

**SCHOOL OF AGRICULTURAL AND VETERINARIAN SCIENCES
SÃO PAULO STATE UNIVERSITY
CAMPUS JABOTICABAL**

**DETECTION AND FUNCTIONAL ASSESSMENT OF STRUCTURAL
VARIANTS USING WHOLE GENOME RE-SEQUENCING DATA IN
NELLORE CATTLE**

Natalia Andrea Marin Garzon

Animal Scientist

M.Sc. Genetics and Animal Breeding

Jaboticabal-SP

2023

**SCHOOL OF AGRICULTURAL AND VETERINARIAN SCIENCES
SÃO PAULO STATE UNIVERSITY
CAMPUS JABOTICABAL**

**DETECTION AND FUNCTIONAL ASSESSMENT OF STRUCTURAL
VARIANTS USING WHOLE GENOME RE-SEQUENCING DATA IN
NELLORE CATTLE**

Student: Natalia Andrea Marín Garzón

Advisor: Prof^ª. Dra. Lucia Galvão de Albuquerque

Co-advisor: Dra. Giovana Vargas

**Thesis presented to the
School of Agricultural and
Veterinarian Sciences–
SãoPaulo State University,
campus of Jaboticabal in
partial fulfillment of
requirements for the degree of
Doctor in Genetics and Animal
Breeding.**

Jaboticabal-SP

2023

M337d Marin-Garzon, Natalia Andrea
Detection and functional assessment of structural variants
using whole genome re-sequencing data in Nellore cattle /
Natalia Andrea Marin-Garzon. -- Jaboticabal, 2024
110 p. : il., tabs.

Tese (doutorado) - Universidade Estadual Paulista (Unesp),
Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal
Orientadora: Lucia Galvão de Albuquerque
Coorientadora: Giovana Vargas

1. Animal breeding. 2. CNV. 3. DNA. 4. Genome sequencing.
5. Zebu cattle. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da
Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal. Dados fornecidos pelo
autor(a).

Essa ficha não pode ser modificada.

Impacto potencial desta pesquisa

O presente estudo permitiu identificar variantes genéticas no genoma de 151 touros representativos da raça Nelore, conhecidas como variações estruturais. Essas variações são compostas por diferentes tipos, dos quais conseguiram-se identificar deleções, duplicações, e inversões de fragmentos de DNA. Também foram identificadas as regiões do genoma onde há maior ocorrência dessas variações na população avaliada, bem com os genes localizados dentre dessas regiões, que estão envolvidos em funções biológicas e moleculares, que são importantes para processos adaptativos, e que também estão associados a características de importância econômica em gado de corte, como a eficiência alimentar. Esses resultados contribuem a desvendar os mecanismos genéticos que promoveram a adaptação ambiental do gado Nelore, causado potencialmente por mudanças estruturais e rearranjos de segmentos de DNA funcionalmente importantes.

Potential impact of this research

The present study allowed the identification of genetic variants in the genome of 151 representative Nelore bulls, known as structural variations. These variations correspond to different types, from which deletions, duplications, and inversions of DNA fragments were identified. Genomic regions with the highest occurrence of these variations in the evaluated population were also identified, as well as the genes located within these regions, which are involved in biological and molecular functions that are important for adaptive processes, and which are also associated with economically important traits in cattle, such as feed efficiency. These results contribute to unveiling the genetic mechanisms driving the environmental adaptation of Nelore cattle, potentially caused by structural changes and rearrangements of functionally important DNA segments.

CERTIFICADO DE APROVAÇÃO

TÍTULO DA TESE: DETECTION AND FUNCTIONAL ASSESSMENT OF STRUCTURAL VARIANTS USING WHOLE GENOME RE-SEQUENCING DATA IN NELLORE CATTLE

AUTORA: NATALIA ANDREA MARIN GARZON

ORIENTADORA: LUCIA GALVÃO DE ALBUQUERQUE

COORIENTADORA: GIOVANA VARGAS

Aprovada como parte das exigências para obtenção do Título de Doutora em Genética e Melhoramento Animal, pela Comissão Examinadora:

Profa. Dra. LUCIA GALVÃO DE ALBUQUERQUE (Participação Virtual)
Departamento de Zootecnia / FCAV UNESP Jaboticabal

Documento assinado digitalmente
 LUCIA GALVAO DE ALBUQUERQUE
Data: 14/11/2023 10:34:52-0300
Verifique em <https://validar.iti.gov.br>

Dr. DIERCLES FRANCISCO CARDOSO (Participação Virtual)
Gensys Tecnologia e Sistemas LTDA / Jaboticabal/SP

Documento assinado digitalmente
 DIERCLES FRANCISCO CARDOSO
Data: 23/11/2023 12:13:15-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. RICARDO VIEIRA VENTURA (Participação Virtual)
Departamento de Nutrição e Produção Animal / FMVZ USP Pirassununga/SP

Documento assinado digitalmente
 RICARDO VIEIRA VENTURA
Data: 07/12/2023 13:22:38-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. LUIS GABRIEL GONZALEZ HERRERA (Participação Virtual)
Departamento de Produção Animal / Universidad Nacional de Colombia - Medellín/Colômbia



Prof. Dr. RAFAEL ESPIGOLAN (Participação Virtual)
Departamento de Zootecnia e Ciências Biológicas / Universidade Federal de Santa Maria - Santa Maria/RS

Documento assinado digitalmente
 RAFAEL ESPIGOLAN
Data: 14/11/2023 11:07:19-0300
Verifique em <https://validar.iti.gov.br>

Jaboticabal, 31 de outubro de 2023

ABOUT THE AUTHOR

NATALIA ANDREA MARÍN GARZÓN- born in April 12th, 1994 in Bello-Antioquia, Colombia. Daughter of Hilda Natalia Garzón and Carlos Marín. Bachelor of Animal Sciences from the National University of Colombia, campus of Medellin-2017. Her studies were granted with the scholarship “Fondo Para La Educación Superior” from Government of Antioquia (2013-2016). During the last year of undergraduate course, she undertook an internship in Genetics and Animal Breeding at São Paulo State University under the supervision of Prof. Dr. Lucia Galvão de Albuquerque and Prof. Dr. Luis Gabriel González Herrera, granted by the Department of Foreign Affairs of National University of Colombia. She started her M.Sc. studies in 2017 in Genetics and Animal Breeding at FCAV-UNESP with a grant from “Coordenação de Aperfeiçoamento de Pessoal de Nível Superior”, under the supervision of Prof. Dr. Lucia Galvão de Albuquerque and Prof. Dr. Ana Fabricia Braga Magalhães, obtaining her title in 2019. During the same year she started her Ph.D. formation in the current graduate program of Animal Sciences at the same university under the supervision of Prof. Dr. Lucia Galvão de Albuquerque and Dr. Giovana Vargas. Her Ph.D. studies were granted by “Coordenação de Aperfeiçoamento de Pessoal de Nível Superior” from 2019 to 2022. In January 2022 she started a position as Research and Development Coordinator at Colombian Association of Zebu Cattle Breeders (ASOCEBU).

I entirely dedicate this work to my mother Hilda, for being my safe harbor every second of my existence, for letting my wings grow and supporting me in the search and reach of my dreams. For being the best example of hope and strength even during the heaviest battles. To her, who with empowerment inspired me to finish my Ph.D.

ACKNOWLEDGMENTS

Firstly, I thank the wise universe that drew the paths I had to take to reach my goals. For making me face difficult situations that surely transform me into a more ethical professional.

I am so grateful to Prof. Dr Lucia for having accepted me in her research group since the beginning of my undergraduate internship. For her teachings and support that were essential for the development of my Ph.D.

I thank ...

All my family, but especially to my parents and my brother for their unconditional support. For motivating me every day to continue reaching my dreams, and for demonstrating every single day of my life, the most pure and sincere feeling of love I have ever received.

My life partner and love, Oscar, who has undoubtedly been my greatest support during this journey. For the sincere love and the loyal friendship, for the advises and for motivate me to be a better person and professional.

My co advisor Dr. Giovana Vargas, who even in the distance was always available to help me through all the difficulties, especially for being a friend during the most challenging moments along this journey.

Professor Luis Gabriel González, who motivated me to embark on the area, who continues teaching me every day and who has played an essential role in the development of my professional career.

Larissa Braga and our virtual meetings, that were essential to guide me in the execution of this study. For being always willingness to help me and discuss results.

To my “Albuquerque” brothers and sisters, especially Lucio, Patricia, Marta, Larissa, Gabby, Bruna, and everyone who helped and taught me from the first day I came to the research group. I always keep in my memories the mornings of extremely dark coffee and good conversations, the “churrasco” and pizza days. I am especially grateful to Lucio, for all the help during the realization of this study.

To my dear friend Valentina, who was my greatest support during the dark days that came with the pandemics of Covid-19, and the lovely friends Brazil gave me: Marisa, Karen, Maria Camila, and Jaque, for the support and loyal friendship.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001"

CONTENT

1. INTRODUCTION	1
2. LITERATURE REVIEW	3
2.1. Structural variants (SVs).	3
2.2. Mechanisms for SVs formation	5
2.3. Bioinformatic algorithms for SV calling.....	6
2.4. Functional relevance of SVs.....	8
3. MATERIAL AND METHODS	9
3.1 DNA samples and resequencing.....	9
3.2. Detection of Structural Variants.....	10
3.3. Filtering of called structural variants.....	11
3.4. Identification of structural variant regions (SVRs).	14
3.5. Statistical analysis and graphical visualization of SVs and SVRs	15
3.6 Functional analysis of highly frequent SVR.....	15
4. RESULTS AND DISCUSSION	16
4.1 Genome wide SV detection: metrics regarding number, length, and chromosomal location	16
4.1.2. Length classification of high-confidence SVs.....	20
4.1.3. Chromosomal location and distribution of high-confidence SVs.	23
4.1.4. Chromosomal extension covered by high-confidence SVs.....	25
4.2. Discovery of potential regions of structural variants (SVR)	29
4.1.2 Average length and autosomal genome extension covered by SVRs ...	34
4.3. Functional annotation and enrichment of SVR.....	41
5. FINAL CONSIDERATIONS.....	51
6. REFERENCES	52
7. SUPPLEMENTARY MATERIAL.....	61
7.1 Supplementary tables.	61
7.2 Supplementary Figures.....	100

List of Tables

Table 1. Statistics summary for the number and length of detected high confidence SV's.....	17
Table 2. Statistics summary for the number and proportion of each SV type according to different size classes.....	21
Table 3. Statistics summary for the number of SVRs, length and proportion of chromosome extension covered by SVR in each of 29 autosomes.....	33
Table 4. Statistics summary for the number, length, and genome coverage rate for each SVR type (present in at least 5% of bulls).....	35
Table 5. Number and proportion of each type of detected SVR according to different size classes.....	36
Table S1. Statistics summary for the number of SV's, length and proportion of autosomal genome extension covered by SV for each the analyzed bulls.....	61
Table S2. Statistics summary for number and length of total structural variants, deletions, duplications, and inversions, across the 29 autosomal chromosomes.....	62
Table S3. Statistics summary for the number. length and proportion of chromosome extension covered by each SVR type along the 29 autosomes.....	66
Table S4. Significant QTLs overlapping the most frequent regions of structural variants.....	70
Table S5. Positional candidate genes overlapping the most frequent SVR.....	76
Table S6. Enrichment analysis of genes overlapping the most frequent SVR: significant Panther Gene Ontology terms (GO) associated to biological process.....	87
Table S7. Enrichment analysis of genes overlapping the most frequent SVR: significant Panther Gene Ontology terms (GO) associated to molecular function.....	96
Table S8. Enrichment analysis of genes overlapping the most frequent SVR: significant Panther Gene Ontology terms (GO) associated to cellular component.....	98
Table S9. Enrichment analysis of genes overlapping the most frequent SVR: significant pathways.....	99

List of figures.

Figure 1. Graphical representation of balanced, unbalanced, and complex structural variation types.....	4
Figure 2. Flowchart of pipeline used to detect structural variants (SV).....	13
Figure 3. Flowchart of pipeline used to detect regions of structural variant (SVR)....	14
Figure 4. Distribution of the detected structural variants according to different size classes.....	21
Figure 5. Number of high-confidence structural variants harbored by each of the 29 autosomal chromosomes.....	24
Figure 6. Average length (Kb) of all predicted high-confidence structural variants across the 29 autosomes.....	26
Figure 7. Average proportion (%) of chromosome extension covered by high-confidence SV (average coverage rate).....	26
Figure 8. Pearson correlation between the number of predicted high-confidence SVs and the chromosome size. N: number of SVs. *** p value < 0.05.....	27
Figure 9. Genome coverage rate: proportion (%) of whole genome extension covered by each type of detected high-confidence SVs in each of the analyzed Nellore bulls.....	28
Figure 10. Pearson correlation between the number of predicted high-confidence SVs and the chromosomal coverage rate for each bull. N: number of SVs. *** p value < 0.05.....	29
Figure 11. Proportion of each type of structural variant regions (SVR).....	30
Figure 12. Distribution of non-redundant SVR frequent in more than 5% of population across the 29 autosomes.....	34
Figure 13. Number of regions of structural variants (SVR) according to different size classes.....	37
Figure 14. Pearson correlation between the number of non-redundant SVRs harbored by each chromosome and their average length.....	38
Figure 15. Chromosomal coverage rate: proportion (%) of chromosome extension covered by regions of structural variants (SVRs).....	39

Figure 16. Pearson correlation between the number of non-redundant SVRs and the chromosome size. N: number of SVRs *** p value < 0.05.....	40
Figure 17. Distribution of the number of the regions of structural variants (SVR) of shared by different proportions of Bulls.....	41
Figure 18. Chromosomal location of the most frequent regions of structural variants (SVR). Blue dots represent regions of deletions, green dots are mixed regions, red dots are the regions of duplications, and black dots the regions of inversions.....	42
Figure 19. Distribution of SVRs overlapping QTLs according to QTL type and trait associated.....	43
Figure 20. Results from PANTHER enrichment analysis. Biological Process and Molecular Function related GO terms.....	45
Figure S1. Distribution of each type of regions of copy number variants (CNVR) across the 29 autosomal chromosomes.....	100
Figure S2. Distribution of regions of inversions across the 29 autosomal chromosomes.....	100
Figure S3a. Average length (Kb) of regions of copy number variants (CNVR) along the 29 autosomal chromosomes.....	101
Figure S3b. Chromosomal coverage rate (%) of regions of copy number variants (CNVR) along the 29 autosomal chromosomes.....	101
Figure S4a. Average length (Kb) of regions of inversions (INVR) along the 29 autosomal chromosomes.....	102
Figure S4b. Chromosomal coverage rate of regions of inversions (INVR) along the 29 autosomal chromosomes.....	102

Detection and functional assessment of structural variants using whole genome re-sequencing data in Nellore cattle.

Abstract- Ongoing advances in genome sequencing technologies have enabled the unravelling of many structural variants (SVs) in livestock genomes. Association of SVs with complex traits are promising targets for animal breeding because of their effects on gene expression. The aims of this study were: *i)* to detect structural variants using whole genome resequencing data of 151 representative Nellore bulls by combining calling algorithms *ii)* to discover non redundant and highly frequent regions of structural variants (SVR) in the analyzed bulls. *iii)* to search for positional candidate genes, and quantitative trait loci (QTL) overlapping the most frequent SVR in the population. *iv)* to assess the functional impact of positional candidate genes overlapping SVR through enriched gene ontology terms (GO terms) related to biological process (BP), molecular function (MF), cellular component (CC), and biochemical pathways. The whole genome re-sequencing (WGS) data from 151 representative Nellore bulls was used to conduct genome-wide structural variation calling, as well as detection of common SVRs for the analyzed bulls. The gene content and nearby QTLs of SVRs were retrieved from publicly available genomic databases, and functional enrichment analysis of positional candidate genes overlapping the most frequent SVRs were conducted using PANTHER. A total of 215,031 high-confidence SV was obtained, most of them corresponding to copy number variants (CNV) (183,032 deletions-DEL, and 14,013 duplications-DUP), and 17,986 inversions (INV). Total structural variation encompasses, on average, 4.81% of the individual autosomal genome extension. A total of 3,752 non-redundant SVR frequent in more than 5% of bulls were obtained, corresponding in more than 97% to regions of copy number variants (CNVR), and 3% to regions of inversions (INVR). All SVR comprises 13.13% of total autosomal genome extension, which were attributed in 11.4% to CNVR and 1.7% to INVR. Among all SVRs, 532 were shared by more than 50% of bulls and overlapped a total of 130 QTL distributed into 6 QTL types: exterior, health, meat and carcass, milk, production, and reproduction, which are related to a total of 50 economically important traits. Most SVR overlapped QTLs related to residual feed intake, structural soundness, multiple birth, clinical mastitis, and milk energy yield. Regarding gene content, 204 SVRs, overlapped a total of 1,164 positional candidate genes, which were significantly overrepresented into GO terms related to BP, MF, CC and one biochemical pathway. Among the significantly enriched genes, we highlight members of the olfactory receptor (OR) gene family, which play essential roles in mechanisms for adaptation to the environment. These genes were mainly found into regions of inversion and mixed events. Similarly, genes from the defensin family (*DEFB*), that play important roles in the innate immune system of multicellular organisms, and which are known to be caused by duplication events that mammalian genomes have undergone. Other important genes were found in this study, such as the members of the secretory phospholipase A2 family, adhesion G protein-coupled receptors, and zinc finger binding proteins. Most of the genes found in this study have been described as potential candidates for feed efficiency indicator traits, which reflects the biochemical mechanisms in which they are involved that have led to the improved fitness. The results of this study provide important knowledge about the mechanisms driving

changes in the genome in Nellore cattle, the contributed to adaptation to environment.

Keywords: DNA; copy number variation; mobile elements; next generation sequencing; sequence gains and losses; zebu cattle

Deteção e avaliação funcional de variantes estruturais utilizando dados de re-sequenciamento do genoma completo em bovinos Nelore.

Resumo-Os avanços contínuos nas tecnologias de sequenciamento do genoma permitiram o desvendamento de diferentes variantes estruturais (SVs) nos genomas dos animais de interesse zootécnico. A associação de SVs com características complexas são alvos promissores para o melhoramento animal devido aos seus efeitos na expressão gênica. Os objetivos deste estudo foram: i). detectar variantes estruturais usando dados de re-sequenciamento do genoma completo de 151 touros Nelore representativos combinando diferentes algoritmos; ii) descobrir regiões não redundantes e altamente frequentes de variantes estruturais (SVR) nos touros analisados. iii). procurar genes candidatos posicionais e locos de características quantitativas (“QTL”) em sobreposição às SVRs mais frequentes na população. iv) avaliar o impacto funcional de genes candidatos posicionais abrigados pelas SVRs através de enriquecimento funcional de termos de ontologia genética (GO terms) relacionados a processo biológico (BP), função molecular (MF), componente celular (CC) e vias bioquímicas. Dados de re-sequenciamento do genoma completo de 151 touros Nelore representativos foram usados para conduzir a detecção de variantes estruturais, bem como para identificar SVRs comuns aos touros analisados. O conteúdo genético e os QTLs próximos de SVRs foram recuperados de bancos de dados genômicos disponíveis publicamente, e a análise de enriquecimento funcional de genes candidatos posicionais que se sobrepõem aos SVRs mais frequentes foi conduzida usando PANTHER. Foram obtidas 215.031 SVs de alta confiança, a maioria correspondendo a variantes de número de cópias (CNV) (183.032 deleções-DEL, e 14.013 duplicações-DUP, e 17.986 inversões (INV). A variação estrutural total abrange, em média, 4,81% da extensão do genoma autossômico individual. Foram obtidos um total de 3.752 SVRs frequentes em mais de 5% dos touros, correspondendo em mais de 97% a regiões de variantes de número de cópias (CNVR) e 3% a regiões de inversões (INVR). Todos os SVRs compreendem 13,13% da extensão total do genoma autossômico, que foram atribuídos em 11,4% a CNVRs e 1,7% a INVRs. Entre todos as SVRs, 532 foram compartilhados por mais de 50% dos touros e sobrepuseram um total de 130 QTL distribuídos em 6 tipos: exterior, saúde, carne e carcaça, leite, produção e reprodução, que estão relacionados a um total de 50 características de importância econômica. A maioria dos QTLs sobrepostos às SVRs estão relacionados a consumo residual, solidez estrutural, nascimentos múltiplos, mastite clínica, e produção de energia do leite. Em relação ao conteúdo gênico, 204 SVRs das mais frequentes se sobrepuseram a um total de 1.164 genes candidatos posicionais, que foram significativamente super-representados em termos GO relacionados a BP, MF, CC e uma via bioquímica. Dentre os genes significativamente enriquecidos, destacamos os membros da família de genes dos receptores olfativos (OR), com funções importantes nos mecanismos de adaptação ao ambiente. Esses genes foram encontrados principalmente dentre de regiões com ocorrência de eventos de inversões e eventos mistos, que podem explicar a prevalência nos genomas ao longo da evolução dos mamíferos. De forma similar, os genes da família das defensinas (DEFB) que desempenham papéis importantes no sistema imunológico inato de organismos multicelulares, e que são conhecidos por serem causados por

eventos de duplicação que sofreram os genomas dos mamíferos. Destacam-se também membros da família da fosfolipase A2 secretora, os receptores acoplados à proteína G de adesão e as proteínas de ligação aos dedos de zinco. A maioria dos genes encontrados neste estudo são candidatos potenciais para características indicadoras de eficiência alimentar, o que reflete os mecanismos bioquímicos nos quais estão envolvidos e que levaram à melhoria do valor adaptativo. Os resultados deste estudo fornecem conhecimento importante sobre os mecanismos que impulsionam as mudanças no genoma de bovinos Nelore, as causas dos processos de adaptação e a caracterização das consequências da variação estrutural em relação à diversidade genética.

Palavras-chave: DNA; elementos móveis; gado zebu, ganhos e perdas de sequência, sequenciamento de próxima geração; variação do número de cópias

1. INTRODUCTION

The widespread developments of molecular techniques and bioinformatic algorithms have enabled the unravelling of genetic mechanisms driving differences between individuals, which are mainly attributed to single-nucleotide variants (SNVs), small insertions and deletions (indels;<50 bp), and structural variations (SVs) (Bickhart, 2014; Fernandes Junior et al., 2020; Ho, Urban & Mill., 2020).

Structural variants are generally defined as inherited variations in DNA sequences larger than 50 Bp that exhibit changes in the copy number, orientation, and/or chromosomal location, in relation to the reference genome (Alkan et al., 2011, Escamarís, Docampo & Rabionet., 2015). The balanced SVs types correspond to the mobile elements (inversions and translocations), which cause rearrangements of DNA sequences (Escamarís, Docampo & Rabionet., 2015; Bickhart., 2014). By other side, small insertions, and deletions (indels), and the copy number variants (CNVs), comprise the unbalanced SVs in which gains or losses of DNA segments are observed. CNVs are the most studied SVs types in livestock, and encompass duplication and deletion events ranging in sizes from 1 Kb to 5 Mb (Ho, Urban & Mills., 2020; Zhang et al., 2009; Freeman et al., 2006).

Compared to SNPs, SVs affect larger fractions of the genome and account for the greatest amount of total polymorphic content among individual genomes (Couldrey, Keehan, Johnson, Tiplady, Winkelman, Littlejohn, & Scott, 2017). SVs may have potentially greater effects than SNPs: from modification of genes to disruptions in regulation of gene expression, leading to recessive disorders (Conrad et al., 2010; Zhang et al., 2009).

A variety of based platforms such as comparative genomic hybridization and SNP arrays have been commonly used for SVs wide detection in livestock (Bickhart et al., 2012). However, array-based approaches lack genome coverage, especially in segmental duplication regions, which are known for being hotspots for CNVs formation (Zhao et al., 2013). The advantages of detecting SV using next generation sequencing data (NGS) include, higher coverage and resolution, more accurate prediction of copy numbers, more precise detection of breakpoints, and higher power

to identify novel SVs (Zhao et al., 2013). In cattle, the study of structural variations has reached an increasingly attention because of their association with economically important phenotypes such as fatty acid profile (Lemos et al., 2018a), meat tenderness (Silva et al., 2016), feed conversion ratio (Santana et al., 2016); growth traits (Liu et al., 2020; Zhou et al., 2016), milk production traits (Zhou et al., 2016), and milk somatic cell score (Durán Aguilar et al., 2017).

Although NGS has become more feasible in livestock, most of studies regarding structural variant wide detection are still being conducted from SNP arrays-based approaches. There are few published studies that explored the structural diversity using whole genome sequencing data all with less than 100 animals (Peripolli et al., 2023; Sun et al., 2023; Braga et al., 2022; Butty et al., 2020; Silva et al., 2016, Santana et al., 2016; Bickhart et al., 2016; Boussaha et al., 2015). Furthermore, most of them have been focused on the assessment of copy number variants (CNVs) rather than the other SVs types.

The detection of balanced and unbalanced structural variants using next generation sequencing data may contribute to the unravelling of the molecular mechanisms underlying genetic diversity, breed adaptation and evolution, especially in Nellore cattle in which they have not been well elucidated so far. Therefore, the aims were: *i*). to detect structural variants using whole genome resequencing data of 151 representative Nellore bulls by combining calling algorithms *ii*) to discover non redundant and highly frequent regions of structural variants (SVR) in the analyzed bulls. *iii*). to search for positional candidate genes, and quantitative trait loci (QTL) overlapping the most frequent SVR in the population. *iv*) to assess the functional impact of positional candidate genes overlapping SVR through enriched gene ontology terms (GO terms) related to biological process (BP), molecular function (MF), cellular component (CC), and biochemical pathways.

2. LITERATURE REVIEW

Genomes are structurally dynamics; they vary from each other in multifactorial ways and the totality of this variation is responsible for the heritability of economically important traits (Conrad et al., 2010). The understanding of the genetic basis of phenotypic diversity within populations is currently one of the main concerns in livestock genetics. Thus, deciphering the nature of genetic variants that affect gene expression may provide a valuable framework for elucidating the causes of phenotypic variation, which contributes to differences in individuals' performance (Stranger et al., 2014).

2.1. Structural variants (SVs).

Structural variants are generally defined as variations in DNA sequences larger than 50 Bp that exhibit changes in the copy number, orientation, and/or chromosomal location between individuals in relation to the reference genome (Alkan et al., 2011, Escamarís, Docampo & Rabionet., 2015). As indicated by the structural consequence, SVs comprise several classes of variants which can be further classified into two types: balanced and unbalanced (Figure 1). (Bickhart & Liu 2014). The balanced types encompass the mobile elements, in which instead of gains or losses of genetic material, there is a DNA rearrangement caused by inversions or translocations. The inversions imply that a DNA fragment is reversed in orientation from the rest of the chromosome, and translocations generate large transfers of DNA segments from one location to another (in the same chromosome or in different chromosomes) (Escamarís, Docampo & Rabionet., 2015 Bickhart & Liu 2014).

By their side, the unbalanced types comprise the copy number variants (CNVs), in which DNA segments can be lost or duplicated (Ho, Urban & Mills., 2020; Zhang et al., 2009; Freeman et al., 2006). CNVs are the most common types of structural variants in the genomes of humans and domestic animals, and they are referred as an intermediate-scale SVs which exhibit sizes ranging from 1 Kb to 5 Mb (Conrad et al., 2010). CNVs necessarily imply changes in the number of base pairs

of DNA sequences, which can be caused by insertion, duplication, and/or deletion events (Ho, Urban & Mills., 2020; Alkan et al., 2011). Insertions refer to the introduction of novel sequences, deletions generate losses of DNA segments, and duplications produce contiguous (tandem) or distant (discontinuous) insertions of existing sequences, in relation to the reference genome (Figure S1).

In addition, there are other complex SVs types which are combinations of the mentioned balanced and unbalanced types, including mobile element insertions, multi-allelic CNVs of highly variable copy number, segmental duplications (SD) also called as low repeat regions, and complex rearrangements (Ho, Urban & Mills., 2020).

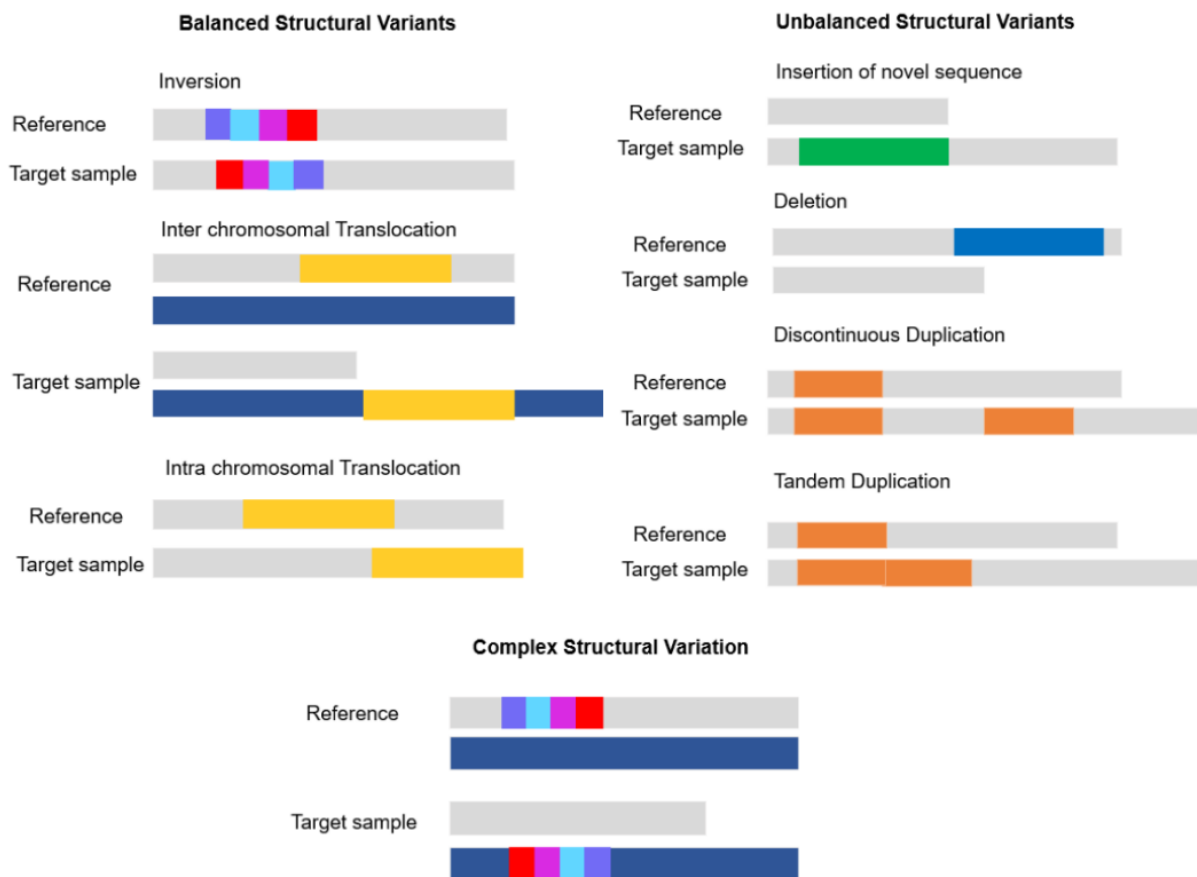


Figure 1. Graphical representation of balanced, unbalanced, and complex structural variation types. Adapted from Escaramís et al., 2015.

2.2. Mechanisms for SVs formation

From molecular point of view, there are different mechanisms by which SVs can occur both during meiosis and mitosis (Escamarís, Docampo & Rabionet., 2015). The main mechanisms causing structural variations include recombination errors such as non-allelic homologous recombination (NAHR), errors during DNA break repair, as in non-homologous end joining (NHEJ), and errors during DNA replication process, such as fork stalling and template switching (FoSTeS).

NAHR occurs during meiosis and mitosis due to recombination events between two DNA segments with similar sequence at non-homologous chromosomes (Bhanuprakash et al., 2018). Segmental duplications (SD) are well known hotspots for NAHR occurrence. They refer to sequences that occur twice or a few times in haploid genomes (Hastings et al., 2009). SD arises from chromosomal rearrangements of duplicated copies, product of tandem repetitions of a DNA segment, or from a duplicative transposition-like process, in which DNA segments are copied and transposed from one chromosome region to another (Freeman et al., 2006; Hastings et al., 2009). NAHR events are more likely to occur when homologous sequences are 10 Kb or larger, have more than 95% homology, and are separated by 50 Kb to 10 Mb. (Escamarís, Docampo & Rabionet., 2015).

The resulting SV from NAHR event depends on the location and orientation of the highly homologous regions: when located in direct orientation within the same chromosome it leads to deletions or duplications; when located on opposite orientation at the same chromosome it causes inversions; and when located in different chromosomes it drives to translocations (Escamarís, Docampo & Rabionet., 2015).

By other side, the non-homologous end-joining (NHEJ) is a DNA repair mechanism that is frequently initiated in response to double-strand breaks (DSBs) in DNA sequence (Bickhart & Liu 2014). During NHEJ events, DSBs are identified, repaired, and joined, often to different regions of the genome than they came from. (Escamarís, Docampo & Rabionet., 2015). In contrast to NAHR, NHEJ can occur in any genomic region, and the consequence is generally the appearance of short

insertions and deletions (Bhanuprakash et al., 2018; Escamarís, Docampo & Rabionet., 2015).

The other mechanism driving SVs formation known as FOSTES, is caused by the interruption of the DNA replication machinery. This leads to the separation of DNA lagging strand from polymerase enzyme and its join to another region of the genome before the restarting of replication process. The DNA fragment in which the polymerase activity was interrupted is duplicated if the lagging strand hybridizes with a DNA segment downstream the problematic region (Bickhart & Liu 2014).

2.3. Bioinformatic algorithms for SV calling.

Small insertions and deletions, and CNVs are the most analyzed SVs types in livestock populations, as they are relatively easier to detect, mainly from array-based approaches. A variety of based platforms such as comparative genomic hybridization and SNP arrays have been used in most studies detecting copy number variants in livestock. Although analyses of a subset of CNVs have provided evidence of linkage disequilibrium with target SNPs (Bickhart et al., 2012), these approaches are quite limited as they lack genome coverage, especially in segmental duplication regions, due to low probe density and resolution and high hybridization noise (Zhao et al., 2013).

Array comparative genomic hybridization (array CGH) relies on the normalization of the signal ratio between a target and reference sample, and conversion into a log₂ ratio (Alkan et al., 2011). Increased log₂ ratio indicates gains in copy number in the target sample, while decrease log₂ ratio indicates copy number loss (Escamarís, Docampo & Rabionet., 2015). Similarly, SNP-based approaches, compare the signal intensities for the target sample to a set of reference hybridizations or to the analyzed population. An increase in the signal intensity is associated to gains in the base pair number, and oppositely, a reduction suggests losses in the base pair number. For that, two metrics are considered: Log R ratio and B allele frequency (BAF). (Cassese et al., 2014; Wang et al., 2007)

With the ongoing developments of next generation sequencing technologies (NGS), followed by a continuous reduction of costs, the livestock genomics

community has been challenged to resequencing the genomes of a growing number of animals, which has resulted in initiatives such as the 1,000 Bull Genomes Project (<http://www.1000bullgenomes.com/>). This has led to an increasing popularity of NGS-based CNV detection. The advantages of detecting CNV using next generation sequencing data (NGS) include higher coverage and resolution, more accurate estimation of copy numbers, more precise detection of breakpoints, and higher power to identify novel CNVs (Zhao et al., 2013).

Characterization of structural variants from NGS data is not an easy process due to the complexity of the genomic data provided. During the past few years, several bioinformatic tools were developed based on different algorithms aiming to obtain more accurate SV calls (Liu & Bickhart., 2012; Alkan et al., 2011). The most used algorithms to identify SVs from NGS data are read depth (RD), read pair (RP) and split read (SR).

Read depth (RD) algorithms analyze the density of reads, mapped into a given interval of the reference sequence. In regions with normal copy numbers, the read depth should be consistent with the average coverage across the genome, and a deviation from the expected average read depth suggest copy number variations Escamarís, Docampo & Rabionet., 2015; Liu & Bickhart., 2012). This approach has only power in detecting CNVs (deletion and duplication events). CNVnator (Abyzov et al. 2011) is a commonly used tools for CNV wide detection in livestock species, which is based on RD algorithm.

By other side, structural variants can be identified through read-pair (RP) methods by assessing abnormal distances or orientations between the paired-end reads, compared to the expected normal genomic structure from the reference genome (Alkan et al., 2011). The information obtained from the paired-end mapping allows the detection of structural variants such as insertions, deletions, duplications, inversions, and translocations (Escamarís, Docampo & Rabionet., 2015).

Finally, split read (SR) methods have particularly great power for identifying complex SVs involving small insertions, deletions, or rearrangements in the DNA sequence. The algorithm works by the direct identification of sequence reads that span the breakpoints of SVs (Escamarís, Docampo & Rabionet, 2015).

2.4. Functional relevance of SVs.

The structural variations have been described as the major force driving the complex genomes evolution. Functionally relevant gene families have been expanded due to duplication mechanisms (Niimura & Nei, 2007). This suggests the key roles of SVs in organisms' evolution, and adaptative responses to challenging environments, that enabled natural selection, domestication, and speciation (Lye & Purugganan., 2019; Radke & Lee., 2015; Lee et al., 2013).

SVs represent an important source of genetic and phenotypic variation that provides genomic structural information complementary to SNP data (Hastings et al., 2009; Hou et al., 2011). Although SNPs are more abundant, SVs affect larger fractions of the genome and account for the greatest amount of total polymorphic content among individual genomes (Couldrey et al., 2017). In addition, SVs may cause potentially greater effects than SNPs, from modification of genes to disruptions in regulation of gene expression that led to the manifestation of recessive alleles (Conrad et al., 2010; Zhang et al., 2009).

The association of genetic variants with certain phenotypes through genome-wide association studies have been mostly based on SNP arrays. The most known SV types are CNVs, which have been associated with human disorders such as Parkinson, Alzheimer, mental retardation, autism, Schizophrenia, HIV susceptibility, and others (Zhang et al., 2009). In livestock, the study of copy number variants has reached an increasingly attention because of their association with economically important phenotypes. A classic example of CNVs affecting phenotypes in livestock species are the two duplications overlapping the *KIT* (KIT proto-oncogene receptor tyrosine kinase) and the *ASIP* genes, which are responsible for white coat color in pigs and sheep, respectively. In chickens, duplications within the *SOX5* gene (SRY-box 5 gene) have been associated with pea-comb phenotypes, and in cattle, a deletion in the *ED1* gene has been known to cause the anhidrotic ectodermal dysplasia. Recently, some authors have conducted genome-wide CNV based studies in cattle for economically important traits including milk production traits (Zhou et al., 2016), milk somatic cell score (Durán Aguilar et al., 2017), growth traits

(Zhou et al., 2016). meat tenderness (Henrique et al., 2016), and fatty acid profile (Lemos et al., 2018a).

Although NGS has become more feasible in livestock, most of studies regarding structural variant wide detection are still being conducted from SNP arrays-based approaches. Few published studies have explored the structural diversity from whole genome sequencing data, most of them using data from less than 200 animals, and focused on the assessment of copy number variants (CNVs) rather than the other SVs types (Peripolli et al., 2023; Sun et al., 2023; Braga et al., 2022; Butty et al., 2020; Bickhart et al., 2016; Santana et al., 2016; Silva et al., 2016; Boussaha et al., 2015, and Letaief et al., 2015).

3. MATERIAL AND METHODS

3.1 DNA samples and resequencing

A total of 151 Nelore bulls comprised the analyzed population for the current study. The ancestors were selected based on their contributions to the genetic diversity of the Brazilian Nelore population. For that, pedigree information of 2,688,124 related individuals, with a total 9,811 sires, from which 6,040 were founders, and 915,371 dams was used. The selected bulls for DNA re-sequencing were the ones less related to the others, those having the higher number of progenies and that have greater contributions for the genotyped population, which consists in more than 60,000 genotyped animals from five Brazilian Nelore breeding programs (DeltaGen <http://deltagen.com.br/>, Nelore Qualitas (<https://qualitas.agr.br/>), Cia do Melhoramento, PAINT (<https://www2.crvlagoa.com.br/paint>), and Instituto de Zootecnia (<http://www.iz.sp.gov.br/>), (FAPESP# 2009/16118-5, FAPESP #2017/10630-2 and #2018/20026-8). For further information regarding the selected ancestors see Fernandes Junior et al., 2020.

Genomic DNA were extracted from semen samples (100 µL) using the DNeasy Blood and Tissue DNA Extraction Kit (Qiagen, Valencia, CA, USA). Each purified

DNA sample was quality-controlled by agarose gel electrophoresis (1%). Extracted DNA was quantified based on the absorbance using the NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific, Santa Clara, CA, USA, 2007).

The whole genome resequencing of representative bulls was performed using Illumina sequencing by synthesis technology. From all DNA samples, 52 were resequenced using the HiSeq X Ten and the other 99 with NovaSeq 6000 System platform (Illumina Inc., San Diego, USA). DNA libraries were constructed with end-linked adapters, and the resulting libraries were resequenced by paired-end sequencing (sequencing of both ends of DNA fragments), which produced an average of 360 million reads per bull with a length of 150 Bp (2x150), and an average genome coverage of 14.5x ranging from 7.8 to 26.3x. Poor quality reads were trimmed using SeqClean software with default parameters. The SAMtools software (H. Li et al., 2009) were used to index reads, and the alignment to the bovine reference genome ARS-UCD1.2 (<http://www.ncbi.nlm.nih.gov/genome/?term=bos+taurus>) was performed using the software HiSat2 (Kim et al., 2015), in which 97.9% of reads were successfully paired. Further details regarding the resequencing and alignment can be found in Fernandes Junior et al., 2020 and Fernandes Junior et al., 2021.

3.2. Detection of Structural Variants

Detection of structural variants (SV) were conducted adapting the workflow proposed by Letaief et al., (2017a) in French dairy and beef cattle (Figure 2). SV calling for each single sample was performed using Parliament2 (Zarate et al., 2021), a software that integrates multiple bioinformatics tools for high-quality structural variation events calling. The consensus strategy of Parliament2 provides information about size, type, location and calling method for each SV event. We used BAM and BAI files containing the whole genome nucleotide sequences of 151 Nellore bulls, as well as BAM and BAI files containing the bovine reference genome sequence (ARS-UCD1.2). For this dataset, we executed Parliament2 setting the default parameters for each sample through single Docker image across the DNAnexus platform. A total of three bioinformatic tools were integrated through Parliament2:

CNVnator (Abyzov et al., 2011), Lumpy (Layer et al., 2014), and Delly (Rausch et al., 2012). Lumpy uses signals from split reads and discordant paired end reads to predict breakpoints of deletions, duplications, and inversions.

After computing Parliament2, we obtained VCF files containing the SV called from each bioinformatic approach, for each of the 151 Nellore bulls, consisting in a total of 6 VCF files, described as follows: 2 VCF files containing CNVnator and Lumpy detected variants, and a total of 4 VCF files containing each Delly detected variant type (duplications, deletions, insertions, and inversions).

3.3. Filtering of called structural variants

The filtering of detected SVs was performed at software, individual and population level, in order to remove unreliable raw SVs and reduce the false discovery rate. For the following analyses, different sets of utilities integrated by BCFtools (Danecek et al., 2021) and Bedtools (Quinlan & Hall, 2010) were used.

All called SVs were firstly filtered at the bioinformatic tool level, following specific criteria as suggested for each SV caller.

i). CNVnator: we only retained the called structural variants that were significant for the t-statistical test ($P_value < 0.05$), for which the null hypothesis statement is that the mean read signal depth in a SV region is the same as the mean signal depth in the sample.

ii). Delly: only called SV that passed the quality filters (i.e., flagged PASS), and supported for more than 3 paired-end events ($PE\ support > 3$), were kept. After filtering, VCF containing the remaining DEL, DUP and INS variants were joined using the option *concat* of BCFtools to obtain a single VCF containing the copy number variants (CNVs), while the VCF containing filtered INV variants was maintained separated.

iii). Lumpy: SVs with a read count supported by more than 1 split-read event or more than 3 paired-end events ($SR > 1$ or $PE > 3$) were maintained.

The SVs that passed the described criteria were compared at the individual level (i.e., within the same sample), to identify a high-confidence set of SVs for each bull (composed only by SVs common to all bioinformatic tools). As each of the three

callers used has a different power to detect a specific SV type, such as CNVnator which only detects CNVs, the within sample analysis was performed for each SV type separately:

i. Identification of high confidence CNVs: the consensus files containing the filtered variants from CNVnator and Lumpy, and the concatenated and filtered file containing duplications, deletions, and insertions from Delly, were compared against each other using the option *intersect* of Bedtools by fitting a minimal reciprocal overlap threshold of 90%.

ii. Identification of high confidence INVs: the consensus filtered file containing the SVs detected from Lumpy, and the one containing the inversions detected from Delly, were compared against each other using the option *intersect* of Bedtools by fitting a minimal reciprocal overlap threshold of 90%.

The establishment of a reciprocal overlap threshold of 90% means, that if at least 90% of the sequence of a certain SV detected by one caller, is the same as the sequence of a SV predicted by the others, that SV must be deemed as high-confidence SV for a certain bull.

After executing the within-sample filtering, we obtained two VCF files for each bull, one containing the high-confidence inversions (INVs), and the other one containing the high confidence copy number variants (CNVs). All VCF files containing each type of high confidence SVs, were joined using the option *concat* of BCFtools, to obtain two files containing the high confidence CNVs and INVs, respectively, for all bulls. CNVs and INVs were then submitted to additional filtering, where SVs located in sexual chromosomes and with a length shorter than 1 Kb and higher than 5 Mb were removed. The remaining CNVs and INVs, were considered for downstream analysis.

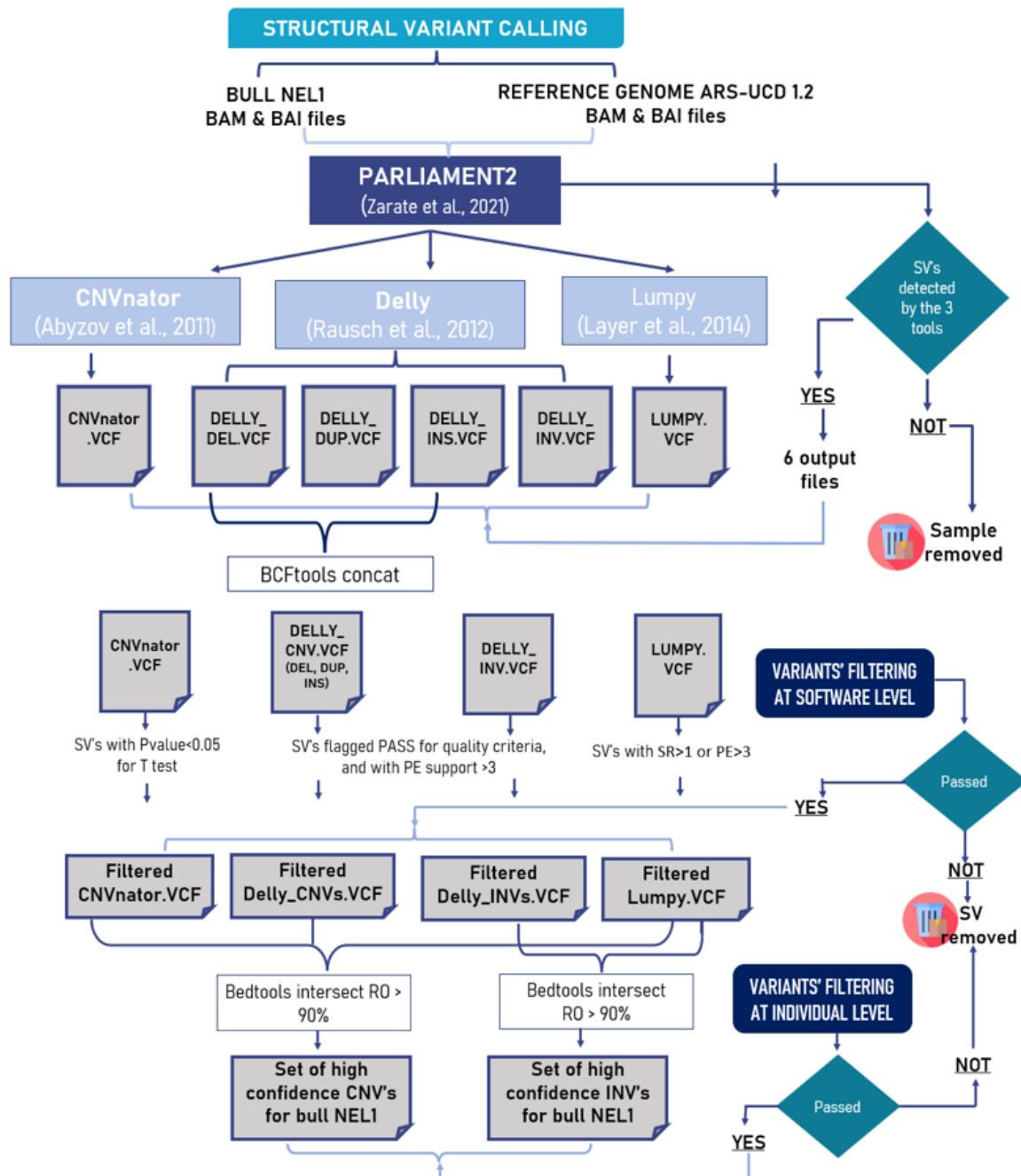


Figure 2. Flowchart of pipeline used to detect structural variants (SV). CNV: copy number variants; RO reciprocal overlap. DEL: deletions; DUP: duplications; INV: inversions. First, SV were detected separately by each tool within each sample through the integration software Parliament2. Second, it was applied filtering at software level following suggested criteria for each tool. Only SVs that passed all filters were retained and compared within sample to retain common CNVs and INVs detected by the different tools. CNVs should share at least 90% of their sequence between the three tools, and INV should share at least 90% of their sequence between Delly and Lumpy calls. Own work.

3.4. Identification of structural variant regions (SVRs).

The join of structural variants into regions is required for population-wide analysis, as it allows the detection of DNA segments having different structural variation events (Redon et al., 2006). In the present study, the regions of structural variants (SVR) were built by collating, separately, the two types of high-confidence SVs (CNVs and INVs), along all bulls that overlapped in at least 1 Bp. For this, we used the utility *merge* of Bedtools fitting operations such as count and collapse, that allow the identification of samples sharing a certain region and the occurring SV event. Regions that were not shared by at least 5% of the bulls (N=7), were removed. As a result, we obtained two sets, one containing the regions of copy number variants (CNVRs), and the other one with the regions of inversions (INVRs), which were then joined in a single file using the option *concat* of BCFtools, to facilitate the descriptive analysis of SVRs (Figure 3).

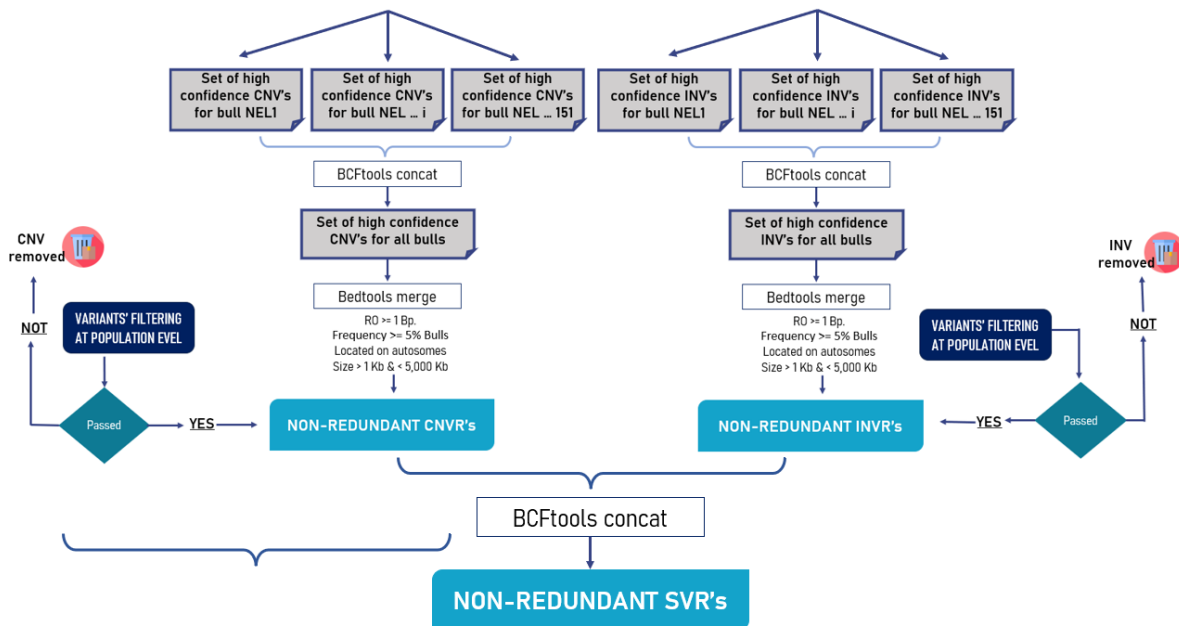


Figure 3. Flowchart of pipeline used to detect regions of structural variant (SVR). CNV: copy number variants; RO reciprocal overlap. DEL: deletions; DUP: duplications; INV: inversions. SVR: regions of structural variants; CNVR: regions of copy number variants; INVR: regions of inversions. All sets of detected high confidence CNVs and INVs were merged across all samples with a minimal reciprocal overlap of 1 Bp. Own work.

3.5. Statistical analysis and graphical visualization of SVs and SVRs

All reported descriptive statistics and plots about the results were calculated using functions and utilities of the R package Tydiverse (Wickham et al., 2019) through the Integrated Development Environment Rstudio. The association of different metrics such as the number, distribution and (average) length of the predicted SVs and SVRs, with chromosome size and genome coverage rate, were assessed through pairwise Pearson correlation method using the R package Performance Analytics (Peterson et al., 2014).

3.6 Functional analysis of highly frequent SVR

The functional implications of predicted SVR were assessed by searching genes and QTLs overlapping the most frequent SVR in the 151 bulls (SVR present in more than 50% of bulls, $N > 75$) using the GALLO package of R (Fonseca et al., 2020). Genes and QTLs were retrieved from the Ensembl Genes database (<https://www.ensembl.org/>, last accessed 05/11/2021) and Animal Genome database (<https://www.animalgenome.org/cgi-bin/QTLdb/BT/index>, last accessed 05/11/2021), respectively. Only significant QTLs ($p_value < 0.05$) and protein coding genes overlapping the most frequent SVR were reported in this study.

Functional enrichment analysis for positional candidate genes overlapping the most frequent SVR were performed through PANTHER 17.0 software. PANTHER statistical overrepresentation test was conducted using, as analyzed list, the set of positional candidate genes harbored by the most frequent SVR (a total of 1,164 genes), and as reference list the *Bos taurus* genes contained in PANTHER databases. The annotation data sets were PANTHER GO-slim Molecular Function, PANTHER GO-slim Biological Process, PANTHER GO-slim Cellular Component, PANTHER Pathway and Reactome Pathway. A Binomial test was applied in order to test whether the genes overlapping the SVR were enriched for the gene ontology categories and biochemical pathways. This assesses the probability of observing a certain number of genes related to a specific biological mechanism within the found gene set given all genes in *Bos taurus* reference genome. As different biological

functions were tested at the same time, the Bonferroni correction method for multiple testing was used, which allows reducing the probability of false positive associations of genes and gene ontologies and/or pathways. Only genes with significant associations (P value < 0.05) with gene ontologies (GO) or pathways were reported. Tidyverse and ggplot2 R packages were used to graphically represent the QTL type and trait ID of significant QTLs, and the significantly enriched GO and/or pathways, through pie charts and bar plots.

4. RESULTS AND DISCUSSION

4.1 Genome wide SV detection: metrics regarding number, length, and chromosomal location

4.1.1 Number of high confidence SVs:

A total of 215,031 high-confidence structural variants were obtained for the analyzed bulls that passed all the described filtering criteria. Statistics summary for the number and length of high confidence SVs, including CNVs (deletions and duplications) and inversions, are presented in Table 1. Although insertions were detected by the SV callers, after applying filtering criteria and executing the reciprocal overlapping into the CNV set, there were not insertions remaining, as they were detected with low quality and corresponded to small insertions (average length of 1 Bp), which are not considered as structural variation.

Table 1. Statistics summary for the number and length of detected high confidence SVs.

SV type		N	Length (kb)			
			Mean (Sd).	Median	Max	Total
CNV	DEL	183,032	22.68 (167.31)	1.91	4,916.08	4,150,681
	DUP	14,013	250.61 (560.12)	63.98	4,906.46	3,511,807
INV		17,986	613.03 (851.34)	167.45	4,989.55	11,025,935
TOTAL		215,031	86.91 (365.11)	2.21	4,989.55	18,688,423

SV: structural variant. CNV: copy number variants; DEL: deletions; DUP: duplications; INV: inversions. N: number of SVs; SD: standard deviation; Max: maximum; Total: sum of all SV lengths. kb: kilo base pairs.

From the total number of detected SVs, 197,045 corresponded to CNVs, and the number deletions was 13.06 times higher than of duplications (Table 1). This result is similar to the reported by Silva et al. (2016), who detected 11.8 times more deletions than duplications in Nellore cattle using whole genome sequencing data. It also agrees with the findings of Santana et al. (2016), in which the number of loss events was higher than the number of gain events for the detected CNVs from SNP data in Nellore cattle, although the ratio presented here is higher than the reported by the latter authors (1.18 times more losses than gains). Similarly, our result is congruent to the ratio reported by Braga et al. (2023) in Gyr cattle (1.96 times more deletions than duplications), who detected CNVs from WGS using CNVnator and Delly software, and agrees with the findings of Hu et al. (2020) in different cattle breeds, including Nellore, where the number of deletions was higher than the number of duplications. It is interesting that the authors pointed out that, the proportion of observed deletions was higher in the *Bos indicus* than in *Bos taurus* analyzed animals. In contrast, the ratio of deletions over duplications observed in this study, diverges from the reported by Peripolli et al. (2023) who used whole genome sequencing data from three Brazilian locally adapted taurine breeds, where

the number of observed duplications were 1.85 times higher than the observed deletions.

The remarkably elevated number of deletions may be related to mechanisms causing structural variations within genome. The non-allelic homologous recombination (NAHR) has been known as the main mechanism causing deletions and translocation events (Arlt et al., 2012b). The findings of our study supported by the previously reported in Nellore and other *Bos indicus* breeds, may suggest that evolution and diversity of Nellore cattle genomes have been caused by the deleted DNA fragments, but this can only be known by investigating the genomes of a higher number of ancestors. Besides, it is plausible that the alignment of sequences to the *Bos taurus* reference genome led to bias. The comparison of the copy number likely excluded specific sequences present in *Bos taurus* that are no longer present in the indicine cattle genomes, leading algorithms to consider those sequences as deleted fragments (Hu et al., 2020). Thus, a better characterization of structural variation in indicine cattle requires the availability of a reliable *Bos indicus* reference genome.

Regardless the algorithm used to detect the structural variants from short reads sequencing, it is often expected to find out a higher number of deletions than duplications, as these latter are more difficult to identify. Read pair approach detects the duplicated copies into a mapped read that are located at a shorter distance than the fragment length, and when the length of the duplicated sequence is larger than the insert size, the algorithm tends to fail in the accurately detection of that duplication (Teo et al., 2012). By other side, the read depth (RD) algorithm is more likely to easily detect deletions, as they are signaled by a decrease in the read depth or coverage at a certain genomic region (Bickhart & Liu, 2014), while duplications may not always result in a straightforward increase in read depth, once the duplicated region may have variable copy numbers and irregular coverage, making detection more challenging (Teo et al., 2012). Furthermore, increases in the read depth or coverage not always are an indicative of duplicated copies. These could be caused by alignment artifacts that cannot be differed from a truly duplication through the RD approaches (Mills et al., 2011).

Combining different SV calling algorithms should reduce the false discovery rate compared to the use of a single approach. Furthermore, the application of a stringent merge strategy and filtering criteria, as conducted in this study, may ensure a more reliable identification of copy number variants (Letaief et al., 2017b). The increase in the accuracy of detection of all types of SVs can be achieved through long-read sequencing (up to 80 kb), in which single reads span complex structural variants, but its use in large scale continues to be limited because of the high cost (Couldrey et al., 2017).

In relation to the balanced structural variants, inversions represented 8,36% of the total predicted high-confidence SVs (Table 1). The proportion of inversions observed in the present study, was lower than the reported by Boussaha et al., (2015), who used WGS data from 380 animals of different *Bos taurus* dairy and beef cattle breeds, in which the inversions represented more than 30% of the total number of detected SVs. Our results also differ from the reported by Chen et al., (2017), where the proportion of inversions was higher than 20% of the total number of detected SVs, in a study using WGS data of 308 Holstein animals. However, our results agree with the reported by the latter authors in Jersey cattle, in which inversions represented 9% of the total number of detected SVs. By other side, the detected number of inversions here, is higher than the percentage of inversions in Holstein cattle reported by Lee et al., (2022) (~4%), who used RD and SR approaches to detect SVs from WGS data.

There is still a lack in knowledge regarding the characterization of balanced structural variants, such as the inversions, in Nellore cattle. This is because most of the reported studies in this breed are based on SNP genotyping data (Lemos et al., 2018b; Santana et al., 2016, Silva et al. 2016), from which is not possible to identify mobile elements but only copy number gains and/or copy number losses (Alkan et al., 2011). Nevertheless, there is an increasingly interest on the creation of structural variation maps in cattle, with a higher number of published studies in taurine dairy and beef cattle using whole genome sequencing data, than in indicine cattle.

In evolutionary genomics, chromosomal inversions are known as powerful forces contributing to local adaptation and diversification, as they have the potential

to protect inverted sequences from recombination, allowing the maintenance of favorable alleles combinations (Wellenreuther & Bernatchez, 2018). Despite of their evolutionary importance, there is not enough information that enable the characterization of this type of SV in cattle, and most of the studies continue focusing on deletion and duplication events. This is likely because the detection of SVs is often conducted through the read depth (RD) approach, which only allows the identification of CNVs, once balanced rearrangements (inversions and translocations) do not cause changes in read depth signature (Rasekh et al., 2017; Escamarís, Docampo & Rabionet., 2015). Furthermore, the complexity of the nature of these variants makes their interpretation and discovery more challenging when compared to the other SV types. The inversions and translocations' breakpoints usually lie within complex repeats that reduce mappability, for which the reliable detection requires the use of bioinformatic tools based on a combination of algorithms such split read and paired reads (Rasekh et al., 2017).

4.1.2. Length classification of high-confidence SVs.

The high confidence SVs have an average length of 86.9 Kb (± 365.1) ranging from 1 Kb to 4,989.5 Kb (Table 1). Although the CNVs, particularly, the deletions, represent the highest number of the detected SVs, on average, they are the smallest (22.68 ± 167.31 Kb), while inversions are the longest ones (613.03 ± 851.34 Kb), followed by duplications (250.61 ± 560.12 Kb) (Table 1). Analysis of distribution of SVs according to different size classes revealed that most CNVs are 1-5 Kb long (73%) while inversions are distributed in all size classes, but most of them being 200 kb long (48%) (Figure 4). From the predicted deletions, the highest proportion are distributed into the classes 1-5 kb and 5-10 kb (77,2 and 16%, respectively), whereas predicted duplications are distributed in all size classes, where the lowest proportion are 5-10 kb long (7,2%) (Table 2).

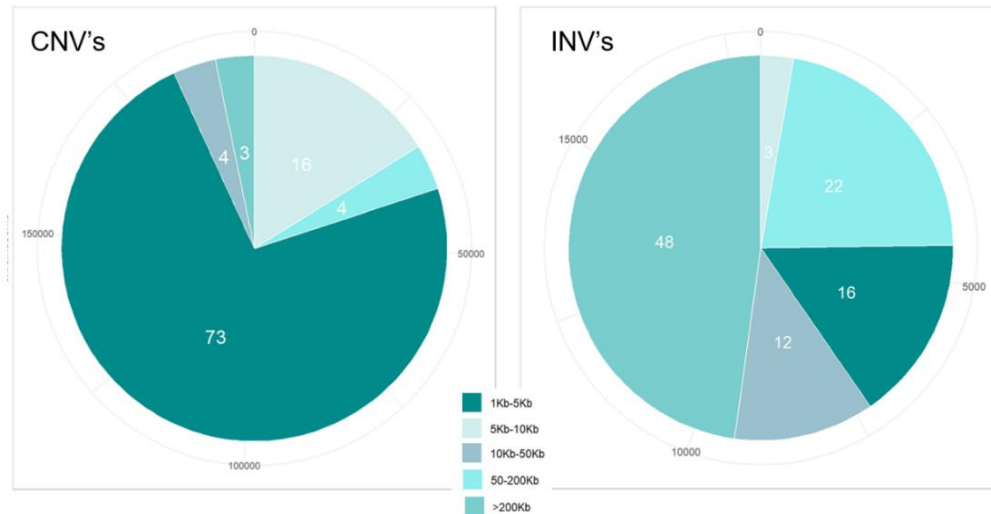


Figure 4. Distribution of the detected structural variants according to different size classes. Copy number variants (CNVs - left) and inversions (INVs - right).

Table 2. Statistics summary of the number and proportion of each SV type according to different size classes.

Size class	SV type	N	Percentage (%) *
1-5 Kb	DEL	141,259	77
	DUP	2,935	21
	INV	2,807	16
5-10 Kb	DEL	30,788	17
	DUP	1,013	7
	INV	493	3
10-50 Kb	DEL	4,593	2
	DUP	2,504	18
	INV	2,126	12
50-200 Kb	DEL	3,322	2
	DUP	4,225	30
	INV	3,962	22
>200 Kb	DEL	3,070	2
	DUP	3,336	24
	INV	8,598	48

DEL: deletions, DUP: duplications; INV: inversions; N: number of SVs; Kb: kilo base pair. *Percentage regarding the number of each SV type.

The distribution of high-confidence CNVs in the established size classes agrees with the results of Bickhart et al. (2012), who observed significant shorter CNVs (<15 Kb) in a Nellore individual, than the average length of the CNVs detected in different taurine individuals using WGS data. In contrast, our results differ from the reported by Lemos et al., (2018b), in Nellore cattle, who observed that most of the CNVs detected from SNP data had length ranging from 10 to 500 Kb, and that less than 30% of the CNVs were 1-10 Kb-long. The results presented here also differ from those reported by Silva et al., (2016) in Nellore cattle, in which the highest number of CNVs predicted from SNP data ranged from 100 to 900 Kb.

It is important to remark that the differences in the CNV length between the reported in the present study, and the results reported in literature, are mostly attributed to the data nature used, and consequently, to the bioinformatic algorithm for CNV calling. WGS data provides higher resolution compared to SNP data, as allows the identification of sequence reads spanning the entire genome. WGS based algorithms enables the identification of smaller structural variants, specially CNVs with a greater power (Ho et al., 2020). In contrast, the CNV detection using SNP genotyping data, often relies on the analysis of changes in the signal intensity at SNP probes. As there is a wide offer of commercial SNP arrays, they all have different probe densities, which directly impact CNV results (Bickhart & Liu, 2014; Alkan et al., 2011). SNP based detection methods, exhibit a good performance detecting large events, but compared to WGS based algorithms, they tend to lack power in detecting smaller CNVs (<10 Kb), because of the low SNP probe density and resolution (Alkan et al., 2011).

In general, the inversions are the largest structural variants detected, with an average length of 613.03 ± 851.34 Kb (median of 167.45 Kb) (Table 1). The length of inversions observed in this study is higher than the reported by Chen et al., (2017) in Holstein (median ~2.53 Kb) and in Jersey (median ~1.34 Kb) cattle. Considering the observed average length of inversions, 48% are distributed into the size class >200 Kb, 22% into the size class 50-200 Kb, while a low proportion (3%) are distributed into the size class 5-10 Kb (Figure 4, Table 2). These results differ from the reported by Boussaha et al., (2015) in European dairy and beef cattle animals,

in which more than 50% of the detected inversions were distributed into the size class corresponding to 1-10 Kb.

It is outstanding that the combination of tools and SV calling algorithms enabled the detection of large inversions, as they tend to be harder to identify. There are few available and reliable information regarding inversions and translocations in livestock species, especially in cattle, limiting comparisons. More studies of inversions are available in humans than in cattle, as there is an increasingly interest in assessing the impact of reversed chromosome regions on phenotypes and diseases (Puig et al., 2015). Accordingly, the vast number of inversions in the human genomes, available the InvFEST public database (Martínez-Fundichely et al., 2014), are lower than 100 Kb long, from which, variants with lengths lower than 10 Kb are considered reliable, and those larger than 10 Kb are probably false positives (Puig et al., 2015).

Nevertheless, genomic studies on inversions reviewed by Wellenreuther & Bernatchez, (2018), showed that the average length of chromosomal inversion events in different species of plants and animals is 8.4 mb, ranging from 130 kb to 100 mb. According to studies in evolutionary genomics in different species including invertebrates, and mammals, large chromosomal inversions are more favored by selection than shorter inversions, exhibiting lower rates of recombination between a larger number of genetically distant loci (Wellenreuther & Bernatchez, 2018; Dobigny et al., 2017).

4.1.3. Chromosomal location and distribution of high-confidence SVs.

The predicted high-confidence structural variants were not uniformly distributed across the 29 autosomal chromosomes. The top five chromosomes harboring the highest number of predicted SVs are BTA1 (N= 14,996 SVs), BTA2 (N=13,773 SVs), BTA5 (N=12,834 SVs), BTA4 (N=12,16 SVs) and BTA10 (N=10,372 SVs), while the top five chromosomes harboring the smallest number of SVs are BTA25 (N=2770 SVs), BTA19 (N=3,145 SVs), BTA22 (N=3,676), BTA27 (N=4,035), and BTA24 (4,607). The distribution of each SV type along the 29 autosomes is shown in Figure 5.

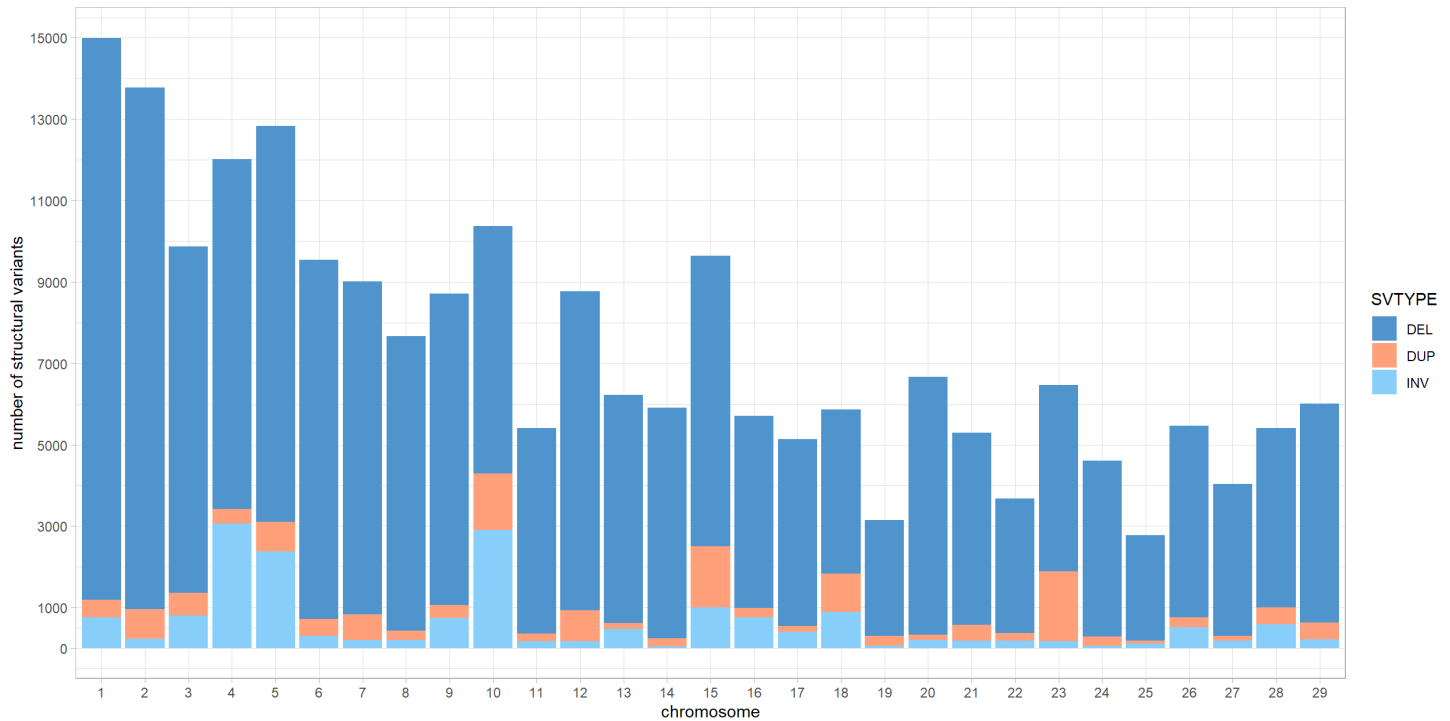


Figure 5. Number of high-confidence structural variants harbored by each of the 29 autosomal chromosomes. DEL: deletions, DUP: duplications. INV: inversions.

Regarding each SV type, the BTA1, BTA2, and BTA5 carry the highest number of deletions, whereas BTA23, BTA15 and BTA10 carry the highest number of duplications, and BTA4, BTA10 and BTA5 harbor the highest number of inversions (Figure 5). Our findings about the distribution of SV along the chromosomes are consistent with the observed by Peripolli et al., (2023) by analyzing NGS data from three Brazilian locally adapted taurine breeds, in which the largest number of structural variants were found on BTA1 and 15, although the authors only predicted CNVs (duplication and deletions). The chromosomal locations observed here are also similar to those reported by Letaief et al., (2017b) in different French dairy and beef cattle breeds, in which the largest number of SV was observed on chromosomes 1, 6 and 5, and the smallest number on chromosome 25. The authors also combined different tools for the SVs calling. The use of different bioinformatic algorithms, when analyzing NGS data, the filtering and validation of SVs at software, at individual, and population level, can lead to the identification of high confidence

structural variants. In contrast, chromosomal location of SV was different to the reported in Nellore cattle from SNP and NGS data by Silva et al., (2016) in which most predicted structural variations were located on BTA19, BTA11, BTA2 and BTA18 for SNP-based called SVs, and on BTA12, BTA15, BTA1, and BTA2 for NGS-based called SVs. Notice that the mentioned study was only focused on CNV assessment.

4.1.4. Chromosomal extension covered by high-confidence SVs.

BTA10 was ranked as the fifth, regarding the total number of SVs that carries, but the average length of its harbored SVs (728.36 kB) was notably high in relation to the average length of SVs located on the other chromosomes, followed by SVs located on BTA9, 23, 5, and 15, with average lengths ranging from 102.98 Kb to 143.48 Kb (Figure 6). However, Pearson correlation did not indicate statistical evidence regarding the association between the number of SVs and the average length of SVs harbored by each chromosome (Pearson correlation score $\rho = 0.197$, p value = 0.3053). Our findings were similar to the reported by Peripolli et al. (2023), in which the longest CNV, corresponding to deletion events, was observed on chromosome 10. The BTA10 was the one having the largest proportion of average chromosome size covered by structural variation (average coverage rate=50.4%), followed by BTA 23, 9, 15, and 28 with average coverage rates ranging from 7.6 to 11% (Figure 7). The lowest average coverage rate was observed on BTA25, which was also the chromosome harboring the lowest number of structural variants and the smallest ones, and was followed by BTA17, 19, 16, 14, 8, and 3 with average coverage rates ranging from 0.9 to 1.1%. These results differ from the reported by Letaief et al. (2017), where the highest percentage of sequence covered by copy number variations in French dairy and beef cattle breeds was found on chromosome 27, but agrees in that the less covered sequence by structural variants was located on BTA25.

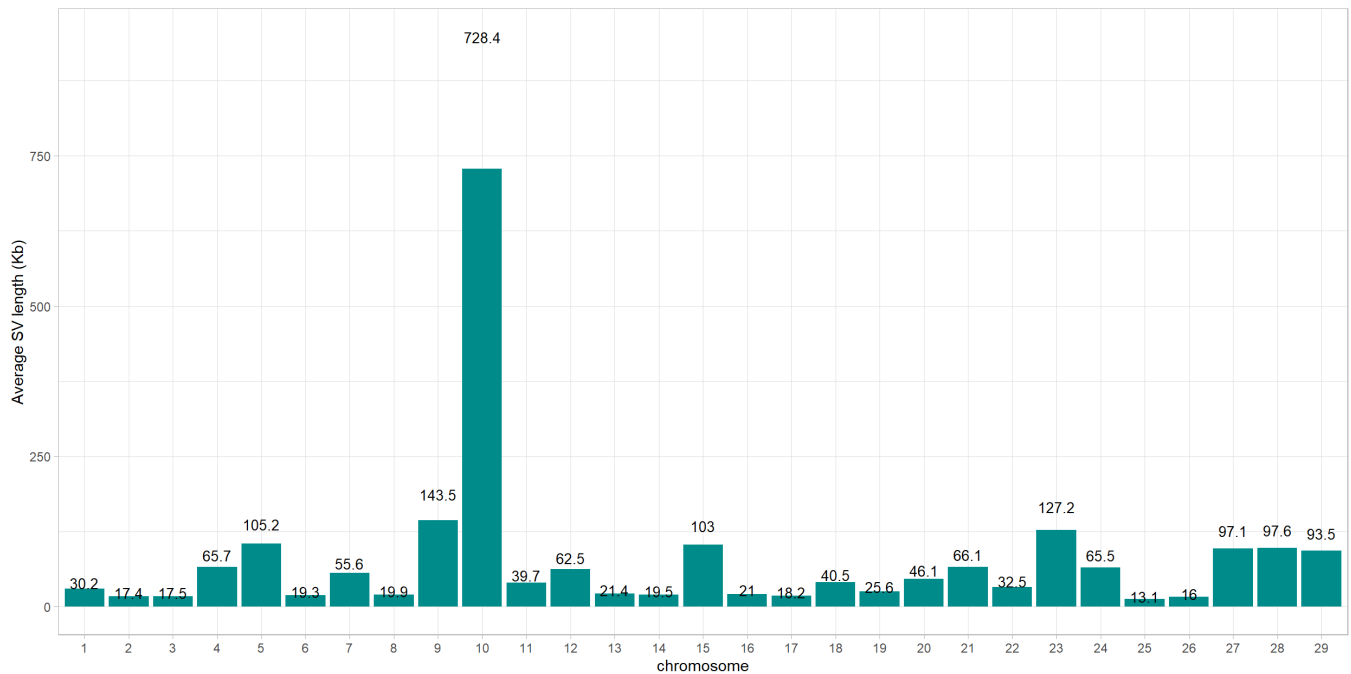


Figure 6. Average length (Kb) of all predicted high-confidence structural variants across the 29 autosomes

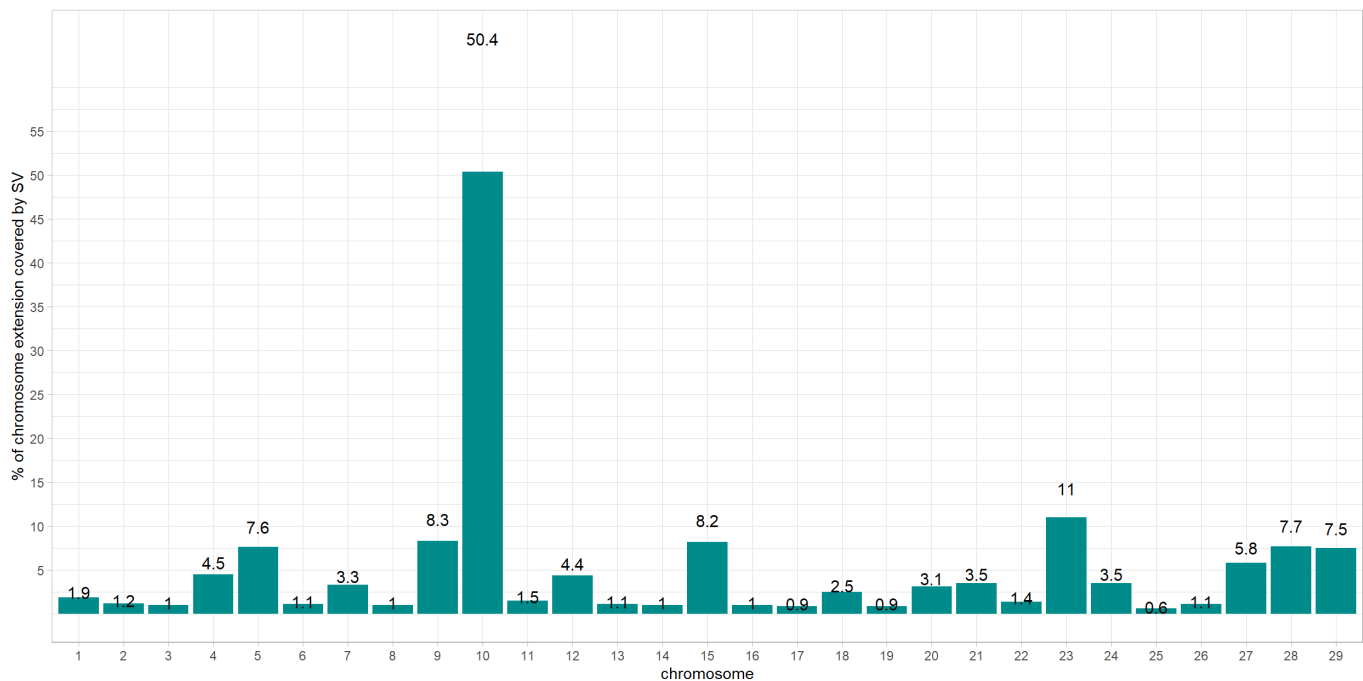


Figure 7. Average proportion (%) of chromosome extension covered by high-confidence SV (average coverage rate).

Pearson correlation analysis revealed a significant high and positive association between the number of structural variants and the chromosome size (Pearson correlation score $\rho = 0.87$, p value $< 6.8e-10$, Figure 8.), which agrees with the findings of Letaief et al. (2017) but different from the results reported by Boussaha et al. (2015) in French dairy cattle. In contrast, no-statistical evidence regarding correlations between the chromosome size and the average proportion of chromosome extension covered by SVs (average coverage rate) was found in this study (Pearson correlation score $\rho = 0.04$, p value $= 0.833$).

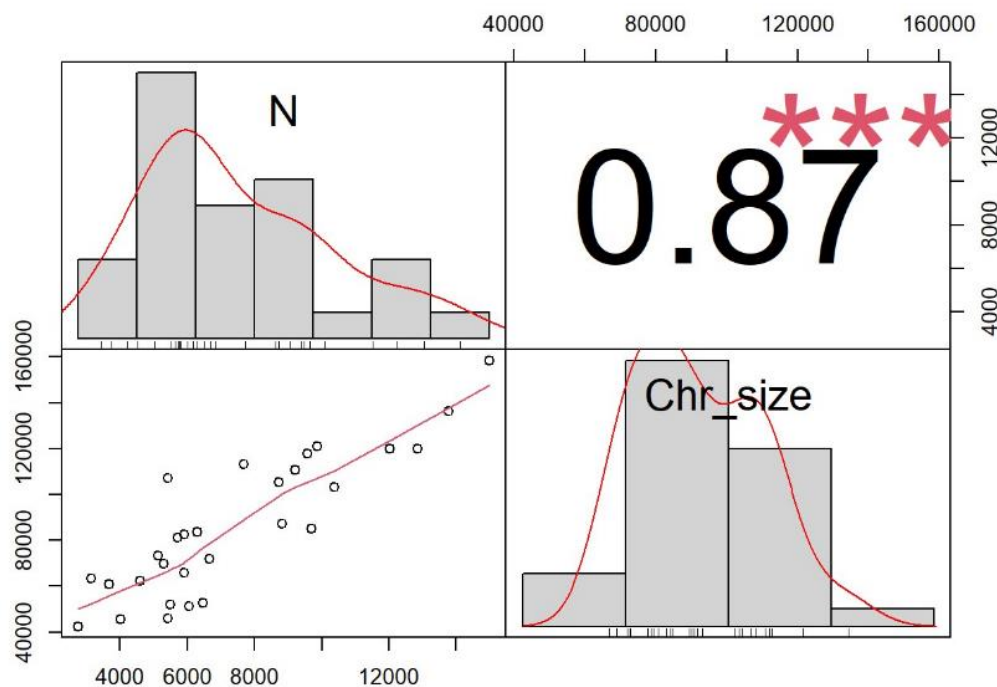


Figure 8. Pearson correlation between the number of predicted high-confidence SVs and the chromosome size. N: number of SVs. *** p value < 0.05 .

At the individual level, an average number of 1,437 SVs (± 894.7) were predicted for each bull, with a minimum number of 89 and a maximum of 3,428 SVs. The descriptive statistics about the number and length of each SV type, observed in the analyzed bulls are presented in Table S1. Structural variation encompasses, on average, 4.81% (± 4.38) of the Nellore bulls total genome extension. The largest genome coverage rate of SV observed in a bull was of 20.27% and was largely

attributed to inversion events, and the lowest genome coverage rate was 0.03% (Figure 9, Table S1). A significant, moderate, and positive correlation was found between the number of SVs predicted per bull in each chromosome, and the chromosomal coverage rate (Pearson correlation score $\rho = 0.36$, p value $< 2.2e-16$, Figure10.)

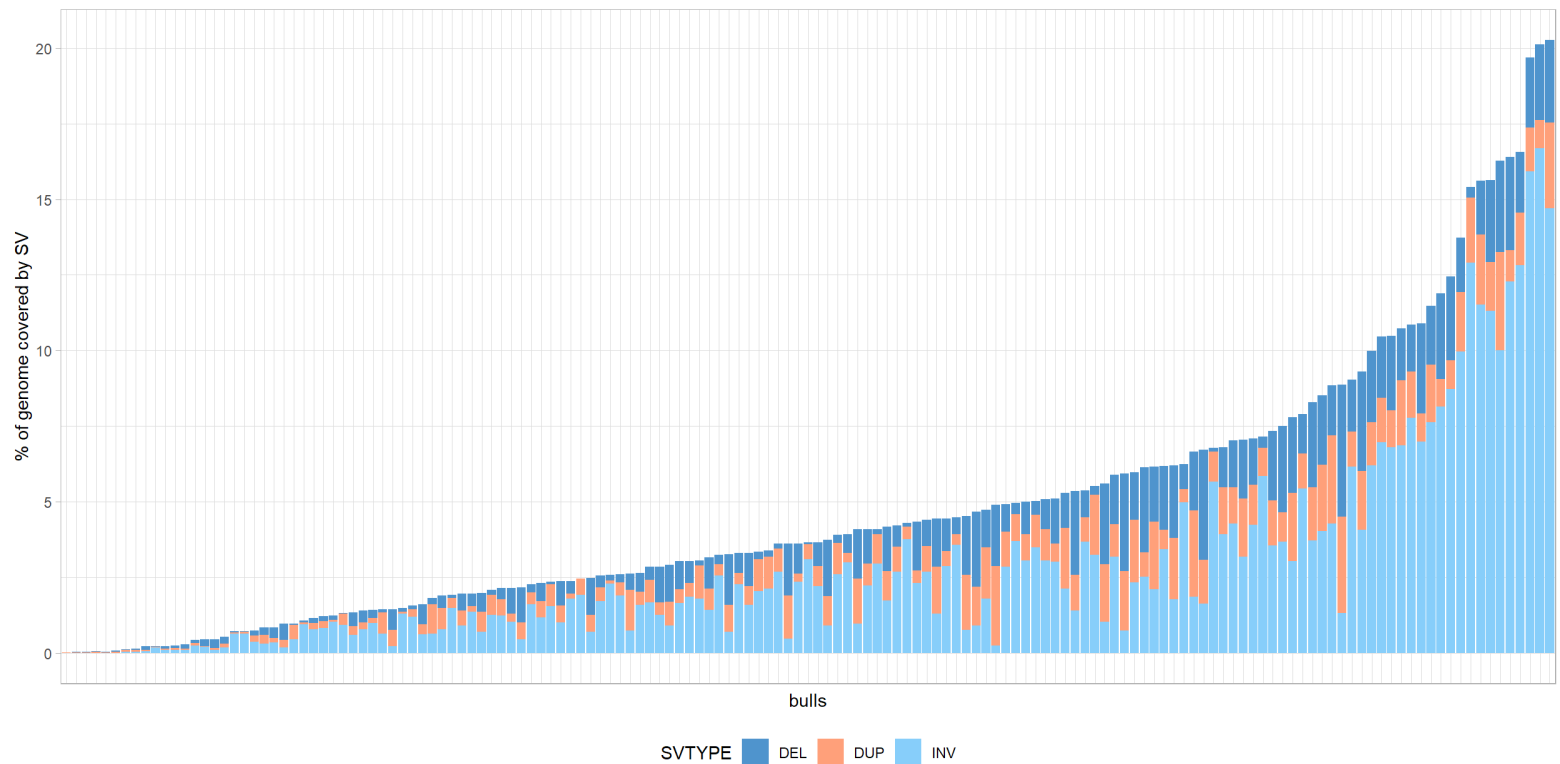


Figure 9. Genome coverage rate: proportion (%) of whole genome extension covered by each type of detected high-confidence SVs in each of the analyzed Nellore bulls. *Each bar represents a bull. DEL: deletions, DUP: duplications, INV: inversions.*

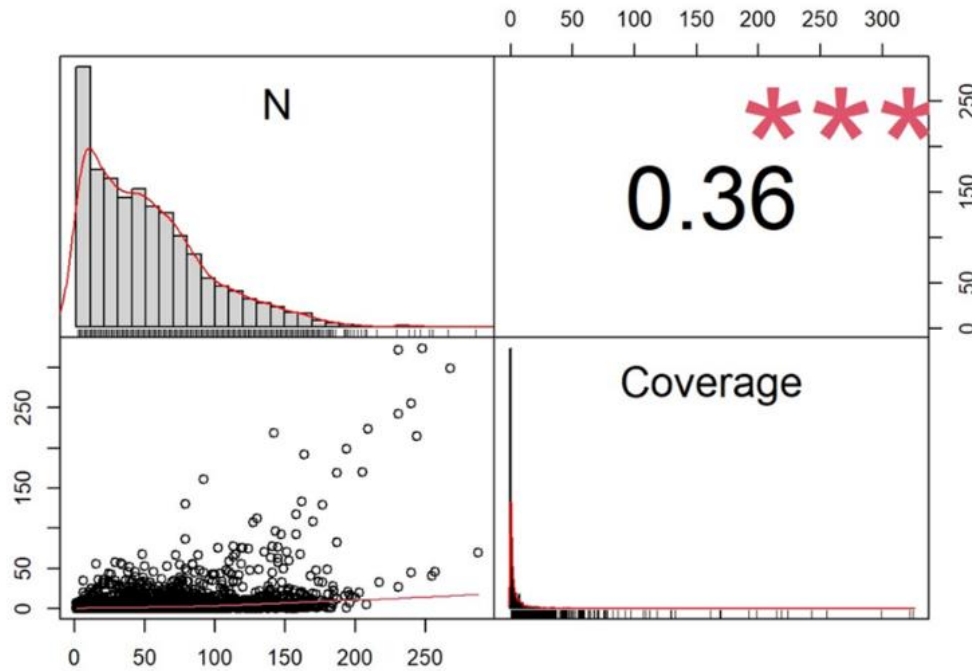


Figure 10. Pearson correlation between the number of predicted high-confidence SVs and the chromosomal coverage rate for each bull. N: number of SVs. *** p value < 0.05.

4.2. Discovery of potential regions of structural variants (SVR)

4.2.1 Number and chromosomal location and distribution of SVRs:

High-confidence structural variants identified in the genome of bulls were used in the discovery of genomic regions, in which different types of structural variants can occur at the same time: in a certain DNA segment it is likely to occur a deletion event in one bull and a duplication event in another bull (across bulls), or it can be also possible to occur both types of variations in the same bull (within animal).

The main purpose of translate structural variants into regions (SVR), is to characterize the genetic architecture underlying populations history, to assess the potential impact of structural variations on gene function and expression, and to integrate information into genomic databases allowing a more comprehensive

understanding of differences between species, breeds, and populations (Redon et al., 2006).

Due to the different nature of SV types, two data sets were formed, one containing the detected CNVs (encompassed by duplication and deletion events), and the other one containing just the inversions. This allowed the proper identification of the occurring SV event into a certain region. Thus, SVRs were constructed from those two data sets, by merging the SVs found in the individual genome of bulls that reciprocally overlapped in at least one base pair, resulting in two types of SVRs: CNVRs (regions of copy number variants) and INVRs (regions of inversions). In addition, only SVRs frequent in at least 5% of analyzed Nellore bulls were considered for further analysis. A total of 3,752 highly frequent SVRs were identified, from which 97% are composed by CNVRs (N=3,641) and 3% correspond to INVRs (N=111) (Figure 11).

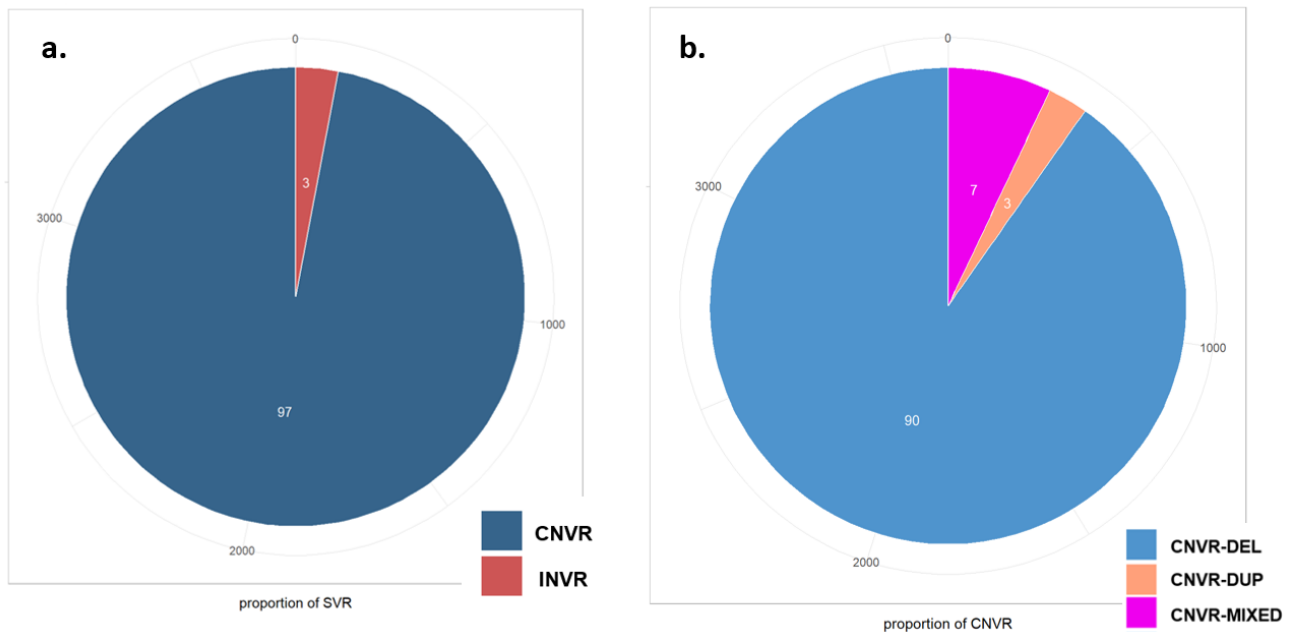


Figure 11. Proportion of each type of structural variant regions (SVR). **a.** Proportion of SVRs (regions of structural variants) corresponding to CNVR (regions of copy number variants) and INVR (regions of inversions). **b.** Proportion of CVR's corresponding to CNVR-DEL (regions of deletions), CNVR-DUP (regions of duplications), and CNVR-MIXED (regions of deletions and duplications).

From the total number of detected CNVRs, the vast majority 90% (N=3,286) correspond to deletion events (CNVR-DEL), a small proportion (7%, N=256) to mixed events in which both deletions and duplications occurred (CNVR-MIXED), and in the less proportion (3%, N=99) to only duplications (CNVR-DUP) (Figure 11b).

The distribution of events occurring into the detected CNVRs observed in this study, is different from the reported by Lemos et al., (2018b) in Nellore cattle, in which the CNVRs were composed in more balanced proportion by the three types of events (38.5% gains, 33% mixed, and 28.5% losses). The results also diverge from the reported CNVRs in Silva et al., 2016 in Nellore cattle, where the 65% of the CNVRs corresponded to mixed events, 31.5% to duplication, and a small proportion (2.89%) to deletion events. Although the CNVRs detected in Nellore cattle by Santana et al., (2016) were attributed in higher proportion to copy number losses than copy number gains, the proportion of deletion events corresponding to the CNVRs observed in this study are higher than the reported by those authors (41,6% deletion events, 35% duplications, and 23% mixed events).

Using information of WGS in Gyr cattle, Braga et al., 2023 found that the CNVRs detected from CNVnator calls were composed in more than 50% by deletions, around 36% by duplications, and 11% by complex events (duplications and deletions). However, the results presented here are more consistent with the reported by these authors, in which the CNVRs detected from Delly calls were attributed in 70% to deletions, 7.8% to duplications, and 22.2% to complex events. Correspondingly, our findings agree with the results of Hou et al., (2012) in different breeds including taurine, indicine and African cattle, in which most of the detected CNVRs were composed by loss events (61.3%), but the representation of regions of deletions observed in this study is still higher.

A considerable higher number of deletions than duplications composing a CNVR was expected, given the total number of DEL calls obtained. Nevertheless, is still surprising the small fraction of duplications representing the common events for all bulls. Interestingly, this finding is similar with the reported by Letaief et al., (2017) in French dairy and beef cattle, in which only 1.6% of CNVRs corresponded to

duplications, 4,1% to complex events (deletions and duplications), and more than 95% were comprised by regions of deletions. Our results are also in concordance with the results reported in a recent study in Simmental cattle combining WGS data from short and long reads, where only 4,2% of the CNVRs were constituted by duplication regions, 39.6% by mixed events, and more than 56% by deletions (Sun et al., 2023). It is important to outline, that the number of the detected duplications in this study corresponded only to tandem duplications, as the algorithms and bioinformatic tools used to conduct the SV calling do not enable the identification of discontinuous duplications. Further factors that have likely driven the small number of called duplications relative to deletions, were already discussed, and are basically referred as the greatest ability of SV callers to identify sequence losses in short reads sequencing data (Chen et al., 2017; Teo et al., 2012), the difficulty to differentiate duplications from alignment artifacts (Mills et al., 2011), and potential molecular mechanisms such as NAHR that potentially leads to more deletions (Arlt et al., 2012b).

The highest number of highly frequent SVRs was harbored by BTA1 (N=261 SVR), followed by BTA6 and BTA2 with 211 SVR each one. BTA25, which carries the lowest number of SVs, has the lowest number of SVR as well (47 SVR), followed by BTA 19 and 22 with 67 SVR each one (Table 3, Figure 12).

Regarding CNVRs, CNVR-DEL and CNVR-MIXED were uniformly distributed across the 29 autosomes, but CNVR-DUP were not observed into BTA20, BTA24 and BTA28 (Figure S1). The BTA1 harbored the highest number of CNVRs (N=254), most of them corresponding to deletions (N=235 CNVR-DEL), followed by mixed (N=12 CNVR-MIXED), and duplication events (N=7 CNVR-DUP) (Table S3, Figure S1). This result is in concordance with previous observations of CNVR distribution in Nellore cattle, reported by Lemos et al., (2018b) and Silva et al. (2016), where the highest number of CNVRs were located on BTA1, but differs from theirs, in that the smallest proportion of CNVR were located on BTA 29 and BTA 27, respectively. Our finding is similar with the results of Letaief et al. (2017), where the highest and the lowest number of CNVRs were distributed on BTA1 and BTA25, respectively, and a

similar distribution was observed by Peripolli et al., 2023 in two of the three locally adapted taurine breeds.

Table 3. Statistics summary for the number of SVRs, length and proportion of chromosome extension covered by SVR in each of 29 autosomes.

Chr	Chr size (kb)	N	Length (Kb)				Chromosomal coverage rate (%)
			Mean	Sd	Max	Total	
1	158,530	261	49.02	339.27	4,661.89	12,794.97	8.07
2	136,230	211	117.22	622.57	6,620.64	24,732.87	18.16
3	121,010	177	73.18	316.3	2,458.36	12,952.17	10.7
4	120,000.6	196	68.83	390.78	4,462.06	13,491.11	11.24
5	120,090	207	66.77	328.65	3,102.13	13,821.45	11.51
6	117,810	211	35.11	195.18	2,080.17	7,408.53	6.29
7	110,680	190	71.92	315.48	2,552.96	13,664.86	12.35
8	113,320	164	65.34	410.55	3,885.85	10,715.99	9.46
9	105,450	165	103.22	500.86	4,633.04	17,031.79	16.15
10	103,310	117	108.21	450.42	2,693.91	12,660.45	12.25
11	106,980	108	192.82	742.09	4,852.18	20,824.23	19.47
12	87,220	149	93.76	455.91	3,494.42	13,970.59	16.02
13	83,470	106	87.24	492.17	4,396.62	9,247.47	11.08
14	82,400	108	91.65	469.05	3,919.73	9,898.47	12.01
15	85,010	177	82.34	344.39	2,771.55	14,574.51	17.14
16	81,010	113	41.7	242.81	2,378.82	4,711.80	5.82
17	73,170	105	49.56	242.82	1,657.28	5,203.79	7.11
18	65,820	92	116.46	380.62	2,993.72	10,714.26	16.28
19	63,450	67	99.1	424.04	3,213.91	6,640.01	10.46
20	7,1970	122	124.22	794.4	8,377.91	15,154.75	21.06
21	69,860	91	157.09	738.07	4,906.43	14,294.91	20.46
22	60,770	67	88.96	429.23	3,086.26	5,960.54	9.81
23	52,500	90	83.86	476.29	4,393.12	7,547.36	14.38
24	62,320	88	117.19	413.8	2,532.45	10,312.71	16.55
25	42,350	47	46.95	192.26	1,019.71	2,206.56	5.21
26	51,990	74	171.39	655.8	3,876.00	12,682.51	24.39
27	45,610	72	114.33	501.51	3,832.58	8,231.99	18.05
28	45,940	79	130.32	489.32	2,704.58	10,295.38	22.41
29	51,100	98	159.21	912.09	8,530.07	15,602.19	30.53

SVR: structural variant region; Chr: chromosome; N: number of SVR; Sd: standard deviation; max: maximum; Total: sum of all SVRs length; Chromosomal coverage rate (%): proportion of chromosome extension covered by SVR.

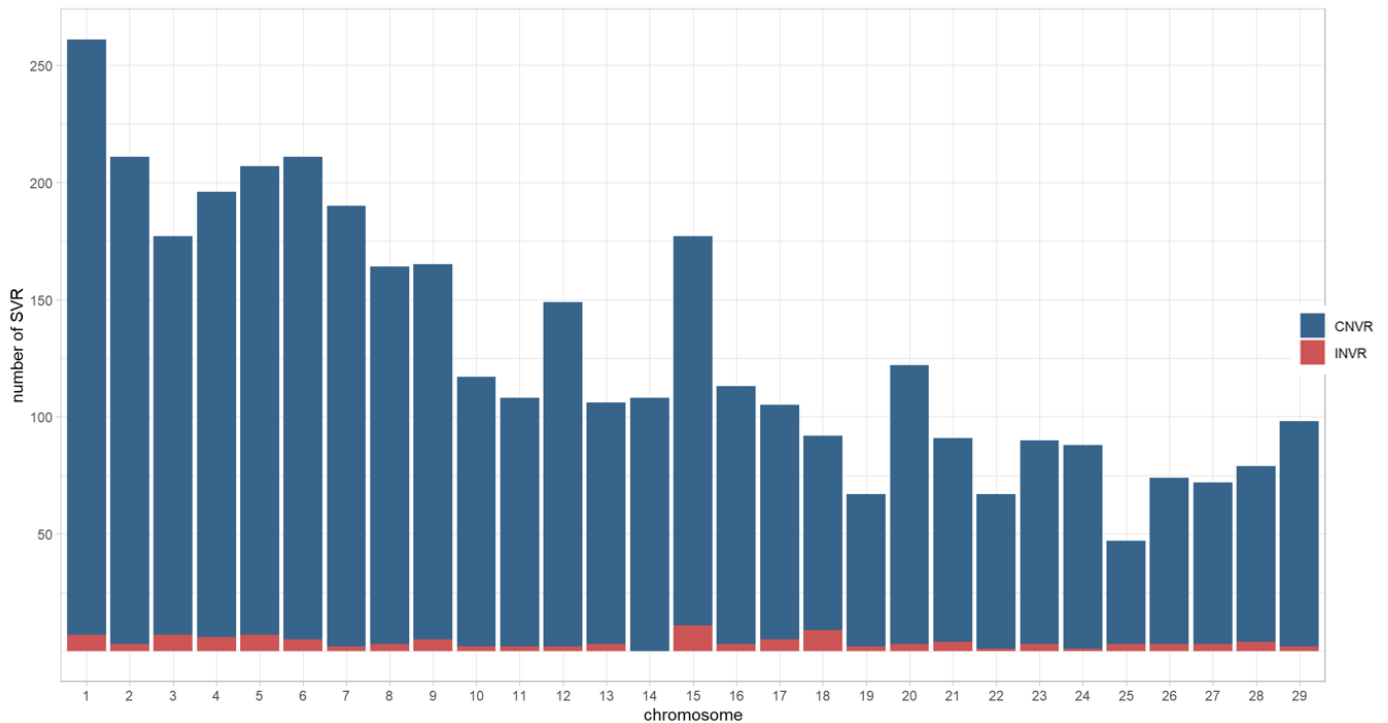


Figure 12. Distribution of non-redundant SVR frequent in more than 5% of population across the 29 autosomes. CNVR: regions of copy number variants; INVR: regions of inversions.

Regarding SVRs composed by inversion events, BTA15 and BTA18 harbored the highest number of INVR, with 11 and 9 INVR, respectively, while BTA22 and BTA24 carried the lowest number of INVR (N=1 INVR). INVR were distributed along the autosomes, except the BTA14 which did not harbor any INVR (Table S3, Figure S2).

4.1.2 Average length and autosomal genome extension covered by SVRs

The 3,752 SVRs, exhibit an average length of 89.91Kb (± 469.09) and a median of 2.08 Kb. The average length of INVR is 396.44 Kb (± 951.03 Kb) (Table 3). Regarding CNVRs, the length tend is similar with the single events, in which region of deletions exhibit the highest number but the shortest length (25.95 ± 232.35 Kb), followed by regions of duplications (36.47 ± 91.23 Kb), and the regions of mixed events ($799 \pm 1,244$ Kb) which represent the longest CNVR type (Table 3).

The longest SVR found has a length of 8,530.07 Kb, and the join of all SVRs length comprises a total of 337,348.3 Kb which represents 13.12% of whole autosomal genome extension (attributed to CNVRs in 11.40% and INVR in 1.71%) (Table 4.)

Table 4. Statistics summary for the number, length, and genome coverage rate for each SVR type (present in at least 5% of bulls).

SVR TYPE	N	Length (Kb)					Chromosomal coverage rate (%)
		Mean	Sd	Median	Max	Total	
CNVR- DEL	3,286	25.92	232.35	1.87	4,852.18	85,188.62	3.31
CNVR-DUP	99	36.47	91.23	5.03	759.28	3,610.60	0.14
CNVR-MIXED	256	799.0	1,244.38	199.56	8,377.91	204,544.38	7.95
ALL CNVR	3,641	80.57	443.25	2.028	8,377.91	293,343.60	11.40
INVR	111	396.44	951.03	74.00	8,530.07	44,004.73	1.71
ALL SVR	3,752	89.91	469.09	2.08	8,530.07	337,348.3	13.12

SVR: region of structural variants; CNVR: regions of copy number variants; CNVR-DEL: region of copy number variants composed by deletions; CNVR-DUP: region of copy number variants composed by duplications; CNVR-MIXED: region of copy number variants composed by deletions and duplications; INVR: regions of structural variants corresponding to inversions; N: number of SVR; Sd: standard deviation; max: maximum; Total: sum of all SVRs length.

This result is in concordance with the previously reported by Lemos et al.(2018b) where the predicted regions with duplication and deletion events (CNVR) from SNP genotyping data, represented the 13% of the genome coverage. It is remarkably that the latter authors analyzed a Nellore population from which the analyzed bulls belong to, thus, similar results regarding the genome coverage of structural variations were expected. However, some of our findings have differed from the previously reported by these authors, mainly due to differences in number of analyzed animals, the data nature used, and the combination of different strategies and algorithms for structural variants calling.

The total genomic coverage of CNVR found in this study was higher than the reported coverage rate in other Nellore cattle populations. For example, Santana et al., (2016) observed a total coverage of CNVR of 9% for predictions made from SNP

data, and Silva et al. (2016) reported an autosomal coverage of CNVR of 1.5% for predictions made from SNP and NGS data. When comparing our results with the reported for other taurine and indicine breeds, our observed coverage rate was also higher than those from Bickhart et al. (2012) (2.1%), Bickhart et al. (2016) (3.1%), Keel et al., (2017) (6.7%), and Hu et al., (2020) (1.5%), and Sun et al., (0.18%), in which CNVRs reported were those shared by diverse cattle breeds including Nellore, Brahman, Gir, and different *Bos taurus* breeds.

Likewise, SV distribution according to size classes, most of the detected SVRs (73%) have lengths ranging from 1 to 5 Kb, 14% are 5-10 Kb-long, 4% are 10-50 Kb-long, 3% 50-200 Kb long, and 6% are larger than 200 Kb. (Table 5). CNVRs are mainly distributed into the classes 1-5 Kb and 5 to 10 Kb (Figure 13a), while INVR are more balanced distribution into three size classes (10 to 50 Kb, 50 to 200 and larger than 200, Figure 13b).

Table 5. Number and proportion of each type of detected SVR according to different size classes.

SIZE CLASS	SVR TYPE	N	Percentage (%)
1-5Kb	CNVR	2,699	71.93
	INVR	31	0.83
	ALL SVR	2,730	73
5-10Kb	CNVR	540	14.39
	INVR	4	0.11
	ALL SVR	544	14
10-50Kb	CNVR	132	3.52
	INVR	13	0.35
	ALL SVR	145	4
50-200 Kb	CNVR	92	2.45
	INVR	24	0.64
	ALL SVR	116	6
>200 Kb	CNVR	178	4.74
	INVR	39	1.04
	ALL SVR	217	6

SVR: structural variant regions; CNVR: copy number variation region; INVR: structural variant regions corresponding to inversion events; N: number of SVR; Proportion %: proportion of each SVR type in relation to the total number of detected SVR.

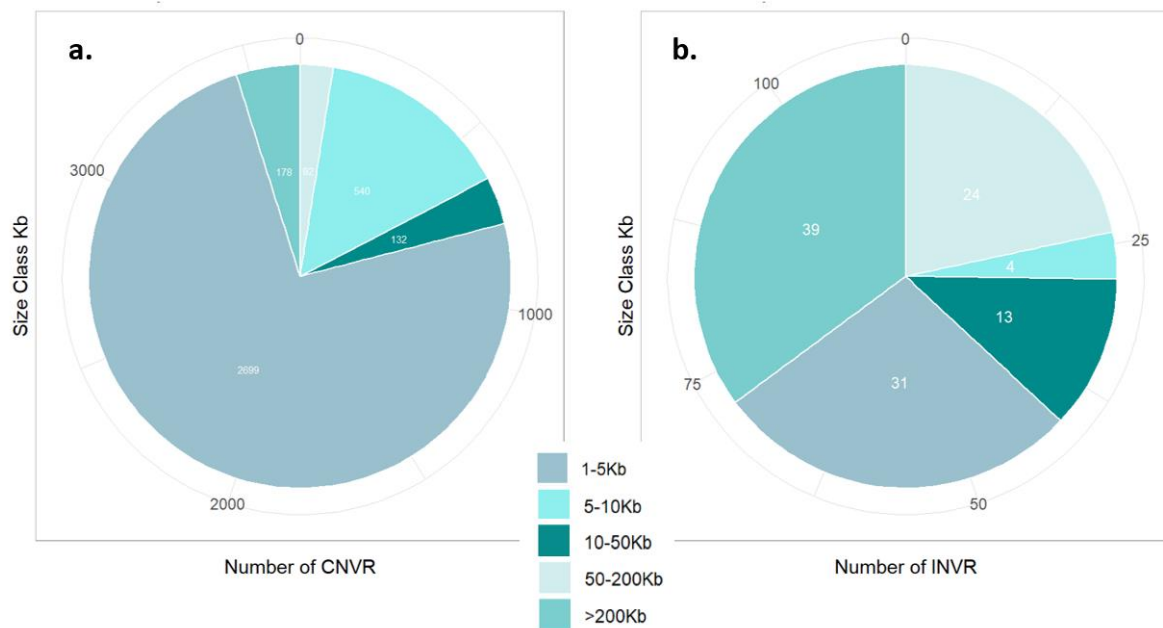


Figure 13. Number of regions of structural variants (SVR) according to different size classes. CNV: regions of copy number variants (left). INVR: regions of inversions (right).

The top five longest CNVRs are located on BTA11, 26, 21, 2 and 24, with average length ranging from 186.99 to 116.22 Kb (Figure S3a). Otherwise, the top five shortest CNVR are harbored by BTA6, 16, 25, 17, and 1, with average lengths ranging from 34.23 to 49.5 Kb. It is remarkable that although BTA 1 is ranked as first, regarding the number, the harbored CNVR are of short lengths. This supports the results of Pearson correlation test, which revealed that there is a trend that the higher the number of SVRs carried by each chromosome, the shorter the SVRs are (Pearson correlation coefficient $\rho = -0.37$, p value = 0.04976, Figure 14). By their side, the top five longest regions of inversions are carried by BTA29, 10, 7, 12, and 20, while the shortest INVR are located on BTA22, 2, 16, 1, and 6, respectively (Figure S4a).

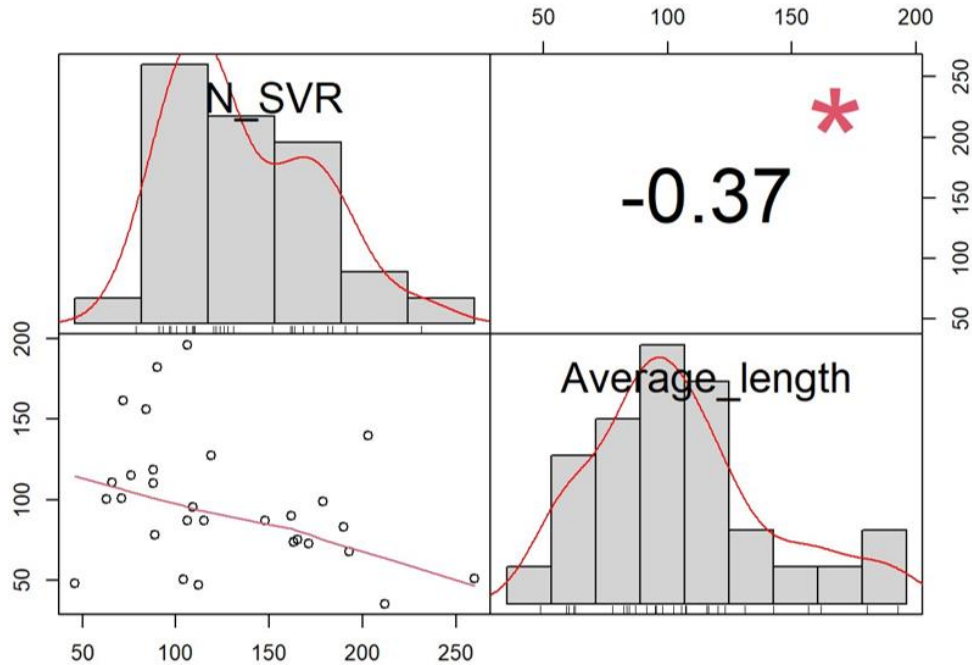


Figure 14. Pearson correlation between the number of non-redundant SVRs harbored by each chromosome and their average length.

Regarding the proportion of total chromosome extension that is covered by SVRs, the BTA29 is the most covered by regions with structural variation (chromosomal coverage rate of 30.53%), followed by chromosomes 26, 28, 20, and 21, with chromosomal coverage rates ranging from 22.46 to 24.39%. By the other side, BTA25 exhibits the lowest chromosomal coverage rate (5.21%) followed by BTA6, 16, 17, and 1, with chromosomal coverage rates ranging between 5.82 to 8.07%, (Figure 15, Table 3).

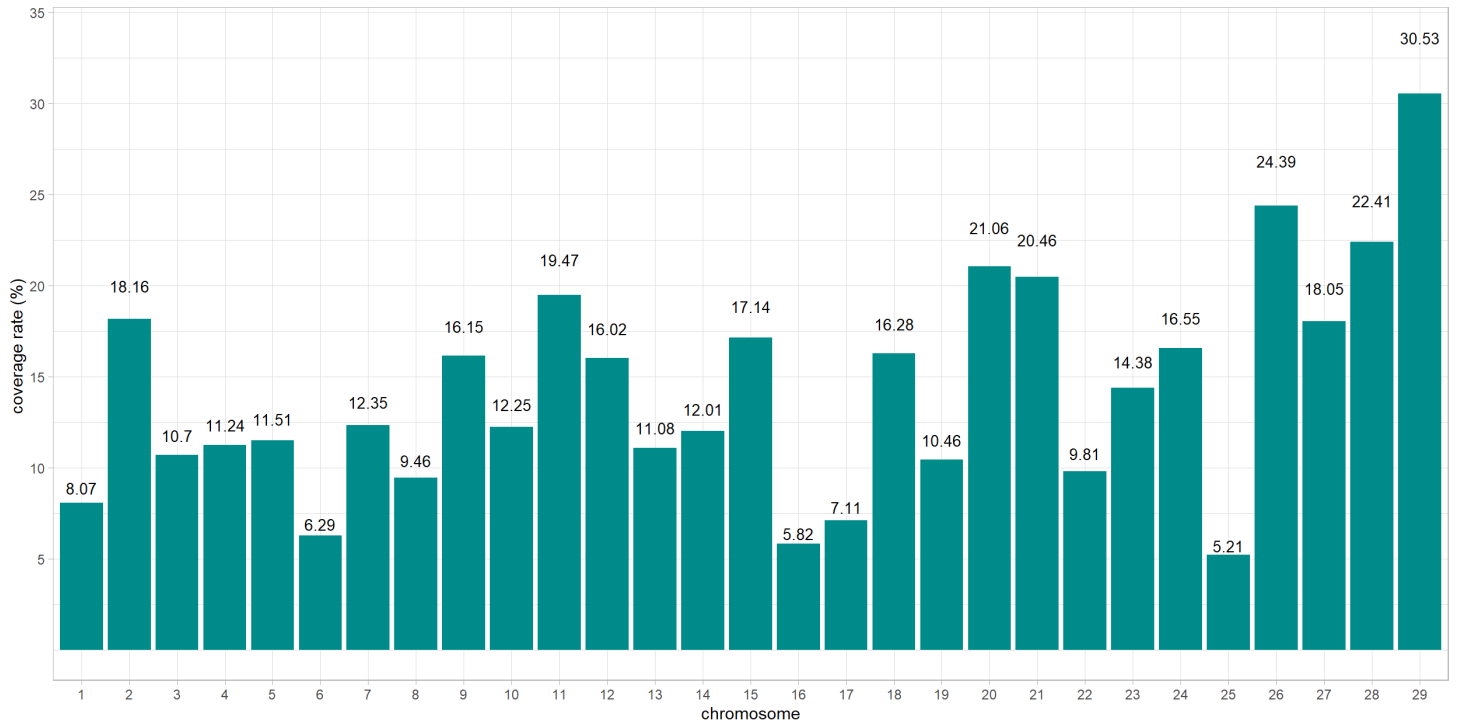


Figure 15. Chromosomal coverage rate: percentage (%) of chromosome extension covered by regions of structural variants (SVRs.)

BTA26 is the most covered chromosome by CNVR (24.34%), while BTA25 is the least covered (4.68%) (Figure S3b). The maximum percentage of chromosomal coverage is higher than maximum percentage reported by Lemos et al. (2018b) (24.26%) in Nellore cattle, and differ from their results regarding the chromosomal distribution of coverage rates, as the most covered chromosomes they observed were BTA19, 23 and 25, and the least covered by CNVR were BTA9, BTA13, and BAT8. The results also differ from the reported by Silva et al. (2016) in Nellore cattle, where CNVRs located on BTA24 were observed encompassing the greatest total extension, and CNVR on BTA19 comprises the least. In other hand, the greatest chromosomal coverage by regions of inversions is observed o BTA29, while BTA22 are badly covered by INVR (Figure S4b).

Although the BTA29 does not exhibit the highest number of SVRs and neither the SVRs located there have the longest length, but there are apparently two regions of inversions of large size that are responsible for that coverage rate. Furthermore,

BTA29 is the shorter one, that is why regions with structural variations encompasses a large proportion of its extension. Our results were coincident to those from Letaief et al., (2017) in French cattle, where chromosomes 25 and 1 were the less covered by structural variants.

Pearson correlation analysis revealed a significant, strong, and positive association between the number of SVRs and the chromosome size (Pearson correlation score $\rho = 0.93$, p value = 4.844×10^{-13} , Figure 16.), which is concurrent with the findings of Letaief et al., 2017 and Silva et al., 2016. In contrast, no-statistical evidence regarding correlations between the chromosome size and the proportion of chromosome extension covered by SVR (chromosomal coverage rate) was found in this study.

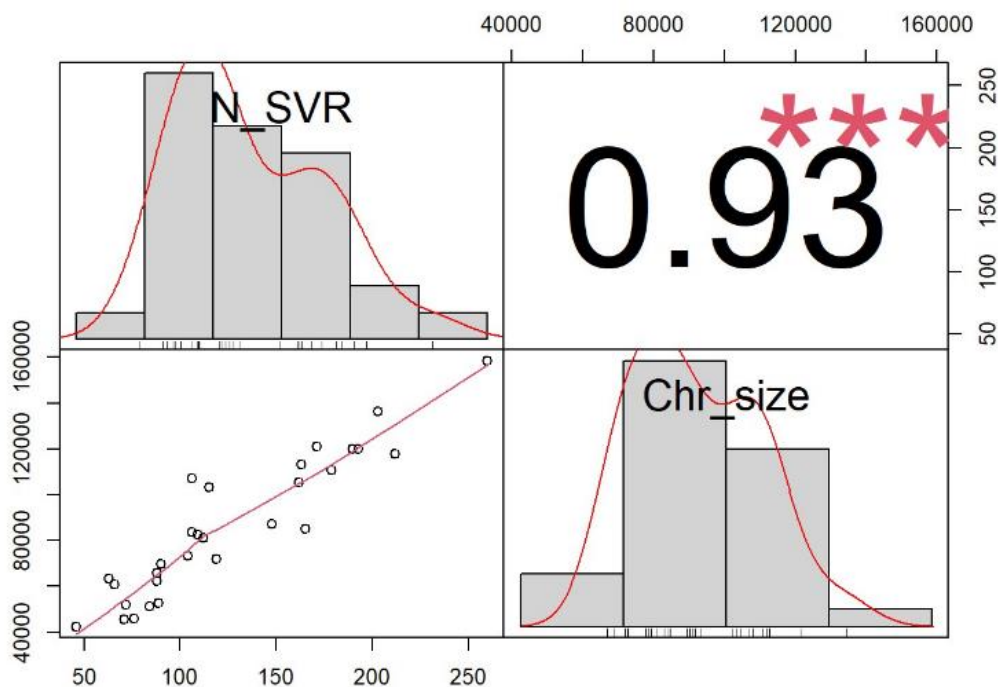


Figure 16. Pearson correlation between the number of non-redundant SVRs and the chromosome size. N: number of SVRs *** p value < 0.05.

4.3. Functional annotation and enrichment of SVR.

Most of the identified SVRs were common for less than 50% of bulls (Figure 18), but a considerable high number of SVRs were shared by more than half of the studied population (N=532 SVRs). To assess the functional impact on genetic mechanisms underlying Nellore cattle structural variations, genes and QTLs overlapping SVRs shared by more than 50% (N > 75 bulls) were retrieved from publicly available genomic dataset, as described in material and methods section. Figure 18 illustrates the chromosomal location of the most frequent SVRs (532 SVRs).

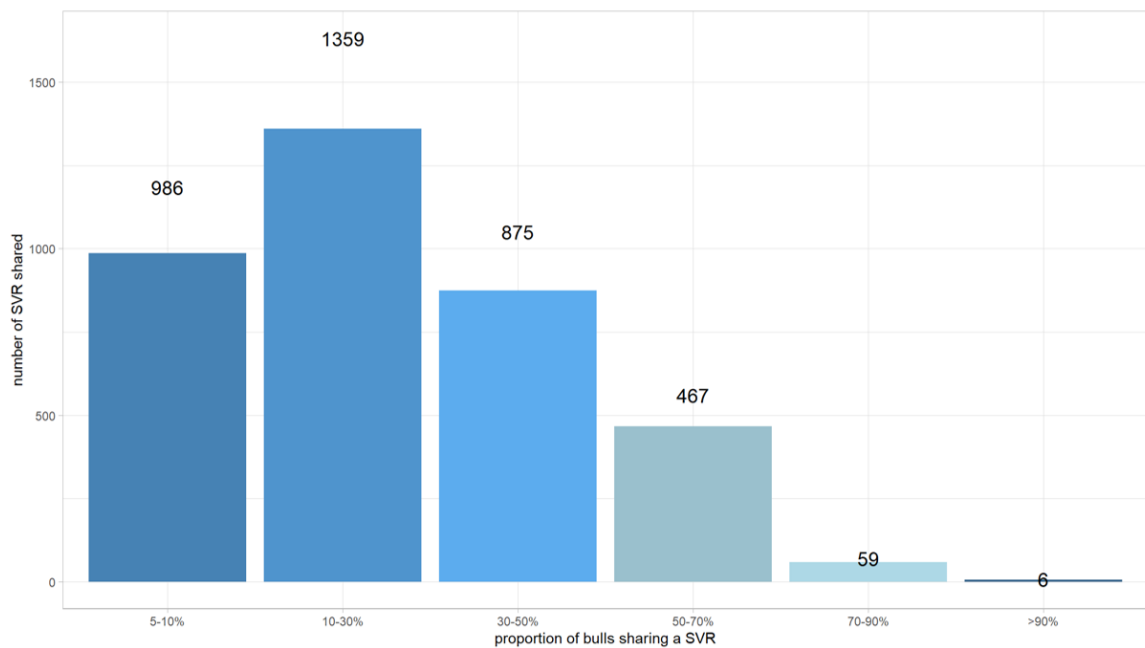


Figure 17. Distribution of the number of the regions of structural variants (SVR) of shared by different proportions of Bulls.

Overall, from the 532 most frequent SVRs, 96 overlapped a total of 130 QTLs, which were distributed in 6 QTL types: exterior, health, meat and carcass, milk, production, and reproduction, which are related to a total of 50 traits (Table S4). Figure 19 represents the number of SVRs overlapping QTLs, according to QTL type and trait associated.

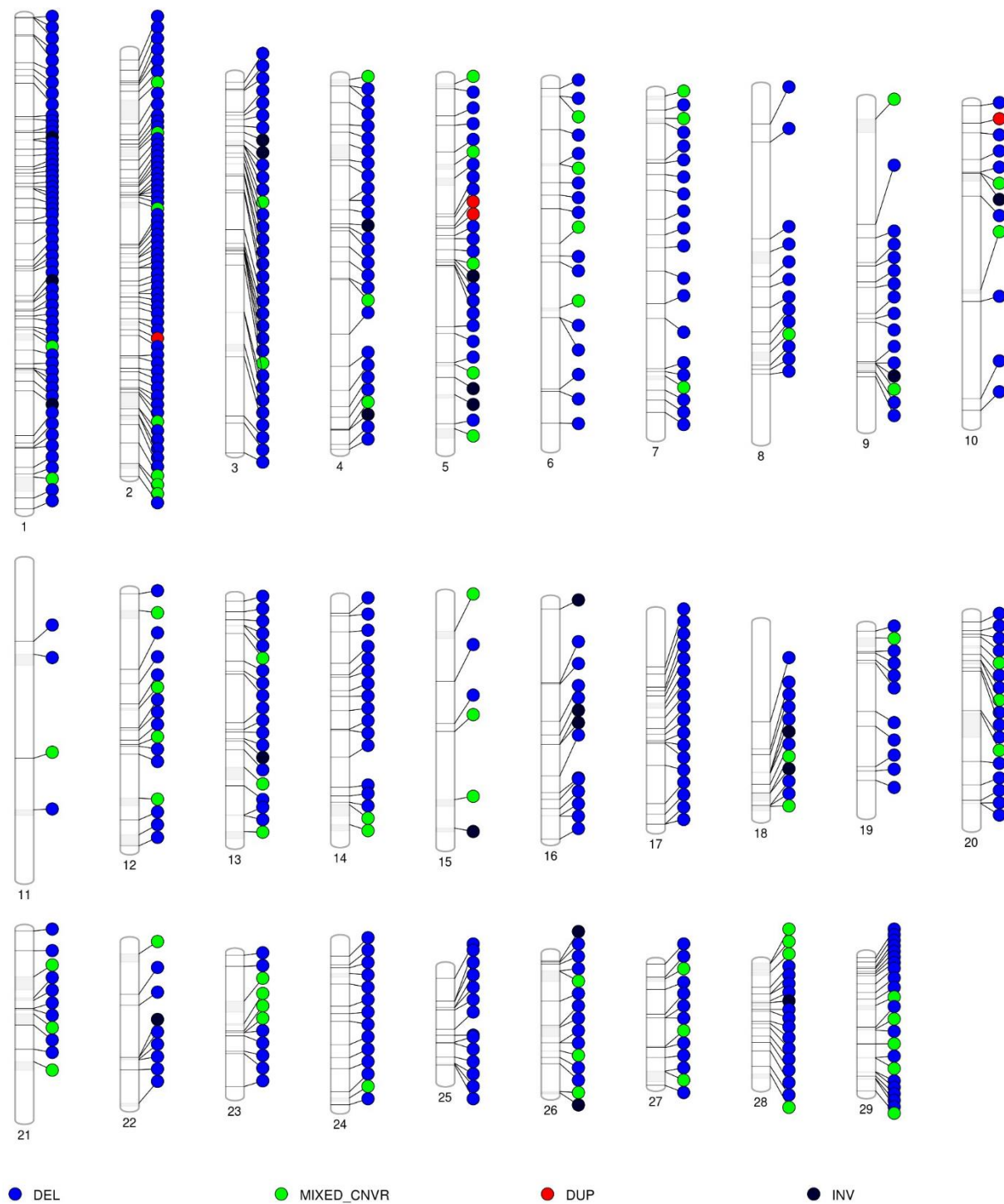


Figure 18. Chromosomal location of the most frequent regions of structural variants (SVR). Blue dots represent regions of deletions, green dots are mixed regions, red dots are the regions of duplications, and black dots the regions of inversions.

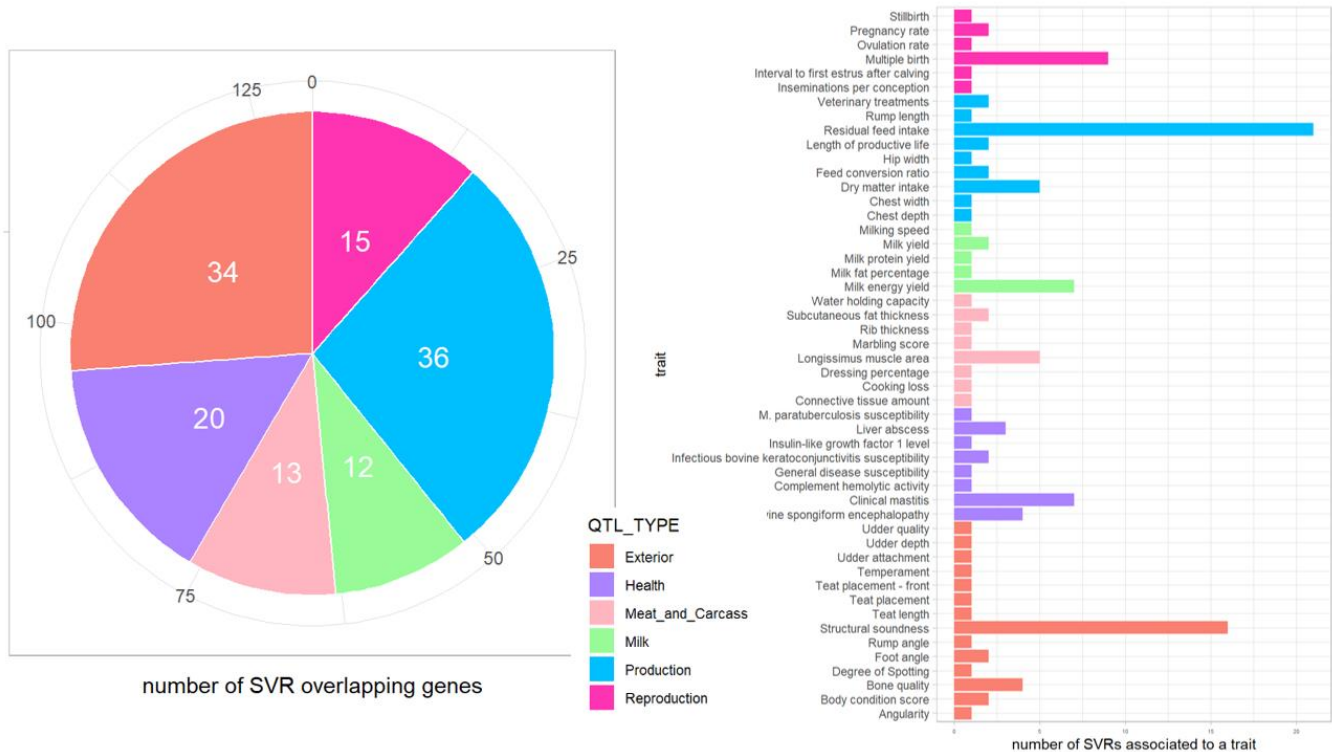


Figure 19. Distribution of SVRs overlapping QTLs according to QTL type and trait associated.

The highest number of SVRs overlapped QTLs related to residual feed intake (21 SVRs), followed by structural soundness (16 SVRs), multiple birth (9 SVRs), clinical mastitis, and milk energy yield (7 SVRs each one). The SVRs overlapping QTLs were distributed in all chromosomes except on BTA24. The BTA6 was the chromosome harboring the largest number of QTLs overlapping SVRs (11 QTLs), followed by BTA7 (8 QTLs) and BTA3, 4, 5 and 23 with 7 QTL each one. In addition, the region named as CNVR1799 (BTA6: 72924569-74154761) that was present in 75 bulls overlapped a total of 6 QTLs, distributed in two QTL types: production (residual feed intake), and exterior (teat placement, foot angle, udder quality, udder depth, and degree of spotting). Regarding the INVR, there were 4 regions that overlapped QTLs: INVR27 (BTA1:39477267-39530842) overlapping a QTL related to structural soundness, INVR131 (BTA3:20943678-22580427) overlapping a QTL related to milk yield, INVR549 (BTA15:78781811-80138873) overlapping a QTL

related to liver abcess, and INVR557 (BTA16:3117468-3121435) overlapping a QTL related to structural soundness. Most of 96 SVRs were linked to just one QTL.

Regarding the gene annotation, from the 532 most frequent SVRs, a total of a 205 SVRs overlapped known genes (Table S5). A total of 1,164 positional candidate genes surrounding the SVRs were annotated. The most annotated region of copy number variants named as CNVR3428 (BTA13: 57357647-57361337) harbored a total of 84 genes, followed by the region CNVR4347 (BTA18:61867768-61878669) which harbored 58 genes, and the regions, CNVR1601 (BTA5:112687401-112706629), CNVR4532 (BTA20: 33443255-33514932), and CNVR4703 (BTA21: 31126645-31188035), harboring 51 genes each one (Table S5).

From the SVR that overlapped positional candidate genes, there were in total 7 regions of inversions, from which the named as INVR549 (BTA15:78970795-78971743) was the most annotated harboring 44 genes, followed by INVR131 (BTA3: 20943678-22580427) that harbored 24 positional candidate genes (Table S5). All the SVRs carrying known genes were distributed in all chromosomes, and the largest number of annotated genes was harbored by BTA21 (116 genes), followed by BTA5, 2, 13 and 18 with 109, 105, 100 and 78 annotated genes, respectively.

PANTHER overrepresentation analysis through Binomial test and Bonferroni correction method showed that from the 1,164 positional candidate genes that overlapped the most frequent SVRs, 695 were significantly enriched for three gene ontology categories and one biochemical pathway (Figure 20). The three gene ontologies were biological process (BP) with 632 genes grouped into 15 significant GO terms (Table S6), molecular function (MF) with 139 genes grouped into 4 significant GO terms (Table S7), and cellular component (CC) with 236 genes grouped into 2 significant GO terms (Table S8). In addition, 26 genes were significantly associated to a biochemical pathway (Table S9).

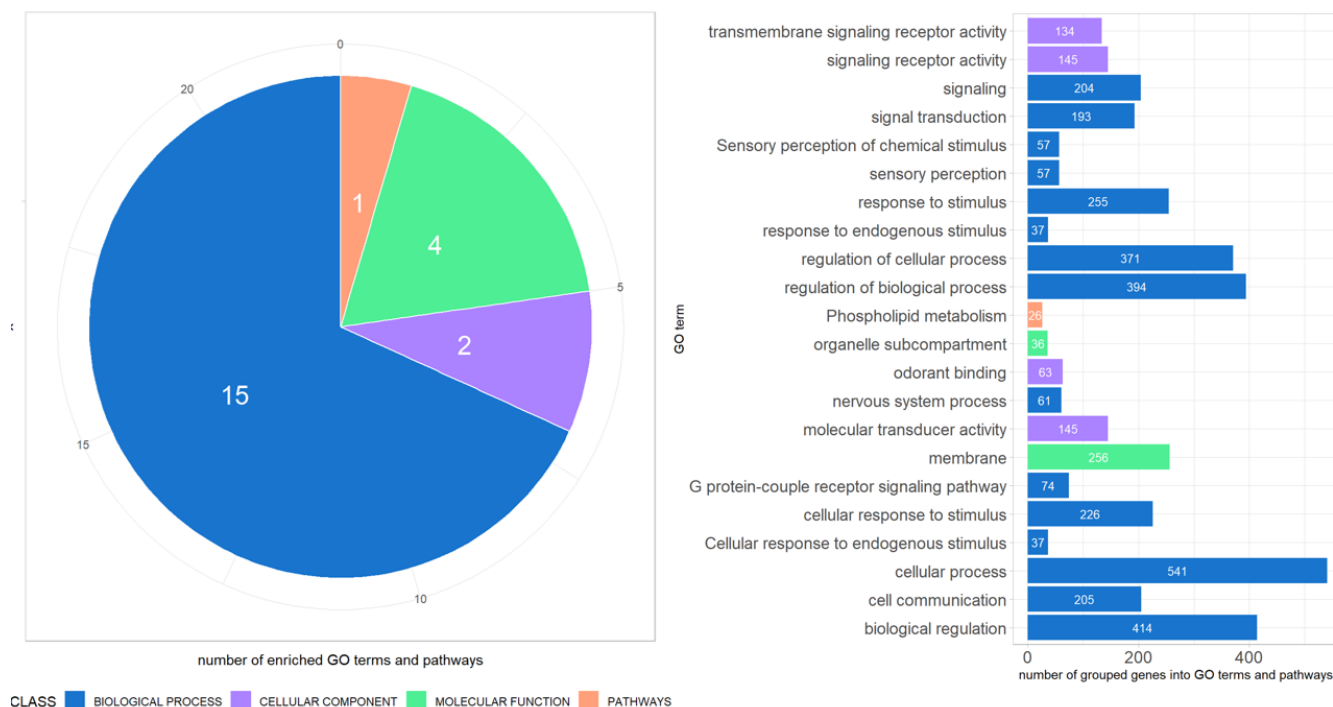


Figure 20. Results from PANTHER enrichment analysis. Biological Process and Molecular Function related GO terms.

Regarding Biological Process related GO terms, most of the enriched genes were significantly associated to cellular process (GO:0009987, N=49), followed by biological regulation (GO:0065007, N=386), regulation of biological process (GO:0050789, N=369), regulation of cellular process (GO:0050794, N=352), and response to stimulus (GO:0050896, N=236). Enriched genes were also associated to biological process related to cell signaling and response, including cell communication (GO:0007154, N=195), signaling (GO:0023052, N=194), signaling transduction (GO:0007165, N=184), G protein-couple receptor signaling pathway (GO:0007186, N=72), nervous system process (GO:0050877, N=61), sensory perception of chemical stimulus (GO:0007606, N=57), sensory perception (GO:0007600, N=57), cellular response to stimulus (GO:0051716, N=31), cellular response to endogenous stimulus (GO:0071495, N=32), and response to endogenous stimulus (GO:0009719, N=31) (Table S6).

For Molecular Function, 63 genes were significantly associated to odorant binding (GO:0005549), 129 to transmembrane signaling receptor activity (GO:0004888), 130 to signaling receptor activity (GO:0038023), and 139 to molecular transducer activity (GO:0060089) (Table S7). With respect to cellular component, 32 genes were significantly associated with organelle sub compartment (GO:0031984), and 230 to membrane (GO:0016020) (Table S8). Finally, a total of 26 genes were significantly associated to the phospholipid metabolism pathway (Table S9).

It is outstanding that from the enriched set, there are several genes (91 genes) belonging to the olfactory receptor (OR) family, which overlapped SVRs located on BTA3, 5, 7, 15, 23 and 29. Most of these OR genes are harbored by regions of inversions, where 39 OR genes are located into the region named as INVR549 (BTA15:78781811-80138873; Table S5) that is shared by 76 bulls and overlaps a health type QTL (liver abscess, Table S4). The regions INVR131 (BTA3:20943678-22580427; Table S5), and INVR240 (BTA5:58727321-58939413; Table S5) harbored 2 OR genes each one, and are shared by 83 and 86 bulls, respectively. Additionally, INVR131 overlaps a milk type QTL related to milk yield (Table S4). The other enriched OR genes are mainly harbored by copy number regions of mixed events (deletions and duplications), such as the region CNVR1946 (BTA7:8714162-10096757; Table S5), that harbored 19 OR genes. This region is common for 101 bulls, and overlaps a production type QTL related to residual feed intake (Table S4). Similarly, 15 OR genes are located into the region CNVR5606 (BTA29:27137256-27598509; Table S5) that is observed in 87 bulls, and 7 OR genes are carried by the region CNVR4958 (BTA23:29208078-29318128; Table S5) common for 76 bulls. Other 5 enriched OR genes overlapped the mixed CNV region CNVR1465 (BTA5:58631317-58908758; Table S5) which is located nearby a reproduction QTL related to multiple birth, and that is common for 102 bulls. The *OR5M10C* gene is located into the region CNVR4931 (BTA23:17751563-22144687; Table S5) which overlaps two reproduction QTLs (multiple birth, and Interval to first estrus after calving; Table S4), and one health type QTL related to health complement hemolytic

activity (Table S4). Finally, the gene *OR5AE3* is harbored by a deletion region named as CNVR2043 (BTA7:40993138-41026889) that is common for 82 bulls.

Some authors have reported significant associations of SNPs located in the vicinity of OR genes with economically important traits in beef cattle. For example, the gene *OR5M10* which was observed in this study within the region INVR549, was reported by Higgins et al., (2018) nearby the SNP rs136789347, which was significantly associated with the average daily gain (ADG) in Irish beef cattle breeds. Here, the *OR5M10* gene was significantly enriched for 12 of the 15 GO terms related to Biological Process and for all GO terms related to Molecular Function (Table S6 and S7). Seabury et al., (2017) also reported significant associations of SNPs surrounding OR genes with average daily gain and residual feed intake in three different U.S beef cattle breeds, and in Nellore cattle, Olivieri et al., (2016) found 3 OR genes harbored by a genomic region located on BTA15 that explained more than 1% of additive genetic variance of dry matter intake (DMI).

The mixed region CNVR1465 (BTA5:58631317-58908758) carried the genes *OR6C1*, *OR6C1N* and *OR6C1S*, which were also found by Santa et al. (2018) overlapping CNV regions significantly associated with feed conversion rate in Nellore cattle, in a genotyped-based study. Those genes were significantly enriched for the MF GO terms transmembrane signaling receptor activity (GO:0004888), signaling receptor activity (GO:0038023), and molecular transducer activity (GO:0060089) (Table S7). The mentioned mixed CNV region, overlapped a reproduction QTL related to multiple birth trait. Although most of studies found in literature reveal the functional role of OR gene family on feed efficiency related traits, some genes belonging to the gene family have also been associated to reproductive performance in cattle (Melo et al., 2017). OR genes are thought to be involved in sperm chemotaxis and in the development of female germ cells and oocyte (Driedich et al., 2012).

Similarly, the gene set *OR7A112*, *OR7A78*, *OR7A95*, *OR7A97* and *OR7A99* harbored by the mixed region CNVR1946, were also reported by Braga et al. (2023) overlapping CNVRs in Brazilian dairy Gyr cattle, from SNP and NGS data. In this study, those genes were significantly associated to 8 of the 15 BP related GO

terms, to 3 MF related GO terms, and to the CC GO term membrane (GO:0016020). Among those genes, the *OR795* gene was also reported by Rafter et al., (2021) as a flanked gene into a deletion region significantly associated with meat carcass traits in Charolais, Holstein and Limousin cattle.

The olfactory receptor genes comprise the largest superfamily in the mammalian genome. They belong to the G-protein-coupled chemoreceptors (GPCRs), and are essentially expressed on countless olfactory sensory neurons within the nasal cavity (Connor et al., 2018; Olender et al., 2008). These protein coding genes play roles in the reception of a wide range of chemical molecules in the environment (K. Lee et al., 2013). Odorant mechanisms are essential for animals as they send signals to the brain in order to locate feed sources, danger, or potential mating (Lee et al., 2013). Besides the nasal cavity, olfactory receptors can express in other organs where they bind to nutrients and metabolites to trigger physiological responses such as changes in gut motility, ventilation rate and cellular migration (Connor et al., 2018). The activity of OR is mediated by the central nervous system and is influenced by hormonal, nutritional, and metabolic mechanisms regulating energy balance that trigger behaviors such as feed searching and intake (Julliard et al., 2017; Fleisher et al., 2015; Palouzier-Paulignan et al., 2012).

The structural conformation of OR gene family has been well studied in several vertebrate species. There is evidence showing that OR gene family have evolved due to duplication mechanisms which has guarantee their maintenance and survival in the vertebrates' genomes (Bickhart & Liu, 2014; Niimura & Nei, 2007). Lee et al. (2013) observed a large variation in OR genes copy number in Korean Holstein and Angus cattle. The authors found via PCR that 40% of the tested OR loci exhibited CNVs, which suggest that copy number variations of OR genes in cattle are highly frequent and that their genetic diversity confers differences in olfaction capacity and then, in fitness. Although duplications have been described as the main mechanism driving maintenance of the OR genes, it results particularly interesting that regions of inversions detected in this study harbored such a great number of genes belonging to the olfactory receptor superfamily. Due to the functional relevance of OR genes for adaptation and fitness, it is plausible that under

selection pressure the desirable alleles for OR genes were protected against recombination for being widely located on regions of inversion events, thus playing important roles in the adaptation of Nellore to environmental conditions.

Another important gene family found overlapping the most frequent SVR is the B defensin gene family (*DEFB*), from which 6 genes (*DEFB1*, *DEFB10*, *DEFB119*, *DEFB124*, *DEFB13*, and *DEFB7*) were significantly associated to 11 biological process GO terms, mainly related to cellular communication, signaling and response to stimulus (Table S6). Defensins are protein coding genes that have key functions in innate immune system of multicellular organisms. Besides of playing important roles in antimicrobial activity, these molecules act in diverse biological mechanisms such as cell signaling, ion channel transport, toxic functions, and enzyme inhibitor activities in animals (Islam et al., 2023). As well as olfactory receptor genes, the immunity-related genes are known to be more likely duplicated than other genes. Our results are in concordance to those from Butty et al., (2020). These authors observed members of the *DEFB* family into their predicted CNV regions in a Holstein population from genotypes and NGS data. Similarly, in a population of Holstein and Jersey cattle, Chen et al. (2017) identified that genes belonging to expanded gene families such as *DEFB* were significantly more associated to SV regions than other genes. Particularly, the gene *DEFB1*, that was found within the mixed copy number region named as CNVR5333 (27: 6182772-7283835), was also reported in the two mentioned studies. The *DEFB1* were significantly related to GO term cellular response to endogenous stimulus, signaling receptor activity and molecular transducer activity.

Important genes in the vicinity of regions of structural variants, have been previously reported in literature for their association with economically important phenotypes. For example, members of the secretory phospholipase A2 family of genes (*PLA2G2A*, *PLA2G2D1*, *PLA2G2E*, *PLA2G5*), which were harbored by the mixed copy number region named as CNVR717 (BTA2:132530986-133810156), were also found into CNV regions significantly associated to meat carcass traits in Nellore cattle by Berton et al., 2021. By other side, four members of Adhesion G Protein-Coupled Receptors (*ADGRF1*, *ADGRF2*, *ADGRF4*, and *ADGRF5*), were

harbored by the mixed region CNVR4931 (BTA23:17751563-22144687), which were linked to one health QTL related to complement hemolytic activity, and to a two reproduction QTLs related to multiple birth, and interval to first estrus after calving (Table S4). Those genes were previously reported by J. Li et al., (2021), for being significantly associated to feed efficiency indicator traits in Canadian crossbred cattle, such as residual feed intake, daily dry matter intake, average daily gain, and metabolic daily gain. A large set of genes belonging to the zinc fingers binding protein family were also found nearby the regions of structural variants. Particularly the mixed regions CNVR4338 (BTA18:57180983-60174708), and CNVR4347 (BTA18:61815370-63400235) harbored various members of that gene family, such as the genes *ZNF524*, *ZNF581* and *ZNF677*, which were previously associated with dry matter intake, average daily gain, and feed efficiency in an experimental population of Nellore cattle by Olivieri et al., (2016). Zinc finger proteins usually regulate gene expression by binding to specific DNA sequences via their DNA-binding zinc finger domains and recruiting cofactors that mediate chromatin remodeling and transcriptional repression or activation. In a transcriptomic study in *Bos indicus* cattle, Samuel & Dinka, (2020) observed that ZNF genes are likely involved in the regulation of expression of olfactory receptor genes, thus ZNF genes may also play roles in cell signaling and communication through the sense of olfaction, which is essential for environmental adaption.

Our findings suggest that genes harbored by the regions of structural variants are more involved in functions related to biological and cell regulation, response to stimulus, and cell communication and signaling. Those functions are essential drivers of environmental adaptation of genomes. The reported literature in Nellore cattle, supports that most of the described genes are potential candidates for feed efficiency indicator traits, which reflects the biochemical mechanisms in which they are involved that have led to improve fitness.

5. FINAL CONSIDERATIONS.

The ongoing developments on NGS-based algorithms have allowed highly confident predictions of structural variants. Here, we detected and assessed deletion, duplication, and inversion events at individual and population level, providing valuable information regarding the structural diversity of the Nellore cattle genome. This contributes to unravel highly prevalent structural variants from the whole genome sequences of a large set of representative ancestors of Brazilian Nellore population. Particularly, this can be the first ongoing study reporting and analyzing mobile elements such as inversions in Nellore cattle, which still lacks information.

Using complementary bioinformatic tools and applying exigent filtering criteria within and between software and sample, we detected a total of 215,031 high-confidence structural variants corresponding to 183,032 deletions-DEL, and 14,013 duplications-DUP, and 17,986 inversions (INV). Moreover, 3,752 non-redundant SVR that were frequent in more than 5% of bulls were identified, corresponding in more than 97% to regions of copy number variants (CNVR), and 3% to regions of inversions (INVR). From CNVR, 90% corresponded to regions of deletions (CNVR-DEL), 3% to regions of duplications (CNVR-DUP) and 7% to mixed regions, in which occur both duplication and deletion events (CNV-MIXED). Among the SVR, 532 were shared by more than 50% of the analyzed key ancestors, and harbored 1,164 positional candidate genes and QTLs with known functional implications in cattle.

Our results contribute to the unravelling of genetic mechanisms driving environmental adaptation of Nellore cattle, caused by structural changes and rearrangements of functionally important genes. As an example, the genes belonging to the highly preserved gene families, as the highlighted olfactory receptors-OR, the defensins, the secretory phospholipase A2 family, the adhesion G protein-coupled receptors, and the zinc fingers binding proteins, that overlapped genomic regions in which different structural variation events are occurring, and which have been mainly described in the literature for being potential candidates for feed efficiency indicator traits.

6. REFERENCES

- Abyzov, A., Urban, A.E., Snyder, M., & Gerstein, M. (2011). CNVnator : An approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Research*, 21, 974–984. <https://doi.org/10.1101/gr.114876.110.974>
- Alkan, C., Coe, B. P., & Eichler, E. E. (2011b). Genome structural variation discovery and genotyping. *Nature Reviews Genetics*, 12(5), 363–376. <https://doi.org/10.1038/nrg2958>
- Almeida Santana, M. H., Junior, G. A. O., Cesar, A. S. M., Freua, M. C., da Costa Gomes, R., da Luz e Silva, S., Leme, P. R., Fukumasu, H., Carvalho, M. E., Ventura, R. V., Coutinho, L. L., Kadarmideen, H. N., & Ferraz, J. B. S. (2016). Copy number variations and genome-wide associations reveal putative genes and metabolic pathways involved with the feed conversion ratio in beef cattle. *Journal of Applied Genetics*, 57(4), 495–504. <https://doi.org/10.1007/s13353-016-0344-7>
- Antunes de Lemos, M. V., Berton, M. P., Ferreira de Camargo, G. M., Peripolli, E., de Oliveira Silva, R. M., Ferreira Olivieri, B., Cesar, A. S. M., Pereira, A. S. C., de Albuquerque, L. G., de Oliveira, H. N., Tonhati, H., & Baldi, F. (2018). Copy number variation regions in Nellore cattle: Evidences of environment adaptation. *Livestock Science*, 207, 51–58. <https://doi.org/10.1016/j.livsci.2017.11.008>
- Arlt, M. F., Wilson, T. E., & Glover, T. W. (2012a). Replication stress and mechanisms of CNV formation. *Current Opinion in Genetics & Development*, 22(3), 204–210. <https://doi.org/10.1016/j.gde.2012.01.009>
- Arlt, M. F., Wilson, T. E., & Glover, T. W. (2012b). Replication stress and mechanisms of CNV formation. *Current Opinion in Genetics and Development*, 22(3), 204–210. <https://doi.org/10.1016/j.gde.2012.01.009>
- Berton, M. P., de Antunes Lemos, M. V., Chud, T. C. S., Stafuzza, N. B., Kluska, S., Amorim, S. T., ... & Baldi, F. (2021). Genome-wide association study between copy number variation regions and carcass-and meat-quality traits in Nellore cattle. *Animal Production Science*, 61(8), 731-744
- 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., Tassell, C. P. Van, Sonstegard, T. S., Eichler, E. E., Liu, G. E., & Hughes, H. (2012). Copy number variation of individual cattle genomes using next-generation sequencing. *Genome Research*, 22, 778–790. <https://doi.org/10.1101/gr.133967.111.778>

- Bickhart, D. M., & Liu, G. E. (2014). The challenges and importance of structural variation detection in livestock. In *Frontiers in Genetics* (Vol. 5, Issue FEB). Frontiers Research Foundation. <https://doi.org/10.3389/fgene.2014.00037>
- Chen, L., Chamberlain, A. J., Reich, C. M., Daetwyler, H. D., & Hayes, B. J. (2017). Detection and validation of structural variations in bovine whole-genome sequence data. *Genetics Selection Evolution*, 49(1). <https://doi.org/10.1186/s12711-017-0286-5>
- Connor, E. E., Zhou, Y., & Liu, G. E. (2018). The essence of appetite: Does olfactory receptor variation play a role? In *Journal of Animal Science* (Vol. 96, Issue 4, pp. 1551–1558). American Society of Animal Science. <https://doi.org/10.1093/jas/sky068>
- Conrad, D. F., Pinto, D., Redon, R., Feuk, L., Gokcumen, O., Zhang, Y., Aerts, J., Andrews, T. D., Barnes, C., Campbell, P., Fitzgerald, T., Hu, M., Ihm, C. H., Kristiansson, K., Macarthur, D. G., Macdonald, J. R., Onyiah, I., Wing, A., Pang, C., ... Hurles, M. E. (2010). Origins and functional impact of copy number variation in the human genome. *Nature*, 464(7289), 704–712. <https://doi.org/10.1038/nature08516>
- Couldrey, C., Keehan, M., Johnson, T., Tiplady, K., Winkelman, A., Littlejohn, M. D., & Scott, A. (2017). Detection and assessment of copy number variation using PacBio long-read and Illumina sequencing in New Zealand dairy cattle. *Journal of Dairy Science*, 100(7), 5472–5478. <https://doi.org/10.3168/jds.2016-12199>
- Couldrey, C., Keehan, M., Johnson, T., Tiplady, K., Winkelman, A., Littlejohn, M. D., Scott, A., Kemper, K. E., Hayes, B., Davis, S. R., & Spelman, R. J. (2017). Detection and assessment of copy number variation using PacBio long-read and Illumina sequencing in New Zealand dairy cattle. *Journal of Dairy Science*, 100(7), 5472–5478. <https://doi.org/10.3168/jds.2016-12199>
- Dobigny, G., Britton-Davidian, J., & Robinson, T. J. (2017). Chromosomal polymorphism in mammals: an evolutionary perspective. *Biological Reviews*, 92(1), 1–21. <https://doi.org/10.1111/brv.12213>
- Durán Aguilar, M., Román Ponce, S. I., Ruiz López, F. J., González Padilla, E., Vásquez Peláez, C. G., Bagnato, A., & Strillacci, M. G. (2017). Genome-wide association study for milk somatic cell score in holstein cattle using copy number variation as markers. *Journal of Animal Breeding and Genetics*, 134, 49–59. <https://doi.org/10.1111/jbg.12238>
- Escaramís, G., Docampo, E., & Rabionet, R. (2015). A decade of structural variants: description, history, and methods to detect structural variation. *Briefings in functional genomics*, 14(5), 305–314.

- Fernandes Júnior, G.A., de Oliveira, H.N., Carneiro, R. *et al.* Whole-genome sequencing provides new insights into genetic mechanisms of tropical adaptation in Nelore (*Bos primigenius indicus*). *Sci Rep* **10**, 9412 (2020). <https://doi.org/10.1038/s41598-020-66272-7>
- Fleischer, J., R. Bumbalo, V. Bautze, J. Strotmann, and H. Breer. 2015. Expression of odorant receptor Olfr78 in enteroendocrine cells of the colon. *Cell Tissue Res.* 361:697–710. doi:10.1007/s00441-015-2165-0
- Fonseca, P. A. S., Suárez-Vega, A., Marras, G., & Cánovas, Á. (2020). GALLO: An R package for genomic annotation and integration of multiple data sources in livestock for positional candidate loci. *GigaScience*, 9(12). <https://doi.org/10.1093/gigascience/giaa149>
- Freeman, J. L., Perry, G. H., Feuk, L., Redon, R., McCarroll, S. A., Altshuler, D. M., Aburatani, H., Jones, K. W., Tyler-smith, C., Hurles, M. E., Carter, N. P., Scherer, S. W., & Lee, C. (2006). Copy number variation: New insights in genome diversity. *Genome Research*, 16(617), 949–961. <https://doi.org/10.1101/gr.3677206.16>
- Georges, M., Charlier, C., & Hayes, B. (2019). Harnessing genomic information. *Nature Reviews Genetics*, 20, 135–156. <https://doi.org/10.1038/s41576-018-0082-2>
- Goddard, M. E., & Hayes, B. J. (2009). Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nature Reviews Genetics*, 10(6), 381–391. <https://doi.org/10.1038/nrg2575>
- Hastings, P. J., Lupski, J. R., Rosenberg, S. M., & Ira, G. (2009). Mechanisms of change in gene copy number. *Nature Reviews Genetics*, 10(AuguST), 551–564. <https://doi.org/10.1038/nrg2593>
- Hayes, B., & Goddard, M. (2010). Genome-wide association and genomic selection in animal breeding. *Genome*, 53, 876–883. <https://doi.org/10.1139/G10-076>
- Hayes, B. J., Bowman, P. J., Chamberlain, A. J., & Goddard, M. E. (2009). Invited review : Genomic selection in dairy cattle : Progress and challenges. *Journal of Dairy Science*, 92(2), 433–443. <https://doi.org/10.3168/jds.2008-1646>
- Henrique, V., Correia, L., Regitano, D. A., Geistlinger, L., Morosini, S., Zimmer, R., & Coutinho, L. L. (2016). Genome-Wide Detection of CNVs and Their Association with Meat Tenderness in Nelore Cattle. *PLOS ONE*, 11, 1–22. <https://doi.org/10.1371/journal.pone.0157711>
- Higgins, M. G., Fitzsimons, C., McClure, M. C., McKenna, C., Conroy, S., Kenny, D. A., McGee, M., Waters, S. M., & Morris, D. W. (2018). GWAS and eQTL analysis identifies a SNP associated with both residual feed intake and

- GFRA2 expression in beef cattle. *Scientific Reports*, 8(1).
<https://doi.org/10.1038/s41598-018-32374-6>
- Ho, S. S., Urban, A. E., & Mills, R. E. (2020). Structural variation in the sequencing era. In *Nature Reviews Genetics* (Vol. 21, Issue 3, pp. 171–189). Nature Research. <https://doi.org/10.1038/s41576-019-0180-9>
- 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(1).
<https://doi.org/10.1186/1471-2164-13-376>
- Hu, Y., Xia, H., Li, M., Xu, C., Ye, X., Su, R., Zhang, M., Nash, O., Sonstegard, T. S., Yang, L., Liu, G. E., & Zhou, Y. (2020). Comparative analyses of copy number variations between *Bos taurus* and *Bos indicus*. *BMC Genomics*, 21(1).
<https://doi.org/10.1186/s12864-020-07097-6>
- Islam, S., Akhand, M. R. N., & Hasan, M. (2023). Evolutionary trend of bovine β -defensin proteins toward functionality prediction: A domain-based bioinformatics study. *Heliyon*, 9(3).
- Julliard, A. K., D. Al Koborssy, D. A. Fadool, and B. Palouzier- Paulignan. 2017. Nutrient sensing: another chemosensitivity of the olfactory system. *Front. Physiol.* 8:468. doi:10.3389/fphys.2017.00468
- Keel, B. N., Keele, J. W., & Snelling, W. M. (2017). Genome-wide copy number variation in the bovine genome detected using low coverage sequence of popular beef breeds. *Animal Genetics*, 48(2), 141–150.
- Kim, D., Langmead, B., & Salzberg, S. L. (2015). *HISAT: a fast spliced aligner with low memory requirements*. 12(4), 357–360.
<https://doi.org/10.1038/nmeth.3317>.HISAT
- Lee, K., Nguyen, D. T., Choi, M., Cha, S. Y., Kim, J. H., Dadi, H., Seo, H. G., Seo, K., Chun, T., & Park, C. (2013). Analysis of cattle olfactory subgenome: The first detail study on the characteristics of the complete olfactory receptor repertoire of a ruminant. *BMC Genomics*, 14(1). <https://doi.org/10.1186/1471-2164-14-596>
- Lee, Y.-L., Costa, G., Moreira, M., Karim, L., Druet, T., Oget-Ebrad, C., Veerkamp, R. F., Groenen, M. A. M., Georges, M., & Bouwman, A. C. (2022). *High-resolution structural variation catalogue in a large-scale whole genome sequenced bovine family cohort data*. <https://doi.org/10.21203/rs.3.rs-2007387/v1>
- Lemos, M. V., Peripolli, E., Berton, M. P., Loise, F., Feitosa, B., Olivieri, B. F., Stafuzza, N. B., Tonussi, R. L., Kluska, S., Lucas, H., Chiaia, J., Mueller, L.,

- Ferrinho, A. M., Simone, A., Prereira, C., Oliveira, H. N. De, Albuquerque, L. G. De, & Baldi, F. (2018). Association study between copy number variation and beef fatty acid profile of Nellore cattle. *Animal Frontiers*, 59, 203–223. <https://doi.org/https://doi.org/10.1007/s13353-018-0436-7>
- Letaief, R., Rebours, E., Grohs, C., Meersseman, C., Fritz, S., Trouilh, L., Esquerré, D., Barbieri, J., Klopp, C., Philippe, R., Blanquet, V., Boichard, D., Rocha, D., & Boussaha, M. (2017a). Identification of copy number variation in French dairy and beef breeds using next - generation sequencing. *Genetics Selection Evolution*, 49(77), 1–15. <https://doi.org/10.1186/s12711-017-0352-z>
- Letaief, R., Rebours, E., Grohs, C., Meersseman, C., Fritz, S., Trouilh, L., Esquerré, D., Barbieri, J., Klopp, C., Philippe, R., Blanquet, V., Boichard, D., Rocha, D., & Boussaha, M. (2017b). Identification of copy number variation in French dairy and beef breeds using next - generation sequencing. *Genetics Selection Evolution*, 49(77), 1–15. <https://doi.org/10.1186/s12711-017-0352-z>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Li, J., Mukiibi, R., Wang, Y., Plastow, G. S., & Li, C. (2021). Identification of candidate genes and enriched biological functions for feed efficiency traits by integrating plasma metabolites and imputed whole genome sequence variants in beef cattle. *BMC Genomics*, 22(1). <https://doi.org/10.1186/s12864-021-08064-5>
- Martínez-Fundichely, A., Casillas, S., Egea, R., Ràmia, M., Barbadilla, A., Pantano, L., Puig, M., & Cáceres, M. (2014). InvFEST, a database integrating information of polymorphic inversions in the human genome. *Nucleic Acids Research*, 42(D1). <https://doi.org/10.1093/nar/gkt1122>
- Melo, T. P., Camargo, G. M. F., Albuquerque, L. G., & Carneiro, R. (2017). Genome-wide association study provides strong evidence of genes affecting the reproductive performance of Nellore beef cows. *PLOS ONE*, 12(5), 1–14. <https://doi.org/10.1371/journal.pone.0178551>
- 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., ... Peterson, J. L. (2011). Mapping copy number variation by population-scale genome sequencing. *Nature*, 470(7332), 59–65. <https://doi.org/10.1038/nature09708>

- Montaldo, H. H., & Meza-herrera, C. A. (1998). Use of molecular markers and major genes in the genetic improvement of livestock. *Electronic Journal of Biotechnology*, 1(2), 1–7.
- Niimura, Y., & Nei, M. (2007). Extensive gains and losses of olfactory receptor genes in mammalian evolution. *PLoS ONE*, 2(8).
<https://doi.org/10.1371/journal.pone.0000708>
- Olender, T., Lancet, D., & Nebert, D. W. (2008). Update on the olfactory receptor (OR) gene superfamily. *Human Genomics*, 3(1), 1479–7364.
<https://doi.org/10.1186/1479-7364-3-1-87>
- Olivieri, B. F., Mercadante, M. E. Z., Cyrillo, J. N. D. S. G., Branco, R. H., Bonilha, S. F. M., De Albuquerque, L. G., De Oliveira Silva, R. M., & Baldi, F. (2016). Genomic regions associated with feed efficiency indicator traits in an experimental nellore cattle population. *PLoS ONE*, 11(10).
<https://doi.org/10.1371/journal.pone.0164390>
- Ornitz, D.M., Itoh, N. Fibroblast growth factors. *Genome Biol* 2, reviews3005.1 (2001). <https://doi.org/10.1186/gb-2001-2-3-reviews3005>
- Palouzier-Paulignan, B., M. C. Lacroix, P. Aimé, C. Baly, M. Caillol, P. Congar, A. K. Julliard, K. ucker, and D. A. Fadool. 2012. Olfaction under metabolic influences.
- Peterson, B. G., Carl, P., Boudt, K., Bennett, R., Ulrich, J., Zivot, E., ... & Wuertz. D. (2014). PerformanceAnalytics: Econometric tools for performance and risk analysis. *R package version*. 1(3).
- Puig, M., Casillas, S., Villatoro, S., & Cáceres, M. (2015). Human inversions and their functional consequences. *Briefings in Functional Genomics*, 14(5), 369–379. <https://doi.org/10.1093/bfgp/elv020>
- Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6), 841–842.
<https://doi.org/10.1093/bioinformatics/btq033>
- Radke, D. W., & Lee, C. (2015). Adaptive potential of genomic structural variation in human and mammalian evolution. *Briefings in functional genomics*, 14(5), 358–368.
- Rafter, P., Gormley, I. C., Purfield, D., Parnell, A. C., Naderi, S., & Berry, D. P. (2021). Genome-wide association analyses of carcass traits using copy number variants and raw intensity values of single nucleotide polymorphisms in cattle. *BMC Genomics*, 22(1). <https://doi.org/10.1186/s12864-021-08075-2>
- Rasekh, M. E., Chiatante, G., Miroballo, M., Tang, J., Ventura, M., Amemiya, C. T., Eichler, E. E., Antonacci, F., & Alkan, C. (2017). Discovery of large genomic

inversions using long range information. *BMC Genomics*, 18(1).

<https://doi.org/10.1186/s12864-016-3444-1>

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., Global variation in copy number in the human genome. *Nature*. 2006 Nov 23;444(7118):444-54. doi: 10.1038/nature05329. PMID: 17122850; PMCID: PMC2669898.

Salmon Hillbertz N.H.C., Isaksson M., Karlsson E.K., Hellmén E., Pielberg G.R., Savolainen P., et al. Duplication of FGF3, FGF4, FGF19 and ORAOV1 causes hair ridge and predisposition to dermoid sinus in Ridgeback dogs. *Nat Genet*. 2007;39:1318–20.

Samuel, B., & Dinka, H. (2020). In silico analysis of the promoter region of olfactory receptors in cattle (*Bos indicus*) to understand its gene regulation.

Nucleosides, Nucleotides and Nucleic Acids, 39(6), 853–865.

<https://doi.org/10.1080/15257770.2020.1711524>

Seabury, C. M., Oldeschulte, D. L., Saatchi, M., Beever, J. E., Decker, J. E., Halley, Y. A., Bhattarai, E. K., Molaei, M., Freetly, H. C., Hansen, S. L., Yampara-Iquise, H., Johnson, K. A., Kerley, M. S., Kim, J. W., Loy, D. D., Marques, E., Neibergs, H. L., Schnabel, R. D., Shike, D. W., ... Taylor, J. F. (2017). Genome-wide association study for feed efficiency and growth traits in U.S. beef cattle. *BMC Genomics*, 18(1). <https://doi.org/10.1186/s12864-017-3754-y>

Silva, J. M., Giachetto, P. F., da Silva, L. O., Cintra, L. C., Paiva, S. R., Yamagishi, M. E. B., & 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(1). <https://doi.org/10.1186/s12864-016-2752-9>

Singh, U., Deb, R., Rahman, R., Alex, R., Kumar, S., Chakraborty, S., Dhama, K., & Sharma, A. (2014). Molecular markers and their applications in cattle genetic research: A review. *Biomarkers and Genomic Medicine*, 6(2), 49–58. <https://doi.org/10.1016/j.bgm.2014.03.001>

Stranger, B. E., Forrest, M. S., Dunning, M., Ingle, C., Beazley, C., Thorne, N., Redon, R., Bird, C. P., Grassi, A., Lee, C., Tyler-Smith, C., Carter, N., Scherer, S. W., Tavaré, S., Deloukas, P., Hurles, M., & Dermitzakis, E. T. (2014).

- Relative Impact of Nucleotide and Copy Number Variation on Gene Expression Phenotypes. *Science*, 315(2007), 848–853.
<https://doi.org/10.1126/science.1136678>
- Sun, T., Pei, S., Liu, Y., Hanif, Q., Xu, H., Chen, N., Lei, C., & Yue, X. (2023). Whole genome sequencing of simmental cattle for SNP and CNV discovery. *BMC Genomics*, 24(1). <https://doi.org/10.1186/s12864-023-09248-x>
- Syvänen, A. (2001). Accessing genetic variation: genotyping single nucleotide polymorphisms. *Nature Reviews Genetics*, 2(December), 930–942.
- Teo, S. M., Pawitan, Y., Ku, C. S., Chia, K. S., & Salim, A. (2012). Statistical challenges associated with detecting copy number variations with next-generation sequencing. In *Bioinformatics* (Vol. 28, Issue 21, pp. 2711–2718).
<https://doi.org/10.1093/bioinformatics/bts535>
- Wellenreuther, M., & Bernatchez, L. (2018). Eco-Evolutionary Genomics of Chromosomal Inversions. In *Trends in Ecology and Evolution* (Vol. 33, Issue 6, pp. 427–440). Elsevier Ltd. <https://doi.org/10.1016/j.tree.2018.04.002>
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., ... Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, 4(43), 1686. <https://doi.org/10.21105/joss.01686>
- Wickham, Hadley. (2009). *Ggplot2: elegant graphics for data analysis*. Springer.
- Yudav, A. K., Tomar, S. S., Jha, A. K., & Singh, J. (2017). Importance of Molecular Markers in Livestock Improvement: A Review. *International Journal of Agriculture Innovations and Research*, 5(4).
- Zarate, S., Carroll, A., Mahmoud, M., Krasheninina, O., Jun, G., Salerno, W. J., Schatz, M. C., Boerwinkle, E., Gibbs, R. A., & Sedlazeck, F. J. (2021). Parliament2: Accurate structural variant calling at scale. *GigaScience*, 9(12), 1–9. <https://doi.org/10.1093/gigascience/giaa145>
- 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, 451–481. <https://doi.org/10.1146/annurev.genom.9.081307.164217>
- 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.
<https://doi.org/doi:10.1186/1471-2105-14-S11-S1>
- Zhou, Y., Utsunomiya, Y. T., Xu, L., Hamidi, E., Bickhart, D. M., Alexandre, P. A., Rosen, B. D., Schroeder, S. G., Carvalheiro, R., Henrique, H., Neves, D. R.,

Sonstegard, T. S., Tassell, C. P. Van, Bento, J., Ferraz, S., 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, 1–9.
<https://doi.org/10.1186/s12864-016-2461-4>

7. SUPPLEMENTARY MATERIAL

7.1 Supplementary tables.

Table S1. Statistics summary for the number of SVs, length and proportion of autosomal genome extension covered by SV for each the analyzed bulls.

ID	N	Length (Kb)				Autosomal genome coverage (%)
		Mean	Sd	Max	Total	
Nel01	2,988	140.03	473.94	4,906.42	418,409.85	16.23
Nel02	2,564	70.48	308.18	3,380.01	180,719.43	7.03
Nel03	2,882	78.95	302.89	2,519.49	227,543.61	8.85
Nel04	2,953	171.42	544.48	3,908.36	506,197.27	19.68
Nel05	3,012	133.54	473.01	4,906.42	402,234.79	15.64
Nel06	2,893	178.82	561.8	4,906.42	517,330.49	20.11
Nel07	3,026	79.13	331.45	4,906.41	239,433.81	9.31
Nel08	2,240	123.1	432.23	2,925.56	275,747.46	10.72
Nel09	2,759	72.69	320.92	4,250.87	200,556.29	7.79
Nel10	2,386	35.16	184.76	4,906.41	83,900.22	3.26
Nel11	2,442	77.24	340.42	4,906.42	188,620.14	7.3
Nel12	2,895	73.77	312.71	2,841.34	401,706.87	15.62
Nel13	2,825	52.54	268.56	3,885.85	353,408.29	13.74
Nel14	2,652	138.32	453.78	4,852.18	232,215.95	9.03
Nel15	2,987	49.61	266.4	4,852.18	280,478.59	10.90
Nel16	2,698	96.83	382.01	4,393.12	320,275.68	12.45
Nel17	2,175	116.64	455.49	3,908.37	74,876.35	2.91
Nel18	2,937	96.03	395.72	3,908.36	279,295.84	10.86
Nel19	2,143	96.4	332.26	3,885.85	158,098.36	6.15
Nel20	2,448	152.2	489.2	4,906.41	128,613.31	5
Nel21	3,048	96.9	387.95	4,906.41	421,594.74	16.39
Nel22	1,355	143.42	469.96	4,906.41	67,228.30	2.61
Nel23	2,098	66.96	289.98	4,906.41	203,155.25	7.89
Nel24	2,529	90.96	346.74	4,906.415	294,977.94	11.47
Nel25	2,810	40.38	176.15	4,906.41	269,830.59	10.49
Nel26	3,173	70.37	293.11	2,498.15	305,861.58	11.89
Nel27	3,424	93.87	354.32	4,396.62	521,128.19	20.26
Nel28	2,200	60.61	223.58	4,906.42	213,181.37	8.28
Nel29	2,969	8.47	38.92	4,916.08	425,806.22	16.55
Nel30	2,708	82.01	334.7	4,906.41	181,338.05	7.05
Nel31	2,823	35.58	179.34	3,885.85	256,768.64	9.98
Nel32	1,373	96.46	330.21	2,349.13	55,439.56	2.15
Nel33	2,170	31.73	197.4	4,906.42	152,706.85	5.94
Nel34	2,868	56.54	264.89	4,250.87	269,229.99	10.47

Nel35	358	73.48	348.08	3,380.01	21,699.68	0.84
Nel36	89	62.84	318.81	367.67	753.48	0.03
Nel37	2,136	48.54	256.29	3,380.01	175,178.94	6.81
Nel38	2,062	73.2	339.9	4,784.88	73,372.54	2.85
Nel39	1,777	43.99	293.28	2,925.56	171,416.29	6.67
Nel40	1,878	64.23	340.33	4,906.42	59,597.25	2.32
Nel42	2,235	67.03	254.87	3,380.01	126,371.71	4.91
Nel44	2,064	97.99	335.91	4,393.12	151,659.96	5.89
Nel45	1,352	21.87	137.34	4,906.42	84,957.65	3.30
Nel46	1,755	47.57	229.83	3,899.00	85,190.99	3.31
Nel47	1,884	27.91	193.21	4,906.42	137,915.81	5.36
Nel48	780	71.31	318.76	4,906.42	34,308.69	1.33
Nel49	1,780	32.83	143.67	4,784.8	114,331.07	4.44
Nel50	1,793	50.72	271.45	2,530.84	120,185.19	4.67
Nel51	2,328	61.69	230.86	2,925.56	228,113.61	8.87
Nel53	534	12.68	48.58	2,260.52	11,679.73	0.45
Nel54	451	79.17	316.11	2,925.56	21,453.29	0.83
Nel56	1,324	70.99	310.28	3,380.01	36,952.01	1.44
Nel57	2,234	80.7	289.79	4,906.41	159,297.40	6.19
Nel58	580	75.38	297.13	1,467.67	19,043.18	0.74
Nel59	2,072	76.01	404.53	4,906.41	105,084.65	4.08
Nel60	2,042	92.14	356.77	3,908.36	125,979.13	4.89
Nel61	446	70.4	349.31	451.45	5,653.91	0.22
Nel62	1,326	24.54	127.99	4,906.42	104,981.96	4.08
Nel63	2,718	42.62	278.34	4,784.85	192,949.66	7.50
Nel64	1,440	91.93	378.29	3,380.01	116,201.35	4.52
Nel65	1,515	26.06	170.76	4,393.12	114,194.05	4.44
Nel66	413	71.24	299.73	4,906.41	31,390.16	1.22
Nel67	1,670	5.17	20.34	3,908.36	153,867.50	5.98
Nel68	2,256	108.3	431.42	4,906.42	158,813.19	6.17
Nel69	553	46.51	254.08	1,467.64	13,570.37	0.53
Nel70	1,828	33.57	193.37	4,784.88	77,915.95	3.03
Nel71	1,879	15.38	78.77	4,829.21	172,729.34	6.72
Nel72	440	59.09	270.29	2,410.09	11,465.21	0.45
Nel73	1,934	20.02	90.55	3,380.01	137,774.24	5.36
Nel74	388	57.61	311.03	367.67	2,004.87	0.08
Nel75	1,462	89.36	397.18	4,906.42	158,327.38	6.15
Nel76	2027	70.97	319.68	4,906.42	94,267.55	3.67
Nel77	1,815	76.58	308.66	4,906.42	60,923.61	2.37
Nel78	461	64.17	346.22	1,197.50	7,088.86	0.28
Nel79	1,892	57.71	267.65	4,906.42	111,799.05	4.35
Nel80	555	69.06	308.86	1,059.84	11,109.46	0.43
Nel81	552	136.56	440.24	2,195.13	31,798.32	1.24
Nel82	426	38.27	246.19	2,174.61	38,066.68	1.48

Nel83	1,715	34.25	206.79	4,906.42	121,714.85	4.73
Nel84	1,211	20.51	154.11	2,925.56	92,743.08	3.61
Nel85	798	95.03	375.61	4,146.59	51,209.07	1.99
Nel86	1,670	103.07	456.98	2,500.29	96,380.50	3.75
Nel87	2,087	108.1	418.78	4,906.42	144,130.87	5.60
Nel88	995	177.4	541.31	4,906.42	135,881.34	5.28
Nel89	1,279	91.19	294.85	4,906.42	48,949.43	1.90
Nel90	1,632	167.86	491.19	3,040.32	55,899.15	2.17
Nel91	1,208	419.48	814.23	2,925.56	24,779.44	0.96
Nel92	2,306	94.92	375.38	4,906.42	219,135.32	8.52
Nel93	177	102.01	454.51	2,498.21	18,243.02	0.71
Nel94	770	107.75	416.91	2,475.82	83,237.77	3.24
Nel96	1,036	176.23	539.12	3,932.79	183,782.91	7.14
Nel97	958	91	294.26	2,500.67	87,360.01	3.39
Nel98	760	167.78	490.87	3,932.79	127,577.25	4.96
Nel99	944	417.86	812.9	4,906.42	395,991.13	15.39
Nel100	706	56.87	263.7	2,139.94	40,147.92	1.56
Nel101	874	114.71	453.29	4,784.87	100,254.86	3.89
Nel102	111	6.9	34.97	367.67	765.5	0.03
Nel103	623	58.7	301.56	3,932.79	36,571.66	1.42
Nel104	866	149.18	453.96	3,380.01	129,193.57	5.02
Nel105	1,190	110.09	470.85	4,906.42	131,006.66	5.09
Nel106	968	146.49	536.09	4,906.42	141,802.27	5.513
Nel107	728	239.4	656.68	4,906.42	174,286.82	6.78
Nel108	1,062	33.83	184.24	2,337.18	35,929.29	1.39
Nel109	1,141	88.52	361.74	2,500.29	101,002.38	3.93
Nel110	1,692	43.26	232.18	4,906.42	73,194.07	2.85
Nel111	277	218.28	603.75	3,380.01	60,463.38	2.35
Nel112	937	53.93	283	2,494.73	50,534.61	1.96
Nel113	680	162.93	440.79	2,548.20	110,790.07	4.31
Nel114	1,506	75.17	302.2	2,475.91	113,204.54	4.40
Nel115	575	136.63	535.95	4,906.41	78,564.40	3.05
Nel116	1,717	62.52	320.67	4,906.41	107,340.62	4.173
Nel117	146	8.86	32.1	367.68	1,293.20	0.05
Nel118	1,763	74.46	336.87	4,146.63	131,268.35	5.10
Nel119	636	77.81	283.05	2,394.25	49,484.06	1.92
Nel120	505	92.06	342.43	4,784.88	46,490.66	1.81
Nel121	579	115.55	440.36	4,906.42	66,901.96	2.60
Nel122	466	201.78	576.65	4,784.88	94,027.42	3.65
Nel123	1,441	75.36	346.14	4,784.88	108,598.89	4.22
Nel124	1,360	46.68	244.06	2,406.78	63,489.57	2.47
Nel127	1,391	66.81	359.96	4,784.881	92,933.32	3.61
Nel128	164	5.77	14.21	126.62	946.98	0.04
Nel129	1,279	47.74	277.4	4,784.88	61,064.10	2.37

Nel130	1,068	73.23	366.94	4,906.42	78,21	3.04
Nel131	1,264	32.81	202.1	3,851.91	41,472.12	1.61
Nel132	626	47.28	280.6	3,119.16	29,600.13	1.15
Nel133	876	75.32	372.42	4,906.42	65,976.61	2.56
Nel134	546	169.8	590.21	4,906.42	92,712.58	3.60
Nel135	1,537	104.3	454.83	4,784.89	160,304.79	6.23
Nel136	190	98.03	322.99	2,551.04	18,625.31	0.72
Nel137	563	10.51	37.76	371.02	5,916.75	0.23
Nel138	186	30.33	350.67	4,784.86	5,641.70	0.22
Nel139	68	7.11	17.06	102.17	483.26	0.02
Nel140	394	15.88	106.23	1,338.08	6,258.44	0.24
Nel141	1,371	84.16	381.91	4,906.42	115,386.03	4.49
Nel142	1,240	65.65	348.02	4,906.42	81,410.86	3.16
Nel143	483	56.64	416.59	4,804.52	27,356.52	1.06
Nel144	551	120.98	418.46	4,784.88	66,657.93	2.59
Nel145	397	216.27	700.13	4,989.55	85,860.41	3.34
Nel146	808	68.58	346.11	4,906.42	55,414.69	2.15
Nel147	574	101.41	417.25	4,784.88	58,210.14	2.26
Nel148	791	85.84	404.83	4,784.88	67,899.52	2.64
Nel149	424	247.93	642.91	4,906.42	105,121.67	4.09
Nel151	1,665	109.43	464.38	4,906.42	182,208.81	7.08
Nel152	271	12.44	42.58	456.67	3,371.68	0.13
Nel154	257	246.9	665.08	4,906.42	63,454.41	2.47
Nel155	483	76.46	322.27	4,906.42	36,931.77	1.44
Nel156	1,296	38.8	200.74	3,540.52	50,288.92	1.95
Nel157	423	127.17	475.14	4,784.89	53,793.92	2.09
Nel158	221	14.57	42.1	367.672	3,220.75	0.12
Nel159	207	163.4	551.41	3,380.01	33,824.66	1.32
Nel160	365	68.21	337.9	4,906.42	246,895.28	0.97

Chr: chromosome; N: number of SVR; Sd: standard deviation; max: maximum; Total: sum of all SVs length.

Table S2. Statistics summary for number and length of total structural variants, deletions, duplications, and inversions, across the 29 autosomal chromosomes.

CHR	DELETIONS			DUPLICATIONS			INVERSIONS			ALL SV			
	N	Length		N	Length		N	Length		N	Length		Average chromosomal coverage (%)
		Mean	Sd		Mean	Sd		Mean	Sd		Mean	Sd	
1	13,813	6.7	50.49	418	805.41	1,156.7	765	30.16	27.25	14,996	30.16	238.45	1.91
2	12,810	11.62	151.35	739	100.01	322.31	224	75.38	321.87	13,773	17.4	170.3	1.18
3	8,521	9.17	84.67	557	70.75	192.75	799	69.47	87.67	9,877	17.52	96.62	0.98
4	8,596	7.98	77.23	360	136.49	253.97	3,060	219.33	126.99	12,016	65.65	137.27	4.53
5	9,735	6.89	45.87	719	155.71	433.16	2,380	492.18	272.55	12,834	105.22	247.21	7.6
6	8,833	5.18	40.94	405	316.04	394.82	306	35.1	49.49	9,544	19.33	110.26	1.09
7	8,183	27.69	193.17	625	362.46	531.56	206	231.49	412.5	9,014	55.56	255.44	3.26
8	7,245	14.79	166.65	228	44.9	59.37	200	175.18	140.16	7,673	19.86	165.86	1
9	7,650	21.65	203.04	311	287.92	766.37	751	1,324.7	1128.6	8,712	143.48	548.5	8.35
10	6,073	167.51	445.61	1,391	360.67	476.81	2,908	2,075.5	512.14	10,372	728.36	965.11	50.43
11	5,060	32.45	277.7	173	85.07	310.33	179	199.94	396.15	5,412	39.67	285.11	1.48
12	7,843	27.95	143.89	749	325.06	364.02	176	482.14	391.28	8,768	62.45	208.27	4.39
13	5,606	6.06	31.69	141	349.09	813.17	472	105.85	119.53	6,219	21.41	141.65	1.1
14	5,656	12.34	102.58	205	214.36	453.09	41	27.42	79.23	5,902	19.46	136.36	1.05
15	7,148	24.78	50.53	1,494	120.6	181.66	1,006	632.42	579.92	9,648	102.98	275.4	8.23
16	4,733	15.65	91.37	222	169.12	511.83	759	11.24	6.99	5,714	21.03	133.95	1.01
17	4,593	11.84	78.94	138	18.57	19.23	407	89.55	161.83	5,138	18.18	89.95	0.89
18	4035	13.16	66.61	951	83.19	135.57	886	119.16	120.6	5,872	40.5	99.79	2.47
19	2,839	13.76	98.4	245	141.22	149.95	61	110.78	77.2	3,145	25.57	109.13	0.95
20	6,336	13.48	163.52	141	205.54	721.44	195	990.37	525.98	6,672	46.09	268.27	3.12
21	4,713	7.86	86.29	397	729.62	1661.53	180	126.84	289.68	5,290	66.08	502.36	3.55
22	3,297	20.1	156.17	190	153.53	418.96	189	126.67	372.99	3,676	32.47	198.42	1.37
23	4,571	59.18	215.96	1,726	278.24	335.43	168	425.32	271.94	6,465	127.18	276.73	10.95
24	4,326	32.48	197.21	224	703.66	621.51	57	64.55	69.08	4,607	65.51	275.88	3.46
25	2,587	7.53	70.74	69	49.02	176.16	114	117.01	59.19	2,770	13.07	78	0.57
26	4,709	12.43	91.2	234	36.54	259.93	521	39.05	105.73	5,464	16	105.8	1.13
27	3,731	41.26	310.23	119	1,370.5	1536.48	185	403.67	430.37	4,035	97.08	470.66	5.8
28	4,408	86.33	371.19	421	153.92	391.69	583	142.5	194.59	5,412	97.64	358.84	7.72
29	5,382	6.65	43.85	421	80.46	214.16	208	2,366.8	1,820.6	6,011	93.49	552	7.48

Table S3. Statistics summary for the number, length and proportion of chromosome extension covered by each SVR type along the 29 autosomes.

Chr	SVR type	N	Mean	Length Sd	Total	Chromosomal coverage rate (%)
1	CNVR-DEL	235	11.67	99.3	2,741.79	1.73
1	CNVR-DUP	7	129.96	280.23	909.73	0.57
1	CNVR-MIXED	12	743.45	1,380.94	8,921.38	5.63
1	INVR	7	31.72	39.47	222.06	0.14
2	CNVR-DEL	185	30.63	223.05	5,667.39	4.16
2	CNVR-DUP	6	7.6	8.15	45.61	0.03
2	CNVR-MIXED	17	1,117.18	1,832.83	18,992.14	13.94
2	INVR	3	9.24	12.41	27.72	0.02
3	CNVR-DEL	153	33.58	233.1	5,137.05	4.25
3	CNVR-DUP	6	97.26	130.1	583.56	0.48
3	CNVR-MIXED	11	412.32	706.25	4,535.48	3.75
3	INVR	7	385.15	589.8	2,696.06	2.23
4	CNVR-DEL	173	32.72	340.61	5,660.01	4.72
4	CNVR-DUP	3	3.06	1.88	9.19	0.01
4	CNVR-MIXED	14	472.94	722.25	6,621.22	5.52
4	INVR	6	200.12	287.3	1,200.69	1
5	CNVR-DEL	182	20.07	193.2	3,652.76	3.04
5	CNVR-DUP	7	41.2	54.83	288.39	0.24
5	CNVR-MIXED	11	738.08	940.83	8,118.85	6.76
5	INVR	7	251.64	413.68	1,761.49	1.47
6	CNVR-DEL	194	15.32	149.67	2,971.78	2.52
6	CNVR-DUP	3	13.85	20.26	41.54	0.04
6	CNVR-MIXED	9	448.65	504.06	4,037.89	3.43
6	INVR	5	71.47	69.78	357.33	0.3
7	CNVR-DEL	165	3.81	6.53	629.19	0.57
7	CNVR-DUP	7	10.58	16.67	74.05	0.07

7	CNVR-MIXED	16	660.22	730.89	10,563.54	9.54
7	INVR	2	1,199.04	1,385.4 6	2,398.07	2.17
8	CNVR-DEL	14 7	29.52	320.27	4,338.92	3.83
8	CNVR-DUP	2	3.32	2.42	6.63	0.01
8	CNVR-MIXED	12	464.32	961.28	5,571.88	4.92
8	INVR	3	266.19	229.49	798.56	0.7
9	CNVR-DEL	14 7	12.32	80.33	1,811.67	1.72
9	CNVR-DUP	5	44.91	66.04	224.54	0.21
9	CNVR-MIXED	8	1,529.14	1,569.5 9	12,233.12	11.6
9	INVR	5	552.49	1,052.3 2	2,762.46	2.62
10	CNVR-DEL	10 7	35.67	233.5	3,817.02	3.69
10	CNVR-DUP	3	12.23	9.27	36.68	0.04
10	CNVR-MIXED	5	1,228.53	1,079.0 8	6,142.67	5.95
10	INVR	2	1,332.04	1,798.5 6	2,664.075	2.58
11	CNVR-DEL	97	139.63	701.38	13,543.88	12.66
11	CNVR-DUP	3	24.15	11.63	72.462	0.07
11	CNVR-MIXED	6	1,034.17	1,131.4 4	6,205.02	5.8
11	INVR	2	501.43	707.15	1,002.86	0.94
12	CNVR-DEL	12 8	31.33	308.57	4,010.03	4.6
12	CNVR-DUP	5	26.97	28.3	134.85	0.15
12	CNVR-MIXED	14	574.11	962.36	8,037.59	9.22
12	INVR	2	894.07	1,262.4 6	1,788.13	2.05
13	CNVR-DEL	95	5.95	25.65	565.24	0.68
13	CNVR-DUP	3	39.75	62.24	119.24	0.14
13	CNVR-MIXED	5	1,523.55	1,878.2 5	7,617.75	9.13
13	INVR	3	315.08	271.95	945.25	1.13
14	CNVR-DEL	97	13.53	106.17	1,312.46	1.59
14	CNVR-DUP	4	1.34	0.31	5.36	0.01
14	CNVR-MIXED	7	1,225.81	1,462.3 1	8,580.66	10.41

15	CNVR-DEL	14 7	14.98	65.46	2,202.48	2.59
15	CNVR-DUP	6	24.73	54.43	148.39	0.17
15	CNVR- MIXED	13	655.71	971.89	8,524.27	10.03
15	INVR	11	336.31	568.24	3,699.36	4.35
16	CNVR-DEL	97	11.45	81.65	1,110.64	1.37
16	CNVR-DUP	5	22.63	23.05	113.14	0.14
16	CNVR- MIXED	8	431.89	815.01	3,455.09	4.27
16	INVR	3	10.98	6.34	32.93	0.04
17	CNVR-DEL	96	47.86	250.6	4,594.13	6.28
17	CNVR-DUP	3	14.75	21.9	44.26	0.06
17	CNVR- MIXED	1	41.79		41.79	0.06
17	INVR	5	104.72	190.71	523.62	0.72
18	CNVR-DEL	66	23.39	101.74	1,543.81	2.35
18	CNVR-DUP	6	74.41	97.37	446.464	0.68
18	CNVR- MIXED	11	593	933.05	6,522.96	9.91
18	INVR	9	244.56	257.18	2,201.04	3.34
19	CNVR-DEL	56	27.6	160.07	1,545.69	2.44
19	CNVR-DUP	1	1.23		1.23	0
19	CNVR- MIXED	8	574.6	1,088.8 2	4,596.82	7.24
19	INVR	2	248.14	100.83	496.28	0.78
20	CNVR-DEL	11 5	3.71	5.08	426.45	0.59
20	CNVR-DUP	4	3,015.56	3,577.1	12,062.26	16.76
20	CNVR- MIXED	3	888.68	743.9	2,666.04	3.7
21	CNVR-DEL	78	13.32	90.74	1,038.63	1.49
21	CNVR-DUP	1	2.56		2.56	0
21	CNVR- MIXED	8	1,586.65	2,080.9 6	12,693.16	18.17
21	INVR	4	140.14	176.55	560.55	0.8
22	CNVR-DEL	59	26.95	186.37	1,589.92	2.62
22	CNVR-DUP	2	79.8	50.11	159.61	0.26
22	CNVR- MIXED	5	841.92	1,329.7 8	4,209.58	6.93
22	INVR	1	1.44		1.44	0
23	CNVR-DEL	71	3.25	2.86	230.80	0.44
23	CNVR-DUP	1	1.69		1.69	0
23	CNVR- MIXED	15	430.09	1,115.3	6,451.38	12.29

23	INVR	3	287.83	438.19	863.49	1.64
24	CNVR-DEL	81	55.78	330.86	4,518.59	7.25
24	CNVR-MIXED	6	932.09	607.05	5,592.52	8.97
24	INVR	1	201.61		201.61	0.32
25	CNVR-DEL	41	27.09	158.91	1,110.77	2.62
25	CNVR-DUP	1	7.72		7.72	0.02
25	CNVR-MIXED	2	432.12	601.35	864.25	2.04
25	INVR	3	74.61	77.16	223.83	0.53
26	CNVR-DEL	64	94.14	500.28	6,024.93	11.59
26	CNVR-DUP	2	6.92	6.28	13.84	0.03
26	CNVR-MIXED	5	1,115.39	1,613.97	5,576.95	10.73
26	INVR	3	355.6	327.41	1,066.79	2.05
27	CNVR-DEL	64	11.26	56.95	720.36	1.58
27	CNVR-MIXED	5	1,051.45	1,614.84	5,257.26	11.53
27	INVR	3	751.46	683.07	2,254.38	4.94
28	CNVR-DEL	66	36.34	265.17	2,398.44	5.22
28	CNVR-DUP	2	1.6	0.51	3.20	0.01
28	CNVR-MIXED	7	877.81	1,090.65	6,144.68	13.38
28	INVR	4	437.27	829.53	1,749.06	3.81
29	CNVR-DEL	80	3.42	5.64	273.77	0.54
29	CNVR-DUP	5	23.33	47.46	116.67	0.23
29	CNVR-MIXED	11	579.29	813.34	6,372.29	12.47
29	INVR	2	4,419.77	5,812.85	8,839.53	17.3

Table S4. Significant QTLs overlapping the most frequent regions of structural variants.

SVR ID	N buls	CHR	SVR START	SVR END	SVRTYPE	QTL TYPE	QTL START	QTL END	QTL ID	TRAIT ID	PUBMED ID	P value
CNVR20	93	1	5962361	5965860	DEL	Production	1948349	13843534	5265	Residual feed intake	18791150	<0.001
INVR27	104	1	39477267	39530842	INV	Exterior	39049021	53876985	3579	Structural soundness	17183116	<0.05
CNVR239	102	1	72748619	72750095	DEL	Health	67679502	112799170	9947	Infectious bovine keratoconjunctivitis susceptibility	17093209	0.0015
CNVR256	105	1	79866184	79867446	DEL	Health	78449771	101673070	1736	Bovine spongiform encephalopathy	15342524	0.0308
CNVR311	127	1	101708575	103826328	MIXED_CNVR	Production	102544700	114132251	1450	Chest width	ISU0026	<0.05
CNVR401	143	1	147428882	152090779	MIXED_CNVR	Milk	151971245	151971245	1514	Milk energy yield	16533362	0.02
CNVR480	139	2	15644871	22265513	MIXED_CNVR	Exterior	22084650	22296189	3486	Teat length	12605852	0.027
CNVR535	124	2	48333206	50338419	MIXED_CNVR	Production	49087578	49299117	5098	Feed conversion ratio	18318789	0.014
CNVR566	77	2	61910398	61911609	DEL	Meat_and_Carcass	57030860	62657792	13188	Connective tissue amount	20826071	0.05
CNVR572	105	2	64775091	64776446	DEL	Production	63726063	79157818	2696	Length of productive life	12613879	0.015
CNVR649	95	2	97564419	97565709	DEL	Exterior	93880919	113596335	3580	Structural soundness	17183116	<0.05
CNVR710	83	2	125354837	125356610	MIXED_CNVR	Milk	122097022	137500220	1516	Milk energy yield	16533362	0.05
CNVR755	110	3	11797918	11800662	DEL	Production	11668354	11856250	5321	Residual feed intake	18791150	<0.15
INVR131	83	3	20943678	22580427	INV	Milk	22171751	22359647	2468	Milk yield	12416823	<0.05
CNVR790	82	3	22791622	23437149	DEL	Meat_and_Carcass	22838783	23026679	13142	Longissimus muscle area	20219139	0.0029
CNVR790	82	3	22791622	23437149	DEL	Meat_and_Carcass	22838783	23026679	13143	Rib thickness	20219139	0.0002
CNVR916	94	3	75972598	75973896	DEL	Production	72058192	87935421	5267	Residual feed intake	18791150	<0.001
CNVR933	120	3	86151722	88610082	DEL	Health	82045815	111422446	2489	Clinical mastitis	11845286	0.014
CNVR970	86	3	109195675	109197080	DEL	Exterior	95545217	111422446	3581	Structural soundness	17183116	<0.05

CNVR985	105	4	1009660	2368583	MIXED_CNVR	Exterior	2211854	2412021	13507	Angularity	20942903	<0.05
CNVR985	105	4	1009660	2368583	MIXED_CNVR	Exterior	2211854	2412021	13508	Body condition score	20942903	<0.05
CNVR1048	84	4	28592368	28593657	DEL	Production	27503055	28964280	64545	Rump length	26420743	0.002
CNVR1048	84	4	28592368	28593657	DEL	Production	27503055	28964280	64547	Hip width	26420743	0.042
CNVR1048	84	4	28592368	28593657	DEL	Production	27503055	28964280	64548	Chest depth	26420743	0.002
CNVR1079	79	4	39382506	39383801	DEL	Production	30625673	48740859	5270	Residual feed intake	18791150	<0.01
CNVR1172	82	4	82517757	82519050	DEL	Health	72370668	91266508	2491	Clinical mastitis	11845286	0.044
CNVR1404	106	5	32557277	35159813	DEL	Exterior	33981009	34158131	19054	Body condition score	21948746	<0.05
CNVR1404	106	5	32557277	35159813	DEL	Health	33981009	34158131	95883	Insulin-like growth factor 1 level	25966067	0.0008
CNVR1404	106	5	32557277	35159813	DEL	Health	35004930	35004934	100994	Liver abscess	27065119	<0.05
CNVR1434	79	5	44064148	44065821	DEL	Exterior	35955928	49240138	3575	Structural soundness	17183116	<0.05
CNVR1449	94	5	50848169	50849280	DEL	Reproduction	50152320	58633846	1760	Multiple birth	15147392	<0.10
CNVR1465	102	5	58631317	58908758	MIXED_CNVR	Reproduction	58627646	65712558	1758	Multiple birth	15147392	<0.01
CNVR1519	77	5	80119727	80120964	DEL	Reproduction	69255014	80449175	1759	Multiple birth	15147392	<0.10
CNVR1697	88	6	26849415	26851941	DEL	Production	26839501	27014808	11449	Residual feed intake	19744142	<0.01
CNVR1707	76	6	32932150	32933444	DEL	Health	31020573	62006085	2493	Clinical mastitis	11845286	0.0006
CNVR1707	76	6	32932150	32933444	DEL	Health	31027585	69412806	1737	Bovine spongiform encephalopathy	15342524	0.0121
CNVR1756	77	6	52406581	52420590	MIXED_CNVR	Milk	50453354	63487430	1519	Milk energy yield	16533362	0.02
CNVR1799	75	6	72924569	74154761	MIXED_CNVR	Production	62759905	88530035	5272	Residual feed intake	18791150	0.0049
CNVR1799	75	6	72924569	74154761	MIXED_CNVR	Exterior	63487430	78318402	1381	Teat placement	14691316	0.029
CNVR1799	75	6	72924569	74154761	MIXED_CNVR	Exterior	63487430	78318402	1382	Foot angle	14691316	0.008
CNVR1799	75	6	72924569	74154761	MIXED_CNVR	Exterior	63487430	78318402	1383	Udder quality	14691316	0.008
CNVR1799	75	6	72924569	74154761	MIXED_CNVR	Exterior	63487430	78318402	1385	Udder depth	14691316	0.008
CNVR1799	75	6	72924569	74154761	MIXED_CNVR	Exterior	63493565	78888150	2426	Degree of Spotting	10589513	0.0001
CNVR1880	76	6	98880369	98886314	DEL	Milk	90283105	109479221	3437	Milking speed	12605852	0.008
CNVR1931	83	7	1595009	2933458	MIXED_CNVR	Production	2082957	2246326	5349	Dry matter intake	18791150	<0.05

CNVR1931	83	7	1595009	2933458	MIXED_CNVR	Health	2246326	2409695	18765	M. paratuberculosis susceptibility	21597988	0.0008
CNVR1946	101	7	8714162	10096757	MIXED_CNVR	Production	7923406	46805275	5274	Residual feed intake	18791150	<0.05
CNVR2059	77	7	45899882	45901104	DEL	Production	42067569	46805275	5277	Dry matter intake	18791150	<0.001
CNVR2059	77	7	45899882	45901104	DEL	Milk	42475992	80867754	3434	Milk fat percentage	12605852	0.004
CNVR2105	76	7	67901454	67902687	DEL	Milk	57996066	73516140	3410	Milk protein yield	12605852	0.0029
CNVR2153	82	7	88765927	88767230	DEL	Exterior	78907323	94754136	3582	Structural soundness	17183116	<0.05
CNVR2163	77	7	95695718	95704297	DEL	Reproduction	95244243	102024065	2423	Multiple birth	11003703	0.049
CNVR2349	75	8	48536156	48537399	DEL	Health	41486835	74759311	2497	Clinical mastitis	11845286	0.0141
CNVR2375	96	8	64006026	64007841	DEL	Exterior	62616650	104596066	3599	Foot angle	17183116	<0.05
CNVR2415	92	8	82188442	82190094	DEL	Reproduction	81703009	108297022	1718	Multiple birth	16026340	0.001
CNVR2511	98	9	6374304	11007347	MIXED_CNVR	Health	6757804	6939465	10111	Clinical mastitis	18487673	<0.005
CNVR2511	98	9	6374304	11007347	MIXED_CNVR	Meat_and_Carcass	8501753	8683414	18466	Cooking loss	21504841	<0.05
CNVR2511	98	9	6374304	11007347	MIXED_CNVR	Meat_and_Carcass	8501753	8683414	18468	Longissimus muscle area	21504841	<0.05
CNVR2609	96	9	52381845	52382892	DEL	Health	45406266	70439207	1745	General disease susceptibility	15328290	<0.01
CNVR2654	76	9	71350749	71352924	DEL	Milk	70439207	94191435	1521	Milk energy yield	16533362	0.03
CNVR2757	76	10	22222947	22224673	DEL	Exterior	22070498	22244419	4633	Udder attachment	10791796	0.0334
CNVR2759	120	10	22638026	25331938	MIXED_CNVR	Exterior	22531387	22705308	4634	Teat placement - front	10791796	0.0122
CNVR2829	78	10	59828018	61289113	MIXED_CNVR	Production	61193892	61367813	5101	Feed conversion ratio	18318789	0.043
CNVR2967	76	11	26893599	26895628	DEL	Production	15038007	51080187	5281	Residual feed intake	18791150	<0.05
CNVR2970	101	11	31279552	35330075	DEL	Exterior	28678460	33083615	3591	Structural soundness	17183116	<0.05
CNVR2970	101	11	31279552	35330075	DEL	Production	30999544	32012975	5358	Residual feed intake	18791150	<0.15
CNVR3035	86	11	66212874	66214572	MIXED_CNVR	Production	65170146	66526836	5356	Residual feed intake	18791150	<0.05
CNVR3067	90	11	83477328	85488107	DEL	Meat_and_Carcass	84155630	84319087	31010	Subcutaneous fat thickness	23872232	0.0212
CNVR3067	90	11	83477328	85488107	DEL	Meat_and_Carcass	84155630	84319087	31013	Longissimus muscle area	23872232	<0.0001

CNVR3234	94	12	51587952	51592472	MIXED_CNVR	Exterior	50663104	80867330	3592	Structural soundness	17183116	<0.05
CNVR3313	79	12	82104408	85598829	DEL	Production	85549541	88089045	5283	Residual feed intake	18791150	0.0285
CNVR3376	83	13	21837559	21838863	DEL	Health	18224379	49021124	1741	Bovine spongiform encephalopathy	15342524	0.0447
CNVR3421	75	13	51711398	51712598	DEL	Production	48989416	55251817	5286	Residual feed intake	18791150	0.0179
CNVR3458	104	13	79063601	81545280	MIXED_CNVR	Production	78002565	81173401	5288	Dry matter intake	18791150	0.0115
CNVR3458	104	13	79063601	81545280	MIXED_CNVR	Meat_and_Carcass	79825796	79825796	30909	Dressing percentage	22843074	0.034
CNVR3458	104	13	79063601	81545280	MIXED_CNVR	Meat_and_Carcass	79825796	79825796	30910	Subcutaneous fat thickness	22843074	0.013
CNVR3458	104	13	79063601	81545280	MIXED_CNVR	Meat_and_Carcass	79825796	79825796	30911	Longissimus muscle area	22843074	0.047
CNVR3473	75	14	5977961	5979027	DEL	Meat_and_Carcass	5853070	6011690	6024	Marbling score	18670061	0.0238
CNVR3527	79	14	37977824	37986411	DEL	Exterior	34602909	41534594	3584	Structural soundness	17183116	<0.05
CNVR3545	81	14	44154951	44156169	DEL	Reproduction	41193562	66564799	1720	Multiple birth	16026340	0.001
CNVR3599	92	14	63713877	63715439	DEL	Reproduction	55350379	66065146	1416	Ovulation rate	15265069	0.014
CNVR3599	92	14	63713877	63715439	DEL	Health	63273438	75447507	2495	Clinical mastitis	11845286	0.0018
CNVR3702	77	15	12445845	15217394	MIXED_CNVR	Production	6395340	17110034	3180	Veterinary treatments	14762087	<0.01
CNVR3702	77	15	12445845	15217394	MIXED_CNVR	Milk	12367579	29415621	1526	Milk energy yield	16533362	<0.01
CNVR3846	76	15	69214267	71636039	MIXED_CNVR	Exterior	71167826	76080761	3593	Structural soundness	17183116	<0.05
CNVR3846	76	15	69214267	71636039	MIXED_CNVR	Exterior	71167826	76080761	3595	Bone quality	17183116	<0.05
CNVR3846	76	15	69214267	71636039	MIXED_CNVR	Health	71587807	71587811	101007	Liver abscess	27065119	<0.05
INVR549	76	15	78781811	80138873	INV	Health	79276621	79276625	96070	Liver abscess	27065119	<0.05
INVR557	117	16	3117468	3121435	INV	Exterior	777973	12162809	3588	Structural soundness	17183116	<0.05
CNVR4150	112	17	30735419	32358325	DEL	Exterior	29210381	31828400	3596	Bone quality	17183116	<0.05
CNVR4173	85	17	39449851	39451014	DEL	Production	37171297	47551783	5291	Residual feed intake	18791150	0.0279
CNVR4186	91	17	43328053	43329329	DEL	Milk	41757793	50742256	1531	Milk energy yield	16533362	0.02
CNVR4215	91	17	64624106	64625414	DEL	Health	57527735	70266728	1738	Bovine spongiform encephalopathy	15342524	0.0015

CNVR4295	76	18	33295173	33298371	DEL	Production	21375890	40333341	5293	Residual feed intake	18791150	<0.05
CNVR4295	76	18	33295173	33298371	DEL	Production	31369509	36643269	3557	Length of productive life	16734691	<0.01
CNVR4295	76	18	33295173	33298371	DEL	Reproduction	31369509	36643269	3558	Pregnancy rate	16734691	<0.01
CNVR4308	78	18	42446551	42447716	DEL	Reproduction	36647950	42681568	2711	Stillbirth	12613879	0.002
CNVR4308	78	18	42446551	42447716	DEL	Reproduction	39038305	42907810	3561	Pregnancy rate	16734691	<0.05
CNVR4420	87	19	28784020	28785724	DEL	Production	19914848	42955864	5297	Residual feed intake	18791150	<0.05
CNVR4432	104	19	47187250	47188669	DEL	Exterior	39945480	53897452	3430	Rump angle	12605852	0.0023
CNVR4432	104	19	47187250	47188669	DEL	Production	45329436	57139404	5299	Residual feed intake	18791150	0.0045
CNVR4466	85	20	2880544	2883706	DEL	Health	1735582	30372685	9948	Infectious bovine keratoconjunctivitis susceptibility	17093209	0.0014
CNVR4472	90	20	7096261	7104874	DEL	Production	7089852	7263410	11473	Dry matter intake	19744142	<0.05
CNVR4639	78	21	587333	588403	DEL	Production	250165	9506289	5300	Residual feed intake	18791150	0.0055
CNVR4703	98	21	31118439	34933659	MIXED_CNVR	Exterior	29928132	34781343	3585	Structural soundness	17183116	<0.05
CNVR4733	76	21	48650283	52078086	MIXED_CNVR	Milk	51692531	51859308	1535	Milk yield	16533362	0.04
CNVR4791	115	22	4536851	7623116	MIXED_CNVR	Exterior	1972879	13141032	3578	Structural soundness	17183116	<0.05
CNVR4810	76	22	19277886	19279359	DEL	Production	13175523	23453812	5302	Residual feed intake	18791150	<0.05
CNVR4860	89	22	46918845	46921641	DEL	Milk	44838170	55281324	1537	Milk energy yield	16533362	0.04
CNVR4906	88	23	5083179	5084574	DEL	Reproduction	3080125	7756064	1763	Multiple birth	15147392	<0.10
CNVR4931	111	23	17751563	22144687	MIXED_CNVR	Reproduction	17388015	27279047	1762	Multiple birth	15147392	<0.05
CNVR4931	111	23	17751563	22144687	MIXED_CNVR	Health	18011123	18142304	19997	Complement hemolytic activity	22155013	<0.05
CNVR4931	111	23	17751563	22144687	MIXED_CNVR	Reproduction	18339075	18470256	16053	Interval to first estrus after calving	21071888	0.018
CNVR4941	120	23	25867738	26691576	MIXED_CNVR	Production	23284592	46634774	5303	Residual feed intake	18791150	<0.05
CNVR4956	82	23	28513221	28765192	MIXED_CNVR	Production	28141561	34297220	1415	Veterinary treatments	10430670	0.0087
CNVR5010	81	23	48933332	48934633	DEL	Production	48471305	52931452	5305	Dry matter intake	18791150	0.0193

CNVR5146	98	25	4291696	4292874	DEL	Production	804787	9905072	5310	Residual feed intake	18791150	0.0232
CNVR5284	108	26	32594550	32603684	DEL	Production	30849510	51415851	5314	Residual feed intake	18791150	<0.05
CNVR5288	123	26	35735429	36116883	MIXED_CNVR	Exterior	35021727	39606580	3576	Structural soundness	17183116	<0.05
CNVR5288	123	26	35735429	36116883	MIXED_CNVR	Exterior	35021727	39606580	3597	Bone quality	17183116	<0.05
CNVR5389	121	27	30794266	31009537	MIXED_CNVR	Health	23371351	35422628	2496	Clinical mastitis	11845286	0.015
CNVR5409	92	27	39944152	39945392	DEL	Exterior	38320350	43398135	3594	Structural soundness	17183116	<0.05
CNVR5409	92	27	39944152	39945392	DEL	Exterior	38320350	43398135	3598	Bone quality	17183116	<0.05
CNVR5453	117	28	14278032	16435472	DEL	Meat_and_Carcass	15708310	15857345	13041	Longissimus muscle area	20812196	0.0156
CNVR5453	117	28	14278032	16435472	DEL	Meat_and_Carcass	15708310	15857345	13042	Water holding capacity	20812196	0.0010
CNVR5501	95	28	38315963	38317727	DEL	Exterior	36804242	44375230	3589	Structural soundness	17183116	<0.05
CNVR5512	77	29	1023317	1024930	DEL	Exterior	677112	1327844	3590	Structural soundness	17183116	<0.05
CNVR5520	87	29	2655292	2660799	DEL	Reproduction	1761665	15469534	3574	Inseminations per conception	16899702	<0.05
CNVR5549	102	29	10696102	10697385	DEL	Exterior	8273379	15469534	1389	Temperament	14691316	0.044
CNVR5597	94	29	26558800	26560399	DEL	Reproduction	21397935	48101385	1722	Multiple birth	16026340	0.004

Table S5. Positional candidate genes overlapping the most frequent SVR.

SVR ID	CHR	START	END	SVR TYPE	N SAMPLE	N GENES	GENES
CNVR21	1	6276365	6278874	DEL	79	1	<i>GRIK1</i>
CNVR137	1	37323674	37325002	DEL	85	1	<i>EPHA3</i>
CNVR188	1	50153285	50154319	DEL	94	1	<i>ALCAM</i>
CNVR199	1	53584204	53587093	DEL	77	1	<i>TRAT1</i>
CNVR213	1	58227809	58230229	DEL	91	1	<i>SIDT1</i>
CNVR230	1	65184016	65186939	DEL	92	1	<i>FSTL1</i>
CNVR260	1	81476362	81477652	DEL	93	1	<i>IGF2BP2</i>
CNVR287	1	93804589	93807311	DEL	85	1	<i>NLGN1</i>
CNVR291	1	95874533	95876738	DEL	104	1	<i>PLD1</i>
CNVR311	1	101708575	103826328	MIXED_CNVR	127	3	<i>RNaseP_nuc</i> <i>SLITRK3</i>
CNVR323	1	111217679	111223193	DEL	92	1	<i>KCNAB1</i>
CNVR327	1	112546128	112547189	DEL	97	1	<i>MME</i>
CNVR329	1	112926505	114402464	DEL	80	4	<i>ARHGEF26</i> <i>DHX36</i> <i>GPR149</i> <i>Metazoa_SRP</i>
CNVR333	1	116630479	116635995	DEL	104	1	<i>IGSF10</i>
CNVR369	1	134116596	134117845	DEL	98	1	<i>EPHB1</i>
CNVR381	1	138180211	138183046	DEL	109	1	<i>CPNE4</i>
CNVR401	1	147428882	152090779	MIXED_CNVR	143	26	<i>C5S_rRNA</i> <i>CBR1</i> <i>CBR3</i> <i>CHAF1B</i> <i>CLDN14</i> <i>COL6A5</i> <i>COL6A6</i> <i>DOP1B</i> <i>DYRK1A</i> <i>ERG</i> <i>ETS2</i> <i>HLCS</i> <i>KCNJ15</i> <i>KCNJ6</i> <i>MGC127133</i> <i>MORC3</i> <i>PIGP</i> <i>PIK3R4</i> <i>RIPPLY3</i> <i>SETD4</i> <i>SIM2</i> <i>SNORA70</i> <i>SNORA72</i> <i>TTC3</i> <i>U6</i> <i>VPS26C</i>
CNVR405	1	154106282	154109126	DEL	91	1	<i>TBC1D5</i>
CNVR436	2	6113942	6115867	DEL	84	2	<i>C2H2orf88</i> , <i>HIBCH</i>
CNVR452	2	9584586	9588726	DEL	105	1	<i>FAM171B</i>
CNVR480	2	15644871	22265513	MIXED_CNVR	139	37	<i>AGPS</i> <i>ATF2</i> <i>ATP5MC3</i> <i>bta-mir-10b</i> <i>bta-mir-4444</i> <i>CCDC141</i> <i>CHN1</i> <i>CHRNA1</i> <i>CWC22</i> <i>CYCT</i> <i>EVX2</i> <i>FKBP7</i> <i>HNRNPA3</i> <i>HOXD1</i> <i>HOXD10</i> <i>HOXD11</i> <i>HOXD12</i> <i>HOXD13</i> <i>HOXD3</i> <i>HOXD4</i> <i>HOXD8</i> <i>HOXD9</i> <i>LNPk</i> <i>MTX2</i> <i>NFE2L2</i> <i>OSBPL6</i> <i>PDE11A</i> <i>PJKV</i> <i>PLEKHA3</i> <i>PRKRA</i> <i>RBM45</i> <i>SESTD1</i> <i>TTC30A</i> <i>TTC30B</i> <i>UBE2E3</i> <i>WIPF1</i> <i>ZNF385B</i>
CNVR500	2	30769455	30770867	DEL	88	1	<i>CSRNP3</i>
CNVR501	2	31422918	33800868	MIXED_CNVR	85	4	<i>bta-mir-11978</i> <i>COBLL1</i> <i>FIGN</i> <i>GRB14</i>
CNVR516	2	42102742	42103918	DEL	100	1	<i>GALNT13</i>
CNVR535	2	48333206	50338419	MIXED_CNVR	124	1	<i>ACVR2A</i>

CNVR566	2	61910398	61911609	DEL	77	1	ZRANB3
CNVR569	2	62972010	62974566	DEL	122	1	MGAT5
CNVR572	2	64775091	64776446	DEL	105	1	NCKAP5
CNVR581	2	67497784	67499090	DEL	80	1	DPP10
CNVR629	2	86600865	87918221	DEL	105	2	PLCL1 SATB2
CNVR631	2	89157705	89180021	DUP	106	1	AOX1
CNVR638	2	90885395	90886550	DEL	91	1	BMPR2
CNVR686	2	109848650	114099525	MIXED_CNVR	102	17	ACSL3 AP1S3 bta-mir-2284y-5 CUL3 DOCK10 FAM124B FARSB KCNE4 MOGAT1 MRPL44 NYAP2 PAX3 SCG2 SERPINE2 SGPP2 SNORD116 WDFY1
CNVR695	2	116603452	118279535	DEL	109	9	7SK bta-mir-2284m CD24 DNER FBXO36 PID1 SLC16A14 SP110 TRIP12
CNVR710	2	125354837	125356610	MIXED_CNVR	83	1	EYA3
CNVR716	2	131894798	131904577	MIXED_CNVR	76	1	SH2D5
CNVR717	2	132530986	133810156	MIXED_CNVR	135	22	AKR7A2 ALDH4A1 CAPZB EMC1 HTR6 IFFO2 MICOS10 MRTO4 NBL1 OTUD3 PLA2G2A PLA2G2D1 PLA2G2D4 PLA2G2E PLA2G5 RNF186 SLC66A1 TAS1R2 TMCO4 U1 U3 UBR4
CNVR767	3	16402266	16404115	DEL	78	1	NUP210L
INVR131	3	20943678	22580427	INV	83	24	ACP6 ANKRD34A ANKRD35 BCL9 CHD1L FMO5 GJA5 GJA8 GNRHR2 GPR89A HJV ITGA10 NUDT17 OR13L12 OR13L2 PDZK1 PEX11B PIAS3 POLR3C POLR3GL PRKAB2 RBM8A RNF115 TXNIP
CNVR790	3	22791622	23437149	DEL	82	3	NOTCH2 PDE4DIP SEC22B
CNVR799	3	28388841	28390422	DEL	89	1	SYCP1
CNVR805	3	31669226	31670436	DEL	99	1	RAP1A
CNVR808	3	32256679	32265298	DEL	95	1	CEPT1
CNVR877	3	57357884	57423242	MIXED_CNVR	76	1	CLCA3
CNVR933	3	86151722	88610082	DEL	120	8	JUN bta-mir-12043 bta-mir- 2285ap FGYY HOOK1 MYSM1 OMA1 TACSTD2
CNVR1013	4	8322121	8323132	DEL	81	1	CDK14
CNVR1025	4	15458189	15459390	DEL	79	1	COL28A1

CNVR1044	4	21800326	26262385	DEL	104	14	AGMO,AGR2, AGR3, AHR, ANKMY2, BZW2, CRPPA, DGKB, ETV1, LRRC72, MEOX2, SNORD31, SOSTDC1, TSPAN13
CNVR1100	4	48267814	48271650	DEL	79	1	HBP1
CNVR1103	4	49258498	49260968	DEL	96	1	NRCAM
CNVR1217	4	92735379	92736874	DEL	109	1	CALU
CNVR1269	4	112827567	113284233	MIXED_CNVR	82	3	GIMAP4 GIMAP7 GIMAP8
INVR220	4	112878459	113341002	INV	105	1	GIMAP5
CNVR1361	5	15682733	15685999	DEL	76	1	MGAT4C
CNVR1390	5	24269568	24271000	DEL	89	1	CEP83
CNVR1392	5	28145439	29865137	MIXED_CNVR	81	26	ASIC1 ATF1 BIN2 CELA1 CERS5 COX14 CSRNP2 DAZAP2 DIP2B FAM186A GALNT6 GPD1 HIGD1C LARP4 LETMD1 LIMA1 METTL7A POU6F1 RACGAP1 SCN8A SLC11A2 SLC4A8 SMAGP SMARCD1 TFCP2 TMPRSS12
CNVR1404	5	32557277	35159813	DEL	106	10	AMIGO2 ANO6 ARID2 ENDOU PCED1B RPAP3 SCAF11 SLC38A1 SLC38A2 SLC38A4
CNVR1440	5	44912667	44921230	DEL	85	1	CPM
CNVR1442	5	47839944	47846278	DUP	117	1	HMGA2
CNVR1449	5	50848169	50849280	DEL	94	1	PPM1H
CNVR1452	5	51908262	51913784	DEL	87	1	TAF12
CNVR1465	5	58631317	58908758	MIXED_CNVR	102	5	OR6C1 OR6C1N OR6C5 OR6C5H OR6C7G
INVR240	5	58727321	58939413	INV	86	2	OR6C1S OR6C3
CNVR1474	5	60914837	60918969	DEL	81	1	CFAP54
CNVR1523	5	82218048	82219804	DEL	79	1	PPFIBP1
CNVR1549	5	91481187	91482483	DEL	85	1	PIK3C2G
INVR254	5	101899361	103060179	INV	104	5	CD163L1 WC1 WC1.3 WC1-12 WC-7
CNVR1600	5	111222125	111230526	DEL	83	1	ENTHD1

							A4GALT ARFGAP3 ARHGAP8 bta-mir-2442 bta-mir-33a CCDC134 CENPM CYB5R3 CYP2D14 DESI1 EFCAB6 FAM118A KIAA0930 MCAT MEI1 MGC127055 MPPED1 NAGA NDUFA6 NFAM1 NUP50 PACSIN2 PARVB PARVG PHETA2 PHF21B POLDIP3 PRR5 RRP7 RTL6 SAMM50 SCUBE1 SEPTIN3 SERHL2 SHISA8 SHISAL1 SMC1B SMDT1 SMIM45 SNU13 SREBF2 SULT4A1 TCF20 TNFRSF13C TSPO TTLL1 TTLL12 U12 UPK3A WBP2NL XRCC6
CNVR1601	5	112681084	115783211	MIXED_CNVR	102	51	
CNVR1622	6	2685238	2694212	DEL	82	1	NAF1
CNVR1637	6	5432990	5800662	MIXED_CNVR	141	1	PDE5A
CNVR1726	6	41154844	41163381	DEL	76	1	KCNIP4
CNVR1773	6	58184175	58185805	DEL	83	1	TMEM156
CNVR1881	6	99699774	99701226	DEL	87	1	WDFY3
							ADAMTS2 CANX CBY3 GRM6 H2AW H2BU1 H3-4 HNRNPH1 MAML1 RNF187 RUFY1 TRIM11 TRIM17 ZFP2 ZNF354C ZNF454 ZNF879
CNVR1931	7	1595009	2933458	MIXED_CNVR	83	17	
CNVR1941	7	6135679	6136939	DEL	90	1	CPAMD8
							OR7A10 OR7A102 OR7A103 OR7A112 OR7A114 OR7A121BP OR7A126 OR7A135 OR7A75 OR7A77 OR7A78 OR7A79 OR7A83 OR7A88 OR7A89 OR7A90 OR7A95 OR7A97 OR7A98 OR7A99
CNVR1946	7	8714162	10096757	MIXED_CNVR	101	20	
CNVR1979	7	23002910	23003919	DEL	88	1	RAPGEF6
CNVR1997	7	26722502	26726361	DEL	85	1	MEGF10
CNVR2043	7	40993138	41026889	DEL	82	1	OR5AE3
CNVR2063	7	50070412	50072356	DEL	95	1	CTNNA1
CNVR2105	7	67901454	67902687	DEL	76	1	SGCD
CNVR2360	8	52731284	56617133	DEL	93	7	CEP78 FOXB2 GNA14 GNAQ PRUNE2 PSAT1 TLE4
CNVR2370	8	60653236	60655397	DEL	75	1	RNF38
CNVR2375	8	64006026	64007841	DEL	96	1	COL15A1
CNVR2392	8	70563946	70565456	DEL	101	1	ENTPD4

CNVR2420	8	84843577	87893268	MIXED_CNVR	93	13	AUH BARX1 bta-let-7d DIRAS2 FAM120A MIRLET7A1 MIRLET7F1 PHF2 PTPDC1 ROR2 SPTLC1 SYK ZNF169 ADGRB3 B3GAT2 bta-mir- 11987 bta-mir-2285cx bta-mir- 30a bta-mir-30f COL9A1 FAM135A LMBRD1 OGFRL1 SDHAF4 SMAP1 U4
CNVR2511	9	6374304	11007347	MIXED_CNVR	98	13	
CNVR2582	9	39956459	39958326	DEL	77	1	METTL24
CNVR2629	9	60221373	60224339	DEL	81	1	BACH2
CNVR2679	9	83933034	83935067	DEL	81	1	ADGB
INVR399	9	84681194	87106609	INV	75	13	GINM1 KATNA1 LATS1 LRP11 NUP43 PCMT1 PPIL4 SAMD5 SASH1 TAB2 ULBP21 UST ZC3H12D
CNVR2682	9	84702539	87229971	MIXED_CNVR	104	1	ULBP17
CNVR2683	9	87476957	87484659	DEL	82	1	IYD
CNVR2686	9	88860011	88865501	DEL	76	1	ESR1
CNVR2715	10	950558	959111	DEL	78	1	MCC
CNVR2759	10	22638026	25331938	MIXED_CNVR	120	2	TRAV16 TRAV24
CNVR2778	10	29826748	29827859	DEL	76	1	FMN1
CNVR2829	10	59828018	61289113	MIXED_CNVR	78	12	ATP8B4 COPS2 DTWD1 FAM227B FGF7 GABPB1 GALK2 HDC SECISBP2L SHC4 SLC27A2 USP8
CNVR2967	11	26893599	26895628	DEL	76	1	CAMKMT
CNVR2970	11	31279552	35330075	DEL	101	2	FSHR NRXN1
CNVR3067	11	83477328	85488107	DEL	90	2	LRATD1 TRIB2
CNVR3170	12	26352588	26354127	DEL	83	1	NBEA
CNVR3179	12	31222379	31230764	DEL	79	1	MTUS2
CNVR3235	12	52282581	52283882	DEL	75	1	MYCBP2
CNVR3306	12	77133433	77134955	DEL	92	1	PCCA
CNVR3313	12	82104408	85598829	DEL	79	18	ABHD13 ANKRD10 ARGLU1 ARHGEF7 bta-mir-2301 CARS2 COL4A1 COL4A2 EFNB2 ING1 IRS2 LIG4 MYO16 NALF1 NAXD RAB20 TEX29 TNFSF13B
CNVR3327	13	1720606	1721781	DEL	89	1	PLCB1
CNVR3352	13	9738091	9746578	DEL	79	1	MACROD2
CNVR3360	13	12482701	12761334	MIXED_CNVR	76	2	ECHDC3 USP6NL
CNVR3385	13	25185101	25188608	DEL	79	1	KIAA1217
CNVR3389	13	26130110	26131120	DEL	78	1	GPR158
CNVR3407	13	42504232	42505535	DEL	83	1	SYNDIG1
CNVR3421	13	51711398	51712598	DEL	75	1	ATRN

							ANGPT4 APCDD1L ATP5F1E AURKA BCL2L1 BMP7 bta-mir- 12010 bta-mir-2285bu-1 bta-mir- 2307 bta-mir-6123 C13H20orf202 C13H20orf85 C13H20orf96 CASS4 CCM2L COX4I2 CSNK2A1 CSTF1 CTCFL CTSZ DEFB116 DEFB119 DEFB121 DEFB122 DEFB122A DEFB123 DEFB124 DEFB127 DEFB129 DUSP15 FAM110A FAM209A FAM210B FKBP1A FOXS1 GCNT7 GNAS HCK HM13 ID1 KIF3B MC3R MYLK2 NELFCD NPEPL1 NRSN2 NSFL1C PCK1 PDRG1 PLAGL2 PMEPA1 POFUT1 PSMF1 RAB22A RAD21L1 RAE1 RBCK1 RBM38 REM1 RSPO4 RTF2 SCRT2 SDCBP2 SIRPB2 SIRPD SLC52A3 SNPH SOX12 SPO11 SRXN1 STX16 TBC1D20 TCF15 TFAP2C TM9SF4 TMEM74B TPX2 TRIB3 TTLL9 TUBB1 VAPB XKR7 ZBP1 ZCCHC3
CNVR3428	13	57357801	61754419	MIXED_CNVR	96	84	
CNVR3447	13	73475362	73476564	DEL	84	1	KCNS1
CNVR3458	13	79063601	81545280	MIXED_CNVR	104	7	ATP9A BCAS1 NFATC2 SALL4 TSHZ2 ZFP64 ZNF217
CNVR3516	14	31280801	31281984	DEL	75	1	CSPP1
CNVR3599	14	63713877	63715439	DEL	92	1	SNX31
							CDH17 CIBAR1 DPY19L4 ESRP1 FSBP GEM INTS8 OTUD6B PDP1 PIP4P2 RAD54B RBM12B RUNX1T1 SLC26A7 TMEM67 TRIQK VIRMA
CNVR3609	14	69476913	73396640	MIXED_CNVR	101	17	ATP6V0D2 bta-mir-2311 CA1 CA13 CA2 CA3 CPNE3 E2F5 LRRCC1 RALYL RBIS RMDN1 WWP1
CNVR3618	14	76147414	78465541	MIXED_CNVR	113	13	CDC82 CEP57 FAM76B JRKL MAML2 MTMR2 SESN3
CNVR3702	15	12445845	15217394	MIXED_CNVR	77	7	SCUBE2
CNVR3764	15	43525742	43527853	DEL	79	1	LRRC4C
CNVR3846	15	69214267	71636039	MIXED_CNVR	76	1	

							FADS2B OR2AH1 OR5AL2 OR5AL8 OR5AL9P OR5AP2 OR5AR1 OR5BE5 OR5G34 OR5M10 OR5M11 OR5M13D OR5M3 OR5T1C OR5T2 OR5T21 OR8H13 OR8H16P OR8I2 OR8J12 OR8J16 OR8J17 OR8J19 OR8J2E OR8J3 OR8J3F OR8K1 OR8K1B OR8K3B OR8K5 OR8K5B OR8K60 OR8K62 OR8K63 OR8K64 OR8K65P OR8K66 OR8K67 OR8U1 OR8U3 OR8U9 OR9G3 OR9G4 OR9G4D
INVR549	15	78781811	80138873	INV	76	44	
CNVR3983	16	27811436	27818375	DEL	76	1	CNIH3
CNVR3987	16	28372632	28377978	DEL	90	1	DNAH14
CNVR4048	16	66990163	66991396	DEL	83	1	HMCN1
CNVR4055	16	72388353	72392394	DEL	83	1	KCNH1
CNVR4057	16	72652971	72661572	DEL	85	1	HHAT
CNVR4150	17	30735419	32358325	DEL	112	1	FAT4
CNVR4215	17	64624106	64625414	DEL	91	1	WSCD2
CNVR4219	17	68057710	68216533	DEL	91	1	TNIP3
							bta-mir-12038 bta-mir-125a C18H19orf84 CLDND2 HAS1 IGLON5 LIM2 MIR99B MIRLET7E NKG7 PPP2R1A SIGLEC10 SPACA6 VSIG10L ZNF175 ZNF350 ZNF432 ZNF613 ZNF614 ZNF677
CNVR4338	18	57180983	60174708	MIXED_CNVR	130	20	
CNVR4347	18	61815370	63400235	MIXED_CNVR	101	58	BRISK1 bta-mir-7865 C19orf85 CCDC106 CDC42EP5 CNOT3 COX6B2 EPN1 EPS8L1 FCAR FIZ1 HSPBP1 ISOC2 KIR2DL5A KIR2DS1 KIR3DL1 KIR3DL2 KIR3DS1 KMT5C LAIR1 LENG1 LENG8 LENG9 MBOAT7 NAT14 NCR1 NDUFA3 NLRP2 OSCAR PPP1R12C PPP6R1 PRPF31 PTPRH RDH13 RPL28 RPS9 SBK2 SHISA7 SSC5D SYT5 TARM1 TFPT TMC4 TMEM150B TMEM190 TMEM86B TNNI3 TNNT1 TSEN34 TTYH1 U2AF2 UBE2S ZNF524 ZNF579 ZNF580 ZNF581 ZNF628 ZNF784
CNVR4388	19	3888233	7102140	MIXED_CNVR	98	8	COX11 HLF MMD PCTP SNORD33 STXBP4 TMEM100 TOM1L1

CNVR4420	19	28784020	28785724	DEL	87	1	STX8
CNVR4422	19	33769029	33770095	DEL	98	1	AKAP10
CNVR4432	19	47187250	47188669	DEL	104	1	MARCHF10
CNVR4435	19	48746803	48748097	DEL	94	1	SMURF2
CNVR4439	19	51935118	51936412	DEL	85	1	RPTOR
CNVR4466	20	2880544	2883706	DEL	85	1	RANBP17
CNVR4474	20	7957267	7958345	DEL	83	1	ARHGEF28
CNVR4478	20	10863070	11932505	MIXED_CNVR	91	1	PIK3R1
CNVR4483	20	13811028	13812670	DEL	84	1	NLN
CNVR4492	20	15772076	17148348	MIXED_CNVR	83	3	DIMT1 IPO11 KIF2A
CNVR4496	20	18317665	18319934	DEL	89	1	NDUFAF2
CNVR4526	20	32704216	32712787	DEL	83	1	OXCT1
CNVR4532	20	33502334	41880241	MIXED_CNVR	116	51	ADAMTS12 AGXT2 AMACR BRX1 bta-mir-12058 bta-mir- 2360 C1QTNF3 C7 C9 CAPSL CARD6 CPLANE1 DAB2 DNAJC21 EGFLAM FYB1 GDNF GOLPH3 IL7R LIFR LMBRD2 MROH2B MTMR12 NADK2 NIPBL NPR3 NUP155 OSMR PDZD2 PRKAA1 PRLR PTGER4 RAD1 RAI14 RANBP3L RICTOR RPL37 RXFP3 SKP2 SLC1A3 SLC45A2 SNORD72 SPEF2 SUB1 TARS1 TTC23L TTC33 U2 UGT3A2 WDR70 ZFR
CNVR4612	20	62792998	62795066	DEL	91	1	ROPN1L
CNVR4613	20	63691500	63693251	DEL	76	1	SEMA5A
CNVR4660	21	7848795	7850554	DEL	85	1	IGF1R
CNVR4679	21	17426356	22332782	MIXED_CNVR	110	48	ABHD2 ACAN AEN AGBL1 ANPEP AP3S2 ARPIN BLM bta- mir-2363-1 bta-mir-9-1 CIB1 CRTC3 DET1 FANCI FES FURIN GPDGP1 HAPLN3 HDDC3 IDH2 IQGAP1 ISG20 KIF7 MAN2A2 MESP1 MESP2 MFGE8 MRPL46 MRPS11 NGRN NTRK3 PEX11A PLIN1 POLG PRC1 RCCD1 RHCG RLBP1 SEC11A SEMA4B TICRR UNC45A VPS33B WDR73 WDR93 ZNF710 ZNF774 ZSCAN2
CNVR4684	21	25087239	25891244	DEL	93	7	ANKRD34C bta-mir-184 CTSH MINAR1 MORF4L1 RASGRF1 TMED3
CNVR4696	21	29096744	29097744	DEL	88	1	PCSK6

CNVR4703	21	31118439	34933659	MIXED_CNVR	98	51	ARID3B bta-mir-631 C21H15orf39 CLK3 COMMD4 COX5A CPLX3 CSK CSPG4 CYP11A1 CYP1A1 CYP1A2 EDC3 ETFA FAM219B FBXO22 GZMB GZMH HMG20A ISL2 ISLR ISLR2 LINGO1 LMAN1L LOXL1 MAN2C1 MPI NEIL1 NRG4 ODF3L1 PEAK1 PML PPCDC PSTPIP1 PTPN9 RCN2 RPP25 SCAMP2 SCAMP5 SCAPER SEMA7A SIN3A SNUPN SNX33 STOML1 STRA6 TMEM266 TSPAN3 UBE2Q2 UBL7 ULK3
CNVR4705	21	35210358	35212183	DEL	77	1	STXBP6
CNVR4733	21	48650283	52078086	MIXED_CNVR	76	7	bta-mir-2365 FBXO33 GEMIN2 LRFN5 PNN SEC23A TRAPPC6B
CNVR4791	22	4536851	7623116	MIXED_CNVR	115	17	CCR4 CMTM6 CMTM7 CMTM8 CNOT10 CRTAP DYNC1LI1 FBXL2 GADL1 GLB1 GPD1L OSBPL10 RBMS3 STT3B SUSD5 TGFB2 TRIM71
CNVR4810	22	19277886	19279359	DEL	76	1	GRM7
CNVR4814	22	23377859	23379252	DEL	91	1	CNTN4
INVR704	22	42362589	42364033	INV	81	1	CFAP20DC
CNVR4853	22	43461307	43462425	DEL	110	1	SLMAP
CNVR4879	22	59170108	60603974	DEL	81	23	ABTB1 ALDH1L1 C22H3orf22 CHCHD6 CHST13 DNAJB8 EEFSEC GATA2 H1-10 KBTBD12 MCM2 MGLL PLXNA1 PODXL2 RAB7A RPN1 RUVBL1 SEC61A1 SLC41A3 TPRA1 TXNRD3 UROC1 ZXDC
CNVR4887	23	933557	934815	DEL	78	1	KHDRBS2
CNVR4906	23	5083179	5084574	DEL	88	1	GFRAL
CNVR4931	23	17751563	22144687	MIXED_CNVR	111	29	AARS2 ADGRF1 ADGRF2 ADGRF4 ADGRF5 ANKRD66 CD2AP CDC5L CLIC5 CYP39A1 ENPP4 ENPP5 HSP90AB1 MEP1A NFKBIE OPN5 OR5M10C OR9G1 PLA2G7 PTCHD4 RCAN2 RUNX2 SLC25A27 SLC35B2 SPATS1 SUPT3H TDRD6 TMEM151B TNFRSF21
CNVR4941	23	25867738	26691576	MIXED_CNVR	120	1	BTNL2

CNVR4956	23	28513221	28765192	MIXED_CNVR	82	4	<i>BOLA BOLA-NC1 bta-mir-11979 JSP.1</i>
CNVR4958	23	29208078	29318128	MIXED_CNVR	79	9	<i>OR12D18 OR12D2 OR12D23 OR12D2E OR12D2F OR12D2H OR105P OR106P OR109</i>
CNVR4969	23	30396882	30401578	DEL	99	1	<i>PGBD1</i>
CNVR4989	23	37062707	37065004	DEL	97	1	<i>CDKAL1</i>
CNVR5060	24	12525699	14130559	DEL	114	2	<i>PIK3C3 SYT4</i>
CNVR5083	24	24436015	24437288	DEL	84	1	<i>KLHL14</i>
CNVR5111	24	43445360	43446582	DEL	91	1	<i>LDLRAD4</i>
CNVR5133	24	58041687	59371493	MIXED_CNVR	89	10	<i>CCBE1 CPLX4 GRP LMAN1 MC4R OACYL PMAIP1 RAX SEC11C ZNF532</i>
CNVR5134	24	60556227	60559438	DEL	76	1	<i>PIGN</i>
CNVR5170	25	15830161	15832135	DEL	81	1	<i>XYLT1</i>
CNVR5187	25	25242174	25250748	DEL	95	1	<i>KATNIP</i>
CNVR5188	25	27594013	27596342	DEL	89	1	<i>SEPTIN14</i>
CNVR5190	25	28171406	28173745	DEL	85	1	<i>TPST1</i>
CNVR5204	25	36151395	36152690	DEL	77	1	<i>PILRA</i>
CNVR5253	26	6613196	10489197	MIXED_CNVR	131	15	<i>A1CF ATAD1 CSTF2T DKK1 LIPF LIPJ LIPK LIPM LIPN MINPP1 PAPSS2 PRKG1 PTEN RNLS SGMS1</i>
CNVR5267	26	17576412	17584971	DEL	95	1	<i>BLNK</i>
CNVR5280	26	27018982	30709942	DEL	131	4	<i>ADD3 bta-mir-2285k-4 SORCS1 XPNPEP1</i>
CNVR5288	26	35735429	36116883	MIXED_CNVR	123	1	<i>ATRNL1</i>
CNVR5314	26	50595409	51803308	MIXED_CNVR	82	5	<i>ADGRA1 BNIP3 INPP5A JAKMIP3 NKX6-2 MCPH1</i>
CNVR5331	27	5507991	5510307	DEL	94	1	<i>5_8S_rRNA bta-mir-2887-1</i>
CNVR5333	27	6182772	7283835	MIXED_CNVR	91	10	<i>DEFB DEFB1 DEFB10 DEFB13 DEFB4A DEFB7 EBD LAP</i>
CNVR5389	27	30794266	31009537	MIXED_CNVR	121	1	<i>UNC5D</i>
CNVR5410	27	40100667	43933251	MIXED_CNVR	112	9	<i>bta-mir-6532 NKIRAS1 NR1D2 RARB RPL15 THRB TOP2B UBE2E1 UBE2E2</i>
CNVR5425	28	591203	3295779	MIXED_CNVR	112	11	<i>ABCB10 ACTA1 CCSAP FMN2 GALNT2 NUP133 PGBD5 RAB4A RHOU TAF5L URB2</i>
CNVR5427	28	3891267	5250113	MIXED_CNVR	97	9	<i>C28H1orf131 DISC1 EGLN1 EXOC8 GNPAT SIPA1L2 SPRTN TRIM67 TSNAX</i>

CNVR5437	28	9398026	9399760	DEL	79	1	ACTN2
CNVR5453	28	14278032	16435472	DEL	117	8	ANK3 BICC1 bta-mir-2403 CCDC6 FAM13C MRLN PHYHIPL SLC16A9
CNVR5470	28	22790966	22793313	DEL	80	1	CTNNA3
CNVR5478	28	24287182	24289948	DEL	102	1	DNAJC12
CNVR5491	28	32788513	32791170	DEL	97	1	KCNMA1
CNVR5505	28	41143208	42971357	MIXED_CNVR	125	19	ADIRF ANXA8L1 ARHGAP22 BMPR1A FRMPD2 GDF10 GDF2 GLUD1 GPRIN2 LDB3 MAPK8 MMRN2 OPN4 PPYR1 RBP3 SHLD2 SNCG SYT15 WAPL
CNVR5512	29	1023317	1024930	DEL	77	1	MED17
CNVR5540	29	6180423	6181717	DEL	100	1	NOX4
CNVR5573	29	18618231	20327355	MIXED_CNVR	139	2	LUZP2 PAK1
CNVR5606	29	27137256	27598509	MIXED_CNVR	87	16	OR8B1 OR8B1AI OR8B1R OR8B1S OR8B1T OR8B1U OR8C26 OR8D2 OR8D6 OR8G3 OR8G3B OR8G3F OR8G47 OR8G5 OR8G5B TMEM225
CNVR5626	29	37880634	38025973	MIXED_CNVR	134	1	PAG7
CNVR5639	29	42809469	42810585	DEL	84	1	SLC22A11

Table S6. Enrichment analysis of genes overlapping the most frequent SVR: significant Panther Gene Ontology terms (GO) associated to biological process.

GO TERM	N GENES	P VALUE	GENES	DEFINITION
Sensory perception of chemical stimulus (GO:0007606)	57	9.90E-06	OR8G5, OR8J12, OR5T2, OR8J17, OR8G3F, OR8U9, OR8I2, OR8C26, OR8J16, OR9G3, OR8K60, OR5M3, OR8K66, OR8K62, OR12D23, OR8B1AI, OR5AL2, OR8K1B, OR5T1C, OR8K5, OR8G3B, OR8K1, OR8B1S, OR8J3F, OR8B1T, OR8J19, OR5AL8, OR8G3, OR5M11, OR9G4, OR8K67, OR8B1R, OR9G4D, OR8K5B, OR2AH1, OR8H13, OR8B1U, OR8K64, OR8K3B, OR8U1, OR12D2H, OR5M13D, OR8J2E, OR8G47, OR8J3, OR5M10C, OR5M10, OR12D2, GNAS, OR8K63, OR8G5B, OR8U3, OR8D6, OR12D2E, OR8D2, OR12D18, OR8B1	The series of events required for an organism to receive a sensory chemical stimulus, convert it to a molecular signal, and recognize and characterize the signal.
Sensory perception (GO:0007600)	57	4.03E-04	OR8G5, OR8J12, OR5T2, OR8J17, OR8G3F, OR8U9, OR8I2, OR8C26, OR8J16, OR9G3, OR8K60, OR5M3, OR8K66, OR8K62, OR12D23, OR8B1AI, OR5AL2, OR8K1B, OR5T1C, OR8K5, OR8G3B, OR8K1, OR8B1S, OR8J3F, OR8B1T, OR8J19, OR5AL8, OR8G3, OR5M11, OR9G4, OR8K67, OR8B1R, OR9G4D, OR8K5B, OR2AH1, OR8H13, OR8B1U, OR8K64, OR8K3B, OR8U1, OR12D2H, OR5M13D, OR8J2E, OR8G47, OR8J3, OR5M10C, OR5M10, OR12D2, GNAS, OR8K63, OR8G5B, OR8U3, OR8D6, OR12D2E, OR8D2, OR12D18, OR8B1	The series of events required for an organism to receive a sensory stimulus, convert it to a molecular signal, and recognize and characterize the signal.
Nervous system process (GO:0050877)	61	2.63E-03	OR8G5, OR8J12, OR5T2, OR8J17, OR8G3F, OR8U9, OR8I2, OR8C26, TNNT3, OR8J16, OR9G3, OR8K60, TNNT1, OR5M3, OR8K66, OR8K62, OR12D23, OR8B1AI, OR5AL2, OR8K1B, OR5T1C, OR8K5, OR8G3B, OR8K1, OR8B1S, OR8J3F, OR8B1T, OR8J19, OR5AL8, OR8G3, OR5M11, OR9G4, OR8K67, OR8B1R, OR9G4D, OR8K5B, OR2AH1, OR8H13, OR8B1U, OR8K64, OR8K3B, OR8U1, OR12D2H, OR5M13D, OR8J2E, CHRNA1, OR8G47, OR8J3, OR5M10C, OR5M10, OR12D2, GNAS, OR8K63, OR8G5B, OR8U3, OR8D6, OR12D2E, OR8D2, OR12D18, PLCB1, OR8B1	An organ system process carried out by any of the organs or tissues of neurological system.
G protein-couple receptor signalling pathway (GO:0007186)	72	1.76E-10	OR8G5, OR8J12, OR5T2, OR8J17, OR8G3F, OR8U9, OR8I2, OR8C26, OR8J16, PPYR1, OR9G3, OR8K60, OR5M3, ADGRF1, OR8K66, OR8K62, GRM7, HTR6, OR8B1AI, GPR149, OR5AL2, OR8K1B, OR5T1C, OR8K5, OR8G3B, OR8K1, OR8B1S, OR8J3F, OR8B1T, OR8J19, OR5AL8, OPN5, OR8G3, OR5M11, OR9G4, ADGRF5, OR8K67, OR8B1R, MC3R, RXFP3, MC4R, GNAQ, OR9G4D, OR8K5B, OR8H13, OR8B1U, PTGER4, OR8K64, OR8K3B, ADGRF4, GRP, PTGER4, OR8U1, OR5M13D, OR8J2E, OR8G47, ADGRF2, OR8J3, OR5M10C, FSHR, OR5M10, GNAS, GNA14, OR8K63, FSHR, ADGRB3, OR8G5B, OPN4, OR8U3, OR8D6, OR8D2, GRM6, TPRA1, OR8B1	The series of molecular signals initiated by a ligand binding to its receptor, in which the activated receptor promotes the exchange of GDP for GTP on the alpha-subunit of an associated heterotrimeric G-protein complex. The GTP-bound activated alpha-G-protein then dissociates from the beta- and gamma-subunits to further transmit the signal within the cell. The pathway begins with receptor-ligand interaction, and ends with regulation of a downstream cellular process. The pathway can start from the plasma membrane, Golgi or nuclear membrane.

Signal transduction (GO:0007165)	184	4.76E-11	OR8G5, OR8J12, TFPT, GRB14, OR5T2, NYAP2, OR8J17, OR7A112, PIK3C3, LAP, RACGAP1, OR7A75, ITGA10, OR8G3F, OR8U9, NR1D2, ABHD2, PRLR, OR8I2, OR8C26, RCAN2, OR8J16, OR7A114, NFAM1, NPR3, BCL2L1, PPYR1, SMURF2, RARB, SEMA5A, OR9G3, OR8K60, EFN2, OR7A88, PRKAA1, OR7A78, OR5M3, ADGRF1, OR8K66, RAPGEF6, GRM7, OR8K62, PLC1, HTR6, TGFB2, SCUBE2, OR8B1A, USP8, GPR149, OR5AL2, PIK3R1, PLXNA1, OR8K1B, OR7A99, INPP5A, OR5T1C, OR8K5, OR8G3B, WIPF1, OR8K1, OR8B1S, DEFB124, NPR3, OR8J3F, OR8B1T, RAD1, BRSK1, ARHGAP8, OR8J19, OR7A90, OR7A102, OR5AL8, OPN5, MAPK8, EPHA3, OR8G3, SUS5, OR5M11, PAK1, OR9G4, ADGRF5, OR7A135, CASS4, OR8K67, BLNK, OR7A103, OR7A97, OR8B1R, MC3R, RXFP3, BCL9, OR7A10, MC4R, PRR5, PRLR, PDE11A, MAML1, RICTOR, RAP1A, GNAQ, TOM1L1, IGF1R, LATS1, NKIRAS1, BNIP3, HJV, LIFR, OR9G4D, OR8K5B, IGF1R, RASGRF1, SEMA4B, OR8H13, OR8B1U, IRS2, PTGER4, OR8K64, OR8K3B, ADGRF4, ACVR2A, GRP, ADGRA1, PTGER4, OR8U1, OR5M13D, SEMA7A, OR8J2E, IL7R, BMPR2 CCR4, CHRN1, OR7A79, FGF7, OR8G47, ADGRF2, OR8J3, OR7A126, OR5M10C, BTNL2, PRKAB2, FSHR, EPS8L1, OR5M10, FYB1, DGKB, OR7A89, SCUBE1, PIK3C2G, GNAS, GDF2, OR7A98, OR7A95, ESR1, GNA14, ROR2, LAP, NFATC2, MAML2, IQGAP1, OR8K63, FSHR, OR7A83, SYK, ADGRB3, OR8G5B, STXBP4, OPN4, BMP7, HCK, THRB, DUSP15, OR8U3, SHC4, OR8D6, PDE5A, BNIP3, TICRR, OR7A77, OR8D2, RHOU, MAML2, NTRK3, GRM6, TPRA1, RPTOR, CPNE3, BMPR1A, PTEN, EPHB1, PLCB1, OR8B1, GRB14, OR109	The cellular process in which a signal is conveyed to trigger a change in the activity or state of a cell. Signal transduction begins with reception of a signal (e.g. a ligand binding to a receptor or receptor activation by a stimulus such as light), or for signal transduction in the absence of ligand, signal-withdrawal or the activity of a constitutively active receptor. Signal transduction ends with regulation of a downstream cellular process, e.g. regulation of transcription or regulation of a metabolic process. Signal transduction covers signaling from receptors located on the surface of the cell and signaling via molecules located within the cell. For signaling between cells, signal transduction is restricted to events at and within the receiving cell.
Cellular response to stimulus (GO:0051716)	31	1.01E-10	GRB14, LAP, NR1D2, ABHD2, SESN3, RARB, TGFB2, PIK3R1, CYP11A1, DEFB124, CYP11A1, PCK1, RAP1A, GNAQ, HJV, IRS2, PTGER4, ACVR2A, PTGER4, BMPR2, PCK1, FGF7, FSHR, GNAS, GDF2, ESR1, GNA14, LAP, FSHR, STXBP4, BMP7, THRB, NTRK3, RPTOR, BMPR1A, TMEM100, GRB14	Any process that results in a change in state or activity of a cell (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of a stimulus. The process begins with detection of the stimulus by a cell and ends with a change in state or activity of the cell.

Cellular process
(GO:0009987)

494

1.23E-03

NADK2, NUDT17, CYP2D14, PDP1, EEFSEC, OR8G5, GATA2, OR8J12, ERG, TFPT, PAPSS2, GRB14, PPM1H, OR5T2, ADD3, ATP8B4, NYAP2, SLC38A4, OR8J17, HDC, CANX, TAF5L, PPFBP1, HIBCH, SNX33, POLDIP3, SKP2, SLC1A3, OR7A112, NAF1, TMEM67, ALCAM, PIK3C3, MEI1, DEFB1, XRCC6, RAD21L1, KATNA1, PDZK1, CIBAR1, A4GALT, EXOC8, ULK3, CRTAP, LAP, RACGAP1, OR7A75, STRA6, TUBB1, PTPRH, ITGA10, OR8G3F, AURKA, AUH, NBEA, SLC4A8, OR8U9, PRUNE2, NR1D2, ABHD2, TM9SF4, ANPEP, PRLR, CTNNA1, PEAK1, OR8I2, RAB7A, CDC5L, OR8C26, CDK14, RBM8A, GJA8, RCAN2, SLC38A1, UNC45A, OR8J16, HIGD1C, HAPLN3, OR7A114, PTPDC1, VPS26C, NFE2L2, COL28A1, TTL1, NFAM1, NPR3, BCL2L1, PPYR1, NRCAM, SMURF2, SESN3, RARB, SUPT3H, SEMA5A, XKR7, OR9G3, TOP2B, GRIK1, OR8K60, LMBRD1, URB2, SEC23A, PTPDC1, EFN2, UNC5D, OR7A88, TNNT1, TBC1D20, NSFL1C, CFAP54, PRKAA1, TPST1, OR7A78, OR5M3, CTNNA1, OR8K66, ADGRF1, TARS1, TXNIP, NDUFAF2, TSEN34, RAPGEF6, GRM7, OR8K62, ID1, ARFGAP3, RANBP17, RPS9, NDUFAF2, PLA2G2A, MTMR2, LIG4, PLCL1, CNOT10, HTR6, DEFB13, NEIL1, TGFB2, LRATD1, RRP7, CAPZB, POLG, MRPL44, SCUBE2, OR8B1A1, PLA2G5, STX8, PPP2R1A, LNP, SEC61A1, KCNMA1, FAT4, USP8, BRX1, TRAPPC6B, GPR149, RPP25, PPP2R1A, FMN2, OR5AL2, DNAH14, SEPTIN14, NUP155, CPLX3, LIMA1, PLD1, LOXL1, TMED3, RUNX2, SIGLEC10, EYA3, PIK3R1, ARHGEF7, THRB, ANO6, H1-10, CERS5, CYP11A1, PLA2G2D1, LAP, PLXNA1, LAP, ATAD1, CDH17, OR8K1B, OR7A99, RTF2, SBK2, INPP5A, OR5T1C, PACSIN2, OR8K5, OR8G3B, WIPF1, RBIS, OR8K1, SDHAF4, OR8B1S, DEFB124, SEC23A, UROC1, SNX31, NPR3, OR8J3F, OR8B1T, RAD1, OXCT1, CD24, EDC3, BRSK1, COX4I2, ARHGAP8, OR8J19, OR7A90, DMT1, CARS2, OR7A102, OR5AL8, RTF2, SCAMP5, OPN5, MAPK8, RHCG, EPHA3, COL15A1, OR8G3, PIGN, PARVG, SUSDS, ISG20, OR5M11, PLEKHA3, GEMIN2, CNOT3, PSTPIP1, SLC38A2, ETS2, ZC3H12D, BIN2, PAK1, OR9G4, ADGRF5, CPLX4, OR7A135, PIGP, CASS4, ACP6, CHAF1B, OR8K67, EBD, CMTM8, ATP5F1E, ZRANB3, BLNK, ASIC1, OR7A103, SATB2, TTC30B, OR7A97, OR8B1R, KCNJ6, MC3R, DOCK10, RXFP3, PIK3R4, GOLPH3, FOXS1, BCL9, GJA5, SLITRK3, OR7A10, MC4R, PRR5, FRMPD2, LDB3, PPCDC, ATP6V0D2, PLA2G2E, COL4A2, PRLR, PDE11A, CYP11A1, CD2AP, MAM1, ATP9A, LRRC4C, EGLN1, STX16, CSNK2A1, PCK1, RICTOR, RAPIA, GNAQ, TOM1L1, IGF1R, LMBRD1, FSTL1, RAB20, LATS1, KCNS1, PHETA2, TNFRSF21, NKIRAS1, BNIP3, CDC42EP5, FGGY, HJV, LIFR, OR9G4D, SCN8A, MCPH1, ATP5MC3, MGC127055, NFE2L2, SCAF11, KIF7, OR8K5B, FOXB2, IGF1R, GLUD1, RASGRF1, PRKRA, TTL9, SEMA4B, SEPTIN3, IPO11, CSK, RUFY1, SCAMP2, OR8H13, ZRANB3, SULF4A1, HIBCH, OR8B1U, DYRK1A, SLC11A2, ANPEP, IRS2, ETFA, PEX11A, CAPZB, COX5A, SGPP2, PHETA2, ADAMTS12, NKX6-2, SGMS1, GPD1L, INTS8, PTGER4, CLK3, OR8K64, KIF2A, OR8K3B, LOXL1, ADGRF4, AARS2, ACVR2A, RHCG, FAM135A, CLDN14, GRP, BLM, ADGRA1, MCM2, DYNC1LI1, FES, DEFB10, PTGER4, OR8U1, SLMAP, NAXD, ALCAM, MRTO4, PLA2G2A, OR5M13D, RAB7A, HSP90AB1, SEMA7A, OR8J2E, LMAN1L, IL7R, MBOAT7, BMPR2, CDKAL1, SMC1B, DISC1, DEFB7, CYCT, COL4A1, MCAT, SCAMP5, PCK1, RBM8A, SEC22B, CCR4, OR7A79, EMC1, CHRNA1, FGF7, FARSB, TMEM86B, OR8G47, ADGRF2, HOOK1, OR8J3, OR7A126, RBIS, OR5M10C, BTNL2, PRKAB2, GLUD1, FSHR, EPS8L1, OR5M10, NOX4, SNCG, FYB1, DGKB, KCNE4, OR7A89, SCUBE1, PIK3C2G, SECISBP2L, DEFB10, GNAS, CWC22, GDF2, OR7A98, OR7A95, ISL2, ESR1, DNAJB8, ENTPD4, GNA14, SYNDIG1, ROR2, LAP, ETFA, CARS2, NFATC2, ADAMTS2, GPD1, WDR73, TNIP3, MAM2, IQGAP1, SMDT1, STRA6, FSHR, OR7A83, OR8K63, RANBP3L, SYK, ADGRB3, OR8G5B, WDFY3, KCNH1, STXBP4, OPN4, PDE4DIP, BMP7, LMAN1, TTC30A, HCK, THRB, DUSP15, SPRTN, OR8U3, SHC4, VPS33B, KIF3B, OR8D6, PEX11B, RAE1, CEP83, XYLT1, PDE5A, CD24, MRPL44, BNIP3, TICRR, OR7A77, GPRIN2, MOGAT1, NUP133, TMEM190, VIRMA, ACAN, OR8D2, PARVB, ACTN2, RHOU, GJA5, CNTN4, NLGN1, HAS1, PDZK1, SNPH, MPI, SPTLC1, CNIH3, SYT5, MAM2, NALF1, SYT15, HNRNPA3, MTMR12, GPD1, NTRK3, SOX12, PTPN9, BCAS1, SEPTIN14, ACSL3, GRM6, PIP4P2, TPRA1, RPTOR, CPNE3, NCKAP5, VAPB, BMPR1A, SYT4, PTEN, CTNNA3, MTX2, EPHB1, PLCB1, RPL15, PRC1, CEPT1, OR8B1, GRB14, ACAN, OR109, PRPF31, RUVBL1, ETV1, NIPBL, KCNJ15

Any process that is carried out at the cellular level, but not necessarily restricted to a single cell. For example, cell communication occurs among more than one cell, but occurs at the cellular level.

Response to stimulus (GO:0050896)	236	2.72E-05	CYP2D14, OR8G5, OR8J12, TFPT, GRB14, OR5T2, NYAP2, OR8J17, CANX, C7, OR7A112, PIK3C3, DEFB1, XRCC6, RAD21L1, ULK3, LAP, RACGAP1, OR7A75, ITGA10, OR8G3F, OR8U9, NR1D2, ABHD2, PRLR, OR8I2, OR8C26, RCAN2, OR8J16, OR7A114, NFE2L2, NFAM1, NPR3, BCL2L1, PPYR1, SMURF2, NRCAM, SESN3, RARB, SEMA5A, OR9G3, OR8K60, EFN2, UNC5D, OR7A88, PRKAA1, OR7A78, OR5M3, ADGRF1, OR8K66, RAPGEF6, GRM7, OR8K62, OR12D23, SLC25A27, LIG4, PLCL1, HTR6, DEFB13, NEIL1, TGFB2, SCUBE2, OR8B1A1, USP8, GPR149, OR5AL2, PIK3R1, THRB, CYP11A1, PLXNA1, LAP, OR8K1B, OR7A99, INPP5A, OR5T1C, OR8K5, OR8G3B, WIPF1, OR8K1, OR8B1S, DEFB124, NPR3, OR8J3F, OR8B1T, RAD1, BRSK1, ULBP17, ARHGAP8, OR8J19, OR7A90, OR7A102, OR5AL8, OPN5, MAPK8, BOLA, EPHA3, OR8G3, SUSP5, ISG20, OR5M11, PAK1, OR9G4, ADGRF5, OR7A135, CASS4, OR8K67, EBD, ZRANB3, BLNK, OR7A103, OR7A97, OR8B1R, MC3R, RXFP3, BCL9, OR7A10, MC4R, PRR5, NDUFA6, PRLR, PDE11A, CYP11A1, MAML1, EGLN1, PCK1, RICTOR, RAP1A, GNAQ, TOM1L1, IGF1R, LATS1, TNFRSF21, NKIRAS1, BNIP3, C9, HJV, LIFR, OR9G4D, MGC127055, NFE2L2, OR8K5B, IGF1R, RASGRF1, OR2AH1, SEMA4B, ZRANB3, OR8H13, OR8B1U, SLC11A2, BOLA-NC1, IRS2, PTGER4, OR8K64, OR8K3B, ADGRF4, ACVR2A, GRP, BLM, ADGRA1, MCM2, DEFB10, FES, PTGER4, OR8U1, OR12D2H, OR5M13D, HSP90AB1, SEMA7A, OR8J2E, IL7R, BMPR2, NDUFA6, DEFB7, PCK1, CCR4, CHRNA1, OR7A79, FGF7, OR8G47, ADGRF2, OR8J3, OR7A126, OR5M10C, BTNL2, PRKAB2, FSHR, EPS8L1, OR5M10, NOX4, FYB1, DGKB, OR12D2, OR7A89, SCUBE1, PIK3C2G, DEFB10, GNAS, GDF2, OR7A98, OR7A95, ESR1, GNA14, ROR2, LAP, NFATC2, DEFB119, TNIP3, MAML2, IQGAP1, ULBP21, OR8K63, FSHR, OR7A83, SYK, ADGRB3, OR8G5B, STXBP4, OPN4, BMP7, HCK, THRB, DUSP15, SPRTN, C7, OR8U3, SHC4, OR8D6, PDE5A, BNIP3, TICRR, OR7A77, OR12D2E, OR8D2, RHOU, CNTN4, SYT5, MAML2, SYT15, OR12D18, NTRK3, GRM6, TPRA1, RPTOR, CPNE3, BMPR1A, SYT4, TMEM100, PTEN, EPHB1, PLCB1, BOLA, OR8B1, GRB14, OR109, NIPBL	Any process that results in a change in state or activity of a cell or an organism (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of a stimulus. The process begins with detection of the stimulus and ends with a change in state or activity of the cell or organism.
--------------------------------------	-----	----------	---	--

Regulation of cellular process (GO:0050794)	352	9.80E-06	ZNF774, EEFSEC, HOXD1, OR8G5, GATA2, ERG, OR8J12, TFPT, GRB14, OR5T2, ADD3, ATF1, NYAP2, OR8J17, PPP6R1, CCSAP, TAF5L, SKP2, RBM38, OR7A112, JUN, PIK3C3, TCF20, SMAP1, ULK3, LAP, RACGAP1, OR7A75, ITGA10, OR8G3F, AURKA, KHDRBS2, MEOX2, OR8U9, NR1D2, ABHD2, PRLR, OR8I2, OR8C26, BARX1, CDK14, SREBF2, PHF2, HOXD8, RCAN2, ZNF613, OR8J16, HIGD1C, HAPLN3, OR7A114, SOSTDC1, RAX, ARPIN, MESP2, NFE2L2, WBP2NL, NFAM1, NPR3, BCL2L1, PPYR1, SMURF2, SESN3, RARB, SEMA5A, OR9G3, GRIK1, OR8K60, SHISA7, EFNB2, ING1, OR7A88, JUN, DKK1, ZNF354C, PRKAA1, OR7A78, OR5M3, ADGRF1, OR8K66, RAPGEF6, GRM7, OR8K62, ID1, WAPL, TFAP2C, PLCL1, HOXD11, CNOT10, ARHGEF28, POFUT1, SMARCD1, HTR6, TGFB2, PMEP1, CAPZB, SCUBE2, OR8B1A, PLA2G5, USP8, GPR149, OR5AL2, CPLX3, PLD1, RUNX2, EYA3, PIK3R1, H1-10, ZNF579, PLXNA1, OR8K1B, OR7A99, INPP5A, OR5T1C, PACSIN2, OR8K5, OR8G3B, HMGA2, WIPF1, OR8K1, POU6F1, OR8B1S, DEFB124, NPR3, OR8J3F, WDFY1, OR8B1T, SIM2, RAD1, BRSK1, CD24, ARHGAP8, OR8J19, SALL4, OR7A90, TNFSF13B, OR7A102, OR5AL8, OPN5, MESP1, MAPK8, ZNF524, EPHA3, OR8G3, SUSD5, OR5M11, ARID3B, CNOT3, ETS2, PAK1, OR9G4, RAB4A, ADGRF5, CPLX4, OR7A135, CASS4, OR8K67, NELFCD, BLNK, OR7A103, SATB2, OR7A97, OR8B1R, MC3R, KCNJ6, DOCK10, HOXD3, RXFP3, FOXS1, BCL9, SLITRK3, OR7A10, MC4R, PRR5, BACH2, TFCP2, CSRN3, TRAT1, SIN3A, PRLR, PDE11A, MAML1, LRRC4C, TNFRSF13C, CSNK2A1, RICTOR, RAP1A, GNAQ, TOM1L1, IGF1R, FSTL1, LATS1, KCNS1, MCC, TNFRSF21, NKIRAS1, BNIP3, ZNF581, CDC42EP5, ZNF614, HJV, LIFR, OR9G4D, NFE2L2, OR8K5B, FOXB2, IGF1R, RASGRF1, PRKRA, ZXDC, SEMA4B, TRIB2, HOXD9, CSK, HOXD4, RUFY1, ZNF432, OR8H13, OR8B1U, DYRK1A, IRS2, CAPZB, TRAT1, GDF10, ZNF879, TRIB3, NKX6-2, SHISA8, PTGER4, CLK3, OR8K64, OR8K3B, ADGRF4, ACVR2A, GRP, ADGRA1, FES, PTGER4, OR8U1, SLMAP, PIAS3, ZNF217, CRT3, HBP1, OR5M13D, SEMA7A, OR8J2E, SHLD2, SP110, IL7R, BMPR2, ZNF175, KCNAB1, LARP4, CCR4, CHRNA1, OR7A79, FGF7, ZFP64, OR8G47, ADGRF2, OR8J3, DPP10, OR7A126, OR5M10C, BTNL2, CSRN2, PRKAB2, FSHR, EPS8L1, OR5M10, SNCG, FYB1, DGKB, ZNF350, KCNE4, OR7A89, MINAR1, SCUBE1, ZNF677, TSHZ2, PIK3C2G, HOXD10, SECISBP2L, ZFP64, GNAS, RIPPLY3, GDF2, OR7A98, ISL2, OR7A95, ESR1, GNA14, SYNDIG1, EVX2, ROR2, ZNF350, LAP, PAX3, NFATC2, RBCK1, TNIP3, HLF, MAML2, IQGAP1, OR8K63, FSHR, OR7A83, SYK, ADGRB3, OR8G5B, SCRT2, STXBP4, OPN4, PDE4DIP, BMP7, PLAGL2, FIZ1, HCK, THRB, DUSP15, HBP1, OR8U3, SHC4, TLE4, OR8D6, AHR, SMARCD1, PDE5A, CD24, BNIP3, TICRR, OR7A77, TAB2, RUNX1T1, GABPB1, ZNF710, ACAN, OR8D2, RHOU, NLGN1, SYT5, CNIH3, MAML2, SYT15, ZNF454, TCF15, KCNIP4, NTRK3, SOX12, GRM6, TPRA1, RPTOR, DAB2, HOXD13, CPNE3, BMPR1A, SYT4, PTEN, EPHB1, PLCB1, OR8B1, GRB14, LDLRAD4, ACAN, GDNF, OR109, RUVBL1, ETV1, KCNJ15, MED17, FKBP1A
--	-----	----------	---

Any process that modulates the frequency, rate or extent of a cellular process, any of those that are carried out at the cellular level, but are not necessarily restricted to a single cell. For example, cell communication occurs among more than one cell, but occurs at the cellular level

Regulation of biological process (GO:0050789)	369	6.00E-05	ZNF774, HOXD1, EESEC, OR8G5, GATA2, OR8J12, ERG, TFPT, GRB14, OR5T2, ADD3, ATF1, NYAP2, OR8J17, PPP6R1, CCSAP, TAF5L, C7, SKP2, OR7A112, RBM38, JUN, PIK3C3, TCF20, SMAP1, ULK3, LAP, RACGAP1, OR7A75, ITGA10, OR8G3F, AURKA, KHDRBS2, MEOX2, OR8U9, NR1D2, ABHD2, PRLR, OR8I2, BARX1, OR8C26, CDK14, GJA8, SREBF2, PHF2, HOXD8, RCAN2, ZNF613, OR8J16, HIGD1C, HAPLN3, OR7A114, SOSTDC1, RAX, ARPIN, MESP2, NFE2L2, WBP2NL, NFAM1, NPR3, BCL2L1, TRIM11, PPYR1, SMURF2, SESN3, RARB, SEMA5A, OR9G3, GRIK1, OR8K60, SHISA7, EFN2, ING1, OR7A88, JUN, NLRP2, DKK1, ZNF354C, PRKAA1, OR7A78, OR5M3, OR8K66, ADGRF1, RAPGEF6, GRM7, OR8K62, ID1 WAPL, TFAP2C, PLCL1, HOXD11, CNOT10, ARHGEF28, SMARCD1, POFUT1, HTR6, TGFB2, PMEPA1, CAPZB, MRPL44, SCUBE2, OR8B1A, PLA2G5, USP8, GPR149, OR5AL2, CPLX3, PLD1, RUNX2, EYA3, PIK3R1, THRB, HMG20A, H1-10, ZNF579, PLXNA1, OR8K1B, OR7A99, INPP5A, OR5T1C, PACSIN2, OR8K5, OR8G3B, HMG2, WIPF1, OR8K1, POU6F1, OR8B1S, DEFB124, NPR3, OR8J3F, WDFY1, OR8B1T, SIM2, RAD1, CD24, EDC3, BRSK1, ARHGAP8, SALL4, OR8J19, OR7A90, TNFSF13B, OR7A102, OR5AL8, OPN5, MESP1, MAPK8, ZNF524, EPHA3, OR8G3, SUSD5, ISG20, OR5M11, ARID3B, CNOT3, ETS2, PAK1, OR9G4, ADGRF5, RAB4A, CPLX4, OR7A135, CASS4, OR8K67, NELFCD, BLNK, OR7A103, SATB2, OR7A97, OR8B1R, KCNJ6, MC3R, DOCK10, HOXD3, RXFP3, FOXS1, BCL9, TRIM17, GJA5, SLITRK3, OR7A10, MC4R, PRR5, BACH2, TRIM17, TFCP2, CSRNP3, TRAT1, SIN3A, PRLR, PDE11A, MAML1, LRRC4C, TNFRSF13C, CSNK2A1, RICTOR, RAP1A, GNAQ, TOM1L1, IGF1R, FSTL1, LATS1, KCNS1, MCC, TNFRSF21, NKIRAS1, DET1, BNIP3, ZNF581, CDC42EP5, C9, ZNF614, HJV, LIFR, OR9G4D, NFE2L2, OR8K5B, FOXB2, IGF1R, RASGRF1, PRKRA, ZXDC, RNF187, SEMA4B, TRIB2, HOXD9, CSK, HOXD4, RUFY1, ZNF432, OR8H13, OR8B1U, DYRK1A, SERPINE2, IRS2, IGF2BP2, CAPZB, TRAT1, GDF10, ZNF879, TRIB3, NKX6-2, SHISA8, PTGER4, CLK3, OR8K64, OR8K3B, ADGRF4, ACVR2A, GRP, ADGRA1, FES, PTGER4, OR8U1, SLMAP, PIAS3, ZNF217, CRT3, HBP1, MRTO4, OR5M13D, SEMA7A, OR8J2E, SHLD2, SP110, IL7R, BMPR2, ZNF175, KCNAB1, CYCT, LARP4, CCR4, OR7A79, CHRNA1, FGF7, ZFP64, OR8G47, ADGRF2, OR8J3, DPP10, OR7A126, OR5M10C, BTNL2, CSRNP2, PRKAB2, FSHR, EPS8L1, OR5M10, SNCG, FYB1, DGKB, ZNF350, KCNE4, OR7A89, SCUBE1, MINAR1, ZNF677, TSHZ2, PIK3C2G, HOXD10, SECISBP2L, ZFP64, GNAS, RIPPLY3, GDF2, OR7A98, OR7A95, ISL2, ESR1, GNA14, SYNDIG1, EVX2, ROR2, ZNF350, LAP, PAX3, NFATC2, RBCK1, TNIP3, HLF, MAML2, IQGAP1, FSHR, OR7A83, OR8K63, SYK, ADGRB3, OR8G5B, SCRT2, STXBP4, OPN4, PDE4DIP, BMP7, PLAGL2, FIZ1, HCK, THRB, DUSP15, C7, HBP1, OR8U3, SHC4, TLE4, OR8D6, AHR, SMARCD1, PDE5A, CD24, MRPL44, BNIP3, TICRR, OR7A77, TAB2, RUNX1T1, GABPB1, ZNF710, ACAN, OR8D2, RHOU, GJA5, NLGN1, CNIH3, SYT5, TRIM11, MAML2, SYT15, ZNF454, TCF15, KCNIP4, NTRK3, SOX12, GRM6, TPRA1, RPTOR, DAB2, HOXD13, CPNE3, BMPR1A, SYT4, PTEN, EPHB1, PLCB1, OR8B1, GRB14, LDLRAD4, ACAN, GDNF, OR109, RUVBL1, ETV1, KCNJ15, MED17, FKBP1A
--	-----	----------	--

Any process that modulates the frequency, rate, or extent of a biological process. Biological processes are regulated by many means; examples include the control of gene expression, protein modification or interaction with a protein or substrate molecule.

Biological regulation (GO:0065007)	386	2.99E-05	ZNF774, HOXD1, EEFSEC, OR8G5, GATA2, OR8J12, ERG, TFPT, GRB14, OR5T2, ADD3, ATP8B4, ATF1, NYAP2, OR8J17, PPP6R1, CCSAP, TAF5L, C7, SKP2, OR7A112, RBM38, JUN, PIK3C3, TCF20, SMAP1, ULK3, LAP, ARHGAP22, RACGAP1, OR7A75, ITGA10, OR8G3F, AURKA, KHDRBS2, MEOX2, SLC4A8, OR8U9, NR1D2, ABHD2, PRLR, OR8I2, BARX1, OR8C26, CDK14, GJA8, SREBF2, PHF2, HOXD8, RCAN2, FURIN, ZNF613, OR8J16, HIGD1C, HAPLN3, OR7A114, SOSTDC1, RAX, ARPIN, MESP2, NFE2L2, WBP2NL, NFAM1, NPR3, BCL2L1, TRIM11, PPYR1, SMURF2, SESN3, RARB, SEMA5A, XKR7, OR9G3, GRIK1, OR8K60, SHISA7, EFNB2, ING1, OR7A88, JUN, NLRP2, DKK1, ZNF354C, PRKAA1, OR7A78, ARHGEF26, OR5M3, OR8K66, ADGRF1, RAPGEF6, GRM7, OR8K62, ID1, WAPL, TFAP2C, PLCL1, FURIN, HOXD11, CNOT10, ARHGEF28, SMARCD1, POFUT1, HTR6, TGFB2, PMEPA1, CAPZB, MRPL44, SCUBE2, OR8B1AI, PLA2G5, USP8, GPR149, OR5AL2, CPLX3, PLD1, RUNX2, EYA3, PIK3R1, THRB, HMG20A, ANO6, H1-10, CYP11A1, ZNF579, PLXNA1, OR8K1B, OR7A99, INPP5A, OR5T1C, PACSIN2, OR8K5, OR8G3B, HMGA2, WIPF1, OR8K1, POU6F1, OR8B1S, DEFB124, NPR3, OR8J3F, WDFY1, OR8B1T, SIM2, RAD1, CD24, EDC3, BRSK1, ARHGAP8, SALL4, OR8J19, OR7A90, TNFSF13B, OR7A102, OR5AL8, OPN5, MESP1, MAPK8, ZNF524, EPHA3, OR8G3, SUSP5, ISG20, OR5M11, ARID3B, CNOT3, ETS2, PAK1, OR9G4, ADGRF5, RAB4A, CPLX4, OR7A135, CASS4, OR8K67, NELFCD, BLNK, OR7A103, SATB2, OR7A97, OR8B1R, KCNJ6, MC3R, DOCK10, HOXD3, RXFP3, FOXS1, BCL9, TRIM17, GJA5, SLITRK3, OR7A10, MC4R, PRR5, BACH2, TRIM17, TFCP2, CSRNP3, ATP6V0D2, TRAT1, SIN3A, PRLR, PDE11A, CYP11A1, MAML1, ATP9A, LRRC4C, TNFRSF13C, CSNK2A1, RICTOR, RAP1A, GNAQ, TOM1L1, IGF1R, FSTL1, LATS1, KCNS1, MCC, TNFRSF21, NKIRAS1, DET1, BNIP3, ZNF581, CDC42EP5, C9, ZNF614, HJV, LIFR, OR9G4D, NFE2L2, OR8K5B, FOXB2, IGF1R, RASGRF1, PRKRA, ZXDC, RNF187, SEMA4B, TRIB2, HOXD9, CSK, HOXD4, RUFY1, ZNF432, OR8H13, OR8B1U, DYRK1A, SERPINE2, IRS2, IGF2BP2, CAPZB, TRAT1, GDF10, ZNF879, TRIB3, NKX6-2, CHN1, SHISA8, PTGER4, CLK3, OR8K64, OR8K3B, ADGRF4, ACVR2A, SIPA1L2, GRP, ADGRA1, FES, PTGER4, OR8U1, SLMAP, PIAS3, ZNF217, CRTC3, HBP1, MRTO4, OR5M13D, HSP90AB1, SEMA7A, OR8J2E, SHLD2, SP110, IL7R, BMPR2, ZNF175, KCNAB1, CYCT, LARP4, CCR4, CHN1, OR7A79, CHRNA1, FGF7, ZFP64, OR8G47, ADGRF2, OR8J3, DPP10, OR7A126, OR5M10C, BTNL2, CSRNP2, PRKAB2, FSHR, EPS8L1, OR5M10, SNCG, FYB1, DGKB, ZNF350, KCNE4, OR7A89, SCUBE1, MINAR1, ZNF677, TSHZ2, PIK3C2G, USP6NL, HOXD10, SECISBP2L, ZFP64, GNAS, RIPPLY3, GDF2, OR7A98, OR7A95, ISL2, ESR1, GNA14, SYNDIG1, EVX2, ROR2, ZNF350, LAP, PAX3, NFATC2, RBCK1, TNIP3, HLF, MAML2, IQGAP1, FSHR, OR7A83, OR8K63, SYK, ADGRB3, OR8G5B, SCRT2, KCNH1, STXBP4, OPN4, PDE4DIP, BMP7, PLAGL2, FIZ1, HCK, THRB, DUSP15, C7, HBP1, OR8U3, SHC4, TLE4, OR8D6, AHR, SMARCD1, PDE5A, CD24, MRPL44, BNIP3, TICRR, OR7A77, TAB2, RUNX1T1, GABPB1, ZNF710, ACAN, OR8D2, RHOU, GJA5, NLGN1, CNIH3, SYT5, TRIM11, MAML2, TBC1D5, SYT15, ZNF454, TCF15, KCNIP4, NTRK3, SOX12, GRM6, TPRA1, RPTOR, DAB2, HOXD13, CPNE3, BMPR1A, SYT4, PCSK6, PTEN, EPHB1, PLCB1, OR8B1, GRB14, LDLRAD4, ACAN, GDNF, OR109, RUVBL1, ETV1, KCNJ15, MED17, FKBP1A	Any process that modulates a measurable attribute of any biological process, quality, or function.
---------------------------------------	-----	----------	--	--

Cell communication (GO:0007154)	195	5.53E-11	OR8G5, OR8J12, TFPT, GRB14, OR5T2, NYAP2, OR8J17, OR7A112, PIK3C3, LAP, RACGAP1, OR7A75, ITGA10, OR8G3F, OR8U9, NR1D2, ABHD2, PRLR, OR8I2, OR8C26, GJA8, RCAN2, OR8J16, OR7A114, NFAM1, NPR3, BCL2L1, PPYR1, SMURF2, SESN3, RARB, SEMA5A, OR9G3, GRIK1, OR8K60, EFNB2, OR7A88, PRKAA1, OR7A78, OR5M3, ADGRF1, OR8K66, RAPGEF6, GRM7, OR8K62, PLCL1, HTR6, TGFB2, SCUBE2, OR8B1A, USP8, GPR149, OR5AL2, CPLX3, PIK3R1, PLXNA1, OR8K1B, OR7A99, INPP5A, OR5T1C, OR8K5, OR8G3B, WIPF1, OR8K1, OR8B1S, DEFB124, NPR3, OR8J3F, OR8B1T, RAD1, BRSK1, ARHGAP8, OR8J19, OR7A90, OR7A102, OR5AL8, OPN5, MAPK8, EPHA3, OR8G3, SUSD5, OR5M11, PAK1, OR9G4, ADGRF5, CPLX4, OR7A135, CASS4, OR8K67, BLNK, OR7A103, OR7A97, OR8B1R, MC3R, RXFP3, BCL9, GJA5, OR7A10, MC4R, PRR5, PRLR, PDE11A, MAML1, RICTOR, RAP1A, GNAQ, TOM1L1, IGF1R, LATS1, NKIRAS1, BNIP3, HJV, LIFR, OR9G4D, OR8K5B, IGF1R, RASGRF1, SEMA4B, OR8H13, OR8B1U, IRS2, PTGER4, OR8K64, OR8K3B, ADGRF4, ACVR2A, GRP, ADGRA1, PTGER4, OR8U1, OR5M13D, SEMA7A, OR8J2E, IL7R, BMPR2, CCR4, CHRNA1, OR7A79, FGF7, OR8G47, ADGRF2, OR8J3, OR7A126, OR5M10C, BTNL2, PRKAB2, FSHR, EPS8L1, OR5M10, SNCG, FYB1, DGKB, OR7A89, SCUBE1, PIK3C2G, GNAS, GDF2, OR7A98, OR7A95, ESR1, GNA14, ROR2, LAP, NFATC2, MAML2, IQGAP1, OR8K63, FSHR, OR7A83, SYK, ADGRB3, OR8G5B, STXBP4, OPN4, BMP7, HCK, THRB, DUSP15, OR8U3, SHC4, OR8D6, PDE5A, BNIP3, TICRR, OR7A77, OR8D2, RHOU, GJA5, NLGN1, SYT5, MAML2, SYT15, NTRK3, GRM6, TPRA1, RPTOR, CPNE3, BMPR1A, SYT4, PTEN, EPHB1, PLCB1, OR8B1, GRB14, OR109	Any process that mediates interactions between a cell and its surroundings. Encompasses interactions such as signaling or attachment between one cell and another cell, between a cell and an extracellular matrix, or between a cell and any other aspect of its environment.
Signaling (GO:0023052)	194	2.30E-11	OR8G5, OR8J12, TFPT, GRB14, OR5T2, NYAP2, OR8J17, OR7A112, PIK3C3, LAP, RACGAP1, OR7A75, ITGA10, OR8G3F, OR8U9, NR1D2, ABHD2, PRLR, OR8I2, OR8C26, GJA8, RCAN2, OR8J16, OR7A114, NFAM1, NPR3, BCL2L1, PPYR1, SMURF2, RARB, SEMA5A, OR9G3, GRIK1, OR8K60, EFNB2, OR7A88, PRKAA1, OR7A78, OR5M3, ADGRF1, OR8K66, RAPGEF6, GRM7, OR8K62, PLCL1, HTR6, TGFB2, SCUBE2, OR8B1A, USP8, GPR149, OR5AL2, CPLX3, PIK3R1, PLXNA1, OR8K1B, OR7A99, INPP5A, OR5T1C, OR8K5, OR8G3B, WIPF1, OR8K1, OR8B1S, DEFB124, NPR3, OR8J3F, OR8B1T, RAD1, BRSK1, ARHGAP8, OR8J19, OR7A90, OR7A102, OR5AL8, OPN5, MAPK8, EPHA3, OR8G3, SUSD5, OR5M11, PAK1, OR9G4, ADGRF5, CPLX4, OR7A135, CASS4, OR8K67, BLNK, OR7A103, OR7A97, OR8B1R, MC3R, RXFP3, BCL9, GJA5, OR7A10, MC4R, PRR5, PRLR, PDE11A, MAML1, RICTOR, RAP1A, GNAQ, TOM1L1, IGF1R, LATS1, NKIRAS1, BNIP3, HJV, LIFR, OR9G4D, OR8K5B, IGF1R, RASGRF1, SEMA4B, OR8H13, OR8B1U, IRS2, PTGER4, OR8K64, OR8K3B, ADGRF4, ACVR2A, GRP, ADGRA1, PTGER4, OR8U1, OR5M13D, SEMA7A, OR8J2E, IL7R, BMPR2, CCR4, CHRNA1, OR7A79, FGF7, OR8G47, ADGRF2, OR8J3, OR7A126, OR5M10C, BTNL2, PRKAB2, FSHR, EPS8L1, OR5M10, SNCG, FYB1, DGKB, OR7A89, SCUBE1, PIK3C2G, GNAS, GDF2, OR7A98, OR7A95, ESR1, GNA14, ROR2, LAP, NFATC2, MAML2, IQGAP1, OR8K63, FSHR, OR7A83, SYK, ADGRB3, OR8G5B, STXBP4, OPN4, BMP7, HCK, THRB, DUSP15, OR8U3, SHC4, OR8D6, PDE5A, BNIP3, TICRR, OR7A77, OR8D2, RHOU, GJA5, NLGN1, SYT5, MAML2, SYT15, NTRK3, GRM6, TPRA1, RPTOR, CPNE3, BMPR1A, SYT4, PTEN, EPHB1, PLCB1, OR8B1, GRB14, OR109	The entirety of a process in which information is transmitted within a biological system. This process begins with an active signal and ends when a cellular response has been triggered.
Cellular response to endogenous stimulus (GO:0071495)	32	1.67E-04	GRB14, LAP, NR1D2, ABHD2, SESN3, RARB, TGFB2, PIK3R1, CYP11A1, DEFB124, CYP11A1, PCK1, RAP1A, GNAQ, HJV, IRS2, PTGER4, ACVR2A, PTGER4, BMPR2, PCK1, FGF7, FSHR, GNAS, GDF2, ESR1, GNA14, LAP, FSHR, STXBP4, BMP7, THRB, NTRK3, RPTOR, BMPR1A, TMEM100, GRB14	Any process that results in a change in state or activity of a cell (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of a stimulus arising within the organism.

response to endogenous stimulus (GO:0009719)	31	1.36E-03	<i>GRB14, LAP, NR1D2, ABHD2, SESN3, RARB, TGFB2, PIK3R1, CYP11A1, DEFB124, CYP11A1, PCK1, RAP1A, GNAQ, HJV, IRS2, PTGER4, ACVR2A, PTGER4, BMPR2, PCK1, FGF7, FSHR, GNAS, GDF2, ESR1, GNA14, LAP, FSHR, STXBP4, BMP7, THRB, NTRK3, RPTOR, BMPR1A, TMEM100, GRB14</i>	Any process that results in a change in state or activity of a cell or an organism (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of a stimulus arising within the organism.
--	----	----------	---	---

Table S7. Enrichment analysis of genes overlapping the most frequent SVR: significant Panther Gene Ontology terms (GO) associated to molecular function.

GO TERM	N GENES	P VALUE	GENES	DEFINITION
odorant binding (GO:0005549)	63	3.45E-18	OR8G5, OR8J12, OR5T2, OR8J17, OR8G3F, OR13L2, OR8U9, OR8I2, OR8C26, OR8J16, OR5AR1, OR9G3, OR8K60, OR5M3, OR8K66, OR5BE5, OR8K62, OR12D23, OR8B1AI, OR5AL2, OR8K1B, OR5T1C, OR8K5, OR8G3B, OR8K1, OR8B1S, OR8J3F, OR8B1T, OR8J19, OR5T21, OR5AL8, OR8G3, OR5M11, OR9G4, OR8K67, OR12D2F, OR8B1R, OR5G34, OR9G4D, OR8K5B, OR8H13, OR8B1U, OR8K64, OR8K3B, OR8U1, OR12D2H, OR5M13D, OR8J2E, OR8G47, OR8J3, OR5M10C, OR5M10, OR12D2, OR8K63, OR8G5B, OR8U3, OR8D6, OR13L12, OR12D2E, OR8D2, OR5AE3, OR12D18, OR8B1	Binding to an odorant, any substance capable of stimulating the sense of smell.
transmembrane signaling receptor activity (GO:0004888)	129	9.31E-06	OR8G5, OR8J12, OR6C1, OR5T2, OR8J17, OR6C1S, OR7A112, OR7A75, OR13L2, OR8G3F, OR8U9, OR6C1N, PRLR, OR8I2, OR8C26, OR8J16, HAPLN3, OR7A114, NFAM1, OR5AR1, PPYR1, OR9G3, GRIK1, OR8K60, UNC5D, OR7A88, OR7A78, OR5M3, ADGRF1, OR8K66, OR5BE5, OR8K62, GRM7, OR12D23, HTR6, TGFB2, OR8B1AI, GPR149, OR5AL2, PLXNA1, OR8K1B, OR7A99, OR5T1C, OR8K5, OR8G3B, OR8K1, OR8B1S, OR8J3F, OR8B1T, OR8J19, OR7A90, OR5T21, OR7A102, OR5AL8, OPN5, EPHA3, OR8G3, OR5M11, OR9G4, ADGRF5, OR7A135, OR8K67, OR7A103, OR12D2F, OR8B1R, OR7A97, MC3R, RXFP3, MC4R, OR7A10, PRLR, OR5G34, IGF1R, LIFR, OR9G4D, OR8K5B, OR6C5H, IGF1R, OR2AH1, OR8H13, OR8B1U, OR6C5, PTGER4, OR8K64, OR8K3B, ADGRF4, ACVR2A, PTGER4, OR8U1, OR12D2H, OR5M13D, OR8J2E, IL7R, BMPR2, CCR4, OR7A79, OR8G47, ADGRF2, OR8J3, OR7A126, OR5M10C, OR6C3, FSHR, OR5M10, OR12D2, OR7A89, OR7A98, OR7A95, ROR2, OR6C7G, OR8K63, FSHR, OR7A83, ADGRB3, OR8G5B, OPN4, OR8U3, OR8D6, OR7A77, OR13L12, OR12D2E, ACAN, OR8D2, OR5AE3, OR12D18, NTRK3, GRM6, TPRA1, TAS1R2, BMPR1A, EPHB1, OR8B1, ACAN, OR109	Combining with an extracellular or intracellular signal and transmitting the signal from one side of the membrane to the other to initiate a change in cell activity or state as part of signal transduction.

signaling receptor activity (GO:0038023)	139	2.91E-06	OR8G5, OR8J12, OR6C1, OR5T2, OR8J17, OR6C1S, OR7A112, OR7A75, OR13L2, OR8G3F, OR8U9, NR1D2, OR6C1N, PRLR, OR8I2, OR8C26, OR8J16, HAPLN3, OR7A114, GFRAL, NFAM1, NPR3, OR5AR1, PPYR1, RARB, OR9G3, GRIK1, OR8K60, UNC5D, OR7A88, OR7A78, OR5M3, ADGRF1, OR8K66, OR5BE5, OR8K62, GRM7, OR12D23, HTR6, TGFB2, OR8B1AI, GPR149, OR5AL2, PLXNA1, OR8K1B, OR7A99, OR5T1C, OR8K5, OR8G3B, OR8K1, OR8B1S, OR8J3F, NPR3, OR8B1T, OR8J19, OR7A90, OR5T21, OR7A102, OR5AL8, OPN5, EPHA3, OR8G3, OR5M11, OR9G4, ADGRF5, OR7A135, OR8K67, OR7A103, OR12D2F, OR8B1R, OR7A97, MC3R, RXFP3, MC4R, OR7A10, PRLR, OR5G34, IGF1R, HJV, LIFR, OR9G4D, OR8K5B, OR6C5H, IGF1R, OR2AH1, OR8H13, OR8B1U, OR6C5, PTGER4, OR8K64, OR8K3B, ADGRF4, ACVR2A, PTGER4, OR8U1, OR12D2H, OR5M13D, OR8J2E, IL7R, BMPR2, CCR4, CHRNA1, OR7A79, OR8G47, ADGRF2, OR8J3, OR7A126, OR5M10C, OR6C3, FSHR, OR5M10, OR12D2, OR7A89, OR7A98, OR7A95, ESR1, ROR2, OR6C7G, OR8K63, FSHR, OR7A83, ADGRB3, OR8G5B, OPN4, THRB, OR8U3, OR8D6, AHR, OR7A77, OR13L12, OR12D2E, ACAN, OR8D2, NLGN1, OR5AE3, OR12D18, NTRK3, GRM6, TPRA1, TAS1R2, BMPR1A, EPHB1, OR8B1, ACAN, OR109	Receiving a signal and transmitting it in the cell to initiate a change in cell activity. A signal is a physical entity or change in state that is used to transfer information in order to trigger a response
molecular transducer activity (GO:0060089)	139	2.91E-06	OR8G5, OR8J12, OR6C1, OR5T2, OR8J17, OR6C1S, OR7A112, OR7A75, OR13L2, OR8G3F, OR8U9, NR1D2, OR6C1N, PRLR, OR8I2, OR8C26, OR8J16, HAPLN3, OR7A114, GFRAL, NFAM1, NPR3, OR5AR1, PPYR1, RARB, OR9G3, GRIK1, OR8K60, UNC5D, OR7A88, OR7A78, OR5M3, ADGRF1, OR8K66, OR5BE5, OR8K62, GRM7, OR12D23, HTR6, TGFB2, OR8B1AI, GPR149, OR5AL2, PLXNA1, OR8K1B, OR7A99, OR5T1C, OR8K5, OR8G3B, OR8K1, OR8B1S, OR8J3F, NPR3, OR8B1T, OR8J19, OR7A90, OR5T21, OR7A102, OR5AL8, OPN5, EPHA3, OR8G3, OR5M11, OR9G4, ADGRF5, OR7A135, OR8K67, OR7A103, OR12D2F, OR8B1R, OR7A97, MC3R, RXFP3, MC4R, OR7A10, PRLR, OR5G34, IGF1R, HJV, LIFR, OR9G4D, OR8K5B, OR6C5H, IGF1R, OR2AH1, OR8H13, OR8B1U, OR6C5, PTGER4, OR8K64, OR8K3B, ADGRF4, ACVR2A, PTGER4, OR8U1, OR12D2H, OR5M13D, OR8J2E, IL7R, BMPR2, CCR4, CHRNA1, OR7A79, OR8G47, ADGRF2, OR8J3, OR7A126, OR5M10C, OR6C3, FSHR, OR5M10, OR12D2, OR7A89, OR7A98, OR7A95, ESR1, ROR2, OR6C7G, OR8K63, FSHR, OR7A83, ADGRB3, OR8G5B, OPN4, THRB, OR8U3, OR8D6, AHR, OR7A77, OR13L12, OR12D2E, ACAN, OR8D2, NLGN1, OR5AE3, OR12D18, NTRK3, GRM6, TPRA1, TAS1R2, BMPR1A, EPHB1, OR8B1, ACAN, OR109	A compound molecular function in which an effector function is controlled by one or more regulatory components.

Table S8. Enrichment analysis of genes overlapping the most frequent SVR: significant Panther Gene Ontology terms (GO) associated to cellular component.

GO_TERM	NGENES	P_VALUE	GENES	DEFINITION
organelle subcompartment (GO:0031984)	32	1.87E-02	ATP8B4, CANX, FURIN, TBC1D20, FURIN, LNPB, SEC61A1, TRAPPC6B, FMN2, RPN1, TTYH1, SEC11A, SCAMP5, PIGN, PLEKHA3, DOP1B, GOLPH3, ATP9A, PHETA2, SCAMP2, SGPP2, PHETA2, SEC11C, LMAN1L, SCAMP5, SEC22B, EMC1, FMN1, LMAN1, MOGAT1, CNIH3, VAPB, PCSK6, CEPT1, TTYH1, FKBP1A	A compartment that consists of a lumen and an enclosing membrane, and is part of an organelle.
membrane (GO:0016020)	230	1.08E-02	OMA1, GRB14, ADD3, ATP8B4, CANX, SGCD, SNX33, C7, SLC1A3, OR7A112, ALCAM, PIK3C3, EPN1, PDZK1, ULK3, STOML1, RACGAP1, OR7A75, STRA6, ITGA10, SLC4A8, NBEA, ABHD2, TM9SF4, ANPEP, LAIR1, PRLR, CTNNA1, GJA8, OSBPL10, FURIN, HAPLN3, OR7A114, GFRAL, NFAM1, BCL2L1, PPYR1, NRCAM, MME, XKR7, GRIK1, LMBRD1, SEC23A, SHISA7, EFNB2, OR7A88, TBC1D20, OR7A78, CTNNA1, ADGRF1, RAPGEF6, SLC41A3, MTMR2, FURIN, TGFB2, PMPA1, STX8, B3GAT2, SLC45A2, SEC61A1, KCNMA1, USP8, FMN2, FCAR, CPLX3, LIMA1, PLD1, TMED3, SIGLEC10, PIK3R1, RPN1, ANO6, CYP11A1, TTYH1, ATAD1, CDH17, OR7A99, PACSIN2, MYCBP2, SEC11A, CLCA3, SEC23A, CD24, ULBP17, COX4I2, OR7A90, OR7A102, SCAMP5, BOLA, PIGN, PSTPIP1, BIN2, RAB4A, CPLX4, OR7A135, CASS4, ATP5F1E, OR7A103, OR7A97, BZW2, MC3R, KCNJ6, PIK3R4, ABHD13, GJA5, OR7A10, MC4R, FRMPD2, NDUFA6, ATP6V0D2, TRAT1, PRLR, CYP11A1, CD2AP, REM1, ATP9A, LRRC4C, TNFRSF13C, STX16, RAP1A, GNAQ, TOM1L1, IGF1R, LMBRD1, GEM, KCNS1, BNIP3, CDC42EP5, C9, UPK3A, HJV, LIFR, SCN8A, ATP5MC3, IGF1R, RASGRF1, OR2AH1, TSPO, OSBPL6, SCAMP2, ENTHD1, SLC11A2, BOLA-NC1, ANPEP, IRS2, DIRAS2, PEX11A, TRAT1, COX5A, SGPP2, SEC11C, SHISA8, PTGER4, ACVR2A, CLDN14, ADGRA1, PTGER4, DPY19L4, ALCAM, HSP90AB1, SEMA7A, LMAN1L, MBOAT7, IL7R, BMPR2, NDUFA6, KCNAB1, SCAMP5, SEC22B, CCR4, OR7A79, EMC1, TMEM86B, DPP10, OR7A126, BTNL2, SIDT1, SIRPB2, MGLL, FSHR, CLDN2, EPS8L1, NKG7, NOX4, FYB1, DGKB, MAN2A2, OR7A89, MINAR1, PIK3C2G, GNAS, OR7A98, OR7A95, GNA14, SYNDIG1, FMN1, SMDT1, STRA6, ULBP21, FSHR, OR7A83, RLBP1, NDUFA3, TARM1, LMAN1, LMBRD2, HCK, C7, SHC4, PEX11B, LIM2, CD24, BNIP3, OR7A77, MOGAT1, GPRIN2, SIRPD, TMEM190, ACAN, ACTN2, GJA5, CNTN4, PDZK1, NCR1, NKG7, SYT5, CNIH3, NALF1, SYT15, MTMR12, OSCAR , KCNIP4, ACSL3, PIP4P2, TPRA1, DAB2, CPNE3, VAPB, BMPR1A, SYT4, PCSK6, TMEM100, CTNNA3, PTEN, MTX2, TSPO, BOLA, CEPT1, GRB14, LDLRAD4, ACAN, OR109, TTYH1, SDCBP2, KCNJ15, FKBP1A	A lipid bilayer along with all the proteins and protein complexes embedded in it and attached to it.

Table S9. Enrichment analysis of genes overlapping the most frequent SVR: significant pathways

PATHWAY	N GENES	P VALUE	GENES	DEFINITION
Phospholipid metabolism	26	1.41E-02	PIK3C3, OSBPL10, MTMR2, PLA2G5, PLD1, PIK3R1, PLEKHA3, RAB4A, ACP6, PIK3R4, PLA2G2E, CSNK2A1, RUFY1, GPD1L, PLA2G2A, MBOAT7, TMEM86B, MGLL, PIK3C2G, GPD1, PCTP, MTMR12, GNPAT, CPNE3, PTEN, CEPT1	Phospholipids contain a polar head group and two long-chain fatty acyl moieties, one of which is generally unsaturated. The head group is a glycerol or serine phosphate attached to a polar group such as choline. These molecules are a major constituent of cellular membranes, where their diverse structures and asymmetric distributions play major roles in determining membrane properties (Dowhan 1997). The four major classes of phospholipids in human plasma membranes are phosphatidylethanolamine, phosphatidylserine, phosphatidylcholine, and sphingomyelin. The first three are derivatives of glycerol while sphingomyelin is a derivative of serine.

7.2 Supplementary Figures.

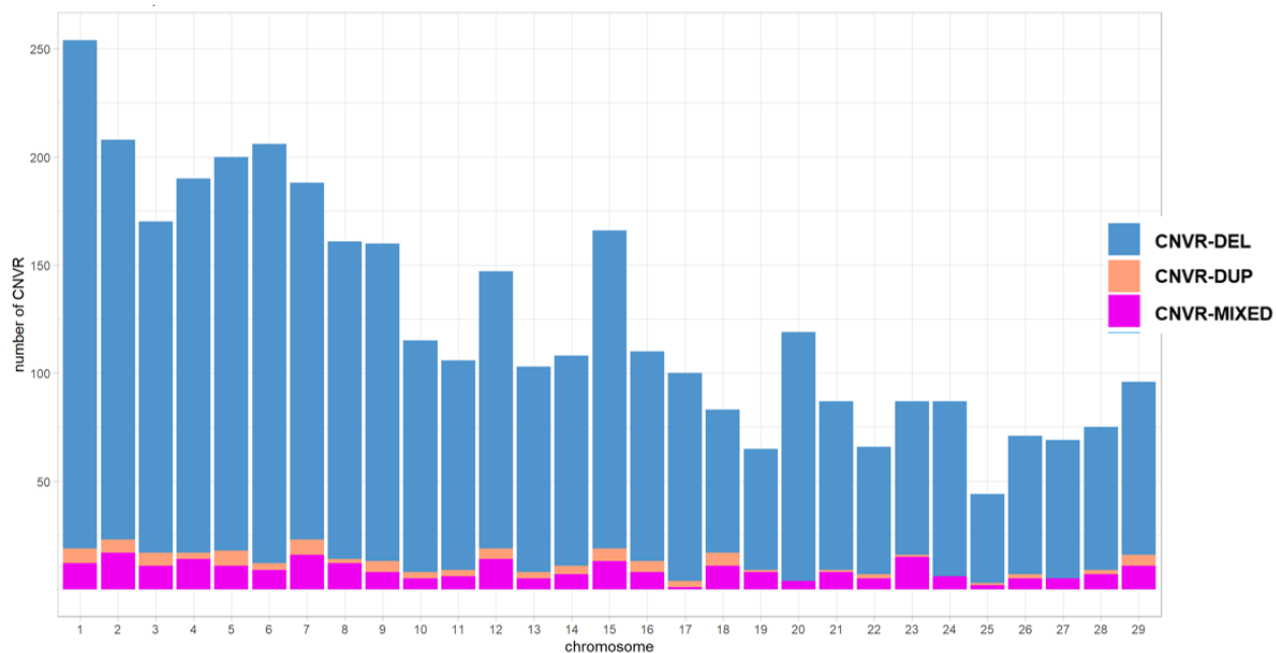


Figure S1. Distribution of each type of regions of copy number variants (CNVR) across the 29 autosomal chromosomes. CNVR-DEL: regions of deletions, CNVR-DUP: regions of duplications, and CNVR-MIXED: regions of deletions and duplications.

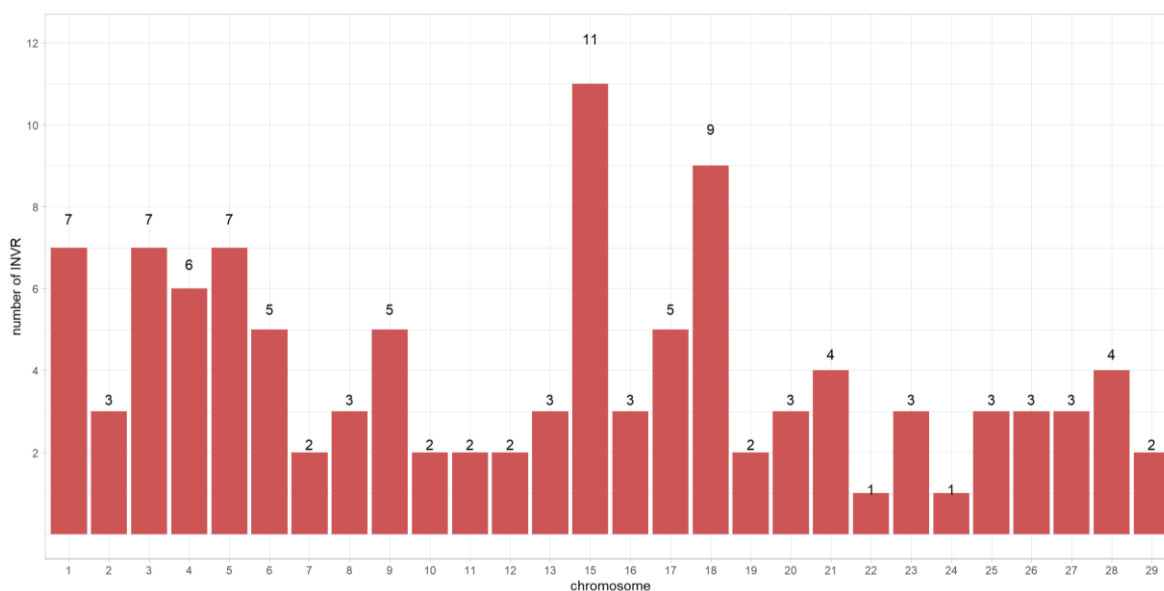


Figure S2. Distribution of regions of inversions across the 29 autosomal chromosomes

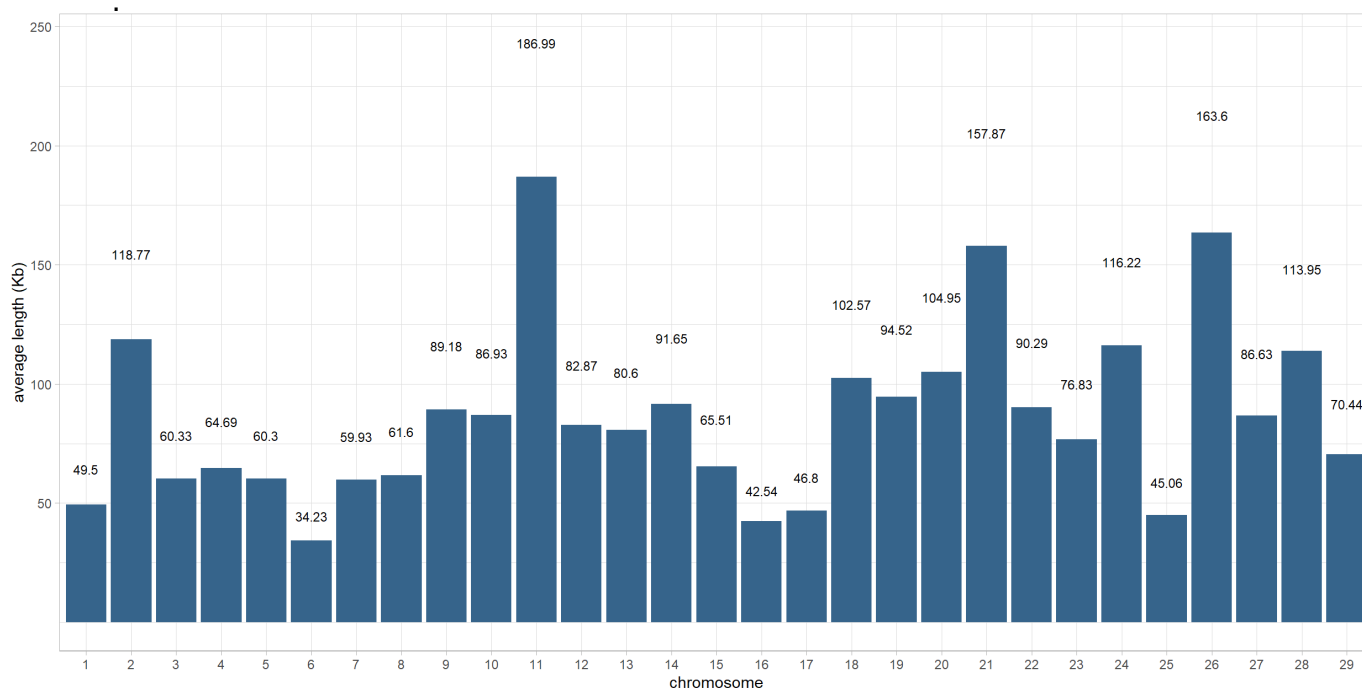


Figure S3a. Average length (Kb) of regions of copy number variants (CNVR) along the 29 autosomal chromosomes.

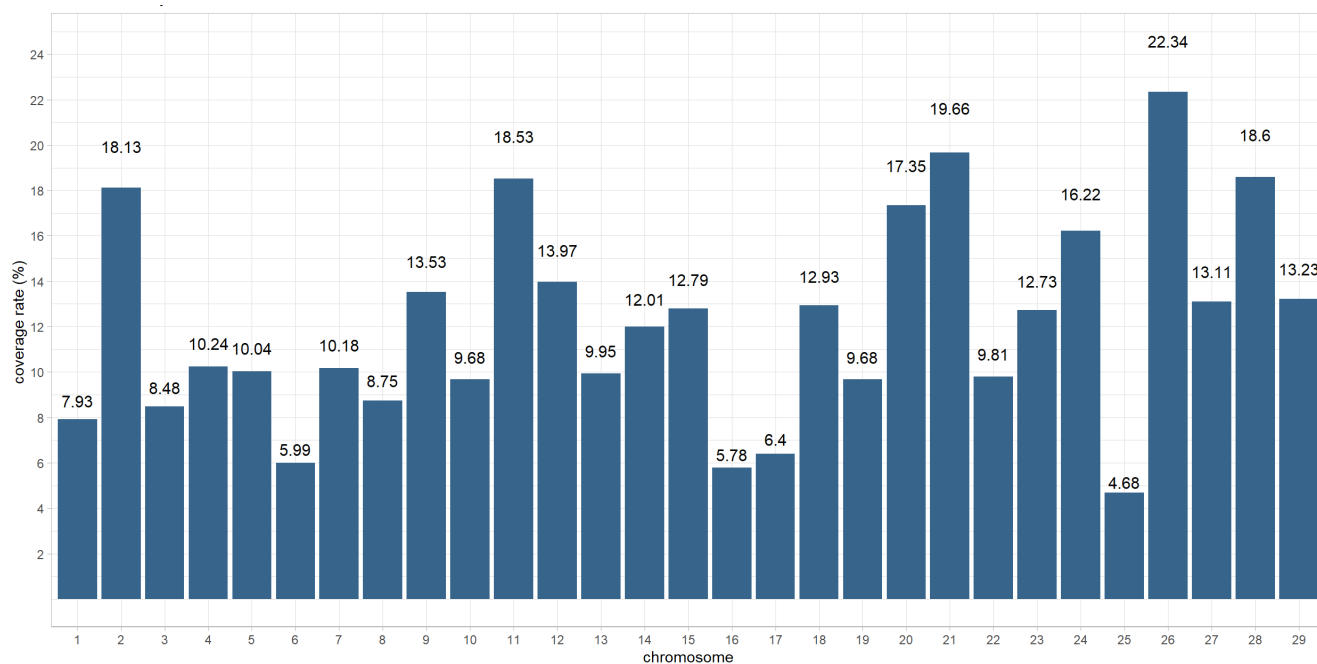


Figure S3b. Chromosomal coverage rate (%) of regions of copy number variants (CNVR) along the 29 autosomal chromosomes.

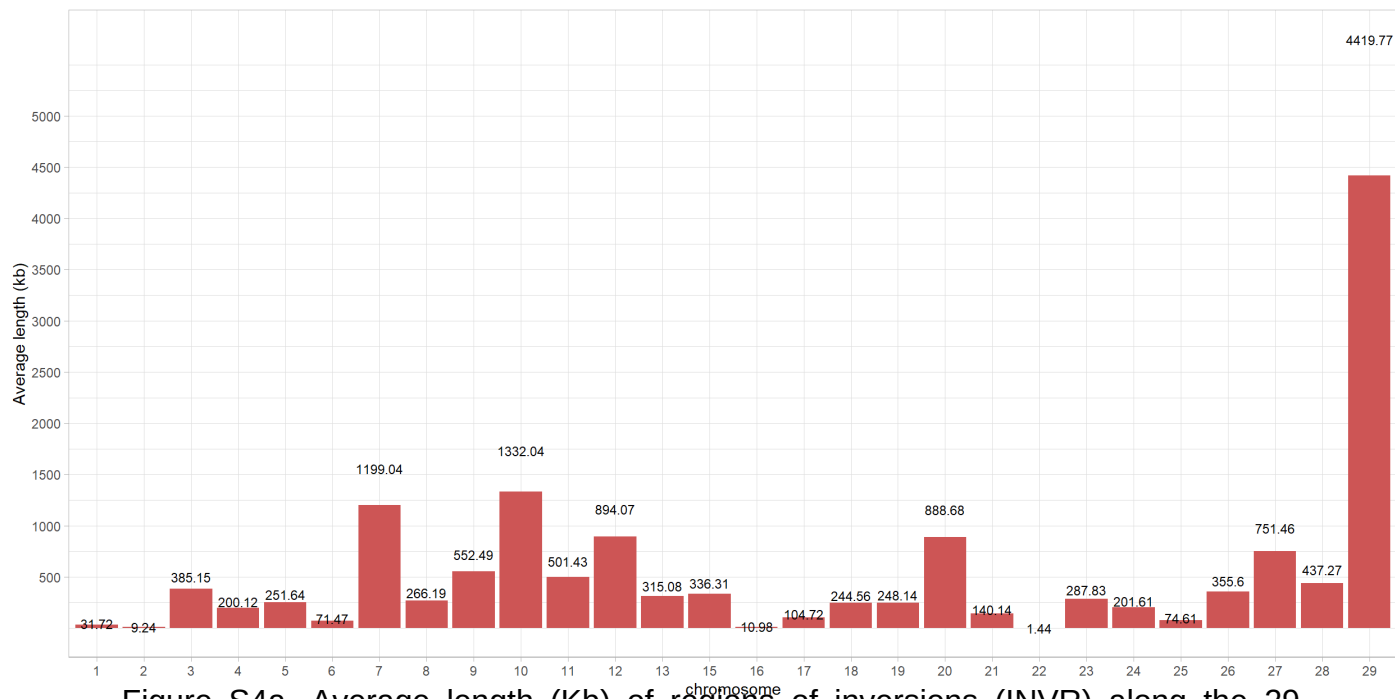


Figure S4a. Average length (Kb) of regions of inversions (INVR) along the 29 autosomal chromosomes.

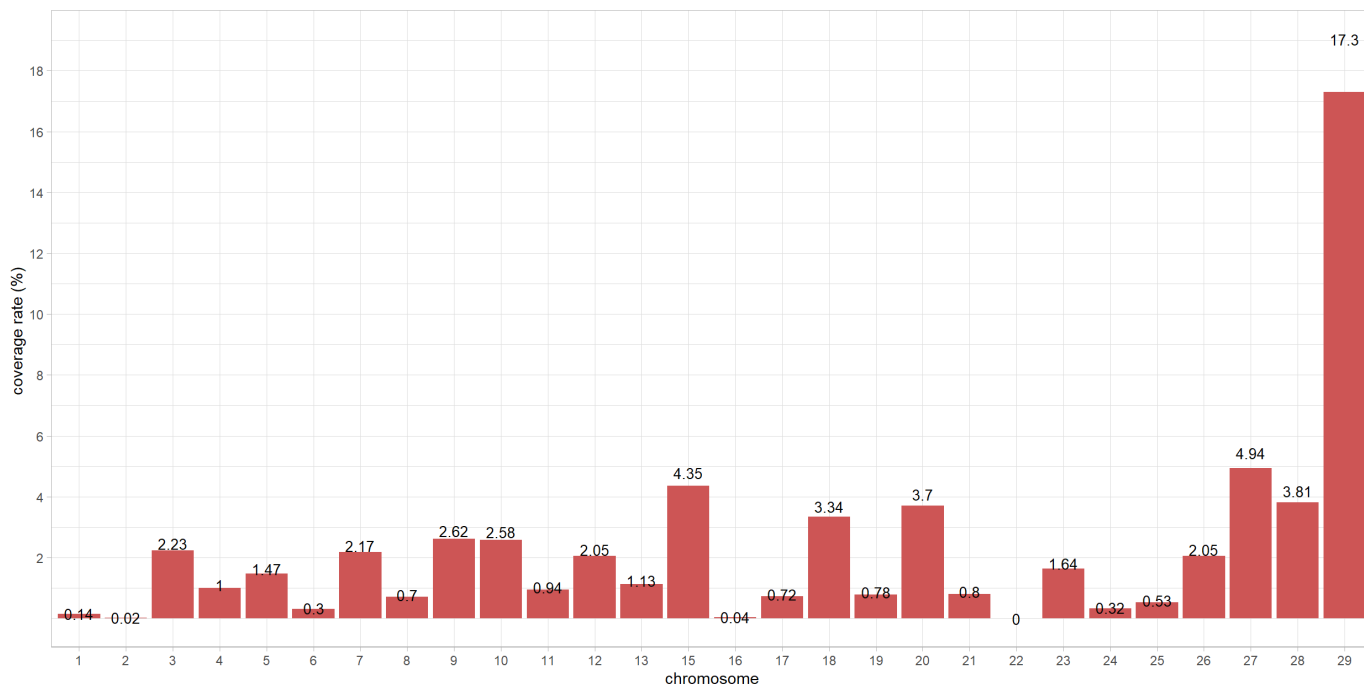


Figure S4b. Chromosomal coverage rate of regions of inversions (INVR) along the 29 autosomal chromosomes.

