65E-01
	GO:0004674	protein serine/threonine kinase activity	AKT1, MAPK12, MAPK11, AATK	4.80E-01	4.78E-01
	GO:0032561	guanyl ribonucleotide binding	ADSSL1, EEFSEC, PDE6G, RHOF	5.10E-01	5.08E-01
	GO:0019001	guanyl nucleotide binding	ADSSL1, EEFSEC, PDE6G, RHOF	5.10E-01	5.15E-01
	GO:0004672	protein kinase activity	AKT1, EPHA4, MAPK12, MAPK11, AATK	5.50E-01	5.49E-01
	GO:0043565	sequence-specific DNA binding	MSX2, PPARA, NR4A2, NFATC1	5.60E-01	5.62E-01
	GO:0046983	protein dimerization activity	CSF1, NR4A2, STAT3	6.00E-01	6.02E-01

	GO:0008092	cytoskeletal protein binding	BAIAP2, CALD1, FSCN1	6.00E-01	6.02E-01
	GO:0003723	RNA binding	NPM1, RPUSD1, EEFSEC, SRP9	6.20E-01	6.18E-01
	GO:0003677	DNA binding	MSX2, PPARA, TRMU, STAT5A, TBX4, STAT5B, NR4A2, SOX8, STAT3, NFATC1	6.30E-01	6.25E-01
	GO:0042802	identical protein binding	ACTB, PTPRG, CSF1	6.70E-01	6.72E-01
	GO:0005525	GTP binding	ADSSL1, EEFSEC, RHOF	7.40E-01	7.44E-01
	GO:0008270	zinc ion binding	PPARA, NPLOC4, TRMU, SQSTM1, ZFP90, ZNF296, NR4A2, ASAP2, HAGHL, ADAM8	9.00E-01	8.99E-01
	GO:0046914	transition metal ion binding	PPARA, NPLOC4, TRMU, POMGNT1, SQSTM1, ZFP90, ZNF296, NR4A2, ASAP2, HAGHL, ADAM8, HPD	9.30E-01	9.30E-01
KEGG_PATHWAY	bta04370	VEGF signaling pathway	AKT1, PTK2, MAPK12, MAPK11, NFATC2, NFATC1	1.30E-03	1.30E-03
	bta04660	T cell receptor signaling pathway	AKT1, MAPK12, GRB2, MAPK11, NFATC2, NFATC1	6.40E-03	6.37E-03
	bta04012	ErbB signaling pathway	AKT1, PTK2, GRB2, STAT5A, STAT5B	1.10E-02	1.14E-02
	bta05200	Pathways in cancer	AKT1, FGF18, WNT7B, PTK2, RASSF5, GRB2, STAT5A, STAT5B, STAT3	2.00E-02	1.97E-02
	bta04010	MAPK signaling pathway	PTPN7, AKT1, FGF18, MAPK12, GRB2, CACNA1H, MAPK11, NFATC2	2.50E-02	2.49E-02
	bta04670	Leukocyte transendothelial migration	ACTB, PTK2, RASSF5, MAPK12, MAPK11	3.70E-02	3.73E-02
	bta04360	Axon guidance	EPHA4, PTK2, PLXNB2, NFATC2, NFATC1	3.90E-02	3.94E-02
	bta04662	B cell receptor signaling pathway	AKT1, GRB2, NFATC2, NFATC1	4.30E-02	4.27E-02
	bta04062	Chemokine signaling pathway	AKT1, PTK2, GRB2, STAT5B, GNG13, STAT3	4.30E-02	4.32E-02
	bta04722	Neurotrophin signaling pathway	AKT1, MAPK12, GRB2, MAPK11, RAPGEF1	4.40E-02	4.39E-02
	bta04664	Fc epsilon RI signaling pathway	AKT1, MAPK12, GRB2, MAPK11	4.90E-02	4.92E-02
	bta05220	Chronic myeloid leukemia	AKT1, GRB2, STAT5A, STAT5B	5.10E-02	5.08E-02
	bta05223	Non-small cell lung cancer	AKT1, RASSF5, GRB2	1.10E-01	1.10E-01
	bta04510	Focal adhesion	AKT1, ACTB, PTK2, GRB2, RAPGEF1	1.50E-01	1.50E-01
	bta04920	Adipocytokine signaling pathway	AKT1, PPARA, STAT3	1.60E-01	1.59E-01

	bta04530	Tight junction	AKT1, ACTB, EPB41L3, CTTN	1.60E-01	1.60E-01
	bta04810	Regulation of actin cytoskeleton	ACTB, FGF18, ENAH, PTK2, BAIAP2	1.70E-01	1.67E-01
	bta05211	Renal cell carcinoma	AKT1, GRB2, RAPGEF1	1.70E-01	1.68E-01
	bta04910	Insulin signaling pathway	AKT1, EXOC7, GRB2, RAPGEF1	1.70E-01	1.68E-01
	bta04310	Wnt signaling pathway	WNT7B, RUVBL1, NFATC2, NFATC1	2.20E-01	2.17E-01
	bta04914	Progesterone-mediated oocyte maturation	AKT1, MAPK12, MAPK11	2.30E-01	2.34E-01
	bta04912	GnRH signaling pathway	MAPK12, GRB2, MAPK11	2.70E-01	2.68E-01
	bta04620	Toll-like receptor signaling pathway	AKT1, MAPK12, MAPK11	2.90E-01	2.89E-01
	bta04650	Natural killer cell mediated cytotoxicity	GRB2, NFATC2, NFATC1	3.30E-01	3.35E-01
	bta00230	Purine metabolism	ADSSL1, ENTPD8, PDE6G	5.10E-01	5.08E-01
	bta04080	Neuroactive ligand-receptor interaction	SSTR5, TSPO, GRM7, GRIK5	5.10E-01	5.14E-01

2. A genome-wide scan of copy number variants using high-density SNPs in Brown Swiss dairy cattle

Published in: Livestock Science 191: 153 - 160 · August 2016. DOI: 10.1016/j.livsci.2016.08.006.

R.T.M.M. Prinsen[a], M.G. Strillacci[a], F. Schiavini[a], E. Santus[b], A. Rossoni[b], V. Maurer[c], A. Bieber[c], B. Gredler[d], M. Dolezal[e], A. Bagnato[a]

[a] Università degli Studi di Milano, Department of Veterinary Medicine, Via Celoria 10, MI 20133. Italy
raphaelle.prinsen@unimi.it, maria.strillacci@unimi.it, fausta.schiavini@unimi.it, alessandro.bagnato@unimi.it

[b] Associazione Nazionale Allevatori Razza Bruna, Località Ferlina, Bussolengo, 204, VR 37012, Italy
enrico.santus@anarb.it, attilio.rossoni@anarb.it

[c] Research Institute of Organic Agriculture, Ackerstrasse 113, 219 Frick, Switzerland
anna.bieber@fibl.org, veronika.maurer@fibl.org

[d] Qualitas AG, Chamerstrasse 56, 6300 Zug Switzerland
Birgit.Gredler@qualitasag.ch

[e] University of Veterinary Medicine, Veterinärplatz 1, 1210 Wien, Austria
marlies.dolezal@gmail.com

Abstract

Detecting genetic variation such as Copy Number Variants (CNVs) in cattle provides the opportunity to study their association with quantitative traits. CNVs are DNA sequences of 50 bp up to several Mb long, which can vary in copy number in comparison with a reference genome. The aim of this study was to investigate CNVs in 1,410 samples of the Brown Swiss cattle breed using Illumina Bovine HD SNP chip information, which includes 777,962 SNPs. After stringent quality control, CNVs were called with the Golden Helix SVS 8.3.1 (SVS) and PennCNV software and were summarized to CNV regions (CNVRs) at a population level (i.e. overlapping CNVs), using BEDTools. Additionally, common CNVRs between the two software were set as consensus regions. Genes within consensus CNVRs were annotated with a GO analysis using the DAVID Bioinformatics Resources 6.7. In order to validate these results, quantitative PCRs were executed on 15 selected CNVRs.

The SVS software identified 25,030 CNVs summarized to 398 CNVRs, which comprised 30 gains, 344 losses and 24 complex CNVRs (i.e. containing both losses and gains), covering 3.92% of the bovine genome. The PennCNV software identified 6,2341 CNVs summarized to 5,578 CNVRs, which comprised 2,638 gains, 2,404 losses and 537 complex CNVRs, covering 7.68% of the bovine genome. The length of these CNVRs ranged from 1,244 bp to 1,381,355 bp. A total of 563 consensus CNVRs were found covering 2.29 % of the UMD 3.1 bovine genome assembly. Of these, 24 were gains, 300 were losses and 239 were complex CNVRs. A total of 775 official gene IDs were annotated in the consensus CNVRs. Among the 537 genes with functional

information, the GO and pathway analysis was reported for those who clustered with a p-value < 0.05. The quantitative PCRs successfully validated 14 (93.33%) of the selected CNVRs.

The result of this study is the first comprehensive genomic analysis of the Brown Swiss breed based on the Illumina Bovine HD SNP chip on such a large number of animals that enriches the CNV map in the bovine genome. These findings also provide valuable information for further CNV studies. Finally, the results of the CNVR map delivers new information for functional, health and productive traits considered in selection programs of the Brown Swiss breed.

Introduction

The discovery of genetic variation in cattle, such as within the Brown Swiss population, is fundamental to associate genomic regions to economically and functionally important traits and to improve the ongoing genomic selection practices. The Brown Swiss dual-purpose breed is known for its excellent milk and meat production. This breed is one of the most represented breeds in Italy but is also bred in other 60 countries all over the world.

Current genomic selection relies solely on SNPs (Meuwissen et al., 2001) and for the Brown Swiss breed this is mainly delivered by the Intergenomics consortium (which shares single-country Brown Swiss genomic evaluations) (Zumbach et al., 2010). Nevertheless other structural variations such as Copy Number Variants (CNVs) may be additional markers, which should be considered in genomic selection evaluation procedures.

CNVs are DNA segments of 50 basepairs (bp) up to several megabases (Mb) large which vary in copy number in comparison with a reference genome (Mills et al., 2011). This variation accounts for roughly 4.6% of the bovine genome (Hou et al., 2011). CNVs have been shown to be associated with complex traits in several species. In horse, sheep and pig coat color is partially determined by CNVs (Clop et al., 2012). Also traits and diseases such as cattle milk production (Xu et al., 2014) and female fertility failure (Kadri et al., 2014) have been shown to be influenced by CNVs. Meyers et al. (2010), demonstrated that osteoporosis in Angus cattle is triggered by a CNV deletion.

Even if CNVs are less frequent than SNPs in terms of absolute numbers, CNVs cover more of the genome and have therefore potentially much bigger phenotypic effects (Yang et al., 2015). CNVs can shape individuals' genomes by changing the gene structure, the gene dosage, and the genes' regulation, which in turn has an impact on the gene expression influencing in this way the final phenotypes (Zhang et al., 2009).

The first CNV detection studies used to employ a single algorithm to identify CNVs. Later, Winchester et al., (2009) showed that the consensus calls between different CNV detection methods is less than 50%. Hence, the detection of reliable CNVs necessitates the usage of at least two algorithms since this reduces the odds to have eventual false positives among the final CNV dataset (Zhao et al., 2013) at the cost of a possibly increased false negative detection rate.

To date, there is still no high-resolution CNV map available in a representative sample across Brown Swiss populations. This study therefore aims to expand the catalogue of CNV regions (CNVRs) in the

bovine genome and delivers a high-resolution map of CNVRs specific to Brown Swiss dairy cattle.

The objective of this study was to investigate genomic structural variation in 1,410 individuals of the Brown Swiss dairy cattle breed using Illumina Bovine HD SNP chip data using two different CNV-detection algorithms: the CNAM of the Golden Helix SVS 8.3.1 (SVS) software and the Hidden Markov Model of the PennCNV software. Consensus CNVRs were validated via quantitative PCRs.

Material and methods

Sampling and genotyping

This study used available Illumina Bovine HD BeadChip information from a total of 1,410 individuals, 496 males and 914 females, belonging to the Brown Swiss dairy cattle breed. Out of these 1,410 samples, 379 samples were genotyped and provided by Braunvieh Schweiz and 812 samples were genotyped and provided by the EC funded LowInputBreeds project. The University of Milan provided the remaining 219 samples genotyped within the EC funded Quantomics project. The Illumina Bovine HD BeadChip features 777,962 loci in the bovine genome with an average SNP spacing of less than 3 Kb. In the present study, the UMD3.1 bovine assembly was used as reference genome.

Quality assurance of CNV raw data

To perform the CNV detection with the SVS and PennCNV software, two metrics were used: the Log R Ratio (LRR), and the B Allele Frequency (BAF) (Peiffer et al., 2006). The CNV detection was performed on a total of 735,239 SNPs on the 29 bovine autosomes.

Quality assurance of LRR raw data and filtering of outlier samples was performed with the SVS software. The complete distribution of the derivative log ratio spread (DLRS) was used to detect and filter outlier samples as described by Pinto et al., (2011). Additionally, samples that had an absolute wave factor greater than 0.05, (as suggested by Diskin et al., 2008) were excluded. A Principal component analysis (PCA) has then been executed to detect the presence of batch effects and correct the LRR values. Therefore, to determine the number of principal components for the correction of the LRR values' variance, a graphical approach was employed.

After checking the quality of the raw data, a total of 294 outlier samples have been identified and excluded from CNV detection (i.e. 1116 used for the CNV mapping).

Copy Number Variation detection

The CNV detection was performed with two different algorithms in order to reduce eventual false positive calls. In particular, the Copy Number Analysis Module (CNAM) implemented in the SVS software (Golden Helix, Bozeman, MT, http://goldenhelix.com) and the PennCNV software (Wang et al., 2008) were used. CNV detection was performed using the univariate analysis in CNAM of the SVS software with the following options: univariate outlier removal, a limit of not more than 100 segments per 10,000 markers with a minimum of 1 marker per segment, and 2,000 permutations per pair with a p-value cutoff of 0.005. Detection with the PennCNV software was performed on the bovine autosomes employing the default PennCNV parameters.

Subsequently, samples were filtered whenever they had a standard deviation of LRR > 0.35 or when they had more than 200 CNVs.

CNVRs and consensus CNVRs definition

CNVRs were defined, as suggested by Redon et al. (2006), within SVS and PennCNV results. More specifically, whenever CNVs overlapped by at least 1 bp, they were summarized as a single CNVR. This has been accomplished using the BEDTools software (-mergeBed command) (Quinlan and Hall, 2010) and CNVRs were catalogued as gain, loss and complex CNVRs (i.e. CNVRs comprising both gain and loss events). A consensus analysis was then obtained with the BEDTools software (-intersectBed command) among the CNVRs identified within the SVS and PennCNV following the approach described by Wain et al., (2009), which identifies CNVRs that fully overlap each other.

Consensus CNVR annotation

The full Ensembl v84 autosomal UMD3.1 gene set was downloaded from the http://www.ensembl.org/biomart/martview/32471e73613a45753e5689f2626f9add website. Annotated bovine genes within consensus CNVRs were found using the intersectBed command of BEDTools (Quinlan and Hall, 2010). A gene ontology (GO) and pathway analysis was performed employing the DAVID online Bioinformatics Resources 6.7 (https://david.ncifcrf.gov/tools.jsp). The following options were selected: a high classification stringency and a false discovery rate correction (FDR).

To further examine the consensus CNVRs, quantitative trait loci (QTL) that overlap with the consensus CNVRs were identified by downloading the QTL list from the animal QTL database (http://www.animalgenome.org/cgi-bin/QTLdb/index, on the UMD3.1 cattle genome assembly database). We filtered the QTL which are larger than 5 Mb. Only QTL overlapping for at least 50% with the consensus CNVRs were considered.

CNVR validation by qPCR

Quantitative PCR (qPCR) experiments were performed to validate 15 randomly chosen consensus CNVRs using samples whose genomic DNA was available. The BTF3 gene was chosen as reference gene for the qPCR experiments (Bae et al., 2010). The primers for the reference gene and for the CNVRs were designed through the Primer Express® Software v3.0.1 (Life Technologies™) with the MGB quantification parameters. The primer and probe sequences were tested for possible hairpins, and self or cross dimer formations with the Oligo Calc calculator (http://www.basic.northwestern.edu/biotools/OligoCalc.html). The sequence similarity was verified utilizing the BLAT alignment tool of the UCSC database (http://genome.ucsc.edu). The qPCR experiments were run in quadruple using the qPCR protocol described by the TaqMan® Copy Number Assays kit (Life tecnologies™) on a 7500 Fast Real-time PCR System instrument (Applied Biosystems). The crossing thresholds cycle (Ct) for every sample tested was analyzed with the CopyCaller™ software (Applied Biosystems).

Results and discussion

Quality assurance of CNV Raw Data

SVS identified 294 samples that did not pass the quality control for DLRS and for Absolute Wave Factors (98 samples both for their DLRS and their Absolute Wave Factor values). Additionally, after CNV mapping, PennCNV identified a total of 85 outlier samples which had more than 200 CNVs and a LRR standard deviation greater than 0.30. Therefor a total of 379 outlier samples have been identified and removed reducing the number of samples to 1,031.

CNVs detection

The CNAM univariate segmentation of the SVS detected 25,030 CNVs, corresponding to 837 CNV gains and 24,193 CNV losses. The HMM of the PennCNV software detected a total of 62,341 CNVs: 19,968 gains and 42,373 losses. Table 1 reports the frequency of CNVs identified (the minimum and maximum, the number of CNV losses and CNV gains), the mean and median values, as well as the CNV coverage related to the bovine UMD 3.1 autosome assembly, for each software.

The PennCNV software identified for all the 1,031 samples CNV gains and a CNV losses, while the SVS software identified for all the 1,031 samples CNV losses but not for all of them (45%) CNV gains. There is a high percentage of CNV singletons among the SVS detection results: on a total of 567 samples with gain CNVs, 379 are singletons, while the remaining 188 samples have a number of CNVs which varies from 2 to 6. The average number of CNVs per sample is 60.47 for the PennCNV (19.37 and 41.10 mean values for gain and loss, respectively), while 24.23 is the average number of CNVs per sample for the SVS results

(1,48 and 23.42 mean values for gain and loss, respectively). As expected both tools detected more losses than gains. This is due to the higher power in methodology to identify in particular homozygous losses, where both copies of a locus are lost. The deletion ratios (calculated as the total number of losses divided by total number of CNV detected) were 96.66% and 67.97% for SVS and PennCNV, respectively. Moreover, as shown in Table 1, even if the number of the CNV gain events is considerably lower both in number as in frequency per sample for the SVS detection results (3.34% of all the SVS CNVs) compared to the PennCNV detection results (32.03 % of all the PennCNV CNVs), the CNVs detected with the SVS software are generally longer. This is also the case for the gain events, which are much less in the SVS results compared to the PennCNV detection results.

CNVRs and consensus CNVRs definition

A total of 398 CNVRs were found after summarizing the CNVs identified with the SVS software (CNVR mean per BTA = 13.72) (Table 2).

In this regards, the total autosomal CNVRs coverage was 9,846,683 bp, which corresponds to 3.92% of the bovine autosomal genome. More specifically, the autosomal CNVRs coverage varied from a minimum of 0.39% on BTA 6, to a maximum of 16.75% on BTA 18. The BTA with the largest number of CNVRs identified was BTA 7, which had 28 (7%) CNVRs. On the opposite, the BTA with the smallest number of CNVRs was BTA 18, which had only 5 (1.3%) CNVRs (Figure 1).

The CNVs identified with the PennCNV software, were summarized to 5,578 CNVRs (Table 2) at a population level (CNVR mean per BTA = 192.34). In this case, the total amount of autosomes covered by CNVRs was equal to 193,063,783 bp, which is equivalent to a coverage of 7.68% on the bovine autosomes genomes. In particular, the sequence covered by CNVRs went from a minimum of 3.63 (BTA 8) to a maximum of 18.84 (BTA 19). The largest number of CNVRs identified is located on BTA 1 and these were 435 (7.08%). On the contrary, the BTA with the smallest number of CNVRs was BTA 28, which had 76 (1.4%) CNVRs (Figure 2).

The consensus analysis across the SVS PennCNV CNVRs allowed the identification of a total of 563 consensus CNVRs (Figure 2, Table 2, Table S1). Among them, there were 24 gains, 301 losses and 238 CNVRs of complex nature (see Table 2). As shown in Table 2 the number of complex consensus CNVRs is much greater than the number of CNVRs of SVS. This does not hold for gain and loss consensus CNVRs. The reason therefore holds in the fact that the average length of complex CNVRs in SVS is much larger than in PennCNV complex regions. Additionally, regions identified as gain in one software and loss in the other one, are classified as complex CNVRs in the consensus map.

The number of CNVRs results considerably smaller in dimension in comparison with the original SVS and PennCNV datasets, not only because the two software identify slightly different regions, but mostly because this consensus analyses is performed with the Wain method (Wain et al., 2009), which only keeps the intersection sequence of the overlapping CNVRs. This procedure aims to report only the most likely

CNVRs as results (i.e. for the two CNVRs that overlap, more than one consensus CNVR can be obtained).

The length of the consensus CNVRs ranged from approximately 1.31 Kb to 4.32 Mb with an average CNVRs size of 102.39 Kb (see Table 2). These consensus CNVRs spanned 13.59 Mb of the bovine autosomes and accounted for approximately 2.9% of the genome bovine sequence. The differences in consensus CNVR numbers per BTA were evident (Figure 1), ranging from 5 (0.9%) to 50 (8.9%) on the BTA27 and BTA1, respectively. The distribution of CNVRs on the chromosomes identified by SVS, PennCNV and the consensus analysis was not uniform. For some chromosomes the proportion of CNVRs identified is similar (11, 13, 14, 16, 21, 25 and 28) for every detection, while for other chromosomes there is a nearly perfect match between the consensus CNVRs and the PennCNV CNVRs (2, 11, 16, 21, 25, 26, 27), and consensus CNVRs and SVS CNVRs (3, 10, 12, 13, 17 and 29).

In addition, given the difference in consensus CNVR proportion described above for each software and for the consensus CNVRs, a comparison among CNV lengths was performed. The PennCNV, SVS and consensus CNVRs were divided according to their length into five classes (for every state) of 1–5 Kb, 5–10 Kb, 10–100 Kb, 100–1000 Kb and >100 Kb. Figure 3 reports the CNVR count (expressed as percentage of the total number of CNVRs of every type of detection) for each class of CNVR length. The majority of CNVRs identified in this study have a length comprised between 10 Kb and 1000 Kb (Figure 3), mainly for the CNVRs found by PennCNV (for every CNVR state). CNVRs detected by CNAM, the highest number of gain CNVRs are

those with a length of 100-1000 Kb, while the complex CNVRs have generally a length of 100-1000 and >1000 Kb (Figure 3).

Figure 4A shows the consensus CNVR details concerning the distribution frequency of the samples that fall within a specific length class (as shown in Figure 3). The consensus CNVR count for each software has been performed using the single software CNV detection characteristics. For both the PennCNV and SVS CNVRs that defined the consensus CNVRs, the most represented CNVRs are loss regions, which mostly have a length that falls within the 10 -100 Kb class. Figure 4B shows the features of the sample count (6 classes) for every CNVR state according to previously defined CNVR length classes. The sample count classes were defined as: 1, 2-5, 5-15, 16-50, 51-500 and >501. The loss consensus CNVRs have a similar sample count distribution for the SVS and PennCNV CNVRs, with most regions falling into the 10-100 Kb class. Furthermore, for the SVS regions, class 1 is the most frequent in all length classes. As for the SVS loss regions, also the SVS gain regions are mostly represented by class 1. What regards the PennCNV regions, class 1 is the most frequent among the first three length classes. More precisely, the most represented sample class is the 2-5 class falling mainly within the 10-100 Kb length class. The highest length and sample classes mainly belong to the SVS complex regions. Lastly, all the sample classes are distributed mostly within the 10-100 and 100-1000 Kb length classes.

To evaluate our results, a comparison between our consensus CNVRs and those identified in previous studies was performed. In particular, only studies that were performed with high-density SNP chips (except for one study on the Brown Swiss breed which used the 50K SNP

chip), as well as recent studies and one study which employed the sequencing method for the CNV detection, were considered (Table 3).
Besides from the breeds analyzed, the CNVRs detection results' mainly rely on the type of detection algorithms employed as well as the technology used. Indeed, the CNVRs found in Bagnato et al., (2015) in the Brown Swiss breed, which uses a 50K BeadChip, overlap for only 18% (27 CNVRs) with the consensus CNVRs found in the study.
As reported in Table 3, the percentage of overlapping CNVRs found in other studies ranges from 5.15% to 26.64%. The high overlap rates (up to 26.64%) occur when the comparison is performed with studies that use a high number of samples (Hou et al., 2012; Sasaki et al., 2016). On the contrary, a low overlap occurs when the comparison is performed with studies that employed a low number of samples or when the comparison is done with consensus CNVRs (which derive from the overlap of CNVs detected with different methods) like in the case of Bagnato et al., (2015) and Wu et al., (2015) (Table 3).

Consensus CNVR annotation

The intersection analysis performed between the bovine gene database (Ensembl Gene 84) and the consensus CNVRs allowed the identification (within or overlapping the consensus CNVRs) of 1,207 Ensemble gene ID, corresponding to 775 genes that have an official bovine gene ID. Of the 563 consensus CNVRs, 218 (38.72%) encompassed one or more genes while 345 (61.28%) did not involve any gene (Table S1). More specifically, of these lasts, 681 (87.87%) were protein-coding genes, 23 (2.94%) were spliceosomal RNAs, 33

(4.22%) were ribosomal RNA genes, 26 (3.32%) were miRNAs and 12 (1.53%) were small nuclear RNAs.

The David database provided the annotation information, according to GO terms and KEGG pathways, for only 537 genes. Table S2 reports the annotation of the statistically significant (p-value < 0.05) clustered and non-clustered genes. Moreover, some of the genes that fall within the consensus CNVRs, could be relevant since these have been proofed to be associated with several economically important traits. For example, the *PRDM16* gene (*PR domain containing 16*) that lies within CNVR_341, is associated with bovine body weight and growth rate (Wang et al., 2012). Another example is the CNVR_475, which contains some of the *BoLA* (*bovine leukocyte antigens*) *Class II complex* genes. These complex genes are expressed on antigen-presenting cells (APCs) (as dendritic cells and macrophages), that present pathogens derived peptides to CD4+ T cells, which then activate macrophages and B cells to generate inflammatory and antibody responses (Behl et al., 2012). Additionally, the *RHOU* gene (*ras homolog family member U*) maps within CNVR_537. This gene encodes for a member of the Rho family of GTPases, which regulates key processes that are needed for mammary gland development (Bray et al., 2011). Lastly, the validated CNVR_280 contains the *SIRPB1* gene that encodes for proteins belonging to the SIRP protein family. *SIRPB1* has been associated with the innate immune system and its interaction with other proteins (such as protein DAP12) seems to promote phagocytosis by macrophages and the migration of neutrophils as part of the regulation of inflammatory response (Laplana et al., 2014).

A study performed by Xu et al (2014), reported a genome wide CNV analysis, revealing variants associated with milk production traits in Holstein cattle. In particular they identified 39 CNVs that were associated with one or more of the considered milk traits. Of these 39 CNVs, 12 overlap with 15 consensus CNVRs (CNVR_84, CNVR_85, CNVR_154, CNVR_259, CNVR_260, CNVR_261, CNVR_262, CNVR_274, CNVR_320, CNVR_364, CNVR_460, CNVR_471, CNVR_486, CNVR_562, CNVR_563) identified in the present study.
Furthermore, the CNVR_274 together with the CNVR_364 and the CNVR_395 have been associated with fat yield, stature and udder depth respectively in a study conducted by Sassi et al. (2016) on Spanish Holstein dairy cattle.
Considering all together, 289 of the 563 identified consensus CNVRs, were found to overlap with 1411 QTLs (Table S4), which were previously reported at the animal QTL database. These QTL are associated with productive, functional (including resistance to diseases) and morphological traits.

CNVR validation by qPCR

Fifteen CNVRs were selected for the validation. Fourteen CNVRs (93%) were confirmed by the qPCR experiments (Supplementary File S3). The invalidated CNVR was not confirmed due to the missing amplification in any sample. The proportion of confirmed positive samples varied from 67% to 100% for every validated CNVR. The table only reports the samples used for the validation and not the samples used as negative controls that varied from 2 to 10 samples per CNVR. Supplementary File S3 also reports the graphical representations of the

CopyCaller output concerning a validated gain (Figure 1), loss (Figure 2) and complex (Figure 3) consensus CNVR. The graphs show the copy number for each CNV identified in a sample, in comparison with the normal state CNV region (which has two copies).

Conclusion

The result of this study is the first CNV genomic analysis on such a large sample of Brown Swiss breed individuals. The detected CNVs are a valuable genomic source of structural variation and enrich the bovine CNV map of the Brown Swiss breed. The genes mapping within the consensus CNVRs could be considered genes to prioritize in further analysis as possible candidate genes involved in functional, productive and health traits. Findings in this study are a valuable source of information for phylogenetic studies and for selection signature analyses. Additionally incorporation of CNVRs in genomic evaluation may be considered in the near future to improve genomic selection programs.

References

Bagnato, A., Strillacci, M.G., Pellegrino, L., Schiavini, F., Frigo E., Rossoni, A., Fontanesi, L., Maltecca, C., Prinsen, R.T.M.M., Dolezal, M.A., 2015. Identification and Validation of Copy Number Variants in Italian Brown Swiss Dairy Cattle Using Illumina Bovine SNP50 Beadchip. Ital J of Anim Sci. 14(3).

Behl DJ, N. K. Verma, Tyagi, N., Mishra, P., Behl, R., Joshi, B.K., 2012. The Major Histocompatibility Complex in Bovines: A Review. ISRN Vet Sci. 872710.

Bickhart, D.M., Xu, L., Hutchison, J.L., Cole, J.B., Null, D.J., Schroeder, S.G., Song, J., Garcia, J.F., Sonstegard, T.S., Van Tassell, C.P., Schnabel, R.D., Taylor, J.F., Lewin, H.A.,

Liu, G.E., 2016. Diversity and population-genetic properties of copy number variations and multicopy genes in cattle. DNA Research 1-10.

Bray, K., Brakebusch, C., Vargo-Gogola, T., 2011. The Rho GTPase Cdc42 is required for primary mammary epithelial cell morphogenesis in vitro. Small GTPases, 2(5), 247-258.

Clop, A.,Vidal, O., Amills, M., 2012. Copy number variation in the genomes of domestic animals. Anim Genet. 43, 503-17.

Diskin, S.J., 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. Nucleic Acids Research. 36, e126.

Hou, Y., Bickhart, D.M., Hvinden, M.L., Li, C., Song, J., Boichard, D.A., Fritz, S., Eggen, 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.

Hou, Y., Bickhart, D.M., Hvinden, M.L., Li, C., Song, J., Boichard, D.A., Fritz, S., Eggen, 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.

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.

Kadri, N.K., Sahana, G., Charlier, C., Iso-Touru, T., Guldbrandtsen, B., Karim L., Nielsen, U.S., Panitz, F., Aamand, G.P., Schulman, N., Georges, M., Vilkki, J., Lund, M.S., Druet, T., 2014. A 660-Kb Deletion with Antagonistic Effects on Fertility and Milk Production Segregates at High Frequency in Nordic Red Cattle: Additional Evidence for the Common Occurrence of Balancing Selection in Livestock. PLoS Genet. 10, e1004049.

Laplana, M., Royo, J.L., García, L.F., Aluja, A., Gomez-Skarmeta, J.L., Fibla, J., 2014. SIRPB1 copy-number polymorphism as candidate

quantitative trait locus for impulsive-disinhibited personality. Genes Brain Behav. 13(7):653-62.

Meuwissen, T.H., Hayes, B.J., Goddard, M.E., 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157, 1819–1829.

Meyers, S.N., McDaneld, 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.L., 2010. A deletion mutation in bovine SLC4A2 is associated with osteopetrosis in Red Angus cattle. BMC Genomics 11, 337.

Mills, R.E., Walter, K., Stewart, C., Handsaker, R.E., Chen, K., Alkan, C., Abyzov, A., Yoon, S.C., Ye, K., Cheetham, R.K., Chinwalla, A., Conrad, D.F., Fu, Y., Grubert, F., Hajirasouliha, I., Hormozdiari, F., Iakoucheva, L.M., Iqbal, Z., Kang, S., Kidd, J.M., Konkel, M.K., Korn, J., Khurana, E., Kural, D., Lam, H.Y., Leng, J., Li, R., 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., McCarroll, S.A., Korbel, J.O., 2011. Mapping copy number variation by population scale genome sequencing. Nature 470, 59-65.

Peiffer, D.A., Le, J.M., Steemers, F.J., Chang, W., Jenniges, T., Garcia, F., Haden, K., Li, J., Shaw, C.A., Belmont, J., Cheung, S.W., Shen, R.M., Barker, D.L., Gunderson, K.L., 2006. High-resolution genomic

profiling of chromosomal aberrations using Infinium whole-genome genotyping. Genome Res. 16, 1136–1148.

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. 8, 512-20.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., Cho, E.K., Dallaire, S., Freeman, J.L., González, J.R., Gratacòs, M., Huang, J., Kalaitzopoulos, D., Komura, D., MacDonald, J.R., Marshall, C.R., Mei, R., Montgomery, L., Nishimura, K., Okamura, K., Shen, F., Somerville, M.J., Tchinda, J., Valsesia, A., Woodwark, C., Yang, F., Zhang, J., Zerjal, T., Zhang, J., Armengol, L., Conrad, D.F., Estivill, X., Tyler-Smith, C., Carter, N.P., Aburatani, H., Lee, C., Jones, K.W., Scherer, S.W., Hurles, M.E., 2006. Global variation in copy number in the human genome. Nature 444, 444-454.

Sasaki, S., Watanabe, T., Nishimura, S., Sugimoto, Y., 2016. Genome-wide identification of copy number variation using high-density single-

nucleotide polymorphism array in Japanese Black cattle. BMC Genet. 17, 26.

Sassi, N.B., González-Recio, O., de Paz-del Rio, R., Rodriguez-Ramilo, S.T., Fernández, A.I., 2016. Associated effects of copy number variants on economically important traits in Spanish Holstein dairy cattle. J. Dairy Sci. 99:6371-6380.

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

Wang, K., Li, M., Hadley, D., Liu, R., Glessner, J., Grant, S.F., Hakonarson, H., Bucan, M., 2007. PennCNV: An integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Res. 17, 1665 – 74.

Winchester, L., Yau, C., Ragoussis, J., 2009. Comparing CNV detection methods for SNP arrays. Brief Funct Genomic Proteomic 8, 353-66.

Wu, Y., Fan, H., Jing, S., Xia, J., Chen, Y., Zhang, L., Gao, X., Li, J., Gao, H., Ren, H., 2015. A genome-wide scan for copy number variations using high-density single nucleotide polymorphism array in Simmental cattle. Anim Genet. 46, 289-98.

Xu, L., Cole, J.B., Bickhart, D.M., Hou, Y., Song, J., VanRaden, P.M., Sonstegard, T.S., Van Tassell, C.P., Liu, G.E., 2014. Genome wide CNV analysis reveals additional variants associated with milk production traits in Holsteins. BMC Genomics 15, 683.

Xu, L., Hou, Y., Bickhart, D.M., Zhou, Y., Hay, E.H.A., Song, J., Sonstegard, T.S., Van Tassell C.P., Liu, G.E., 2016. Population-genetic properties of differentiated copy number variations in cattle. Sci. Rep. 6, 23161.

Yang, Z., Zhuan, B., Yan, Y., Jiang, S., Wang, T., 2015. Integrated analyses of copy number variations and gene differential expression in lung squamous-cell carcinoma. Biol. Res. 48, 47.

Zhang, F., Gu, W., Hurles, M.E., Lupski, J.R., 2009. Copy number variation in human health, disease, and evolution. Annu Rev Genomics Hum Genet 10, 451-81.

Zhang, Q., Ma, Y., Wang, X., Zhang, Y., Zhao, X., 2015. Identification of copy number variations in Qinchuan cattle using BovineHD Genotyping Beadchip array. Mol Genet Genomics. 290, 319-27.

Zhao, M., Wang, Q., Wang, Q., Jia, P., Zhao, Z., 2013. Computational tools for copy number variation (CNV) detection using next-generation sequencing data: features and perspectives. BMC Bioinformatics 14 (Suppl 11), S1.

Zumbach, B., Jorjani, H., Dürr, J., 2010. Brown Swiss genomic evaluation. Interbull Bull 42, 44–51.

Table 1. Characteristics and sizes (bp) of CNVs detected in autosomal chromosomes by the SVS and by the PennCNV software.

Type	CNVs	Mean length	Median length	Size range	Min CNVs per sample	Max CNVs per sample
SVS						
Loss	24,193	155,956	49,855	1,244 – 5,903,844	12	42
Gain	837	297,700	227,903	3,489 – 2,135,206	1	6
All	25,030	160,696	51,897	1,244 – 5,903,844	12	45
PennCNV						
Loss	42,373	52,637	24,907	961 – 991,125	9	171
Gain	19,968	89,640	35,820	997 – 1,915,954	4	178
All	62,341	64,489	27,462	961 – 1,915,954	27	198

Table 2. Descriptive statistics for CNVRs identified by SVS and PennCNV software and for consensus CNVRs.

Type	CNVRs	Mean length	Median length	Size range	CNVR coverage	% CNVR
SVS						
Loss	344	166,479	18,860	1,244 – 5,903,844	57,268,952	2.28
Gain	30	268,167	134,989	4,152 – 1,935,069	8,045,010	0.32
Complex	24	1,381,355	499,127	7,584 – 6,371,690	33,152,521	1.32
All	398	247,403	24,563	1,244 – 6,371,690	98,466,483	3.92
PennCNV						
Loss	2,637	20,459	12,200	1,167 – 571,006	53,952,246	2.15
Gain	2,410	24,122	12,278	1,040 – 1,151,879	58,134,073	2.31
Complex	531	152,499	51,790	1,683 – 4,337,260	80,977,464	3.22
All	5,578	34,611	13,351	1,167 – 4,337,260	193,063,783	7.68
Consensus CNVRs (SVS-PennCNV)						
Loss	301	64,129	14,838	1,039 – 553,753	9,683,625	0.39
Gain	24	191,645	119,876	4,152 – 1,151,879	4,599,501	0.18
Complex	238	182,209	37,803	1,525 – 4,325,190	43,365,836	1.73
All	563	102,396	23,065	1,309 – 4,325,190	57,648,962	2.29

Table 3. Comparison between CNVRs detected in previous studies and those detected in the present study.

Study	Method	Algorithm used	Samples	N° breeds	CNVR	Length (Mb)	N. overlap (%)
Bagnato et al., 2015	SNP chip (54 kb)	CNAM SVS; PennCNV	651	1	150	17.10	29 (5.15)
Hou et al., 2012	SNP chip (777 kb)	PennCNV	674	27	3,346	142.7	150 (26.64)
Jiang et al., 2013	SNP chip (777 kb)	PennCNV	96	1	367	42.74	57 (10.12)
Zhang et al., 2015	SNP chip (777 kb)	PennCNV	6	1	365	13.13	36 (6.39)
Wu et al., 2015	SNP chip (777 kb)	PennCNV; CNVPartition	792	1	263	35.48	55 (9.77)
Sasaki et al., 2016	SNP chip (777 kb)	PennCNV	1,481	1	861	43.65	120 (21.31)
Xu et al., 2016	SNP chip (777 kb)	CNAM SVS	300	8	263	15.64	61 (10.83)
Bickhart et al., 2016	Sequencing		75	8	1,853	87.50	150 (26.64)
This study 2016	SNP chip (777 kb)	CNAM SVS; PennCNV	1,410	1	563	57.65	563

Supporting information

Figure 1. Percentage of the CNVRs per BTA obtained with SVS, PennCNV software and the consensus analyses.

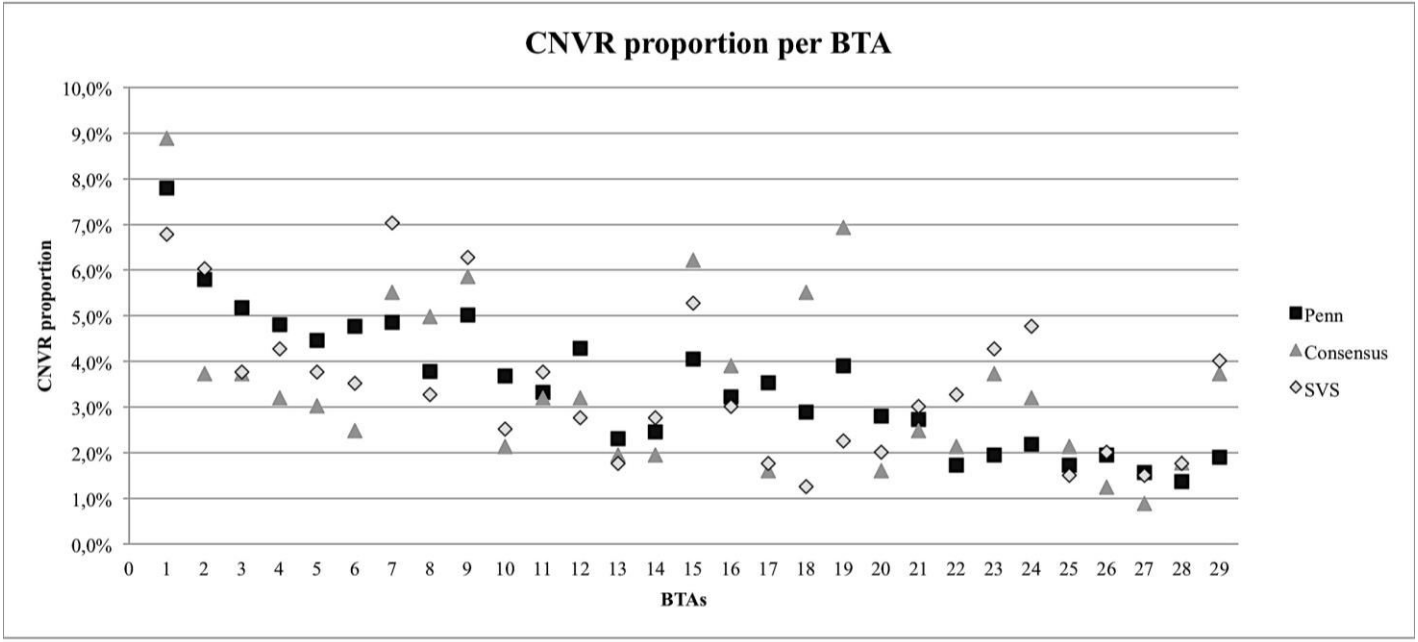


Figure 2. Physical distribution of the consensus gain, loss and complex CNVRs.

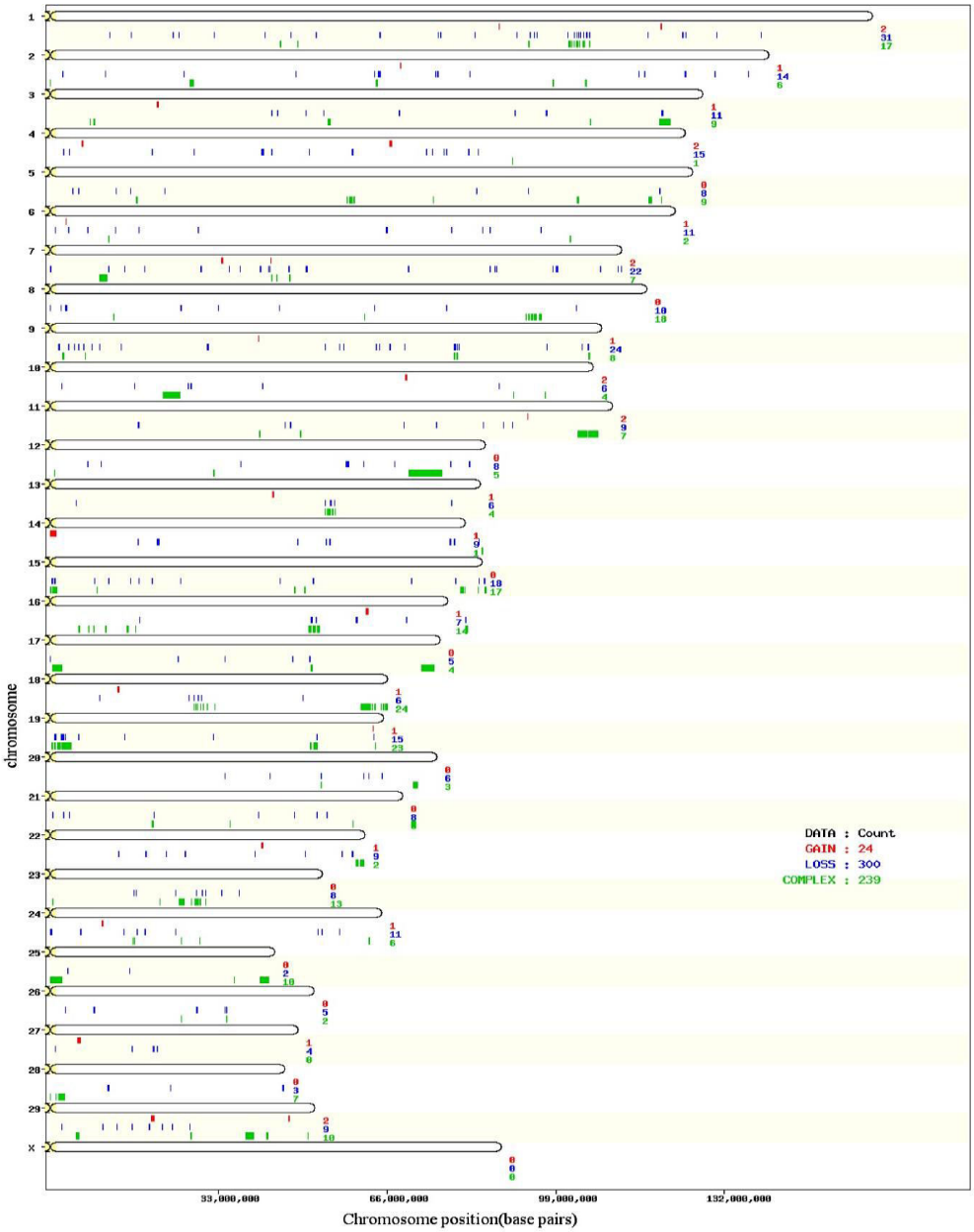


Figure 3. Distributions of CNVR lengths identified with SVS, PennCNV and the consensus analysis.

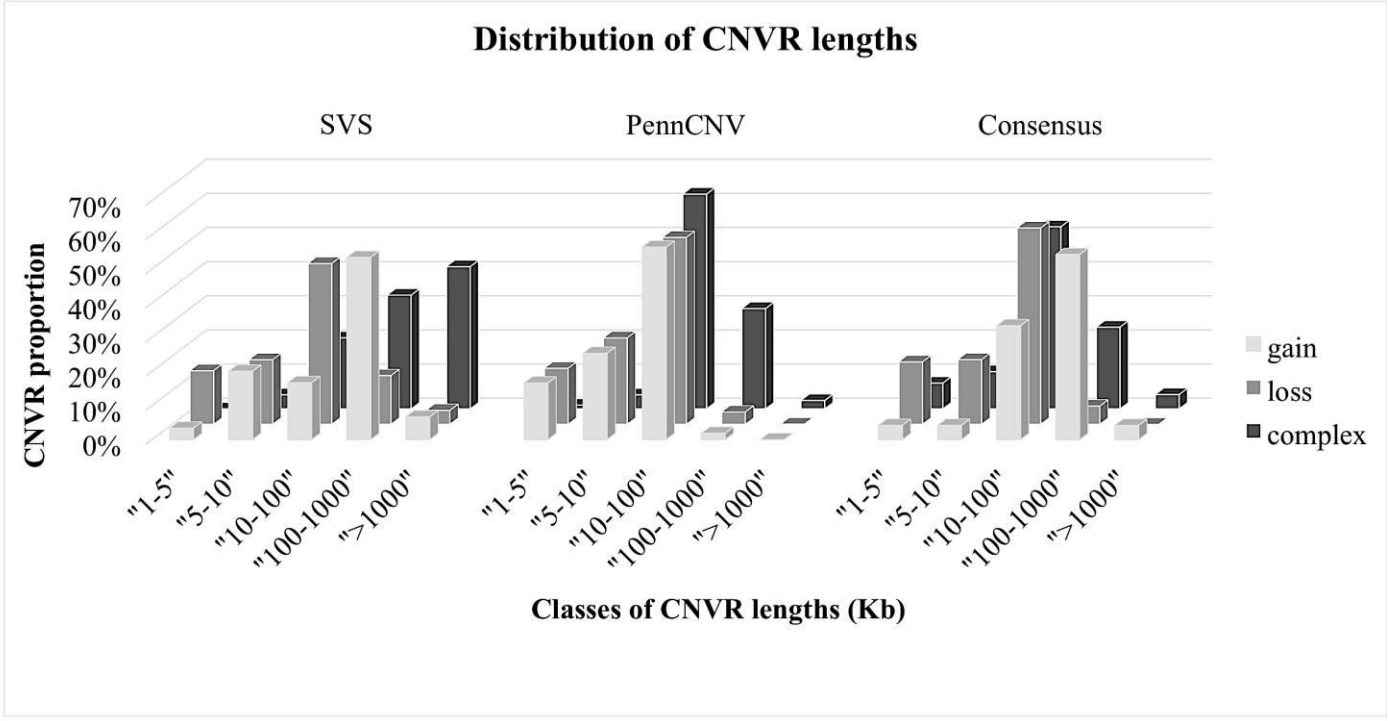
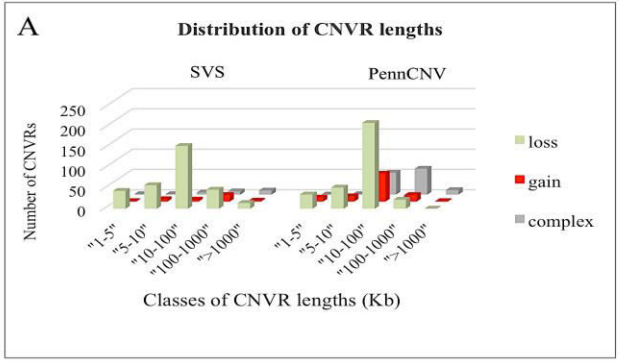
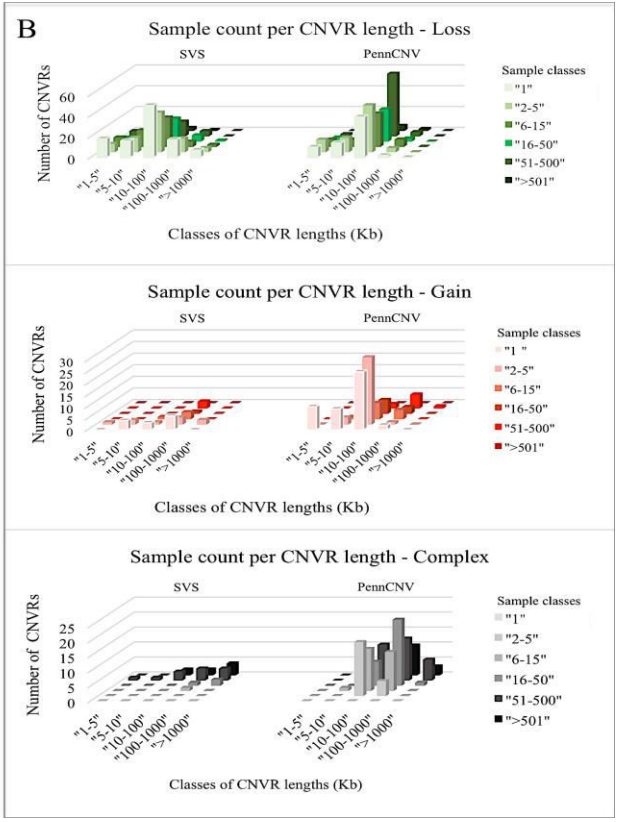


Figure 4. CNVR length and frequency distributions of sample classes. A) CNVR distribution based on classes of CNVR lengths. B) Sample count per classes of samples, in each class of CNVR length, according to the different CNVR states.





Supplementary File S1. Consensus CNVRs and annotated genes found within them.

CNVR_ID	Chr	Start	End	Length (bp)	State - consensus CNVR	Genes within consensus CNVR
CNVR_1	1	11763666	11767713	4047	loss	
CNVR_2	1	16021282	16023240	1958	loss	
CNVR_3	1	24054781	24069666	14885	loss	
CNVR_4	1	24073877	24078545	4668	loss	
CNVR_5	1	25205821	25229453	23632	loss	
CNVR_6	1	32134740	32142914	8174	loss	
CNVR_7	1	42119177	42190693	71516	loss	
CNVR_8	1	45148191	45212691	64500	complex	
CNVR_9	1	47183770	47215572	31802	loss	
CNVR_10	1	48501686	48529650	27964	loss	
CNVR_11	1	52119241	52122517	3276	loss	
CNVR_12	1	64632165	64637164	4999	loss	*B4GALT4*
CNVR_13	1	75987278	76045397	58119	loss	
CNVR_14	1	76468213	76470326	2113	loss	
CNVR_15	1	83218713	83238102	19389	loss	*EPHB3*
CNVR_16	1	87875403	87879555	4152	gain	*PEX5L*
CNVR_17	1	91350265	91395660	45395	loss	
CNVR_18	1	93730576	93819471	88895	complex	
CNVR_19	1	94039763	94050170	10407	loss	
CNVR_20	1	94867796	94889332	21536	loss	
CNVR_21	1	95378415	95382456	4041	loss	
CNVR_22	1	101306706	101327602	20896	loss	
CNVR_23	1	101527889	101537204	9315	complex	
CNVR_24	1	101697459	101732405	34946	complex	
CNVR_25	1	101858606	101865663	7057	complex	
CNVR_26	1	102123303	102158433	35130	complex	
CNVR_27	1	102635348	102651984	16636	complex	
CNVR_28	1	102700027	102722564	22537	loss	
CNVR_29	1	102989935	103002096	12161	complex	*SLITRK3*

CNVR_30	1	103098743	103135856	37113	complex	*SI*
CNVR_31	1	103154830	103168693	13863	loss	*SI*
CNVR_32	1	103221426	103275245	53819	complex	
CNVR_33	1	103432447	103444563	12116	loss	
CNVR_34	1	103629992	103647890	17898	complex	
CNVR_35	1	103834800	103837936	3136	loss	
CNVR_36	1	104249990	104268504	18514	complex	
CNVR_37	1	104473015	104490694	17679	complex	
CNVR_38	1	104522667	104540151	17484	loss	
CNVR_39	1	104553225	104607581	54356	complex	
CNVR_40	1	104647647	104698566	50919	complex	
CNVR_41	1	104988070	105086422	98352	loss	
CNVR_42	1	105108951	105137775	28824	loss	
CNVR_43	1	105590228	105637750	47522	loss	
CNVR_44	1	105673506	105678148	4642	complex	
CNVR_45	1	117105202	117111226	6024	loss	
CNVR_46	1	119468185	119624353	156168	gain	*TM4SF4, WWTR1*
CNVR_47	1	123795324	123949121	153797	loss	
CNVR_48	1	124384563	124420643	36080	loss	
CNVR_49	1	130549225	130552004	2779	loss	
CNVR_50	1	139215807	139221702	5895	loss	
CNVR_51	2	6889	35126	28237	complex	
CNVR_52	2	2519574	2531178	11604	loss	
CNVR_53	2	10849034	10943112	94078	loss	*U1*
CNVR_54	2	26297642	26304783	7141	loss	
CNVR_55	2	27361767	28051401	689634	complex	*5S_rRNA, bta-mir-2353, CERS6, NOSTRIN, STK39*
CNVR_56	2	48125231	48176523	51292	loss	
CNVR_57	2	63500222	63505440	5218	loss	
CNVR_58	2	63864966	64072254	207288	complex	
CNVR_59	2	64237382	64791135	553753	loss	
CNVR_60	2	68598340	68657656	59316	gain	
CNVR_61	2	75472312	75475758	3446	loss	
CNVR_62	2	75913220	75933405	20185	loss	
CNVR_63	2	82198830	82210776	11946	loss	

CNVR_64	2	98480344	98490521	10177	complex	
CNVR_65	2	98508373	98518278	9905	complex	
CNVR_66	2	104824189	105011773	187584	complex	*SHOX*
CNVR_67	2	115322601	115354397	31796	loss	
CNVR_68	2	116347140	116410431	63291	loss	
CNVR_69	2	124330343	124394326	63983	loss	
CNVR_70	2	130200917	130204643	3726	loss	
CNVR_71	2	136749793	136776410	26617	loss	
CNVR_72	3	7957960	7964523	6563	complex	
CNVR_73	3	8635643	8682290	46647	complex	
CNVR_74	3	21066806	21202832	136026	gain	*U1*
CNVR_75	3	43462907	43521501	58594	loss	*AGL*
CNVR_76	3	44555762	44663244	107482	loss	*SNX7*
CNVR_77	3	50167465	50191213	23748	loss	
CNVR_78	3	53625840	53628690	2850	loss	
CNVR_79	3	54369910	54872418	502508	complex	*GBP4, GBP2, U2*
CNVR_80	3	68384020	68432145	48125	loss	
CNVR_81	3	91038851	91047310	8459	loss	
CNVR_82	3	97045586	97049900	4314	loss	
CNVR_83	3	97296240	97349844	53604	loss	
CNVR_84	3	105711958	105720505	8547	complex	*SCMH1*
CNVR_85	3	119450405	119504166	53761	complex	*CSF2RA*
CNVR_86	3	119695073	119700584	5511	complex	
CNVR_87	3	119752388	119775433	23045	loss	
CNVR_88	3	119778431	119783864	5433	loss	*NDUFA10*
CNVR_89	3	119880092	119923307	43215	complex	
CNVR_90	3	120049410	120060932	11522	complex	
CNVR_91	3	120094145	120102241	8096	loss	
CNVR_92	3	120117354	121408443	1291089	complex	*AGXT, ANKMY1, ATG4B, BOK, bta-mir-149, C2orf54, CAPN10, CROCC2, D2HGDH, DTYMK, GAL3ST2, GPC1, GPR35, HDLBP, ING5, KIF1A, MTERF4, MYEOV2, NEU4, OTOS, PASK, PDCD1, PPP1R7, RNPEPL1, RTP5, SEPT2, SNED1, STK25, THAP4*
CNVR_93	4	2637110	2713262	76152	loss	

CNVR_94	4	2714029	2729422	15393	loss	
CNVR_95	4	3913942	3915686	1744	loss	
CNVR_96	4	6223614	6468141	244527	gain	
CNVR_97	4	19985853	20010948	25095	loss	
CNVR_98	4	28243455	28249822	6367	loss	
CNVR_99	4	41429901	41758384	328483	loss	
CNVR_100	4	43481123	43487686	6563	loss	
CNVR_101	4	50796862	50816295	19433	loss	
CNVR_102	4	59156453	59242378	85925	loss	
CNVR_103	4	66487945	66890071	402126	gain	*FKBP14, MTURN, PLEKHA8, SCRN1, WIPF3*
CNVR_104	4	73676903	73785765	108862	loss	
CNVR_105	4	74889140	74940057	50917	loss	*ZNF804B*
CNVR_106	4	77110920	77112875	1955	loss	*CFAP69, STEAP2*
CNVR_107	4	77631664	77634082	2418	loss	*NUDCD3*
CNVR_108	4	81840906	81990298	149392	loss	*RALA*
CNVR_109	4	83936446	83942707	6261	loss	
CNVR_110	4	90630112	90641790	11678	complex	
CNVR_111	5	4513012	4531901	18889	loss	
CNVR_112	5	5626620	5638151	11531	loss	
CNVR_113	5	13035547	13054388	18841	loss	
CNVR_114	5	15832131	15838518	6387	loss	
CNVR_115	5	17005200	17085717	80517	complex	
CNVR_116	5	22514133	22563988	49855	loss	
CNVR_117	5	58140836	58251760	110924	complex	
CNVR_118	5	58682687	58752692	70005	complex	
CNVR_119	5	58966295	59215089	248794	complex	*OR6C1*
CNVR_120	5	59421039	59627471	206432	complex	
CNVR_121	5	75000225	75012411	12186	complex	
CNVR_122	5	83602670	83615323	12653	loss	*ITPR2*
CNVR_123	5	93601025	93611357	10332	loss	
CNVR_124	5	103122325	103454169	331844	complex	*WC1-8, WC1.3*
CNVR_125	5	117256108	117659032	402924	complex	*TTC38*
CNVR_126	5	119318618	119333338	14720	loss	
CNVR_127	5	119716023	119729902	13879	complex	

CNVR_128	6	1072995	1100413	27418	loss	
CNVR_129	6	3190744	3217181	26437	gain	
CNVR_130	6	3626246	3668053	41807	loss	
CNVR_131	6	7463583	7489700	26117	loss	*SYNPO2*
CNVR_132	6	11492050	11528112	36062	complex	
CNVR_133	6	12846571	12856335	9764	loss	
CNVR_134	6	17388856	17400789	11933	loss	
CNVR_135	6	28989049	28999522	10473	loss	
CNVR_136	6	65940020	65961129	21109	loss	
CNVR_137	6	78634839	78653718	18879	loss	
CNVR_138	6	84627027	84736612	109585	loss	
CNVR_139	6	86122072	86150255	28183	loss	
CNVR_140	6	96055154	96059367	4213	loss	
CNVR_141	6	101831059	101921411	90352	complex	
CNVR_142	7	156544	286246	129702	loss	
CNVR_143	7	9771150	9915411	144261	complex	
CNVR_144	7	10044994	11091861	1046867	complex	
CNVR_145	7	11136869	11145155	8286	complex	
CNVR_146	7	11526456	11555132	28676	loss	
CNVR_147	7	14714172	14719678	5506	loss	
CNVR_148	7	18518171	18566035	47864	loss	
CNVR_149	7	29560140	29627140	67000	loss	
CNVR_150	7	33578586	33849742	271156	gain	
CNVR_151	7	35130073	35132132	2059	loss	
CNVR_152	7	37209166	37243797	34631	loss	
CNVR_153	7	41232074	41241356	9282	loss	
CNVR_154	7	42745346	42788788	43442	loss	*OR2AK2*
CNVR_155	7	43087430	43097791	10361	loss	
CNVR_156	7	43292715	43330896	38181	gain	*OR2M3, OR14C36*
CNVR_157	7	43487164	43498441	11277	complex	
CNVR_158	7	44437375	44444959	7584	complex	
CNVR_159	7	46770810	46802262	31452	loss	
CNVR_160	7	46942526	46964290	21764	complex	
CNVR_161	7	47030369	47063805	33436	complex	

CNVR_162	7	50227210	50255419	28209	loss	
CNVR_163	7	70191120	70262424	71304	loss	
CNVR_164	7	86076889	86123236	46347	loss	
CNVR_165	7	87108583	87133023	24440	loss	
CNVR_166	7	87452028	87457363	5335	loss	
CNVR_167	7	98502599	98508931	6332	loss	*CAST*
CNVR_168	7	99113453	99116654	3201	loss	
CNVR_169	7	99322117	99328261	6144	loss	
CNVR_170	7	107712506	107744442	31936	loss	
CNVR_171	7	111189898	111214553	24655	loss	
CNVR_172	7	111913792	111918914	5122	loss	
CNVR_173	8	95630	135066	39436	loss	*5S_rRNA, HIATL1*
CNVR_174	8	2252209	2253891	1682	loss	
CNVR_175	8	3101470	3116170	14700	loss	
CNVR_176	8	3352998	3359075	6077	loss	
CNVR_177	8	12451687	12504399	52712	complex	
CNVR_178	8	25680324	25694771	14447	loss	*ADAMTSL1*
CNVR_179	8	32961137	32968374	7237	loss	
CNVR_180	8	44934828	44944671	9843	loss	
CNVR_181	8	61619394	61655615	36221	complex	*ZCCHC7*
CNVR_182	8	63530836	63544304	13468	loss	
CNVR_183	8	77675650	77687415	11765	loss	
CNVR_184	8	93154633	93156158	1525	complex	
CNVR_185	8	93641657	93645305	3648	complex	
CNVR_186	8	93776052	93832928	56876	complex	
CNVR_187	8	93853057	93875806	22749	complex	
CNVR_188	8	93906676	93916177	9501	complex	
CNVR_189	8	94108258	94115663	7405	complex	
CNVR_190	8	94191547	94200849	9302	complex	
CNVR_191	8	94386951	94405438	18487	complex	
CNVR_192	8	94510963	94525501	14538	complex	
CNVR_193	8	94793275	94810609	17334	complex	
CNVR_194	8	94841235	94867227	25992	complex	
CNVR_195	8	94940218	94943827	3609	complex	

CNVR_196	8	95019722	95027602	7880	complex	
CNVR_197	8	95090752	95220168	129416	complex	
CNVR_198	8	95889216	96060313	171097	complex	*OR13C3*
CNVR_199	8	96129918	96181815	51897	complex	
CNVR_200	8	103007542	103054921	47379	loss	*UGCG*
CNVR_201	9	1799757	1841593	41836	loss	
CNVR_202	9	2596486	2626145	29659	complex	
CNVR_203	9	3626237	3659242	33005	loss	
CNVR_204	9	4889983	4906838	16855	loss	
CNVR_205	9	5586598	5619077	32479	loss	
CNVR_206	9	6581699	6597289	15590	loss	
CNVR_207	9	6964612	6977472	12860	complex	
CNVR_208	9	8241626	8284608	42982	loss	*ADGRB3*
CNVR_209	9	9735054	9802704	67650	loss	
CNVR_210	9	13985299	13992239	6940	loss	
CNVR_211	9	30921439	30972803	51364	loss	
CNVR_212	9	40765749	40866390	100641	gain	
CNVR_213	9	53895094	53902977	7883	loss	
CNVR_214	9	56694798	56733584	38786	loss	
CNVR_215	9	57453142	57466086	12944	loss	
CNVR_216	9	63928374	63933005	4631	loss	
CNVR_217	9	64596890	64604811	7921	loss	
CNVR_218	9	66599290	66620995	21705	loss	
CNVR_219	9	69527371	69536017	8646	loss	
CNVR_220	9	79041730	79059988	18258	loss	
CNVR_221	9	79158631	79204066	45435	complex	
CNVR_222	9	79240560	79261394	20834	complex	
CNVR_223	9	79347692	79369200	21508	loss	
CNVR_224	9	79617946	79640663	22717	complex	
CNVR_225	9	79736610	79759823	23213	complex	
CNVR_226	9	79804912	79814487	9575	loss	
CNVR_227	9	80030283	80057982	27699	loss	
CNVR_228	9	97347618	97357043	9425	loss	
CNVR_229	9	104120889	104131917	11028	loss	*DACT2*

CNVR_230	9	105225717	105323209	97492	loss	
CNVR_231	9	105448095	105462864	14769	complex	
CNVR_232	9	105470625	105514891	44266	loss	*DLL1*
CNVR_233	9	105601001	105695468	94467	complex	*TBP, PSMB1*
CNVR_234	10	2326007	2343720	17713	loss	
CNVR_235	10	16662635	16665097	2462	loss	
CNVR_236	10	22169580	25327628	3158048	complex	*TRAV14DV4, TRAV16, TRAV17, TRAV18, TRAV29DV5, TRDC*
CNVR_237	10	25460016	25519923	59907	complex	
CNVR_238	10	27061809	27098152	36343	loss	
CNVR_239	10	27637415	27659286	21871	loss	
CNVR_240	10	41584872	41596729	11857	loss	
CNVR_241	10	69627476	69744520	117044	gain	
CNVR_242	10	69855316	69862924	7608	gain	*EXOC5*
CNVR_243	10	87873996	87878635	4639	loss	
CNVR_244	10	90679288	90686795	7507	complex	
CNVR_245	10	96944699	96952072	7373	complex	
CNVR_246	11	17341756	17406557	64801	loss	
CNVR_247	11	41021689	41204355	182666	complex	
CNVR_248	11	46023684	46026983	3299	loss	
CNVR_249	11	47034705	47074624	39919	loss	
CNVR_250	11	47167732	47195270	27538	loss	
CNVR_251	11	49071385	49093033	21648	complex	
CNVR_252	11	69357111	69374266	17155	loss	
CNVR_253	11	75666032	75696206	30174	loss	
CNVR_254	11	84769675	84772952	3277	loss	
CNVR_255	11	88702826	88708902	6076	loss	
CNVR_256	11	90570475	90582204	11729	loss	
CNVR_257	11	93443138	93467609	24471	gain	
CNVR_258	11	93527989	93571781	43792	gain	
CNVR_259	11	103328266	104200942	872676	complex	*AGPAT2, bta-mir-126, C9orf69, CAMSAP1, CARD9, CCDC187, DNLZ, EGFL7, FAM69B, GPSM1, INPP5E, KCNT1, LCN9, LHX3, NACC2, NOTCH1, PMPCA, SNORA17, SOHLH1. UBAC1QSOX2, SDCCAG3,*

						SEC16A, SNAPC4,
CNVR_260	11	104227967	104239885	11918	complex	*ABO, U6*
CNVR_261	11	104261327	105200994	939667	complex	*5S_rRNA, ADAMTS13, ADAMTSL2, BRD3, CACFD1, CACNA1B, DBH, FAM163B, MED22, REXO4, RPL7A, SARDH, SLC2A6, SNORD24, SNORD36, snR47, STKLD1, SURF2, SURF4, SURF6, TMEM8C, U6atac, VAV2, WDR5*
CNVR_262	11	105502072	107118341	1616269	complex	*ABCA2, ANAPC2, ARRDC1, C8G, C9orf116, C9orf142, C9orf172, CCDC183, CLIC3, COL5A1, CYSRT1, DPH7, EDF1, ENTPD2, ENTPD8, EXD3, FAM166A, FBXW5, FUT7, GRIN1, LCN1, LCN10, LCN12, LCN15, LCN8, LCNL1, LRRC26, MAMDC4, MAN1B1, MRPL41, MRPS2, NDOR1, NELFB, NOXA1, NPDC1, NRARP, NSMF, OLFM1, PHPT1, PNPLA7, PPP1R26, PTGDS, RNF208, RNF224, RXRA, SAPCD2, SLC34A3, SSNA1, TMEM141, TMEM203, TMEM210, TOR4A, TPRN, TRAF2, TUBB4B, UAP1L1, ZMYND19*
CNVR_263	11	107167457	107282960	115503	complex	*ATHL1, IFITM1, IFITM5, NLRP6, PSMD13, SIRT3*
CNVR_264	12	936045	944850	8805	complex	
CNVR_265	12	7507845	7509154	1309	loss	
CNVR_266	12	10028559	10073042	44483	loss	
CNVR_267	12	32054331	32182960	128629	complex	
CNVR_268	12	37358859	37411432	52573	loss	
CNVR_269	12	58054721	58439701	384980	loss	
CNVR_270	12	61368647	61396071	27424	loss	
CNVR_271	12	67537603	67576208	38605	loss	
CNVR_272	12	70363408	72123747	1760339	complex	
CNVR_273	12	72174261	72272866	98605	complex	
CNVR_274	12	72409908	76735098	4325190	complex	
CNVR_275	12	78494480	78498214	3734	loss	
CNVR_276	12	82144071	82215811	71740	loss	*FGF14*
CNVR_277	13	5135656	5163722	28066	loss	
CNVR_278	13	43654009	43776717	122708	gain	
CNVR_279	13	53858853	53868094	9241	loss	

CNVR_280	13	53932232	53982995	50763	complex	*SIRPB1*
CNVR_281	13	54181983	54934270	752287	complex	*ABHD16B, ARFGAP1, ARFRP1, BHLHE23, BIRC7, bta-mir-124b, bta-mir-1388, C20orf195, CHRNA4, COL20A1, DNAJC5, EEF1A2, GMEB2, HAR1A, KCNQ2, LKAAEAR1, MYT1, ZGPAT, PPDPF, NPBWR2, OPRL1, PRPF6, PTK6, RGS19, SAMD10, SLC17A9, SLC2A4RG, SOX18, SRMS, STMN3, TCEA2, TNFRSF6B, TPD52L2, U6, UCKL1, YTHDF1, ZBTB46, ZNF512B*
CNVR_282	13	54948937	54956566	7629	loss	
CNVR_283	13	55059307	55128121	68814	loss	*NTSR1*
CNVR_284	13	55169927	55409520	239593	complex	*bta-mir-1-1, bta-mir-133a-1, bta-mir-2306, CABLES2, GATA5, LAMA5, RPS21, RBBP8NL, SLCO4A1*
CNVR_285	13	55681971	55697174	15203	loss	*CDH4*
CNVR_286	13	55926682	55952017	25335	complex	
CNVR_287	13	78614384	78628457	14073	loss	
CNVR_288	14	74984	1226863	1151879	gain	*5S_rRNA, U6*
CNVR_289	14	17329297	17336372	7075	loss	
CNVR_290	14	21069733	21121341	51608	loss	*PRKDC*
CNVR_291	14	21284128	21287635	3507	loss	
CNVR_292	14	48454835	48463465	8630	loss	*EXT1*
CNVR_293	14	54123999	54126508	2509	loss	*CSMD3*
CNVR_294	14	54809641	54833610	23969	loss	
CNVR_295	14	78332048	78386690	54642	loss	
CNVR_296	14	79161833	79328333	166500	loss	
CNVR_297	14	84067817	84069606	1789	loss	*COL14A1*
CNVR_298	14	84514336	84628243	113907	complex	*U6*
CNVR_299	15	43737	70433	26696	complex	
CNVR_300	15	112320	236845	124525	complex	*OR5M9, OR9G1*
CNVR_301	15	373904	390683	16779	loss	
CNVR_302	15	519898	531604	11706	loss	
CNVR_303	15	547375	553963	6588	complex	
CNVR_304	15	560099	582084	21985	complex	
CNVR_305	15	657809	695497	37688	complex	

CNVR_306	15	803960	838184	34224	complex	
CNVR_307	15	881109	917040	35931	loss	*OR4C15*
CNVR_308	15	951039	984263	33224	complex	
CNVR_309	15	1031558	1051698	20140	loss	
CNVR_310	15	1076743	1090123	13380	complex	
CNVR_311	15	1277543	1384528	106985	complex	
CNVR_312	15	8686216	8689537	3321	loss	
CNVR_313	15	9172874	9214121	41247	complex	
CNVR_314	15	11534530	11541499	6969	loss	
CNVR_315	15	15764194	15773838	9644	loss	
CNVR_316	15	17347205	17366812	19607	loss	*U6atac*
CNVR_317	15	20057061	20060323	3262	loss	
CNVR_318	15	25530542	25535567	5025	loss	
CNVR_319	15	45144601	45155661	11060	loss	
CNVR_320	15	47873132	48047397	174265	complex	HIST1H4C, OR52N5, OR52N1, U6
CNVR_321	15	49894984	49940983	45999	complex	
CNVR_322	15	51405045	51412532	7487	loss	
CNVR_323	15	51571466	51574083	2617	loss	
CNVR_324	15	70725250	70747191	21941	loss	
CNVR_325	15	79382459	79388462	6003	loss	
CNVR_326	15	79499062	79525723	26661	loss	*U6*
CNVR_327	15	80369812	80842722	472910	complex	*OR8K3, OR8K1, OR5J2*
CNVR_328	15	81298654	81308681	10027	complex	*OR5G3*
CNVR_329	15	83785281	83852224	66943	complex	
CNVR_330	15	84038466	84046192	7726	loss	
CNVR_331	15	85089965	85132419	42454	loss	
CNVR_332	15	85181100	85199311	18211	complex	
CNVR_333	15	85216368	85276403	60035	complex	
CNVR_334	16	5670600	5846840	176240	complex	
CNVR_335	16	7586354	7685266	98912	complex	
CNVR_336	16	8582055	8623933	41878	complex	
CNVR_337	16	10845800	10975736	129936	complex	
CNVR_338	16	15066677	15247073	180396	complex	
CNVR_339	16	16691955	16696640	4685	complex	

CNVR_340	16	17579695	17611127	31432	loss	
CNVR_341	16	50599725	50866828	267103	complex	*ARHGEF16, bta-mir-551a, MEGF6, PRDM16, TP73, TPRG1L, WRAP73*
CNVR_342	16	51003284	51011081	7797	complex	
CNVR_343	16	51118732	51127861	9129	loss	
CNVR_344	16	51159693	51177627	17934	complex	
CNVR_345	16	51191208	51224815	33607	loss	
CNVR_346	16	51397454	51494795	97341	complex	
CNVR_347	16	51499853	51876595	376742	complex	*5S_rRNA, FAM213B, HES5, MMEL1, MORN1, FAAP20, PANK4, PEX10, PLCH2, PRKCZ, RER1, SKI, TTC34*
CNVR_348	16	51923326	51980700	57374	complex	*GABRD, PRKCZ*
CNVR_349	16	52075019	52109994	34975	loss	*CALML6, CFAP74, GNB1, TMEM52, U6*
CNVR_350	16	52243837	52792404	548567	complex	
CNVR_351	16	60051473	60175280	123807	loss	*ACAP3, AGRN, ANKRD65, AURKAIP1, B3GALT6, bta-mir-200a, bta-mir-200b, bta-mir-429, C1orf159, C1orf233, CCNL2, CPSF3L, CPTP, DVL1, FAM132A, HES4, ISG15, KLHL17, MIB2, MMP23B, MRPL20, MXRA8, NOC2L, ATAD3A, CDK11B, PERM1, PLEKHN1, PUSL1, RNF223, SAMD11, SCNN1D, SDF4, SSU72, TMEM240, TMEM88B, TNFRSF18, TNFRSF4, TTLL10, UBE2J2, VWA1*
CNVR_352	16	61954549	62197375	242826	gain	*ABL2, AXDND1, SOAT1*
CNVR_353	16	69756940	69774641	17701	loss	
CNVR_354	16	81343003	81402055	59052	loss	*C1orf106, KIF21B*
CNVR_355	16	81460177	81720984	260807	complex	*CACNA1S, IGFN1, PKP1, TMEM9*
CNVR_356	17	2892	23517	20625	loss	
CNVR_357	17	502830	2437899	1935069	complex	*5S_rRNA, CPE, MAP9, MSMO1, NPY2R, TLL1, U2*
CNVR_358	17	25056695	25119996	63301	loss	*PRAME*
CNVR_359	17	34352319	34354178	1859	loss	
CNVR_360	17	47597456	47612294	14838	loss	
CNVR_361	17	50746686	50962760	216074	loss	
CNVR_362	17	51115979	51370688	254709	complex	
CNVR_363	17	72801495	73023888	222393	complex	*ZNF280B, ZNF280A*

CNVR_364	17	73088087	75148791	2060704	complex	*AIFM3, ARVCF, BCR, bta-mir-130b, bta-mir-185, bta-mir-2323, bta-mir-2893, bta-mir-301b, C22orf15, C22orf39, CABIN1, CCDC116, CCDC188, CDC45, CHCHD10, CLDN5, COMT, CRKL, DDT, DERL3, DGCR14, DGCR2, DGCR8, GGT5, GNAZ, GNB1L, GP1BB, GSC2, GUCD1, HIC2, HIRA, KLHL22, LZTR1, MAPK1, MED15, MIF, MMP11, MRPL40, GSTT4, GSTT2, TANGO2, MZT2B, ZNF74, ADORA2A, TUBA3E, CCDC74B, GSTT1, P2RX6, PI4KA, PPIL2, PPM1F, RAB36, RANBP1, RSPH14, RTN4R, SCARF2, SDF2L1, SEPT5, SERPIND1, SLC25A1, SLC7A4, SMARCB1, SMPD4, SNAP29, SNORA77, SNRPD3, SPECC1L, SUSD2, THAP7, TMEM191B, TRMT2A, SSK1B, TSSK2, TXNRD2, U6, UBE2L3, UFD1L, UPB1, VPREB3, YDJC, YPEL1, ZDHHC8, ZNF70*
CNVR_365	18	9741999	9752287	10288	loss	*CDH13*
CNVR_366	18	13372279	13439223	66944	gain	*BANP, CA5A*
CNVR_367	18	27260474	27289775	29301	loss	
CNVR_368	18	28172582	28194989	22407	loss	
CNVR_369	18	28238841	28250625	11784	complex	
CNVR_370	18	28549564	28569092	19528	complex	
CNVR_371	18	28903336	28915852	12516	complex	
CNVR_372	18	29053999	29064198	10199	loss	
CNVR_373	18	29524967	29527100	2133	complex	
CNVR_374	18	29683997	29690176	6179	loss	
CNVR_375	18	29948741	29983673	34932	complex	
CNVR_376	18	30793331	30829313	35982	complex	
CNVR_377	18	32252914	32255615	2701	complex	
CNVR_378	18	49534810	49547073	12263	loss	
CNVR_379	18	60939823	60958365	18542	complex	
CNVR_380	18	60972689	61054591	81902	complex	*ZNF331*
CNVR_381	18	61224167	61234666	10499	complex	
CNVR_382	18	61309478	61929947	620469	complex	
CNVR_383	18	61988764	62065073	76309	complex	*CACNG7, MYADM, PRKCG*

CNVR_384	18	62102358	62104335	1977	complex	
CNVR_385	18	62195088	62253639	58551	complex	*NLRP9*
CNVR_386	18	62372459	62822909	450450	complex	*BRSK1, COX6B2, EPS8L1, FIZ1, HSPBP1, IL11, ISOC2, KMT5C, NAT14, PPP1R12C, PPP6R1, PTPRH, RDH13, RPL28, SBK2, SHISA7, SSC5D, SYT5, TMEM150B, TMEM190, TMEM86B, TNNI3, TNNT1, U6, UBE2S, ZNF628, ZNF784, ZNF865*
CNVR_387	18	63076581	63122224	45643	complex	*CDC42EP5, LENG8, LENG9, TTYH1*
CNVR_388	18	63131499	63134077	2578	complex	
CNVR_389	18	63512359	63577253	64894	complex	
CNVR_390	18	63763902	63806033	42131	complex	
CNVR_391	18	64804480	64811257	6777	complex	
CNVR_392	18	64919064	64922303	3239	complex	*ZNF550*
CNVR_393	18	65222323	65237934	15611	complex	*ZNF211*
CNVR_394	18	65284002	65298014	14012	complex	*ZNF154*
CNVR_395	18	65743470	65999195	255725	complex	*A1BG, CENPBD1, CHMP2A, MZF1, RPS5, SLC27A5, SNORA30, TRIM28, UBE2M, ZNF132, ZNF274, ZNF446, ZNF584, ZNF8*
CNVR_396	19	391270	605671	214401	complex	
CNVR_397	19	829604	837958	8354	loss	*CA10*
CNVR_398	19	887492	991707	104215	complex	*CA10*
CNVR_399	19	1063665	1067869	4204	loss	*CA10*
CNVR_400	19	1194221	1198623	4402	loss	*CA10*
CNVR_401	19	1517845	1535401	17556	complex	
CNVR_402	19	1769782	1814664	44882	complex	
CNVR_403	19	1835437	1872426	36989	complex	
CNVR_404	19	1984300	2105374	121074	complex	
CNVR_405	19	2270929	2282405	11476	loss	
CNVR_406	19	2321350	2340436	19086	complex	
CNVR_407	19	2403059	2414418	11359	loss	
CNVR_408	19	2484709	2527884	43175	loss	
CNVR_409	19	2538355	2540445	2090	complex	
CNVR_410	19	2551336	2575861	24525	loss	
CNVR_411	19	2599454	2646238	46784	complex	

CNVR_412	19	2662875	2665924	3049	loss	
CNVR_413	19	2698613	2718995	20382	loss	
CNVR_414	19	2816582	2889922	73340	complex	
CNVR_415	19	3025966	3029934	3968	loss	
CNVR_416	19	3044252	3172170	127918	complex	
CNVR_417	19	3232969	3337282	104313	complex	
CNVR_418	19	3475570	3523952	48382	complex	
CNVR_419	19	3607151	3621986	14835	complex	
CNVR_420	19	3634677	3662871	28194	complex	
CNVR_421	19	3689372	3694497	5125	complex	
CNVR_422	19	3766390	3854059	87669	complex	
CNVR_423	19	3981080	3983914	2834	complex	
CNVR_424	19	4237345	4253126	15781	complex	
CNVR_425	19	5616821	5622026	5205	loss	
CNVR_426	19	14601910	14606603	4693	loss	
CNVR_427	19	31961208	31963600	2392	loss	*ARHGAP44*
CNVR_428	19	51035766	51073939	38173	complex	
CNVR_429	19	51587553	51790892	203339	complex	*ALYREF, ANAPC11, ARHGDIA, ARL16, CCDC137, FAM195B, GCGR, HGS, SLC25A10, NPB, NPLOC4, OXLD1, P4HB, PCYT2, PDE6G, PPP1R27, SIRT7, TSPAN10*
CNVR_430	19	51820628	52265051	444423	complex	*7SK, AATK, ACTG1, BAHCC1, BAIAP2, bta-mir-338, bta-mir-3533, CEP131, ENTHD2, FAAP100, FSCN2, NPLOC4, SLC38A10*
CNVR_431	19	52306333	52338200	31867	loss	*RPTOR*
CNVR_432	19	63212668	63276370	63702	gain	*APOH, CEP112*
CNVR_433	19	63344091	63351304	7213	loss	
CNVR_434	19	63673271	63697953	24682	complex	*CACNG4*
CNVR_435	20	34351217	34352688	1471	loss	
CNVR_436	20	43071653	43081692	10039	loss	
CNVR_437	20	53050109	53054891	4782	complex	
CNVR_438	20	53085145	53115549	30404	loss	
CNVR_439	20	61393360	61478204	84844	loss	
CNVR_440	20	62460002	62467603	7601	loss	

CNVR_441	20	65053956	65067807	13851	loss	
CNVR_442	20	71155267	71871033	715766	complex	*5S_rRNA, AHRR, BRD9, CEP72, CLPTM1L, EXOC3, LPCAT1, NKD2, SLC12A7, SLC6A18, SLC6A19, SLC6A3, SLC9A3, TERT, TPPP, TRIP13*
CNVR_443	20	71942837	71992748	49911	complex	*CCDC127, LRRC14B, SDHA, U6*
CNVR_444	21	552613	599530	46917	loss	
CNVR_445	21	2719823	2738444	18621	loss	
CNVR_446	21	3882864	3896865	14001	loss	
CNVR_447	21	20079777	20190573	110796	complex	
CNVR_448	21	20196212	20211419	15207	complex	
CNVR_449	21	20333716	20335407	1691	loss	
CNVR_450	21	35222315	35234076	11761	complex	
CNVR_451	21	40832856	40837384	4528	loss	
CNVR_452	21	47857235	47862491	5256	loss	*MIPOL1*
CNVR_453	21	52305970	52329035	23065	loss	
CNVR_454	21	54164240	54208626	44386	loss	
CNVR_455	21	59279537	59301942	22405	complex	
CNVR_456	21	70810033	71468453	658420	complex	*ADSSL1, AHNAK2, AKT1, BRF1, BTBD6, C14orf79, C14orf80, CDCA4, CEP170B, GPR132, IGHE, INF2, JAG2, MTA1, NUDT14, PACS2, PLD4, SIVA1, TMEM121, U6, ZBTB42*
CNVR_457	21	71483648	71573501	89853	complex	
CNVR_458	22	13515723	13519684	3961	loss	
CNVR_459	22	18777748	18834063	56315	loss	*GRM7*
CNVR_460	22	22888039	22894597	6558	loss	
CNVR_461	22	26467884	26473227	5343	loss	
CNVR_462	22	26604789	26616978	12189	loss	
CNVR_463	22	40116995	40123451	6456	loss	*PTPRG*
CNVR_464	22	41522624	41597682	75058	gain	*FHIT*
CNVR_465	22	49982895	49984849	1954	loss	*DOCK3*
CNVR_466	22	57213784	57217013	3229	loss	*MKRN2*
CNVR_467	22	59193424	59243513	50089	loss	*NUP210*
CNVR_468	22	60032430	60227147	194717	complex	*DNAJB8, EEFSEC, RUVBL1S, EC61A1*
CNVR_469	22	60844745	61379134	534389	complex	*ALDH1L1, C3orf22, CFAP100, CHCHD6, CHST13,*

						KLF15, PLXNA1, SLC41A3, TXNRD3, UROC1, ZXDC
CNVR_470	23	557106	601091	43985	complex	*KHDRBS2*
CNVR_471	23	16476027	16489116	13089	loss	*GLTSCR1L*
CNVR_472	23	16975545	16978195	2650	loss	*ABCC10*
CNVR_473	23	21534544	21539674	5130	complex	
CNVR_474	23	24665722	24669690	3968	loss	*TRAM2*
CNVR_475	23	25335659	25452180	116521	complex	*BOLA-DQA1, BOLA-DQB, BOLA-DQA5*
CNVR_476	23	25626467	26337243	710776	complex	
CNVR_477	23	27665155	27700720	35565	complex	*LA-DQB*
CNVR_478	23	28448873	28469826	20953	complex	*JSP.1*
CNVR_479	23	28486281	28578501	92220	complex	*BOLA, TRIM26*
CNVR_480	23	28642554	28645014	2460	loss	
CNVR_481	23	28671240	28679974	8734	complex	*TRIM31*
CNVR_482	23	28818476	28874842	56366	complex	
CNVR_483	23	29024901	29096015	71114	complex	
CNVR_484	23	29295813	29299187	3374	complex	
CNVR_485	23	29473514	29489851	16337	complex	
CNVR_486	23	29788998	29808818	19820	loss	
CNVR_487	23	30441844	30443202	1358	loss	
CNVR_488	23	30544966	30548742	3776	complex	
CNVR_489	23	33573692	33661207	87515	loss	
CNVR_490	23	37063074	37071377	8303	loss	
CNVR_491	24	54398	399441	345043	loss	*5S_rRNA*
CNVR_492	24	6016877	6019806	2929	loss	
CNVR_493	24	6059474	6071575	12101	loss	
CNVR_494	24	10183509	10326608	143099	gain	
CNVR_495	24	14491426	14515490	24064	loss	
CNVR_496	24	16218919	16223480	4561	complex	
CNVR_497	24	16585921	16595650	9729	complex	
CNVR_498	24	17028792	17040399	11607	loss	
CNVR_499	24	17052607	17063489	10882	loss	
CNVR_500	24	18610795	18676702	65907	loss	
CNVR_501	24	24579328	24582206	2878	loss	
CNVR_502	24	25705529	25727624	22095	complex	

CNVR_503	24	29349786	29365628	15842	complex	
CNVR_504	24	52373257	52390527	17270	loss	
CNVR_505	24	53324484	53328928	4444	loss	
CNVR_506	24	56750126	56759394	9268	loss	*WDR7*
CNVR_507	24	62411069	62448869	37800	complex	
CNVR_508	24	62608951	62618436	9485	complex	*SERPINB10*
CNVR_509	25	25945	45134	19189	complex	
CNVR_510	25	60964	1384741	1323777	complex	*5S_rRNA, ARHGDIG, AXIN1, BAIAP3, C1QTNF8, C25H16ORF13, CACNA1H, CAPN15, CCDC154, CCDC78, CCSMST1, CHTF18, CLCN7, CRAMP1, DECR2, EME2, FAM173A, FAM195A, FBXL16, GNG13, GNPTG, HAGH, HAGHL, HBM, HBQ1, HN1L, IFT140, IGFALS, IL9R, JMJD8, LMF1, LUC7L, MAPK8IP3, METRN, MPG, MRPL28, MRPS34, MSLN, MSLNL, NARFL, NHLRC4, NME3, NME4, FAM234A, HBA1, HBA, HBZ, NPRL3, NUBP2, PDIA2, PIGQ, POLR3K, PRR35, PTX4, RAB11FIP3, RAB40C, RGS11, RHBDF1, RHBDL1, RHOT2, RPUSD1, SNRNP25, SOX8, SPSB3, SSTR5, STUB1, TEKT4, TELO2, TMEM204, TMEM8A, TSR3, UBE2I, UNKL, WDR24, WDR90, WFIKKN1*
CNVR_511	25	1401564	1854547	452983	complex	*ABCA3, ACA64, BRICD5, bta-mir-1225, bta-mir-1842, bta-mir-2382, bta-mir-940, CASKIN1, DNASE1L2, E4F1, ECI1, GFER, HS3ST6, MEIOB, MLST8, NDUFB10, NOXO1, NPW, NTHL1, PGP, PKD1, RAB26, RNF151, RNPS1, RPL3L, RPS2, SLC9A3R2, SNORA64, SNORD60, SYNGR3, TBL3, TRAF7, TSC2, ZNF598*
CNVR_512	25	1881725	1901339	19614	complex	
CNVR_513	25	1968761	2023043	54282	complex	*AMDHD2, ATP6V0C, C16orf59, TBC1D24, NTN3*
CNVR_514	25	2057495	2325172	267677	complex	*FLYWCH2, KCTD5, PDPK1, PRSS21, PRSS22, PRSS27, PRSS33, SRRM2*
CNVR_515	25	3526794	3533841	7047	loss	*CORO7, VASN*
CNVR_516	25	15577684	15585727	8043	loss	
CNVR_517	25	36110341	36119122	8781	complex	

CNVR_518	25	41164302	41537126	372824	complex	*AMZ1, BRAT1, CHST12, EIF3B, FTSJ2, GNA12, GRIFIN, IQCE, LFNG, MAD1L1, NUDT1, SNX8, TTYH3, ADAP1, bta-mir-2389, bta-mir-339a, C7orf50, COX19, CYP2W1, DNAAF5, ELFN1, FAM20C, GET4, GPER1, GPR146, INTS1, MAD1L1, MAFK, MICALL2, PDGFA, PRKAR1B, PSMG3, TMEM184A, UNCX, ZFAND2A*
CNVR_519	25	41585097	42747919	1162822	complex	*5S_rRNA*
CNVR_520	25	42808698	42851121	42423	complex	
CNVR_521	26	3067421	3102104	34683	loss	
CNVR_522	26	8659718	8677467	17749	loss	*ASAH2*
CNVR_523	26	25719640	25757447	37807	complex	
CNVR_524	26	28713575	28802987	89412	loss	
CNVR_525	26	34296519	34299839	3320	loss	
CNVR_526	26	34528805	34582753	53948	loss	*DCLRE1A*
CNVR_527	26	34671883	34682561	10678	complex	
CNVR_528	27	1106270	1122102	15832	loss	*CSMD1*
CNVR_529	27	5444267	5973102	528835	gain	*EBD*
CNVR_530	27	16100122	16171902	71780	loss	
CNVR_531	27	20135051	20347345	212294	loss	
CNVR_532	27	20972939	20980184	7245	loss	
CNVR_533	28	96862	105828	8966	complex	
CNVR_534	28	1195606	1199610	4004	complex	*GALNT2*
CNVR_535	28	1644820	1648309	3489	complex	
CNVR_536	28	1922407	1976561	54154	complex	
CNVR_537	28	2008114	2275570	267456	complex	*RHOU*
CNVR_538	28	2322619	2403958	81339	complex	
CNVR_539	28	2412723	2869287	456564	complex	*OR5L2, OR5AS1, OR5D14*
CNVR_540	28	11338585	11542172	203587	loss	
CNVR_541	28	23677341	23684507	7166	loss	*LRRTM3, CTNNA3*
CNVR_542	28	45649575	45703110	53535	loss	
CNVR_543	29	2371184	2374104	2920	loss	*FAT3*
CNVR_544	29	5155732	5681077	525345	complex	*TRIM77, TRIM64, U6*
CNVR_545	29	10346832	10386009	39177	loss	

CNVR_546	29	13219091	13227360	8269	loss	
CNVR_547	29	16137026	16144790	7764	loss	
CNVR_548	29	19332326	19359233	26907	loss	
CNVR_549	29	19397260	19484418	87158	loss	
CNVR_550	29	19877465	20316317	438852	gain	*LUZP2*
CNVR_551	29	21932178	21981201	49023	loss	
CNVR_552	29	24014819	24025296	10477	loss	
CNVR_553	29	27332982	27434090	101108	loss	
CNVR_554	29	27640611	27663302	22691	complex	
CNVR_555	29	38388712	38415482	26770	complex	
CNVR_556	29	38431365	38459739	28374	complex	
CNVR_557	29	38665259	38713186	47927	complex	
CNVR_558	29	38894641	38920839	26198	complex	
CNVR_559	29	39043214	39613214	570000	complex	*PAG1*
CNVR_560	29	39712182	39766845	54663	complex	
CNVR_561	29	42532912	42567601	34689	complex	
CNVR_562	29	46703510	46837463	133953	gain	*CPT1A, GAL, PPP6R3*
CNVR_563	29	50438384	50444320	5936	complex	*bta-mir-2409, AP2A2*

Supplementary File S2. Sheet 1. Clustered genes.

Category	Term	Gene count	P Value	Genes
Annotation Cluster 1	**Enrichment Score: 1.83**			
GOTERM_MF_FAT	GO:0004842~ubiquitin-protein ligase activity	8	1.1E-3	*MIB2, PPIL2, UBE2M, TRAF7, UBE2J2, UBE2L3, UBE2S, STUB1*
GOTERM_MF_FAT	GO:0019787~small conjugating protein ligase activity	9	2.9E-3	*MIB2, PPIL2, UBE2M, UBE2I, TRAF7, UBE2J2, UBE2L3, UBE2S, STUB1*
GOTERM_MF_FAT	GO:0016881~acid-amino acid ligase activity	9	2.5E-2	*MIB2, PPIL2, UBE2M, UBE2I, TRAF7, UBE2J2, UBE2L3, UBE2S, STUB1*
GOTERM_MF_FAT	GO:0016879~ligase activity, forming carbon-nitrogen bonds	10	2.9E-2	*ADSSL1, MIB2, PPIL2, UBE2M, UBE2I, TRAF7, UBE2J2, UBE2L3, UBE2S, STUB1*
Annotation Cluster 2	**Enrichment Score: 1.63**			
GOTERM_CC_FAT	GO:0000267~cell fraction	15	8.4E-3	*SNAP29, SOAT1, P4HB, ARHGDIG, SEPT2, ARFRP1, CORO7, ACTG1, PGP, DGCR2, TPPP, ENTPD8, TXNRD3, STEAP2, SDF4*
GOTERM_CC_FAT	GO:0005624~membrane fraction	12	3.4E-2	*SNAP29, SOAT1, P4HB, ARHGDIG, PGP, DGCR2, SEPT2, ENTPD8, ARFRP1, TXNRD3, CORO7, STEAP2*
GOTERM_CC_FAT	GO:0005626~insoluble fraction	12	4.2E-2	*SNAP29, SOAT1, P4HB, ARHGDIG, PGP, DGCR2, SEPT2, ENTPD8, ARFRP1, TXNRD3, CORO7, STEAP2*
Annotation Cluster 3	**Enrichment Score: 1.55**			
GOTERM_MF_FAT	GO:0017076~purine nucleotide binding	56	9.1E-3	*SEPT5, D2HGDH, SEPT2, PASK, DTYMK, RHOU, AKT1, ACTG1, PDPK1, NUBP2, RALA, STK39, RAB26, AIFM3, STK25, NLRP9, PI4KA, ABCC10, ARL16, PRKCG, UBE2J2, NDUFA10, NME4, PANK4, MAPK1, KIF1A, NME3, UBE2M, TXNRD3, TXNRD2, RUVBL1, GBP4, STEAP2, EEFSEC, UBE2S, PRKCZ, THAP7, ARFRP1, RAB40C, BRSK1, EPHB3, PTK6, SBK2,*

				ENTPD8, TUBA3E, AATK, ADSSL1, EEF1A2, UBE2L3, PDE6G, SDHA, RHOT2, TSSK1B, ABL2, CLCN7, SRMS
GOTERM_MF_FAT	GO:0000166~nucleotide binding	61	3.2E-2	*SEPT5, D2HGDH, SEPT2, PASK, DTYMK, RHOU, AKT1, ACTG1, PDPK1, NUBP2, RALA, STK39, RAB26, AIFM3, STK25, NLRP9, PI4KA, ABCC10, ARL16, PRKCG, UBE2J2, NDUFA10, NME4, PANK4, MAPK1, KIF1A, NME3, UBE2M, TXNRD3, TXNRD2, RUVBL1, GBP4, STEAP2, EEFSEC, UBE2S, THAP7, PRKCZ, ARFRP1, RAB40C, BRSK1, EPHB3, EIF3B, PTK6, SBK2, ENTPD8, TUBA3E, AATK, ADSSL1, EEF1A2, RNPS1, SKI, SIRT7, UBE2L3, PDE6G, SDHA, NDOR1, RHOT2, TSSK1B, ABL2, CLCN7, SRMS*
GOTERM_MF_FAT	GO:0032553~ribonucleotide binding	50	4.4E-2	*SEPT5, SEPT2, PASK, DTYMK, RHOU, AKT1, ACTG1, PDPK1, NUBP2, RALA, STK39, RAB26, STK25, NLRP9, PI4KA, ABCC10, ARL16, PRKCG, UBE2J2, NDUFA10, PANK4, NME4, MAPK1, KIF1A, NME3, UBE2M, RUVBL1, GBP4, EEFSEC, UBE2S, PRKCZ, THAP7, ARFRP1, RAB40C, BRSK1, EPHB3, PTK6, SBK2, ENTPD8, TUBA3E, AATK, ADSSL1, EEF1A2, UBE2L3, PDE6G, RHOT2, TSSK1B, CLCN7, ABL2, SRMS*
GOTERM_MF_FAT	GO:0032555~purine ribonucleotide binding	50	4.4E-2	*SEPT5, SEPT2, PASK, DTYMK, RHOU, AKT1, ACTG1, PDPK1, NUBP2, RALA, STK39, RAB26, STK25, NLRP9, PI4KA, ABCC10, ARL16, PRKCG, UBE2J2, NDUFA10, PANK4, NME4, MAPK1, KIF1A, NME3, UBE2M, RUVBL1, GBP4, EEFSEC, UBE2S, PRKCZ, THAP7, ARFRP1, RAB40C, BRSK1, EPHB3, PTK6, SBK2, ENTPD8, TUBA3E, AATK, ADSSL1, EEF1A2, UBE2L3, PDE6G, RHOT2, TSSK1B, CLCN7, ABL2, SRMS*

Annotation Cluster 4	Enrichment Score: 1.53			
Category	Term	Count	PValue	Genes
GOTERM_CC_FAT	GO:0005833~hemoglobin complex	4	2.4E-3	*HBA, HBM, HBZ, HBQ1, HBA1*
GOTERM_BP_FAT	GO:0015671~oxygen transport	4	4.8E-3	*HBA, HBM, HBZ, HBQ1, HBA1*
GOTERM_MF_FAT	GO:0019825~oxygen binding	4	5.9E-3	*HBA, HBM, HBZ, HBQ1, HBA1*
GOTERM_BP_FAT	GO:0015669~gas transport	4	9.3E-3	*HBA, HBM, HBZ, HBQ1, HBA1*
GOTERM_MF_FAT	GO:0005344~oxygen transporter activity	3	2.2E-2	*HBA, HBM, HBQ1, HBA1*
Annotation Cluster 5	**Enrichment Score: 1.47**			
Category	Term	Count	PValue	Genes
GOTERM_MF_FAT	GO:0032561~guanyl ribonucleotide binding	16	2.8E-2	*SEPT5, THAP7, ADSSL1, SEPT2, EEF1A2, ARFRP1, RAB40C, ARL16, PDE6G, RHOU, RALA, RHOT2, TUBA3E, RAB26, EEFSEC, GBP4*
GOTERM_MF_FAT	GO:0019001~guanyl nucleotide binding	16	3.0E-2	*SEPT5, THAP7, ADSSL1, SEPT2, EEF1A2, ARFRP1, RAB40C, ARL16, PDE6G, RHOU, RALA, RHOT2, TUBA3E, RAB26, EEFSEC, GBP4*
GOTERM_MF_FAT	GO:0005525~GTP binding	15	4.5E-2	*SEPT5, THAP7, ADSSL1, SEPT2, EEF1A2, ARFRP1, RAB40C, ARL16, RHOU, RALA, RHOT2, TUBA3E, RAB26, EEFSEC, GBP4*
Annotation Cluster 6	**Enrichment Score: 1.35**			
GOTERM_MF_FAT	GO:0043167~ion binding	87	3.4E-2	*FHIT, ZNF584, RNPEPL1, JAG2, ZXDC, RHOU, MKRN2, ZGPAT, GATA5, NUBP2, ZNF446, FIZ1, ZCCHC7, ZNF280A, ZNF280B, STK25, EGFL7, NUDT14, RXRA, DLL1, PRKCG, MMP11, HAGH, RAB11FIP3, CPE, ZNF784, ACAP3, MIB2, ZFAND2A, TRIM31, HAGHL, STEAP2, PMPCA, ING5, ARFGAP1, PRKCZ, TRAF2, ZNF274, THAP4, ADAMTS13, HBM, COMT, PRDM16, CDH4, ZNF331, ADAP1, TRIM64, HBA, FAT3, LPCAT1, PLCH2, LHX3,*

				ENTPD8, TTYH1, ZNF74, TCEA2, TRAF7, SDF4, NTHL1, ZNF70, AGL, MICALL2, ADSSL1, NPLOC4, POLR3K, KLF15, HBA1, SIRT7, DBH, TP73, ZNF628, ITPR2, CDH13, NDOR1, RNF151, MMP23B, ZNF211, HBZ, MZF1, HGS, HBQ1, RHOT2, ZMYND19, TSSK1B, ABL2, CLCN7, TLL1, CACNA1B
GOTERM_MF_FAT	GO:0043169~cation binding	85	4.9E-2	*FHIT, ZNF584, RNPEPL1, JAG2, ZXDC, RHOU, MKRN2, ZGPAT, GATA5, NUBP2, ZNF446, FIZ1, ZCCHC7, ZNF280A, ZNF280B, STK25, EGFL7, NUDT14, RXRA, DLL1, PRKCG, MMP11, HAGH, RAB11FIP3, CPE, ZNF784, ACAP3, MIB2, ZFAND2A, TRIM31, HAGHL, STEAP2, PMPCA, ING5, ARFGAP1, PRKCZ, TRAF2, ZNF274, THAP4, ADAMTS13, HBM, COMT, PRDM16, CDH4, ZNF331, ADAP1, TRIM64, HBA, FAT3, LPCAT1, PLCH2, LHX3, ENTPD8, ZNF74, TCEA2, TRAF7, SDF4, NTHL1, ZNF70, AGL, MICALL2, ADSSL1, NPLOC4, POLR3K, KLF15, HBA1, SIRT7, DBH, TP73, ZNF628, ITPR2, CDH13, NDOR1, RNF151, MMP23B, ZNF211, HBZ, MZF1, HGS, HBQ1, RHOT2, ZMYND19, TSSK1B, ABL2, TLL1, CACNA1B*
Annotation Cluster 7	**Enrichment Score: 1.32**			
GOTERM_BP_FAT	GO:0042424~catecholamine catabolic process	3	5.2E-3	*SLC6A3, COMT, DBH*
GOTERM_BP_FAT	GO:0042420~dopamine catabolic process	3	5.2E-3	*SLC6A3, COMT, DBH*
GOTERM_BP_FAT	GO:0034313~diol catabolic process	3	5.2E-3	*SLC6A3, COMT, DBH*
GOTERM_BP_FAT	GO:0019614~catechol catabolic process	3	5.2E-3	*SLC6A3, COMT, DBH*
GOTERM_BP_FAT	GO:0042417~dopamine metabolic	3	3.5E-2	*SLC6A3, COMT, DBH*

	process			
GOTERM_BP_FAT	GO:0042402~biogenic amine catabolic process	3	3.5E-2	*SLC6A3, COMT, DBH*
Annotation Cluster 8	**Enrichment Score: 0.91**			
GOTERM_BP_FAT	GO:0000271~polysaccharide biosynthetic process	4	2.1E-2	*AKT1, CHST12, EXT1, AGL*
Annotation Cluster 8	**Enrichment Score: 0.77**			
KEGG_PATHWAY	bta04940:Type I diabetes mellitus	5	3.9E-2	*BOLA-DQA1, CPE, BOLA-DQA5, BOLA-DQB, JSP.1*
Annotation Cluster 10	**Enrichment Score: 0.66**			
GOTERM_BP_FAT	GO:0045862~positive regulation of proteolysis	3	4.9E-2	*AKT1, AURKAIP1, STUB1*
Annotation Cluster 11	**Enrichment Score: 0.65**			
GOTERM_CC_FAT	GO:0005669~transcription factor TFIID complex	3	4.5E-2	*GATA5, EDF1, TBP*

Supplementary File S2. Sheet 2. Not clustered genes.

Category	Terms	Count	Genes	P-Value
GOTERM_MF_FAT	GO:0019904~protein domain specific binding	9	*CHMP2A, CARD9, STMN3, BAIAP2, HGS, UBE2I, LUC7L, TNNI3, STUB1*	9.00E-03
GOTERM_MF_FAT	GO:0050660~FAD binding	6	*SDHA, D2HGDH, AIFM3, TXNRD3, TXNRD2, STEAP2*	1.80E-02
GOTERM_MF_FAT	GO:0046914~transition metal ion binding	63	*FHIT, ZNF584, RNPEPL1, ZXDC, MKRN2, ZGPAT, GATA5, NUBP2, ZNF446, FIZ1, ZCCHC7, ZNF280A, ZNF280B, RXRA, PRKCG, MMP11, HAGH, CPE, ZNF784, ACAP3, MIB2, ZFAND2A, TRIM31, HAGHL, STEAP2, PMPCA, ING5, ARFGAP1, PRKCZ, TRAF2, ZNF274, THAP4, ADAMTS13, HBM, PRDM16, ZNF331, ADAP1, TRIM64, HBA, LHX3, ZNF74, TCEA2, TRAF7, NTHL1, ZNF70, MICALL2, NPLOC4, POLR3K, KLF15, SIRT7, HBA1, DBH, TP73, ZNF628, NDOR1, RNF151, ZNF211, MMP23B, HBZ, MZF1, HGS, HBQ1, ZMYND19, ABL2*	4.10E-02
GOTERM_MF_FAT	GO:0050662~coenzyme binding	9	*SDHA, SOAT1, NDOR1, D2HGDH, AIFM3, TXNRD3, TXNRD2, SIRT7, STEAP2*	4.20E-02
GOTERM_MF_FAT	GO:0008270~zinc ion binding	52	*ZNF584, RNPEPL1, ZXDC, MKRN2, ZGPAT, GATA5, ZNF446, FIZ1, ZCCHC7, ZNF280A, ZNF280B, RXRA, PRKCG, MMP11, HAGH, ZNF784, CPE, ACAP3, TRIM31, ZFAND2A, MIB2, HAGHL, PMPCA, ING5, ARFGAP1, PRKCZ, TRAF2, ZNF274, THAP4, ADAMTS13, PRDM16, ZNF331, ADAP1, TRIM64, LHX3, ZNF74, TCEA2, TRAF7, ZNF70, MICALL2, NPLOC4, POLR3K, SIRT7, KLF15, TP73, ZNF628, RNF151, ZNF211, MMP23B, MZF1, HGS, ZMYND19*	4.50E-02
GOTERM_MF_FAT	GO:0003924~GTPase activity	7	*THAP7, EEF1A2, RALA, TUBA3E, GBP4, EEFSEC, RHOU*	4.60E-02
GOTERM_BP_FAT	GO:0007017~microtubule-based process	9	*PRKCZ, THAP7, KIF1A, STMN3, TPPP, RHOT2, RANBP1, TUBA3E, TEKT4*	4.70E-02

GOTERM_BP_FAT	GO:0042058~regulation of epidermal growth factor receptor signaling pathway	3	*CDH13, ZGPAT, PDE6G*	1.20E-02
GOTERM_BP_FAT	GO:0000226~microtubule cytoskeleton organization	6	*PRKCZ, THAP7, STMN3, TPPP, RANBP1, TEKT4*	3.40E-02
GOTERM_CC_FAT	GO:0031301~integral to organelle membrane	6	*RHOT2, RER1, STEAP2, EXT1, LFNG, SEC61A1*	3.90E-02
GOTERM_CC_FAT	GO:0042995~cell projection	14	*SEPT2, PDGFA, SYT5, BAIAP2, PRKCG, RHOU, PEX5L, CTNNA3, AKT1, CDH13, GNB1, TUBA3E, TEKT4, CACNA1B*	1.20E-02
KEGG_PATHWAY	bta00240:Pyrimidine metabolism	9	*NME4, NME3, POLR3K, UPB1, DTYMK, ENTPD8, TXNRD3, TXNRD2, UCKL1*	1.00E-02
KEGG_PATHWAY	bta04330:Notch signaling pathway	6	*NOTCH1, HES5, JAG2, DLL1, LFNG, DVL1*	1.20E-02
KEGG_PATHWAY	bta05223:Non-small cell lung cancer	6	*AKT1, FHIT, MAPK1, PDPK1, RXRA, PRKCG*	2.40E-02
KEGG_PATHWAY	bta04150:mTOR signaling pathway	6	*AKT1, MAPK1, PDPK1, TSC2, MLST8, RPTOR*	2.60E-02

Supplementary Table S3a. Description and details of primers designed for the validation of the 15 consensus CNVRs.

CNVR	Bta	Start	End	Forward	Reverse	Probe
CNVR_58_complex	2	63864966	64072254	TGCATGCACACAGGAATGTTAC	TGCCCCTAAGAAGGAGTCGTT	ACTCTGTTCAGCCCTTC
CNVR_72_complex	3	7957960	7964523	GACTAATGGCAAGAGCCGTGTA	AGGCAGGAACAGAAGGAGGAA	TGAGCATGTCACTTTAA
CNVR_96_gain	4	6223614	6468141	GCCCGGCGGACACTAAG	CCAGCATTATGTCCTTCATCAACA	TCAGGAAGCTGTGGCCA
CNVR_104_loss	4	73676903	73785765	TCCTGCCCAGATACCATATCCTT	CGAGGCAAGCTCTACAGGAAA	TGGCATTCAAATCAC
CNVR_119_complex	5	58966295	59215089	GGAGATAGGATAGAAAGAAATGGAGAAC	ATGGGAGTGATGGAAAATTGAAG	CACTCTTAAATTCCC
CNVR_124_complex	5	103122325	103454169	CTCCCGAAGGAGGAAGAGATAGT	TGCAGGCAGCTGGGATGT	CGTCCAGGGCTTC
CNVR_271_loss	12	67537603	67576208	GCCTGGTGTTGTCATGATGAA	CCCGTACACTGACACCAAAGTG	TTTTGCGCTTGAAGCAG
CNVR_276_loss	12	82144071	82215811	TGTCTGTGATAGGCTTGGCA	CTGAAGCATTTGGGCCCTTT	TCAGAATTCTCAAGGTCACGC
CNVR_280_complex	13	53932232	53982995	TGCGAAATTCTGGAAGAGGAA	GGGTGCCTGGTGCAATTC	CCTGAGGACATGAAGTT
CNVR_315_loss	15	15764194	15773838	GGATCAGGGAAGGCCAATTC	CCCACGACCTCAAGCTCATT	AGGATGACATGAAAGAC
CNVR_338_complex	16	15066677	15247073	TGGGAGTGATTACCTGGTGTTTC	ACAGTGGCAAAATAAGATGCATTG	TAGGGAGACAATGATTCAGT
CNVR_411_complex	19	2599454	2646238	CAGTGAGCCAAAGCCAATCC	AATCCAACTTGCCGGCTAGTATT	CCTCCACAGGAATC
CNVR_539_complex	28	2412723	2869287	ACATTCAGGCTGCCATTTTGT	GAGGCGGGATGTCACAGAAA	TCCAATATCGTCAACCATT
CNVR_544_complex	29	5155732	5681077	CACGGGCGCACCACTT	CCCCCGATGAATGGCTATC	AGCTCCCTGCTCGAC
CNVR_551_loss	29	21932178	21981201	CCCTCGCTATCGGTGTTGAT	GCCAGGTAAAAGACAGAGAATGC	TCTCTTGGGTACAACTC

Supplementary Table S3b. Results of the qPCR validation.

CNVR	Bta	Samples	Confirmed sample rate	Validated region
CNVR_58_complex	2	7	0,71	YES
CNVR_72_complex	3	1	1	YES
CNVR_96_gain	4	11	1	YES
CNVR_104_loss	4	4	1	YES
CNVR_119_complex	5	9	0,67	YES
CNVR_124_complex	5	11	0,82	YES
CNVR_271_loss	12	8	0,88	YES
CNVR_276_loss	12			NO
CNVR_280_complex	13	3	1	YES
CNVR_315_loss	15	4	1	YES
CNVR_338_complex	16	11	1	YES
CNVR_411_complex	19	11	0,73	YES
CNVR_539_complex	28	11	0,73	YES
CNVR_544_complex	29	11	1	YES
CNVR_551_loss	29	11	0,91	YES

Supplementary Figure 1. Graphical representation of the CNVR_96_gain validation qPCR result from Copy Caller.

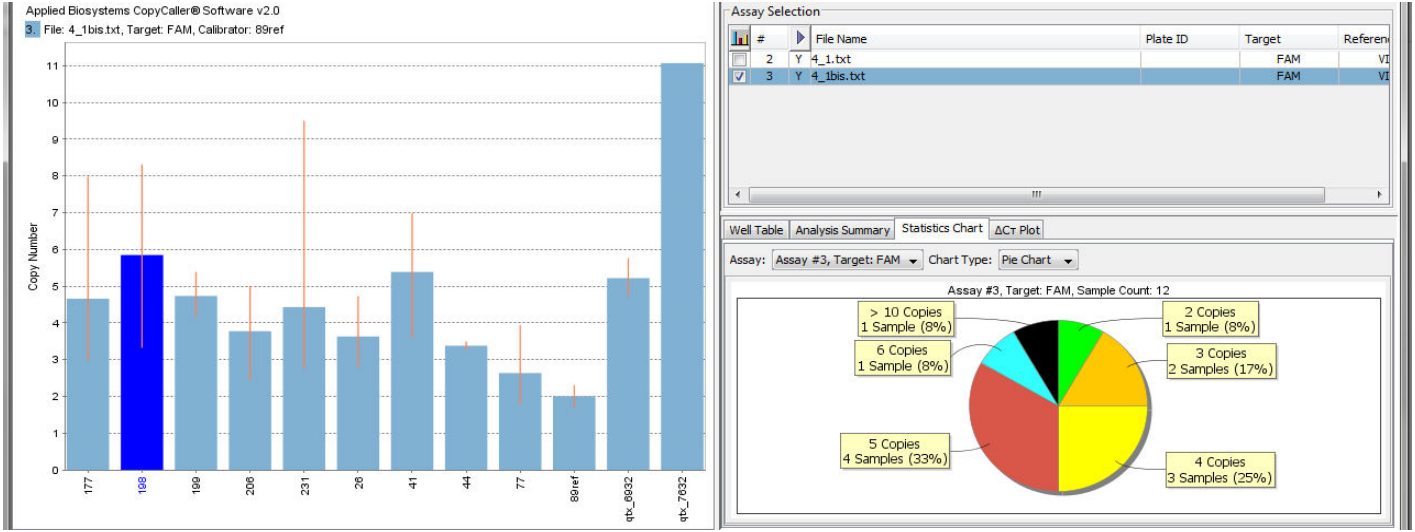


Figure 2. Graphical representation of the CNVR_315_loss validation qPCR result from Copy Caller.

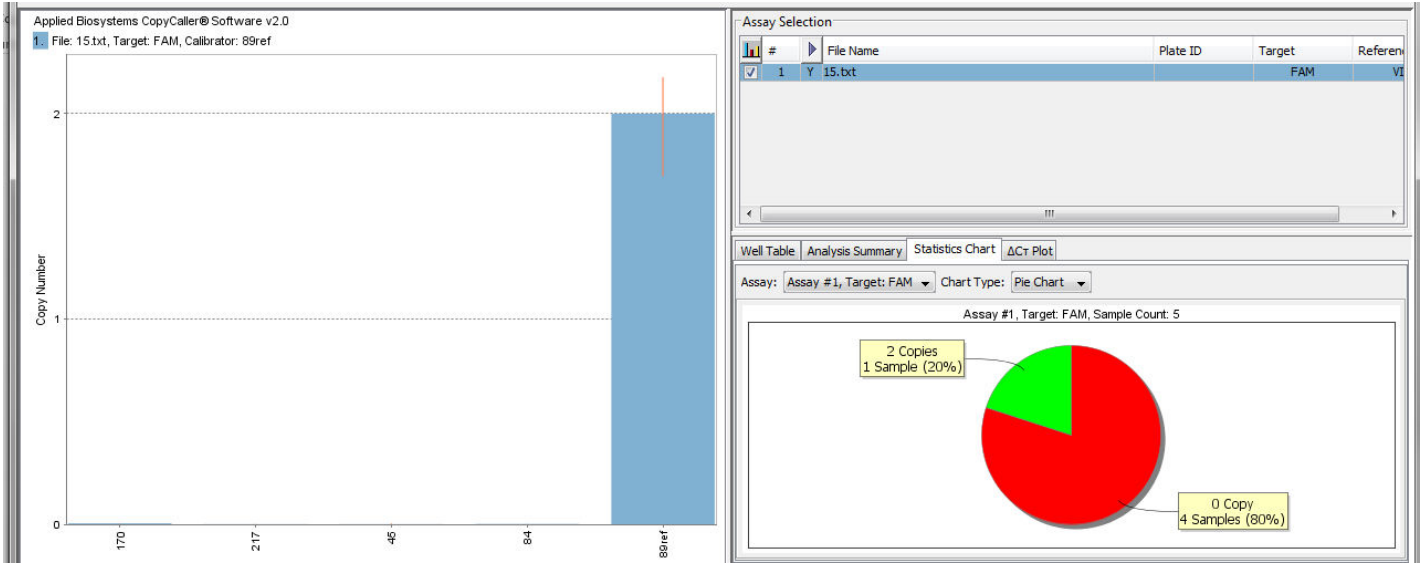
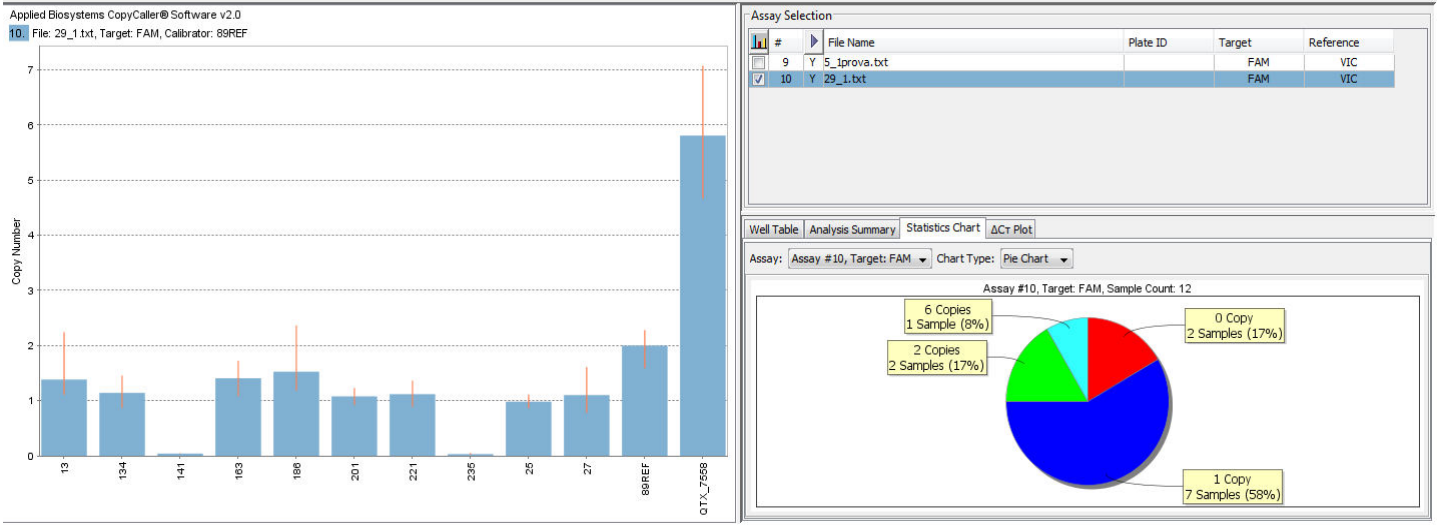


Figure 3. Graphical representation of the CNVR_544_complex validation qPCR result from Copy Caller.



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

Submitted to: Livestock Science

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
raphaelle.prinsen@unimi.it, maria.strillacci@unimi.it, alessandro.bagnato@unimi.it
[b] Associazione Nazionale Allevatori Razza Bruna, Località Ferlina, Bussolengo, 204, VR 37012, Italy
attilio.rossoni@anarb.it
[c] Qualitas AG, Chamerstrasse 56, 6300 Zug Switzerland
Birgit.Gredler@qualitasag.ch
[d] Research Institute of Organic Agriculture, Ackerstrasse 113, 5070 Frick, Switzerland
anna.bieber@fibl.org

Corresponding author: Maria Giuseppina Strillacci, email: maria.strillacci@unimi.it

Abstract

Copy Number Variation (CNV) can be used in association studies to disclose the genetic basis of quantitative traits phenotypic variation. 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. Up to date, no genome-wide association study (GWAS) with CNVs and quantitative traits in such a large Brown Swiss population (i.e. with 1,116 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).

Key Words: CNV; GWAS; Brown Swiss

Introduction

Different genomic features can regulate quantitative traits through the variation of gene expression levels. Copy Number Variation (CNV), a type of genomic structural variation, is one of those genomic features. CNVs are a type of genetic variation in which a sequence of

nucleotides varies in copy number (i.e. gain or loss) in an individual's genome when compared with a reference genome and can range from 50 bp, up to several Mb (Mills et al., 2011).

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 (Zhang et al., 2009). CNVs coverage represents approximately 4,6% of the entire bovine genome (Hou et al., 2011). Additionally CNVs have been shown to be associated with complex traits in several species, including cattle (Zhou et al., 2016, Zhu et al., 2016 and Xu et al., 2014). At present the majority of studies in cattle are focused on the detection of CNVs (Keel et al., 2016, da Silva et al., 2016, Prinsen et al., 2016 and Bagnato et al., 2015).

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 through a CNV-GWAS study, some CNVs to be 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).

Up to now, no study on the relationship between CNVs and economically important traits in the Brown Swiss breed using HD SNP microarray data for the CNV detection is available.

This study describes a genome-wide association analysis (GWAS) between CNVs and eight economically important traits: Milk yield (MY), Fat yield (FY), Protein yield (PY), Somatic cell score (SCS), Fore udder attachment (FUA), Rear udder height (RUH), Rear udder width (RUW) and Overall udder score (UDD), in order to identify candidate CNVs and genes that could be used in cattle genetic improvement programs.

Material and methods

Sampling and genotyping

This study used a total of 1,116 Brown Swiss dairy cattle breed samples, of which 313 males, and 803 females. Of these 1,116 samples, 160 samples were genotyped by Braunvieh Schweiz and 761 samples were genotyped within the EC funded LowInputBreeds project. The last 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.

Phenotypes

The Estimated Breeding Values (EBVs) for the production traits (milk yield - MY, fat yield - FY, 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 were calculated as:

$DEBV = PA + (EBV - PA) / REL_{dau}$

where:

DEBV = Deregressed EBV, *PA* = Parent Average, *EBV* = Estimated Breeding Value and REL_{dau} = Reliability from daughters according to Van Raden et al., (2009) and used as dependent variable in the association analysis.

CNV-association with productive, udder and health traits

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

The SVS 8.4.4 Correlation-trend test was used for the CNV-association with the traits MY, FY, PY, FUA, RUH, RUW, UDD and SCS. 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.

CNVs annotation

Gene content of associated CNVs was determined by downloading the complete Ensembl v.86 autosomal UMD3.1 gene set from the http://www.ensembl.org/biomart/martview/32471e73613a45753e5689f2626f9add website. Annotated bovine genes within significant associated CNVs were found using the intersectBed command of BEDTools (Quinlan and Hall, 2010). CNV overlaps with genes were defined 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 false discovery rate correction (FDR).

Results and discussion

CNV-association and annotation results

Not all the 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.

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

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).

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. Figure 1 and Figure 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.

CNV_ID	BTA	Start Position	End Position	Length (bp)	Gene symbol	P value after FDR correction						
						MY	FY	PY	FUA	RUH	RUW	UDD
CNV_1	1	157863948	157864818	870					3.16E-02			
CNV_2	2	24941791	24944009	2218								3.54E-02
CNV_3	2	27363242	28053884	690642	*NOSTRIN, CERS6*				3.90E-02			
CNV_4	2	27365813	28041802	675989	*NOSTRIN, CERS6*				4.50E-02			
CNV_5	2	64237382	64791135	553753	*NCKAP5*						4.61E-02	
CNV_6	4	43481123	43487686	6563	*MAGI2*		3.81E-02					
CNV_7	5	17129202	17130567	1365			3.80E-02					
CNV_8	6	86150255	86154933	4678			1.34E-02	2.41E-02	6.66E-03			
CNV_9	9	6964612	6977472	12860		2.29E-02	3.13E-02					
CNV_10	9	29175108	29177772	2664		2.18E-02	1.05E-02	3.11E-02	3.55E-02			6.23E-03
CNV_11	9	56016603	56024933	8330					3.10E-02	2.50E-02		
CNV_12	10	22156872	23191564	1034692	*TRAV29DV5, TRDC, TRAC*						3.75E-02	
CNV_13	10	22159956	25173443	3013487	*TRAV29DV5, TRAV14DV4, TRAV17, TRAV18, TRDC, TRAV16, TRAC*						3.75E-02	
CNV_14	10	23125923	24756165	1630242	*TRAC*						4.87E-02	
CNV_15	10	23133923	25173443	2039520	*TRAV14DV4, TRAV17, TRAV18, TRAV16, TRAC*						4.87E-02	
CNV_16	10	23152360	24710421	1558061	*TRAC*						4.87E-02	
CNV_17	10	27063674	27098152	34478							3.21E-02	
CNV_18	10	27064247	27086857	22610							3.21E-02	
CNV_19	10	27067275	27078449	11174							3.99E-02	
CNV_20	10	27077351	27084658	7307							3.99E-02	
CNV_21	10	27084658	27095664	11006							3.21E-02	
CNV_22	10	27095664	27098152	2488							3.73E-02	
CNV_23	11	17341756	17413419	71663			2.31E-02					
CNV_24	11	17359045	17406557	47512			2.31E-02					
CNV_25	12	71173494	72419043	124554		4.17E-02						

CNV_26	12	72432362	73323427	9 891065		2.79E-02	5.01E-03	3.14E-02				
CNV_27	12	72444330	73745994	130166 4		2.79E-02	5.01E-03	3.14E-02				
CNV_28	12	72499272	73673865	117459 3		2.79E-02	5.01E-03	3.14E-02				
CNV_29	12	72504287	73772980	126869 3		2.09E-02	6.37E-03	2.87E-02				
CNV_30	12	72507076	73345563	838487		2.80E-02	7.27E-03					
CNV_31	12	74241667	74359566	117899			4.11E-02					
CNV_32	12	74302958	78490302	418734 4	*OXGR1, DZIP1, DNAJC3, CLDN10, HS6ST3, UGGT2, MBNL2*		4.86E-02					
CNV_33	12	74327005	74787542	460537			2.41E-02					
CNV_34	12	75770032	76191654	421622			3.09E-02	2.54E-02				
CNV_35	12	75824857	76270340	445483			3.83E-02	2.37E-02	3.36E-02	2.09E-02		
CNV_36	12	75855694	75946247	90553			3.83E-02	2.37E-02	3.36E-02	2.09E-02		
CNV_37	17	50956421	51097333	140912							1.51E-02	
CNV_38	17	50958929	51097333	138404							1.51E-02	
CNV_39	18	7448849	7450797	1948	*CDYL2*					1.01E-02		
CNV_40	18	25754843	25756038	1195	*ADGRG3*				8.96E-03			
CNV_41	19	2631478	2637013	5535					7.51E-03			
CNV_42	20	62461185	62467603	6418					6.82E-03	4.97E-02		
CNV_43	20	62466989	62467603	614					3.46E-02	4.97E-02		
CNV_44	23	25417035	25428298	11263					3.46E-02			
CNV_45	23	25426985	25442124	15139					3.46E-02			
CNV_46	23	25429469	25747610	318141	*BOLA-DRB3, BoLA DR-ALPHA*				3.46E-02			
CNV_47	23	25440853	25443589	2736					7.84E-03			
CNV_48	23	25442124	25875898	433774	*LA-DQB, BOLA-DRB3, BoLA DR-ALPHA*				4.74E-03			
CNV_49	23	25846520	26297751	451231	*LA-DQB*				8.05E-03			
CNV_50	23	25860822	25863611	2789	*LA-DQB*				2.87E-02			
CNV_51	23	29041704	29096015	54311		2.15E-02	7.45E-03					

CNV_52	23	29059352	29096015	36663	2.18E-02	7.66E-03	4.78E-02	
CNV_53	23	29078825	29096015	17190	2.18E-02	7.66E-03	4.78E-02	
CNV_54	24	62431830	62448399	16569				3.65E-02
CNV_55	26	28705385	28812479	107094				3.23E-02
CNV_56	29	42548131	42561300	13169		6.52E-03	3.29E-02	3.28E-02

Table 2. Significant CNV list, their position on the genome (chromosome number - BTA and starting and end bp on the chromosome), their length, symbol of the genes mapping in the associated CNVs and the GWAS p values for production traits (milk yield - MY, fat yield - FY, 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).

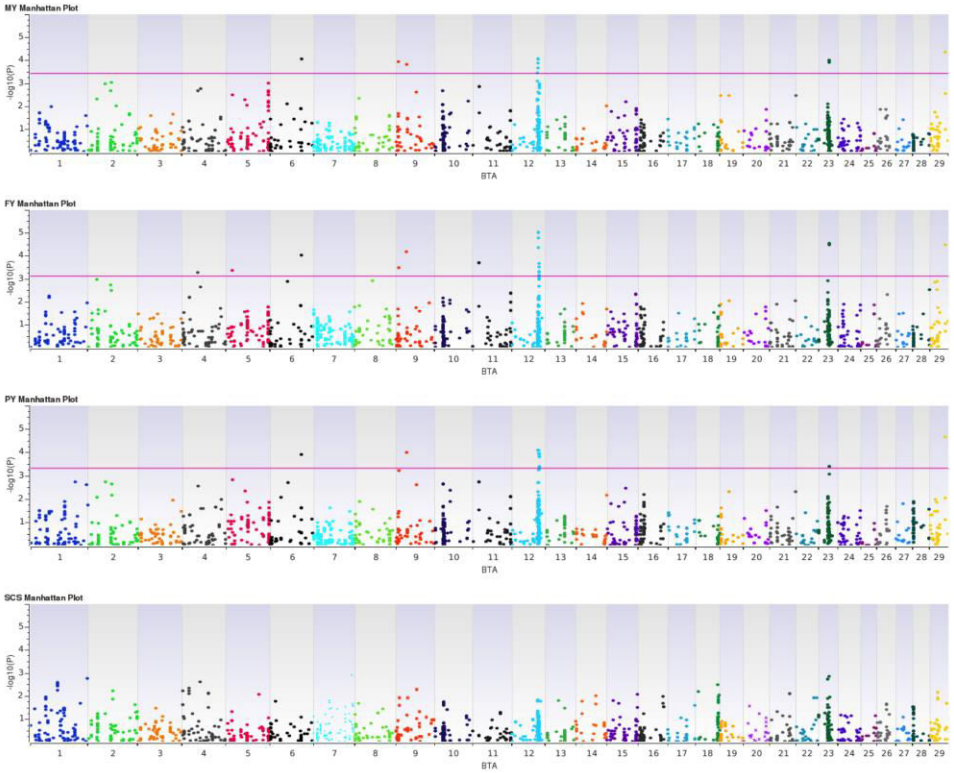


Figure 1. Manhattan Plots for the traits: MY, FY, PY and SCS. FDR 5% significance threshold plotted for all traits except for the SCS trait where no CNV was significant.

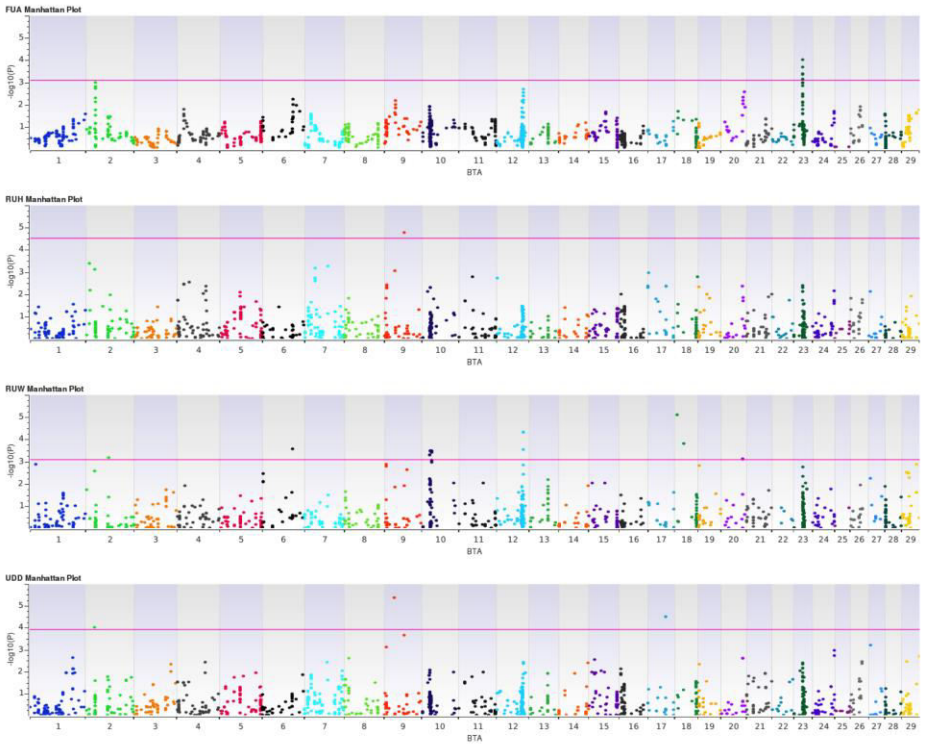


Figure 2. Manhattan Plots for the traits: FUA, RUH, RUW and UDD. FDR 5% significance threshold plotted for all traits.

For the production traits (Figure 1), 11, 22 and 12 CNVs were significantly associated with MY, FY and PY, respectively. In Figure 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 CNVs. On the contrary there are 40 CNVs that do not contain any annotated gene (Table 2). This may be due to the partial annotation, especially in non-human specie.

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 discontiguous 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)J somatic rearrangement of TRA (T cell receptor □□□and TRD (T cell receptor □□□chains expressed by □□ and □□ T-cells (Connelley et al., 2014). The V(D)J genetic recombination mechanism therefore allows the combination of highly diverse T cell receptors which is essential for the recognition of the huge number of antigens to which hosts could be exposed (Nikolich-Zugich et al., 2004). Moreover, as suggested by Connelley et al., (2014), analysis of the duplication events that generated the large number of TRAV/TRDV genes could elucidate how these genes evolved. Indeed, his study presents data indicating that the majority of the TRAV/TRDV gene

expansion is attributable to a series of duplications of multiple homology units. This could explain the CNVs found in this study concerning the genes mentioned above.

Within CNV_32 on chromosome 12, which resulted significantly associated with the FY trait, map the genes *CLDN10* (*Claudin-10*) and *HS6ST3 (Heparan sulfate 6-O-sulfotransferase 3)*. The *CLDN10* gene has been found to be overexpressed in the periprostatic adipose tissue of human prostate cancer patients (Ribeiro et al., 2012). This gene encodes for an essential tight junction element with a role in adipose tissue considering its functions in the stroma arrangement and cellular connections (Günzel et al., 2009).

The *HS6ST3* gene is part of a group of genes involved in the heparin metabolism. In a study based on a cross of 246 Wagyu x Limousin cattle breeds, authors tested the hypothesis that genes involved in heparan sulfate and heparin metabolism also affect the regulation of lipid metabolism in bovine muscle. Heparin is indeed reported to decrease the degradation rate of lipoprotein lipase in adipocytes and to promote the adipocyte differentiation (Jiang et al., 2011). Interestingly its presence has been demonstrated also in human milk (Coppa et al., 2013).

Some genes of the *BoLA* (*bovine leukocyte antigens*) Class II complex (*BOLA-DQ2, BOLA-DQB, BOLA-DRB3)* map in four CNVs (CNV_46, CNV_48, CNV_49 and CNV_50) significantly associated with FUA on BTA23. Class II molecules, expressed on antigen-presenting cells (APCs) (such as dendritic cells and macrophages), present pathogens derived peptides to $CD4^+$ T cells, which activate macrophages and B cells to generate inflammatory and antibody responses (Behl et al.,

2012). The evidence for the association of *BOLA-DRB3* with mammary gland phenotype and with mastitis resistance has been reported in different studies (Ogorevc et al., 2009).

The CNVs significantly associated with our traits were compared with the ones found in the study of Xu et al. (2014). In particular, CNV_25, CNV_26, CNV_27, CNV_28 and CNV_29 of our study (all associated with MY, FY and PY except for CNV_25 which is only associated with MY and FY) overlap with CNV22 (associated with Protein Percentage) and CNV27 (associated with MY and Protein Percentage) of Xu et al., (2014).

Sassi et al. (2016) performed a study in the Spanish Holstein cattle breed, performing a GWAS between CNVs mapped in the same population and production traits. Even if they found several CNVs associated to production traits, none of them overlap with the associated CNVs in this study. This may be due to the different breed used by Sassi et al., i.e. Spanish Holstein, compared to the one we analysed, i.e. Brown Swiss.

Surprisingly there were no significant CNVs associated with the SCS trait at a FDR < 5%. Even if this was the case we comment the CNVs with raw p-value's < 0.05 (i.e. before FDR correction) by comparing our results with the CNVs that were found to be associated with the SCS trait in the study of Durán Aguilar et al., (2016) performed in Holstein cattle. A total of 7 CNVs of this study (with p-value < 0.05) overlap with several CNVRs (copy number variation regions) of the study of Durán Aguilar et al., (2016). Specifically the overlapping CNVs are present on BTA 3 (50.16 – 50.19 Mb), BTA 10 (22.11 - 25.51 Mb), BTA 12 (70.36 – 76.75 Mb), BTA 13 (53.93 – 53.99 Mb), BTA 18 (60.93 – 65.99 MB),

BTA 23 (25.33 – 26.33 Mb) and BTA 23 (28.48 – 29.09 Mb). Of these 7 CNVs, 2 CNVs have a gene mapped within them. The CNV present on BTA 23 (25.33 – 26.33 Mb) overlaps with the *BOLA-DBR3* gene, as described above. The CNV present on BTA 3 (50.16 – 50.19 Mb) overlaps with the *FNBP1L* (*formin binding protein 1 like*) gene. This gene is involved in the cellular autophagy process, which is needed to remove defective organelles, protein aggregates, and intracellular pathogens (Huett et al., 2009).

The other available study considering the association of CNVs with SCS is the one of Sassi et al., (2016), where no common CNV associated with somatic cell count, the trait used in their study, was found.

No other study is available, at best of our knowledge, on GWAS between CNVs and production, functional and health traits in dairy cattle.

Several genes involved in the immune resistance map within significant CNVs on BTA10 and BTA23 that are associated with udder traits extensively studied in cattle for their relation with mastitis resistance. Here, in particular, we find a significant association with FUA and with RUW, two traits selected in dairy cattle for their association with low somatic cell count, thus indirectly related to mastitis resistance. The genetic correlations of RUW with each of the three traits, FUA, udder cleft and udder depth, have been estimated in the Italian Brown to be 0.66, 0.53, and -0.20 respectively (Samoré et al., 2010). It is well known by literature (De Groot et al., 2002; Kadarmideen, 2004; Samoré and Groen, 2006) that a deeper udder (i.e. higher than hocks) with a stronger cleft and a larger rear udder width in order to accommodate a

higher production, is associated to a higher mastitis resistance and higher longevity of milking cows. Here we found genes involved in immune resistance of individuals for the two linear udder traits FUA and RUW. Selection for mastitis resistance was based on the indirect selection using linear udder traits for decades, favoring cows with stronger udder cleft and FUA and larger RUW. These findings suggest that selection for udder traits have been contemporarily also based on selecting more efficient individuals for immune response.

In this study, we specifically identified CNVs associated with traits that are indirect indicators of mastitis resistance, a complex disease trait. As reported by Redon et al. (2006) CNV variation in human may affect gene expression for disease complex traits. For CNV_12 up till CNV_16 on BTA10 Prinsen et al., (2016) reported a complex (i.e. gain and loss) region. This, in association to findings of Connelley et al. (2014), suggests that the increasing number of copies of the TRA/TRD locus (associated to FUA and RUW in this study) may have facilitated the selection for higher expression of immune resistance involved loci, thus mastitis resistance and its relation with udder linear traits. The same holds true for the *BOLA* complex gene annotated in the associated CNVs here identified on BTA23 (Table 2).

Conclusion

The result of this study is a first genomic analysis on such a large sample of Brown Swiss cows associating CNVs to production, udder and health traits.

The association found between udder traits and the results of gene mapping in the significant CNVs provides an genomic evidence that

indirect selection for mastitis resistance using udder traits is generating an indirect response in immune resistance expression capability. At present, selection for mastitis resistance through indirect selection using udder traits and SCS is widely applied in dairy cattle populations and in selection indexes. The usefulness of morphological traits as criteria of selection has been always discussed. Their relation with quantitative productive functional and health traits was highlighted through the estimation of genetic correlations, but it was somehow difficult to express the genomic and biological basis of these relations. The general hypothesis for a more functional udder was based on the capacity of a deep udder with a strong fore attachment and cleft to last several lactations, being less in touch with manure and dirt in the farm and then to be less affected by infective agents. The associations of CNVs with FUA and RUW and the genes mapping within them in this study, suggest that the selection for improved udder morphology may be also driven by a genomic basis that favors the immune resistance and reaction of individuals. The genomic mechanism proposed in literature is a duplication of regions containing genes for immune resistance and to their gene expression.

Based on the results of this study the selection for mastitis resistance operated in the Brown Swiss population seems to reflect the increase in immune resistance and reaction capability at a genomic level. The genes mapping in the CNVs here identified, can be a first indication to further investigate their role in metabolic pathways related to mastitis resistance.

Conflict of interest statement

The authors of the present study declare that they do not have any potential conflict of interest including any financial, personal or other relationships with other people or organizations within three years of beginning the work submitted that could inappropriately influence the present work.

Acknowledgements

This work was supported by the EU projects Quantomics (Grant agreement no. 222664) and LowInputBreeds (Contract no. 222623) and by funds of Braunvieh Schweiz and Qualitas AG.

References

Bagnato, A., Strillacci, M.G., Pellegrino, L., Schiavini, F., Frigo E., Rossoni, A., Fontanesi, L., Maltecca, C., Prinsen, R.T.M.M., Dolezal, M.A., 2015. Identification and Validation of Copy Number Variants in Italian Brown Swiss Dairy Cattle Using Illumina Bovine SNP50 Beadchip. Ital J of Anim Sci. 14(3).

Behl, D.J., Verma, N. K., Tyagi, N., Mishra, P., Behl, R., Joshi, B. K., 2012. The Major Histocompatibility Complex in Bovines: A Review. ISRN Vet Sci. 2012: 872710.

Connelly, T.K., Degnan, K., Longhi, C.W., Morrison, W.I., 2014. Genomic analysis offers insights into the evolution of the bovine TRA/TRD locus. BMC Genomics, 15:994.

da Silva, J.M., Giachetto, P.F., da Silva, L.O., Cintra, L.C., Paiva, S.R., Yamagishi, M.E., Caetano, A.R., 2016. Genome-wide copy number variation (CNV) detection in Nelore cattle reveals highly frequent variants in genome regions harboring QTLs affecting production traits. BMC Genomics. 17:454.

de Almeida Santana, M. H., Oliveira Junior, G. A., Mello Cesar, A. S., Castelani Freua, M., da Costa Gomes, R., da Luz e Silva, S., Leme, P. R., Fukumasu, H., Carvalho, M. E., Vieira Ventura, R., Lehmann Coutinho, L., Kadarmideen, H. N., Sterman Ferraz J. B., 2016. Copy number variations and genome-wide associations reveal putative genes and metabolic pathways involved with the feed conversion ratio in beef cattle. J. Appl. Genetics. 57(4): 495-504.

De Groot, B.J., Keown, J.F., Van Vleck, L.D., Marotz, E.L., 2002. Genetic parameters and responses of linear type, yield traits and somatic cell scores to divergent selection for predicted transmitting ability for type in Holsteins. J. Dairy Sci. 85:1578-1585.

Durán Aguilar, M., Román Ponce, S.I., Ruiz López, F.J., González Padilla, E., Vásquez Peláez, C.G., Bagnato, A., Strillacci, M.G. 2016. Genome-wide association study for milk somatic cell score in Holstein cattle using copy number variation as markers. J Anim Breed Genet. ISSN 0931-2668.

VanRaden, P. M., Van Tassell, C.P., Wiggans, G. R., Sonstegard, T. S., Schnabel, R. D., Taylor, J. F., Schenkel F. S., 2009. Invited review:

Reliability of genomic predictions for North American Holstein bulls. J. Dairy Sci. 92 (1) :16–24.

Günzel, D., Stuiver, M., Kausalya, P.J., Haisch, L., Krug, S.M., Rosenthal, R., Meij, I.C., Hunziker, W., Fromm, M., Muller, D. 2009. Claudin-10 exists in six alternatively spliced isoforms that exhibit distinct localization and function. J Cell Sci. 122:1507–1517.

Hou, Y., Liu, G.E., Bickhart, D.M., Cardone, M.F., Wang, K., Kim, E.S., 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.

Huett, A., Ng, A., Cao, Z., Kuballa, P., Komatsu, M., Daly, M.J., Podolsky, D.K., Xavier, R.J., 2009. A novel hybrid yeast-human network analysis reveals an essential role for FNBP1L in antibacterial autophagy. J Immunol. 182(8): 4917-30.

Jiang, Z., Michal, J.J., Wu, X,L., Pan, Z., MacNeil, M.D., 2011. The heparan and heparin metabolism pathway is involved in regulation of fatty acid composition. Int J Biol Sci. 7(5):659-63.

Kadarmideen, H.N., 2004. Genetic correlations among body condition score, somatic cell score, milk production, fertility and conformation traits in dairy cows. Anim. Sci. 79:191-201.

Keel, B.N., Keele, J.W., Snelling, W.M., 2016. Genome-wide copy number variation in the bovine genome detected using low coverage sequence of popular beef breeds. Anim Genet. DOI: 10.1111/age.12519

Long, Y., Su, Y., Ai, H., Zhang, Z., Yang, B., Ruan, G., Xiao, S., Liao, X., Ren, J., Huang, L., Ding, N., 2016. A genome-wide association study of copy number variations with umbilical hernia in swine. Anim. Genet. 47(3): 298-305.

Mills, R.E., Walter, K., Stewart, C., Handsaker, R.E., Chen, K., Alkan, C., Abyzov, A., Yoon, S.C., Ye, K., Cheetham, R.K., Chinwalla, A., Conrad, D.F., Fu, Y., Grubert, F., Hajirasouliha, I., Hormozdiari, F., Iakoucheva, L.M., Iqbal, Z., Kang, S., Kidd, J.M., Konkel, M.K., Korn, J., Khurana, E., Kural, D., Lam, H.Y., Leng, J., Li, R., 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., McCarroll, S.A., Korbel, J.O.; 1000 Genomes Project. 2011. Mapping copy number variation by population-scale genome sequencing. Nature. 470(7332): 59–65.

Nikolich-Zugich, J., Slifka, M.K., Messaoudi, I., 2004. The many important facets of T-cell repertoire diversity. Nat Rev Immunol. 4(2): 123-32.

Ogorevc, J., Kunej, T., Razpet, A., Dovc, P., 2009. Database of cattle candidate genes and genetic markers for milk production and mastitis. Animal Genetics. 40(6): 832–851.

Prinsen, R.T.M.M., Strillacci, M.G., Schiavini, F., Santus, E., Rossoni, A., Maurer, V., Bieber, A., Gredler, B., Dolezal, M., Bagnato, A., 2016. A genome-wide scan of copy number variants using high-density SNPs in Brown Swiss dairy cattle. Livestock Science. 191: 153–160.

Quinlan, A.R., Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 26(6): 841-2.

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., Cho, E.K., Dallaire, S., Freeman, J.L., González, J.R., Gratacòs, M., Huang, J., Kalaitzopoulos, D., Komura, D., MacDonald, J.R., Marshall, C.R., Mei, R., Montgomery, L., Nishimura, K., Okamura, K., Shen, F., Somerville, M.J., Tchinda, J., Valsesia, A., Woodwark, C., Yang, F., Zhang, J., Zerjal, T., Zhang, J., Armengol, L., Conrad, D.F., Estivill, X., Tyler-Smith, C., Carter, N.P., Aburatani, H., Lee, C., Jones, K.W., Scherer, S.W., Hurles, M.E., 2006. Global variation in copy number in the human genome. Nature. 444(7118): 444-54.

Ribeiro, R., Monteiro, C., Catalán, V., Hu, P., Cunha, V., Rodríguez, A., Gómez-Ambrosi, J., Fraga, A., Príncipe, P., Lobato, C., Lobo, F., Morais, A., Silva, V., Sanches-Magalhães, J., Oliveira, J., Pina, F., Lopes, C., Medeiros, R., Frühbeck, G., 2012. Obesity and prostate

cancer: gene expression signature of human periprostatic adipose tissue. BMC Medicine. 10(1): 108.

Samoré, A.B., Groen, A.F., 2006. Proposal of an udder health genetic index for the Italian Holstein Friesian based on first lactation data. Ital. J. Anim. Sci. 5:359-370.

Samore, A.B., Rizzi, R., Rossoni, A., Bagnato, A., 2010. Genetic parameters for functional longevity, type traits, somatic cell scores, milk flow and production in the Italian Brown Swiss. Ital. J. Anim. Sci. 9:145-152.

Sassi, N.B,. González-Recio, Ó., de Paz-del Río, R., Rodríguez-Ramilo, S.T., Fernández, A.I., 2016. Associated effects of copy number variants on economically important traits in Spanish Holstein dairy cattle. J. Dairy Sci. 99:1–10.

Völker, M., Backström, N., Skinner, B.M., Langley, E.J., Bunzey SK, Ellegren H, Griffin DK., 2010. Copy number variation, chromosome rearrangement, and their association with recombination during avian evolution. Genome Res. 20(4):503-11.

Xu, L., Cole, J.B., Bickhart, D.M., Hou, Y., Song, J., VanRaden, P.M., Sonstegard, T.S., Van Tassell C.P., Liu G.E., 2014. Genome wide CNV analysis reveals additional variants associated with milk production traits in Holsteins. BMC Genomics. 15:683.

Zhang, F., Gu, W., Hurles, M.E., Lupski, J.R., 2009. Copy number variation in human health, disease and evolution. Annu Rev Genomics Hum Genet.10:451-81.

Zhou, Y., Utsunomiya, Y.T., Xu, L., Hay el, H.A., Bickhart, D.M., Alexandre, P.A., Rosen, B.D., Schroeder, S.G., Carvalheiro, R., de Rezende Neves, H.H., Sonstegard, T.S., Van Tassell, C.P., Ferraz, J.B., Fukumasu, H., Garcia, J.F., Liu, G.E., 2016. Genome-wide CNV analysis reveals variants associated with growth traits in Bos indicus. BMC Genomics. 17:419.

Zhu, C., Fan, H., Yuan, Z., Hu, S., Ma, X., Xuan, J., Wang, H., Zhang, L., Wei, C., Zhang, Q., Zhao, F., Du, L., 2016. Genome-wide detection of CNVs in Chinese indigenous sheep with different types of tails using ovine high-density 600K SNP arrays. Sci Rep. 6:27822.

General Discussion

The aim of genetic improvement in livestock production is to modify the genetic frequency of genes related to economically important traits in order to amplify their phenotypic influence and therefore the farmers income. In the last years, selection aimed at optimization of qualitative and quantitative aspects of milk production and functional characteristics (e.g. longevity, fertility, mastitis resistance and improved udder traits).

The first study describes a first CNV detection on a large number of bull samples of the Italian Brown Swiss cattle breed based on whole genome SNP genotyping data, using the two software packages PennCNV and SVS7. The use of two software was helpful to reduce false discovery rates, which are very common in CNV studies. The detection of a specific CNV with two software increased the reliability of the results and also provided more affordable indications on the actual dimensions of the CNV identified. The GO and pathway analysis identified genes that have shown differential expression in association with milk production, carcass, meat, reproduction and growth traits in cattle. The results of this study enrich the bovine CNV map delivering new information for association studies with economically interesting traits.

The second study describes the first CNV genomic analysis on a very large population of Brown Swiss breed individuals based on HD SNP genotyping data. The detected CNVs are a valuable genomic font of

structural variation and enrich the 50K based bovine CNV map of the Brown Swiss breed. The genes mapping within the consensus CNVRs could be considered genes to select in further analysis as possible genomic features to prioritize. The findings of this study are also a valuable source of information for phylogenetic studies and for selection signature analysis. Moreover, the incorporation of CNVRs in routineary genomic evaluations could be considered in the near future to improve genomic selection programs.

The result of the third study is a first high-density genomic association analysis on a large sample of the Brown Swiss cattle breed among CNVs and production, udder and health traits. The association found with the udder traits and the genes mapping in the CNV regions may provide a first molecular evidence that indirect selection for udder traits is related to immune resistance. Nowadays selection for mastitis resistance through indirect selection using udder traits and SCS is widely applied in dairy cattle populations and in economic selection indexes. The convenience in using morphological traits as selection criteria has always been discussed. Their relation with quantitative productive, functional and health traits has been highlighted through the estimation of the genetic correlations, but it was somehow hard to explain the genomic and biological basis of these relations. The general hypothesis for a more functional udder was related to the capacity of a deep udder with a strong fore attachment and cleft is better conformed to the last several lactations and less in contact with manure and dirt in the farm. However, no specific hypothesis has been related with the morphological traits. The associations of CNVs with

FUA and RUW and the genes mapping within them in this study, indicate that the selection for improved udder morphology is driven by a genomic basis that favors the immune resistance and reaction of individuals. In literature, the proposed genomic mechanism is a duplication of regions containing genes responsible for immune resistance. Based on the results of this study the selection for mastitis resistance operated in the Brown Swiss population seems to reflect the increase in immunological resistance and reaction competence at a genomic level. The genes mapping in the identified CNVs could be a first indication to further investigate their role in metabolic pathways related to mastitis resistance.

Extra – curricular activities

Conferences, congresses and workshops

• Conference: ASPA 2015: presented poster: A high-resolution CNVR map in Brown Swiss dairy cattle according to PIC and LD with SNPs 09-06-2015 / 12-06-2015, Milan, Italy

• Workshop: Ruminomics Regional Workshop 05-09-2015 / 06-09-2015, Lodi, Italy

• Workshop: Principi Genetici della Produzione di Latte 29-09-2015, Milan, Italy

• Workshop: Elsevier Publishing Campus Come pubblicare con successo 13-04-2015, Milan, Italy

• Workshop: How to Prepare and Present Scientific Papers Workshop 28-08-2016 Belfast, UK

• EAAP 2016 Annual Meeting: presented poster: A genome-wide association study using CNVs for production traits in Brown Swiss dairy cattle, 29-08-2016 / 02-09-2016 Belfast, UK – Winner of a scholarship to attend the meeting

• Veterinary and Animal Science Days 15-07-2015 / 17-07-2015 – UNIMI. Presentation and abstract: A high-resolution CNV map across Brown Swiss cattle populations (Published in: International Journal of Health, Animal Science and Food Safety).

• Veterinary and Animal Science Days 08-06-2016 / 10-06-2016 – UNIMI. Presentation and abstract: Genome-wide association studies using copy number variants in Brown Swiss Dairy cattle (Published in: International Journal of Health, Animal Science and Food Safety).

Publications

• Emanuela Tullo, Erika Frigo, Attilio Rossoni, Raffaella Finocchiaro, Marco Serra, Nicoletta Rizzi, Antonia Bianca Samorè, Fabiola Canavesi, Maria Giuseppina Strillacci, **Raphaelle Prinsen**, Alessandro Bagnato. Genetic parameters of fatty acids in Italian Brown Swiss and Holstein cows, in Animal Science 2013.

• Alessandro Bagnato, Laura Pellegrino, Maria G. Strillacci, Fausta Schiavini, Erika Frigo, Dinesh Velayutham**, Raphaelle T. M. M. Prinsen**, Marlies A. Dolezal. Identification and validation of copy number variants in Italian Brown Swiss dairy cattle using Illumina Bovine SNP50 Beadchip, 2015. Italian Journal of Animal Science.

• Erika Frigo, Antonia B Samorè, Liliana Reghenzani, Nicola Bergomi, Maria G Strillacci, Fausta Schiavini, **Raphaelle T.M.M. Prinsen**, Maria C Cozzi, Marco Serra, Attilio Rossoni and Alessandro Bagnato. Variation of milk components in the Italian Brown cattle Journal of Dairy Research, 2015.

• **R.T.M.M. Prinsen**, M.G. Strillacci, F. Schiavini, E. Santus, A. Rossoni, V. Maurer, A. Bieber, B. Gredler, M. Dolezal, A. Bagnato. A genome-wide scan of copy number variants using high-density SNPs in Brown Swiss dairy cattle, 2016. Livestock Science.

• Cozzi Maria Cristina, Valiati Paolo, Cherchi Raffaele, Gorla Erica, **Prinsen Raphaelle**, Longeri Maria, Bagnato Alessandro, Strillacci Maria Giuseppina. Mitochondrial DNA genetic diversity in six Italian donkey breeds (Equus asinus), 2016. Mitochondrial DNA Part A.

• **R.T.M.M. Prinsen**, A. Rossoni, B. Gredler, A. Bieber, A. Bagnato, M.G. Strillacci. A genome wide association study between CNVs and quantitative traits in Brown Swiss cattle. Livestock Science.

Extra- curricular courses

• Population Genetics Introductory course – Vienna – VETMEDUNI

• Introduction to Bioinformatics – Perugia - Polo d'Innovazione di Genomica, Genetica e Biologia S.C. a R.L (2015)

• Functional Genomics - Milano –UNIMI

• Bioinformatics and Comparative Genomics Milano –UNIMI

• Selezione Genomica Milano –UNIMI

• Systems Biology Milano –UNIMI

• Zootecnia e miglioramento Genetico Milano –UNIMI

Acknowledgements

I would like to thank the members of my research group for their suggestions, feedbacks and discussions over the years.

I would also like to thank my amazing family for giving me the strength to complete this journey.