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

IDENTIFICATION OF CANDIDATE LETHAL HAPLOTYPES AND RECOMBINATION EVENTS IN NELLORE CATTLE

PATRÍCIA IANA SCHMIDT
Animal Scientist

2020

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

IDENTIFICATION OF CANDIDATE LETHAL HAPLOTYPES AND RECOMBINATION EVENTS IN NELLORE CATTLE

Patrícia Iana Schmidt
Advisor: Prof.ª Dra. Lucia Galvão de Albuquerque
Co-Advisor: Prof.ª Dra. Ana Fabrícia Braga Magalhães

Dissertation presented to the School of Agricultural and Veterinarian Sciences – São Paulo State University, Campus of Jaboticabal in partial fulfillment of requirements for the degree of master in Genetics and Animal Breeding.

S353i

Schmidt, Patríca Iana
Identification of candidate lethal haplotypes and recombination events in Nellore cattle / Patríca Iana Schmidt. -- Jaboticabal, 2020
112 p. : il., tabs.

Dissertação (mestrado) - Universidade Estadual Paulista (Unesp), Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal
Orientadora: Lucia Galvão de Albuquerque
Coorientadora: Ana Fabrícia Braga Magalhães

1. Beef cattle. 2. Expected frequencies. 3. Haplotype phasing. 4. Homozygous individuals. 5. Paternal meiosis. 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.

unesp

UNIVERSIDADE ESTADUAL PAULISTA

Câmpus de Jaboticabal

CERTIFICADO DE APROVAÇÃO

TÍTULO DA DISSERTAÇÃO: IDENTIFICATION OF CANDIDATE LETHAL HAPLOTYPES AND RECOMBINATION EVENTS IN NELLORE CATTLE

AUTORA: PATRÍCIA IANA SCHMIDT
ORIENTADORA: LUCIA GALVÃO DE ALBUQUERQUE
COORIENTADORA: ANA FABRÍCIA BRAGA MAGALHÃES

Aprovada como parte das exigências para obtenção do Título de Mestra em GENÉTICA E MELHORAMENTO ANIMAL, pela Comissão Examinadora:

Profa. Dra. LUCIA GALVÃO DE ALBUQUERQUE
Departamento de Zootecnia / FCAV / Unesp - Jaboticabal

Pesquisador Dr. ROBERTO CARVALHEIRO (VIDEOCONFERÊNCIA)
Departamento de Zootecnia / FCAV / UNESP - Jaboticabal

Pesquisador Dr. DANIEL JORDAN DE ABREU SANTOS (VIDEOCONFERÊNCIA)
University of Maryland / College Park-Maryland/EUA

Jaboticabal, 30 de julho de 2020

ABOUT THE AUTHOR

PATRICIA IANA SCHMIDT - born on May 19, 1994 in Concordia, Santa Catarina, Brazil, daughter of Irdes Kochenborger Schmidt and Volnei Antonio Schmidt. She started the undergraduate in Animal Science in May 2013 at the Federal University of Pelotas, Campus of Capão do Leão. She was a scientific initiation scholarship holder from 2015 to 2018, under the supervision of Prof. Dra. Arione Augusti Boligon. During her undergraduate she also did internships at beef cattle farms. She obtained a bachelor's degree in Animal Science in February 2018. She started her Master of Science (MSc.) studies in the Graduate Program in Genetics and Animal Breeding at FCAV-UNESP, Campus of Jaboticabal in March 2018, under the supervision of Prof. Dra. Lucia Galvão de Albuquerque and Dra. Ana Fabricia Braga Magalhães. Her M.Sc studies have been granted with a scholarship from CAPES, from March 5, 2018 to February 28, 2019. She was a FAPESP scholarship holder from March 1, 2019 to July 31, 2020. She did an internship during November 1, 2019 to April 30, 2020 at University of Maryland and at Animal Genomic Improvement Laboratory (AGIL/USDA), under the supervision of Dr. Li Ma, Dr. Paul VanRaden and Dr. John B. Cole, supported by FAPESP.

“Life is not easy for any of us. But what of that? We must have perseverance and above all confidence in ourselves. We must believe that we are gifted for something and that this thing must be attained.”
Marie Curie

“If you know you are on the right track, if you have this inner knowledge, then nobody can turn you off... no matter what they say.”
Barbara McClintock

I entirely dedicate this work to my parents, Irdes e Volnei, and my brother Lucas, by the unconditional love, for being my safe harbor, for allowing me to fly and supporting me to reach my dreams. To them, who understood my absence and were always present making the distance insignificant.

I offer to women, because their role in society and in science deserves to be empowered.

ACKNOWLEDGMENTS

Firstly, I thank the God that lit up the paths I had to take, for making me face difficult situations that surely transformed me into a better person.

I am so grateful to

To my advisor, Prof. Dra. Lucia for the opportunity to work in her research group since the beginning of my final internship. For your guidance, inspiration, patience, and understanding in both research and personal life. I particularly want to thank her for giving me the opportunity to join her laboratory and work on these exciting projects.

To my supervisor at University of Maryland, Prof. Dr. Li Ma, and my supervisors at USDA/ARS (United States Department of Agriculture/Agricultural Research Service), Dr. Paul VanRaden and Dr. John Cole, for welcoming me into their research group, for the excellent opportunity and for giving me support and attention for the development of my projects. To all researchers at AGIL (Animal Genomic Improvement Laboratory) and Doctor Ma's graduate students, it was a pleasure to meet them.

To my undergraduate' advisor, Prof. Dra. Arione, who showed Genetic and Animal Breeding area to me, gave me excellent opportunities on academic area, taught me, was patient and keeps helping me always. She was an amazing professor, an inspiration for me and became a great friend.

To all professors from graduate program in Genetics and Animal Breeding – FCAV - UNESP for all help, classes and who were always available to help me.

To all my family, but especially to my parents, Irdes e Volnei and my brother Lucas, for their unconditional support. For encouraging me to leave my comfort zone and go into the search of my professional realization. For motivating me every day, for believing in me, for never measuring efforts to help me and for demonstrating every single day of my life, the most pure and sincere feeling of love I have ever received.

To my partner and colleague, Gabriel, for putting all his knowledge within my reach, for helping in the development of this work, for the loyal friendship, for the advises and for motivate me to be a better person and professional, for the sincere love, affection and patience, without your presence it would be all more difficult.

To my co-advisor, Prof. Dra. Ana Fabrícia and my special friend Larissa, who were always available to help me since of my final internship, for professional encouragement, for the opportunities to work on other projects, for never measuring efforts to help me, for attention and especially, for loyal friendship.

To my great friends and co-workers Diogo, Lucio and Malane, for all the constructive conversations, for advisement, for the special moments of happiness and especially for the loyal and true friendship.

To my “Albuquerque” sisters and brothers Gerardo, Dani, Leonardo, Salvatore, Natália, Ana Cristina, Samuel, Andrés, Marta, Gabi, Bruna, Isabelle and also to the other friends and co-workers from graduate program in Genetics and Animal Breeding, everyone who helped and taught me from the first day I came to the research group. Thanks for the classes, the coffee breaks and good conversations, and for the “churrasco” and pizza days.

To my roommate Aiina, for all help and support during my internship in Maryland. For friendship and for making me feel better during this challenged period.

This study was financed in part by São Paulo Research Foundation (Grants #2009/16118-5, #2017/10630-2, #2018/20026-8, #2018/17812-1 and #2019/10123-9).

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

SUMMARY

IDENTIFICAÇÃO DE CANDIDATOS A HAPLÓTIPOS LETAIS E EVENTOS DE RECOMBINAÇÃO EM BOVINOS NELORE

Resumo - Com dados genômicos,alelos recessivos letais podem ser descobertos a partir de haplótipos comuns na população, mas que nunca são homozigotos em animais vivos; além disso, informações genômicas também permitem a caracterização de padrões e taxas de recombinação que são importantes para a compreensão da diversidade genética em todo o genoma. Os objetivos do presente estudo foram identificar haplótipos letais, com base nas frequências populacionais esperadas, e construir um mapa de recombinação para identificar as regiões de *hotspots* para o entendimento da diversidade genética dessa população de gado Nelore. As informações genealógicas compreenderam 2.688.124 animais e o software conflict.f90 foi usado para corrigir erros mendelianos e imputar SNP's ausentes usando genótipos parentais. Um total de 4.447 animais Nelore foram genotipados com um painel de alta densidade (777.962 marcadores SNP) e 4.041 com um painel contendo 74.677 marcadores. As coordenadas genômicas dos marcadores foram baseadas na montagem do genoma *Bos taurus* ARS-UCD1.2. Os haplótipos foram construídos usando o método de janelas deslizantes implementado no software findhap.f90 v3. O número esperado de indivíduos homozigotos foi calculado por meio de dois métodos: Simples - assumindo acasalamento aleatório e usando o número de indivíduos genotipados dividido por 4 e multiplicado pelo quadrado da frequência de portadores; e Acasalamento - utilizou o padrão de acasalamento real, considerando o número de acasalamentos do touro portador e do avô materno portador dividido por 4. As taxas de recombinação foram medidas por um método indireto, extraindo pares de progênie-pai do pedigree de bovinos Nelore. Ambos, pai e filhos, estavam genotipados e foram faseados para inferir eventos de recombinação para uma meiose paterna. As regiões de *hotspots* foram definidas como intervalos SNP com taxa de recombinação > 2,5 desvios padrão acima da média. Vinte e seis haplótipos apresentaram alta frequência esperada, mas nenhum homozigoto foi observado. Dois haplótipos no cromossomo 1:56408787-56947331 e em 21:22003502-22770526 se sobrepõem a defeitos previamente conhecidos, Deficiência de Uridina Monofosfato Sintetase e síndrome de Braquispina, respectivamente. Além disso, os candidatos haplótipos letais no cromossomo 7:52418587-53136816 e no cromossomo 12:27930543-28993509 correspondem a possíveis assinaturas de seleção encontradas anteriormente em uma população semelhante de bovinos Nelore. Para as análises funcionais, usamos escores SIFT para classificar as mutações como deletérias ou tolerantes. Encontramos 55 genes candidatos responsáveis por abrigar as possíveis mutações deletérias e 11 genes com mutações tolerantes. Extraímos 21.391 eventos de recombinação e 659 meioses paternas com um número médio de recombinações por meiose de 32,4 para machos Nelore. Foram encontradas 520 regiões de *hotspots*, principalmente nos cromossomos 1, 6 e 11, com maiores taxas de recombinação. Encontramos 52 genes candidatos subjacentes às regiões de hotspot e às vias relacionadas aos termos do GO associadas. Foram encontradas algumas vias como degradação da lisina, metabolismo do piruvato, miocardite viral, fatores de transcrição basal e termos de GO relacionados aos processos de transcrição e tradução. A detecção de haplótipos letais, bem como a caracterização de eventos de recombinação em uma

população, podem ajudar a fornecer um conhecimento importante sobre a diversidade genética em todo o genoma para uma melhoria adicional do ganho genético.

Palavras-chave: bovinos de corte, faseamento de haplótipos, frequências esperadas, indivíduos homozigotos, meiose paterna

IDENTIFICATION OF CANDIDATE LETHAL HAPLOTYPES AND RECOMBINATION EVENTS IN NELLORE CATTLE

Abstract - With genomic data, lethal recessives may be discovered from haplotypes that are common in the population but are never homozygous in live animals. In addition, it also allows the characterization of patterns and rates of recombination that are important for the understanding of genetic diversity throughout the genome. The objectives of the present study were to identify lethal haplotypes, based on expected population frequencies and to build a recombination map for identify the hotspots regions for the understanding of genetic diversity of this Nellore cattle population. Pedigree information comprised 2,688,124 animals and the conflict.f90 software was used to correct Mendelian errors and fill missing SNP using parental genotypes. A total of 4,447 Nellore animals were genotyped with a high-density panel (777,962 SNP markers) and 4,041 with a panel containing 74,677 markers. Map locations are from the ARS-UCD1.2 *Bos taurus* genome assembly. The haplotypes were constructed using the sliding windows method implemented in findhap.f90 software v3. Expected numbers of homozygous individuals were calculated through two methods: Simple - assuming random mating and using the number of individuals genotyped divided by 4 and multiplied by the square of the carrier frequency; and Mating - using the actual mating pattern for calculating the number of carrier service sire × carrier maternal grandsire matings divided by 4. Recombination rates were measured by an indirect method, extracting progeny-sire pairs from pedigree of Nellore cattle. Both, sire and offspring were genotyped and phased in order to infer about recombination events for a paternal meiosis. Hotspot regions were defined as SNP intervals with recombination rate > 2.5 standard deviations above the mean. Twenty-six haplotypes had high expected frequency but no homozygotes observed. Two haplotypes on chromosome 1:56,408,787-56,947,331 and on 21:22,003,502-22,770,526 overlaps with previously known defects: Deficiency of Uridine Monophosphate Synthase and Brachyspina syndrome, respectively. Furthermore, the candidate lethal haplotypes on chromosome 7:52,418,587-53,136,816 and on chromosome 12:27,930,543-28,993,509 match with potential signatures of selection found previously in a similar population of Nellore cattle. For the functional analyses, we used SIFT scores to classify mutations as deleterious or tolerant. We found 55 candidate genes responsible for harboring the deleterious mutations and 11 genes with tolerant mutations. We extracted 21,391 crossover events and 659 paternal meiosis with an average number of crossovers per meiosis of 32.4 for Nellore males. There were found 520 hotspots regions, especially in chromosomes 1, 6 and 11, with the highest recombination rates. We have found 52 candidate genes underlying hotspot regions and associated GO terms related pathways. Some pathways as Lysine degradation, Pyruvate metabolism, Viral myocarditis, Basal transcription factors and GO terms related to transcription and translation processes, were found. The detection of lethal haplotypes, as well as the characterization of recombination events in a population, provide important information regarding genetic diversity throughout the genome for a further improvement of genetic gain.

Keywords: beef cattle, expected frequencies, haplotype phasing, homozygous individuals, paternal meiosis

Chapter 1 – General Considerations

1.1. INTRODUCTION

The use of artificial insemination has allowed great diffusion of individuals of superior genetic material but also results in the using of a relatively small number of sires in Brazilian Nellore breeding programs (ZAVAREZ et al., 2015). The use of genetically superior and proven animals as parents of the next generation may also increase mating between relatives and reduce effective population size (Ne) (DAETWYLER et al., 2014). The effective size of the population is inversely proportional to the rate of inbreeding (WRIGHT, 1922; FARIA et al., 2009).

Using Nellore data from 1994 to 1998, Faria et al. (2009), found an Ne size of 68 unrelated animals. In addition, they reported that the number of contributing ancestors decreased with time and the most important ancestor accounted for 14% of the Nellore animals born in Brazil. Zavarez et al. (2015), also with Nellore, used high-density SNP chips to characterize autozigosity and concluded that the massive use of a few bulls and artificial insemination generated homozygous segments in the genomes of more than 70% of the studied animals. These homozygous segments indicate the occurrence of inbreeding, increase the probability of expression of deleterious recessive alleles, and also can reduce the animal adaptive value and the genetic diversity (WRIGHT, 1922).

For understanding genetic diversity throughout the genome, Weng et al. (2019) reported that is important to characterizing patterns of recombination rates. Meiotic recombination, through crossing-over, plays an important role in the genetic diversity and the genome evolution of organisms with sexual reproduction (MOURESAN et al., 2019). During this process, new genetic variation are created by introducing new combinations of existing alleles of different loci in the same

chromosome and deleterious mutations that might otherwise accumulate are removed (GONEN et al., 2017; MOURESAN et al., 2019; SHEN et al., 2018); in addition to introducing new combinations of genetic variants, resulting in offspring with traits that differ from those in their parents and contributing to genetic diversity (BARTON & CHARLESWORTH, 1998; COOP & MYERS, 2007; MA et al., 2015; WANG et al., 2016; SHEN et al., 2018).

Initially, pedigree analysis (BLUNN & HUGHES, 1938) or/and strategies that required genotypes from affected embryos (CHARLIER et al., 2008) were used to provide evidence of recessive effects. After the availability of large-scale genotyping technologies, dense molecular markers offered the possibility of identifying lethal recessive variants (CHARLIER et al., 2008). However, most genetic variants are not expressed independently in the genomes, but are connected to each other, located closely along a single chromosome (BANSAL et al., 2008). The detection of haplotypes common in the population is a tool to locate lethal recessive alleles (VANRADEN et al., 2011) and, therefore, more efficient in capturing deleterious variation compared to individual SNPs. The detection and understanding of the genetic basis controlling the occurrence and accumulation of lethal haplotypes, as well as the characterization of recombination events into a population, will give important knowledge regarding genetic diversity throughout the genome for a further improvement of genetic gain.

1.2. LITERATURE REVIEW

1.2.1. Lethal haplotypes

Lethal recessive alleles were often discovered from reports of abnormal calves and subsequent breeding tests to confirm inheritance. Samples of affected tissues of animals affected were collected by various diseases that cause economic impacts on production, such as complex vertebral malformation (CVM), congenital muscular dystonia (CMD) and crooked tail syndrome (CTS) and the corresponding disease sites were mapped using programs for detect overlapping homozygous chromosomal segments among the affected cattle (CHARLIER et al., 2008; KHATIB et al., 2009). After detecting homozygous segments and using data from Belgian Blue and Italian Chianina cattle, Charlier et al. (2008) re-sequenced at least one affected animal for each disease. The animals were slaughtered, specific tissues related to each disease were collected, the pathogenic relevance of each mutation was established by aligning the protein sequence and, finally, all individuals were genotyped to identify the mutations of diseases in whole population.

A methodology to find lethal haplotypes that only requires genotype data from phenotypically normal individuals and not from the affected embryos that die, was proposed by VanRaden et al. (2011), based on the construction and identification of haplotypes in low frequency in homozygous state, although high frequencies could be expected due to the excessive use of ancestors with these haplotypes (SAHANA et al., 2013). This approach, therefore, is different of the previous strategy (LANDER & BOTSTEIN, 1987; CHARLIER et al., 2008; KHATIB et al., 2009; HUQUN et al., 2010). With this method it is possible to discover lethal defects without using any phenotype. If the numbers of genotyped individuals are large, expected numbers of

homozygous haplotypes will also be sufficiently large so that their complete absence likely is not by chance. This methodology has been used, for example, by Pausch et al. (2015) and Hoff et al. (2017) in beef cattle and Häggman and Uimari (2016) in Yorkshire pigs.

Some US dairy cattle breed associations (Holstein, Jersey, and Brown Swiss) are reporting carrier status since August 2011 for all animals with BovineSNP50 genotypes. Currently, haplotype tests are used routinely in dairy cattle (Table 1) to identify animals with desirable traits of economic importance, discover new genetic disorders, and track carrier status of genotyped animals. Recessive haplotypes, affecting fertility and stillbirth in dairy cattle (haplotypes BH1–BH2, HH1–HH5, and JH1–JH2), were described in detail by VanRaden et al. (2011, 2013, 2014). A recessive mutation that causes embryo death (haplotype HH6) was also discovered by Fritz et al. (2018). Carrier status for Holstein haplotypes with mutations for Brachyspina (haplotype HH0), bovine leucocyte adhesion deficiency (BLAD; haplotype HHB), complex vertebral malformation (CVM; haplotype HHC), deficiency of uridine monophosphate synthase (DUMPS; haplotype HHD), mulefoot (syndactyly; haplotype HHM), polledness (haplotype HHP), red coat color (haplotypes HBR, HDR, and HHR), and cholesterol deficiency (HCD) were reported by the Council on Dairy Cattle Breeding (Bowie, MD - https://www.uscdcb.com/). Brown Swiss haplotype tests for spinal dismyelination (SDM; haplotype BHD), spinal muscular atrophy (SMA; haplotype BHM), and Weaver Syndrome (haplotype BHW) were also found. A haplotype that affects conception rate in Ayrshires (AH1) was reported by Cooper et al. (2014), and an Ayrshire haplotype that affects fertility (AH2) was discovered by Null et al. (2017).

Within international Angus populations several defects that are effectively lethal such as, Neuropathic hydrocephalus, Arthrogryposis Multiplex and Osteoporosis have been propagated in recent decades (TESSELING & PARNELL, 2013). These alleles can reach high frequencies due to their propagation by the extensive use of popular sire lines via artificial insemination, or due to linkage to beneficial alleles at strongly selected loci (KADRI et al., 2014). In a population such as the U.S. registered Angus breed, a high proportion of latent recessive lethal variants could still be segregating without having been early detected by breeders, particularly because most of the reproductive process is unmonitored and the extent of the impact that these alleles have on fertility and fitness in livestock is unknown (HOFF et al., 2017).

Considering that Brazil is the world's largest exporter of beef and has the largest commercial cattle herd in the world (ABIEC, 2019) and around 80% of the Brazilian herd has Bos indicus contribution, with the Nellore breed corresponding to 90% of these Zebu animals (ABCZ, http://www.abcz.com.br), it is important to obtain more information to assist in the management of the mating and selection in these herds. Based on reports from the other beef and dairy cattle breeds, we can assume that there may be lethal recessive alleles in the Nellore population as well. By identifying the animals that carry the lethal recessive alleles, breeders can anticipate problems caused by disadvantageous variants, through the implementation of selection actions, avoiding the mating of carriers and, potentially validation by gene editing.

Table 1* - List of haplotypes currently tracked in the U.S. genomic evaluation system, the frequency of the minor (less common) haplotype, and the location in base pairs (bp) of the haplotype based on the ARS-UCD1.2 genome assembly (ROSEN et al., 2020). Locations are exact (single bp) when the causative mutation is known and approximate (range of bp) when it is unknown, a duplication, a deletion, or multiple mutations. The frequency of carriers in the population is generally twice the haplotype frequency because carriers have 1 defective and 1 normal haplotype

Breed	Haplotype	Functional/ *gene* name	Haplotype frequency (%)	Chromosome	ARS-UCD region (bp)	Reference
Ayrshire	AH1	PIRM/ *UBE3B*	11.16	17	63,668,380	Cooper et al. (2014), Venhoranta et al. (2014)
	AH2	*RPAP2*	10.5	3	51,086,099 – 51,119,146	Null et al. (2017)
Brown Swiss	BH2	*TUBD1*	6.65	19	10,833,921	Schwarzenbacher et al. (2016)
	BHD	SDM/*SPAST*	1.31	11	13,246,972 – 14,736,876	Hafner et al. (1993), Thomsen et al. (2010)
	BHM	SMA/*KDSR* (*FVT1*)	3.24	24	61,620,374	El-Hamidi et al. (1989), Krebs et al., 2007
	BHP	Polledness/ *POLLED*	1.22	1	2,578,598	Medugorac et al. (2012), Rothammer et al. (2014)
	BHW	Weaver/ *PNPLA8*	0.58	4	49,339,002 – 49,461,342	McClure et al. (2013), Kunz et al., 2016
Holstein	HBR	Black/red coat color/ *MC1R* (*MSHR*)	0.75	18	14,705,501	Lawlor et al. (2014)
	HCD	Cholesterol deficiency/ *APOB*	2.28	11	77,872,709	Kipp et al. (2015), Charlier (2016), Menzi et al. (2016), Schütz et al. (2016)
	HDR	Dominant red coat	0.03	3	9,361,962	Capitan et al. (2014), Lawlor et al. (2014), Dorshorst et al., 2015

	HH0	color Brachyspina/ *FANCI*	1.65	21	20,775,563	Agerholm et al. (2006), Charlier et al. (2012)
	HH1	*APAF1*	1.28	5	62,810,245	Adams et al. (2012)
	HH2	—	1.21	1	93,501,204 – 95,581,556	VanRaden et al. (2011), McClure et al. (2014)
	HH3	*SMC2*	2.64	8	93,753,358	Daetwyler et al. (2014), McClure et al. (2014)
	HH4	*GART*	0.23	1	1,997,582	Fritz et al. (2013)
	HH5	*TFB1M*	2.39	9	91,847,117 – 91,937,003	Cooper et al. (2013), Schütz et al. (2016)
	HH6	*SDE2*	0.44	16	29,015,336 – 29,059,673	Fritz et al. (2018)
	HHB	BLAD/ *ITGB2*	0.21	1	144,770,078	Shuster et al. (1992)
	HHC	CVM/ *SLC35A3*	1.1	3	43,261,946	Agerholm et al. (2001)
	HHD	DUMPS/ *UMPS*	0.01	1	69,151,931	Shanks et al. (1984)
	HHM	Mulefoot/ *LRP4*	0.05	15	76,807,960	Eldridge et al. (1951), Duchesne et al. (2006)
	HHP	Polledness/ *POLLED*	0.88	1	2,578,598	Medugorac et al. (2012), Rothammer et al. (2014)
	HHR	Red coat color/ *MC1R* (*MSHR*)	3.29	18	14,705,965	Joerg et al. (1996)
Jersey	JH1	*CWC15*	9.21	15	15,449,431	Sonstegard et al. (2013)
	JHP	Polledness/ *POLLED*	2.04	1	2,578,598	Medugorac et al. (2012), Rothammer et al. (2014)

*Table constructed and provided by the following researchers J.B. Cole, P.M. VanRaden, D.J. Null, J.L. Hutchison, and S.M. Hubbard from Animal Genomics and Improvement Laboratory - United States Department of Agriculture (AGIL – USDA)

1.2.2. Recombination events

Characterizing pattern and rates of recombination is important for the understanding of genetic diversity throughout the genome (WENG et al., 2019). In eukaryotes, meiotic recombination is a fundamental biological process in which parental genetic materials are exchanged during meiosis for egg or sperm formation, introducing new combinations of genetic variants, resulting in offspring with traits that may differ from those in their parents and contributing to genetic diversity (BARTON & CHARLESWORTH, 1998; COOP & MYERS, 2007; MA et al., 2015; WANG et al., 2016; SHEN et al., 2018).

Recombination rates vary widely between and within chromosomes, species, sex and possibly age; that is, this variation is under both genetic and environmental controls (OLLIVIER, 1995; BROMAN et al., 1998; KOREN et al., 2002; MYERS et al., 2005). In addition, some studies have suggested that maternal age has also a significant effect on crossover frequency (SHERMAN et al., 1994; KONG et al., 2004). As recombination is under genetic and environmental control, it is possible to manipulate it by performing environmental modifications or via gene editing through the evidence of favorable alleles (JENKO et al., 2015). These manipulations may increase response to selection in traits of economic interest, if used in breeding programs, releasing greater genetic variability to each generation (BATTAGIN et al., 2016).

In diploid organisms, the recombination events at meiosis can be discovered by comparing the haplotypes of an individual and its parents (TAO et al., 2018) and can be measured both directly and indirectly. The direct methods measure the recombination frequency by examining proteins related to some specific stages of

meiosis, such as Single-Sperm Genome-Wide Genotyping (LU et al., 2012; WANG et al., 2012) or Molecular Assay (PAN et al., 2011). Indirect methods build recombination maps based on the identification of local patterns in Linkage Disequilibrium (MOURESAN et al., 2019) or using genomic information from pedigree data (MA et al., 2015; SHEN et al., 2018). Mouresan et al. (2019), used the persistency of linkage disequilibrium (LD) phase among closely related populations to estimate the recombination rate for seven populations of autochthonous beef cattle in Spain. This method brings similar results as previous methods; however, has several limitations since differences in the mutation rate and/or selection events can locally affect the persistency of LD phase and it requires populations to be close enough that the recombination rate is well conserved.

Using a pedigree method, MA et al., (2015) extracted a total of 185,917 three-generation families that included an offspring, both parents, and two grandsires per family genotyped by various SNP chips from the large US Holstein cattle pedigree with over half million genotyped cattle. In each family, these authors phased the genotypes of the two parents and the offspring, and inferred recombination events for a paternal meiosis from the sire/offspring pair and for a maternal meiosis from the dam/offspring pair. They reported, for the first time, that bulls have more recombination than cows, contrary to the common perception that females have more recombination than males in many species of mammals (MA et al., 2015). Similar results were reported by Shen et al. (2018) in Jersey, Ayrshire and Brown Swiss breeds, in which males recombination map was >10% longer than females recombination map.

Recombination events do not occur totally randomly throughout the genomic DNA, instead they are more likely to take place in some regions of the genome than in others (KAUPPI et al., 2004). So, the term “hotspot” has been used to indicate a locus or genomic region occurring a greater than average frequency of meiotic recombination (LICHTEN & GOLDMAN, 1995). In mammals, hotspots display a length of 1-2 kb, and are usually surrounded by larger regions that do not undergo recombination which largely define the inheritance patterns in each generation (PAIGEN & PETKOV, 2010). Although recombination events are known to be distributed nonrandomly across the genome, it has been difficult to detect the precise location at which they occur. In humans, the pedigree approach has allowed the identification of genomic regions of 1 to 10-kb in length in which recombination events seem to be grouped, i.e. there are recombination hotspots (PAIGEN & PETKOV, 2010). However, this approach displays limitation in detecting low-frequency events (KAUPPI et al., 2004).

The pedigree approach has also been used to detect hotspots location in different cattle breeds. Shen et al. (2018) defined a recombination hotspot region as SNP intervals with recombination rate > 2.5 standard deviations above the mean. The authors found variability in the number of hotspots regions among different breeds (Jersey, Ayrshire and Brown Swiss). Moreover, males displayed a total of 233 hotspots shared by all those breeds while females showed only 44 hotspot regions. Although each breed showed specific features in the distribution of recombination rate and hotspot regions, they presented similar global recombination patterns across the genome. The authors also reported that hotspots regions, based on a GWAS, showed a significant association near the *PRDM9* gene.

Studies in several mammals including mice, dogs, humans and swine performed by Dietrich et al. (1996), Neff et al., (1999), Kong et al., (2002) and Tortereau et al., (2012), respectively, showed higher recombination rates in females compared to males. In more inbred species or with asexual reproduction, the effective amount of recombination may be much lower, increasing the likelihood of deleterious alleles becoming fixed, manly if there is a mutation nearby that benefits the species. (HADANY & FELDMAN, 2005). Failures in recombination may lead to a potentially deleterious outcome, as aneuploidy (LIPKIN et al., 2002). When advantageous new alleles arise and spread within a population, deleterious alleles at neighboring loci also can spread and to fix in areas of low recombination (HARTFIELD & OTTO, 2011). In contrast, higher recombination rates increase the likelihood of beneficial alleles to establish and hinder the establishment fixation of deleterious alleles within a lineage (HARTFIELD & OTTO, 2011). Although most of the studies are done on humans and mice, studies on other mammalian species, such as cattle, can provide comparative information to better understand recombination events.

1.3. OBJECTIVES

1.3.1. General objective

The objectives of the present study were to identify lethal haplotypes in a Nellore population, based on expected population frequencies and to identify deleterious and tolerant candidate mutations through functional analyzes; to build a recombination map and to identify the hotspots regions for the understanding of genetic diversity of this Nellore cattle population.

1.3.2. Specific objectives

- To build haplotypes using sliding window method
- To identify lethal haplotypes in Nellore cattle, based on expected population frequencies
- to detect crossover haplotypes, in order to refine the map location and include more carriers of the lethal mutation
- To identify, through functional analyses, deleterious and tolerant mutations in candidate genes
- To detect recombination events using Pedigree Method
- To build recombination map for Nellore males
- To identify hotspots regions
- To identify, through functional analyzes, candidate genes in hotspots regions

1.4. REFERENCES

ADAMS, H.A., T. SONSTEGARD, P.M. VANRADEN, D.J. NULL, C. VAN TASSELL, AND H. LEWIN. Identification of a nonsense mutation in APAF1 that is likely causal for a decrease in reproductive efficiency in Holstein dairy cattle. **Journal of dairy science**, v. 99, n. 8, p. 6693-6701, 2016.

AGERHOLM, J.S., C. BENDIXEN, O. ANDERSEN, AND J. ARNBJERG. Complex vertebral malformation in Holstein calves. **Journal of veterinary diagnostic investigation**, v. 13, n. 4, p. 283-289, 2001.

AGERHOLM, J.S., F. MCEVOY, AND J. ARNBJERG. Brachyspina syndrome in a Holstein calf. **Journal of veterinary diagnostic investigation**, v. 18, n. 4, p. 418-422, 2006.

BANSAL, V.; HALPERN, A. L.; AXELROD, N.; BAFNA, V. An MCMC algorithm for haplotype assembly from whole-genome sequence data. **Genome Research**, v. 18, p. 1336–1346, 2008.

BARTON, N. H. & CHARLESWORTH, B. Why sex and recombination? **Science**, 281:1986-1990, 1998.

BATTAGIN M., GORJANC G., FAUX A. M., JOHNSTON S. E., & HICKEY J. M. Effect of manipulating recombination rates on response to selection in livestock breeding programs. **Genetics Selection Evolution**, 48:44, 2016.

BLUNN, C. T.; HUGHES, E. H. Hydrocephalus in swine: a new lethal defect. **Journal of Heredity**, v. 29, n. 5, p. 203-208, 1938.

BROMAN K. W., MURRAY J. C., SHEFFIELD V. C., WHITE R. L. & WEBER J. L. Comprehensive human genetic maps: individual and sex-specific variation in recombination. American Journal of Human Genetics, 63:861–9, 1998.

CAPITAN, A., P. MICHOT, F. GUILLAUME, C. GROHS, A. DJARI, S. FRITZ, S. BARBEY, P. OTZ, E. BOURNEUF, D. ESQUERRÉ, Y. GALLARD, C. KLOPP, AND D. BOICHARD. Rapid discovery of mutations responsible for sporadic dominant genetic defects in livestock using genome sequence data: Enhancing the value of farm animals as model species. **Proc. 10th World Congr. Genet. Appl. Livest. Sci. Commun**, v. 182, 2014.

CHARLIER, C. The role of mobile genetic elements in the bovine genome. In: **Abstract W636 in Proc. Plant Anim. Genome XXIV Conf. San Diego, CA**. 2016.

CHARLIER, C., J.S. AGERHOLM, W. COPPIETERS, P. KARLSKOV-MORTENSEN, W. LI, G. DE JONG, C. FASQUELLE, L. KARIM, S. CIRERA, N. CAMBISANO, N. AHARIZ, E. MULLAART, M. GEORGES, AND M. FREDHOLM. A deletion in the bovine FANCI gene compromises fertility by causing fetal death and brachyspina. **PloS one**, v. 7, n. 8, 2012.

CHARLIER, C.; COPPIETERS, W.; ROLLIN, F.; DESMECHT, D.; AGERHOLM, J. S.; CAMBISANO, N.;CARTA, E.; DARDANO, S.; DIVE, M.; FASQUELLE, C.; FRENNET, J. C; HANSET, R.; HUBIN, X.; JORGENSEN, C.; KARIM, L.; KENT, M.; HARVEY, K.; PEARCE, B. R.; SIMON, P.; TAMA, N.; NIE, H.; VANDEPUTTE, S.; LIEN, S.; LONGERI, M.; FREDHOLM, M.; HARVEY, R. J.; GEORGES, M. Highly effective SNP-based association mapping and management of recessive defects in livestock. **Nature genetics**, v. 40, n. 4, p. 449, 2008.

COOP G. & MYERS S. R. Live hot, die young: transmission distortion in recombination hotspots. **PLoS Genetics**, 3:e35, 2007.

COOPER, T.A., G.R. WIGGANS, D.J. NULL, J.L. HUTCHISON, J.B. COLE. Genomic evaluation, breed identification, and discovery of a haplotype affecting fertility for Ayrshire dairy cattle. **Journal of dairy science**, v.97, p. 3878–3882, 2014.

COOPER, T.A., G.R. WIGGANS, P.M. VANRADEN, J.L. HUTCHISON, J.B. COLE, AND D.J. NULL. Genomic evaluation of Ayrshire dairy cattle and new haplotypes affecting fertility and stillbirth in Holstein, Brown Swiss and Ayrshire breeds. In: **ADSA-ASAS Joint Annual Meeting, poster**. 2013.

DAETWYLER, H. D.;CAPITAN, A.; PAUSCH, H.; STOTHARD, P.; BINSBERGEN, R.; BRONDUM, R. F.; LIAO, X.;DJARI, A.; RODRIGUEZ, S. C.;GROHS, C.;ESQUERRÉ, D.; BOUCHEZ, O.; ROSSIGNOL, M. N.;KLOPP, C.; ROCHA, D.; FRITZ, S.;EGGEN, A.; BOWMAN, P. J.; COOTE, D.; CHAMBERLAIN, A. J.; ANDERSON, C.; VANTASSELL, C. P.; HULSEGGE, I.; GODDARD, M. E.; GULDBRANDTSEN, B.; LUND, M. S.; VEERKAMP, R. F.; BOICHARD, D. A.; FRIES, R.; HAYES, B. J. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. **Nature genetics**, v. 46, n. 8, p. 858, 2014.

DIETRICH W. F., MILLER J., STEEN R., MERCHANT M. A., DAMRON-BOLES D., HUSAIN Z., DREDGE R., DALY M. J., INGALLS K. A. & O'CONNOR T.J. A comprehensive genetic map of the mouse genome. **Nature**, 380:149–52, 1996.

DUCHESNE, A., M. GAUTIER, S. CHADI, C. GROHS, S. FLORIOT, Y. GALLARD, G. CASTE, A. DUCOS, AND A. EGGEN. Identification of a doublet missense

substitution in the bovine LRP4 gene as a candidate causal mutation for syndactyly in Holstein cattle. **Genomics**, v. 88, n. 5, p. 610-621, 2006.

ELDRIDGE, F.E., W.H. SMITH, AND W.M. MCLEOD. Syndactylism in Holstein-Friesian cattle: Its inheritance, description and occurrence. **Journal of Heredity**, v. 42, n. 5, p. 241-250, 1951.

EL-HAMIDI, M., H.W. LEIPOLD, J.G.E. VESTWEBER, AND G. SAPERSTEIN. Spinal muscular atrophy in Brown Swiss calves. **Journal of Veterinary Medicine Series A**, v. 36, n. 1-10, p. 731-738, 1989.

FARIA, F. J. C.; FILHO, A. V.; MADALENA, F. E.; JOSAHKIAN, L. A. Pedigree analysis in the Brazilian Zebu breeds. **Journal of Animal Breeding and Genetics**, v. 126, n. 2, p. 148-153, 2009.

FRITZ, S., HOZE, C., REBOURS, E., BARBAT, A., BIZARD, M., CHAMBERLAIN, A., ESCOUFLAIRE, C., VANDER JAGT, C., BOUSSAHA, M., GROHS, C. AND ALLAIS-BONNET, A. An initiator codon mutation in SDE2 causes recessive embryonic lethality in Holstein cattle. **Journal of dairy science**, v. 101, n.7, p.6220-6231, 2018.

FRITZ, S., A. CAPITAN, A. DJARI, S. C. RODRIGUEZ, A. BARBAT, A. BAUR, C. GROHS, B. WEISS, M. BOUSSAHA, D. ESQUERRÉ, C. KLOPP, D. ROCHA, AND D. BOICHARD. Detection of haplotypes associated with prenatal death in dairy cattle and identification of deleterious mutations in GART, SHBG and SLC37A2. **PloS one**, v. 8, n. 6, 2013.

HADANY L. & M. W. FELDMAN. Evolutionary traction: the cost of adaptation and the evolution of sex. **Journal of Evolutionary Biology**, 18:309–314, 2005.

HAFNER, A., E. DAHME, G. OBERMAIER, P. SCHMIDT, AND G. DIRKSEN. Spinal dysmyelination in new-born Brown Swiss x Braunvieh calves. **Journal of Veterinary Medicine, Series B**, v. 40, n. 1-10, p. 413-422, 1993.

HÄGGMAN, J.; UIMARI, P. Novel harmful recessive haplotypes for reproductive traits in pigs. **Journal of Animal Breeding and Genetics**, v. 134, n. 2, p. 129-135, 2017.

HARTFIELD M. & OTTO S. P. Recombination and hitchhiking of deleterious alleles. **Evolution: International Journal of Organic Evolution**, 65:2421-2434, 2011.

HOFF, J. L.; DECKER, J. E.; SCHNABEL, R. D.; TAYLOR, J. F. Candidate lethal haplotypes and causal mutations in Angus cattle. **BMC genomics**, v. 18, n. 1, p. 799, 2017.

HOZÉ, C., ESCOUFLAIRE, C., MESBAH-UDDIN, M., BARBAT, A., BOUSSAHA, M., DELOCHE, M.C., BOICHARD, D., FRITZ, S. AND CAPITAN, A. A splice site mutation in CENPU is associated with recessive embryonic lethality in Holstein cattle. **Journal of dairy science**, v. 103, n.1, p.607-612, 2020.

HUQUN, S. FUKUYAMA, H. MORINO, H. MIYAZAWA, T. TANAKA, T. SUZUKI, M. KOHDA, H. KAWAKAMI, Y. OKAZAKI, K. SEYAMA, K. HAGIWARA. A quantitatively-modeled homozygosity mapping algorithm, qHomozygosityMapping, utilizing whole genome single nucleotide polymorphism genotyping data. In: BMC bioinformatics. **Bio Med Central**, 2010. p. S5.

JENKO, J., GORJANC, G., CLEVELAND, M. A., VARSHNEY, R. K., WHITELAW, C. B. A., WOOLLIAMS, J. A., & HICKEY, J. M. Potential of promotion of alleles

by genome editing to improve quantitative traits in livestock breeding programs. **Genetics Selection Evolution**, v. 47, n. 1, p. 55, 2015.

JENKO, J., MCCLURE, M.C., MATTHEWS, D., MCCLURE, J., JOHNSSON, M., GORJANC, G. AND HICKEY, J.M. Analysis of a large dataset reveals haplotypes carrying putatively recessive lethal and semi-lethal alleles with pleiotropic effects on economically important traits in beef cattle. **Genetics Selection Evolution**, v.51, n.1, p.9, 2019.

JOERG, H., H.R. FRIES, E. MEIJERINK, AND G.F. STRANZINGER. Red coat color in Holstein cattle is associated with a deletion in the MSHR gene. **Mammalian genome: official journal of the International Mammalian Genome Society**, v. 7, n. 4, p. 317, 1996.

KADRI NK, SAHANA G, CHARLIER C, ISO-TOURU T, GULDBRANDTSEN B, KARIM L, NIELSEN US, PANITZ F, AAMAND GP, SCHULMAN N, GEORGES M, VILKKI J, LUND MS, DRUET T. A 660-Kb deletion with antagonistic effects on fertility and milk production segregates at high frequency in Nordic Red cattle: additional evidence for the common occurrence of balancing selection in livestock. **PLoS genetics**, v. 10, n. 1, 2014.

KHATIB, H., C. MALTECCA, R.L. MONSON, V. SCHUTZKUS, J.J. RUTLEDGE. Monoallelic maternal expression of STAT5A affects embryonic survival in cattle. **BMC genetics**, v. 10, n. 1, p. 13, 2009.

KIPP, S., D. SEGELKE, S. SCHIERENBECK, F. REINHARDT, R. REENTS, C. WURMSER, H. PAUSCH, R. FRIES, G. THALLER, J. TETENS, J. POTT, M. PIECHOTTA, AND W. GRÜNBERG. A new Holstein haplotype affecting calf survival. **Interbull Bulletin**, n. 49, 2015.

KONG A., GUDBJARTSSON D. F., SAINZ J., JONSDOTTIR G. M., GUDJONSSON S. A. & RICHARDSSON B. A high-resolution recombination map of the human genome. **Nature Genetics**, 31:241–7, 2002.

KONG A., BARNARD J., GUDBJARTSSON D. F., THORLEIFSSON G., JONSDOTTIR G., SIGURDARDOTTIR S. & LAMB N. E. Recombination rate and reproductive success in humans. **Nature Genetics**, 36:1203, 2004.

KOREN A., BEN-AROYA S. & KUPIEC M. Control of meiotic recombination initiation: a role for the environment? **Current Genetics**, 42:129–39, 2002.

KREBS, S., I. MEDUGORAC, S. RÖTHER, K. STRÄSSER, AND M. FÖRSTER. A missense mutation in the 3-ketodihydrosphingosine reductase *FVT1* as candidate causal mutation for bovine spinal muscular atrophy. **Proceedings of the National Academy of Sciences**, v. 104, n. 16, p. 6746-6751, 2007.

KUNZ, E., S. ROTHAMMER, H. PAUSCH, H. SCHWARZENBACHER, F. SEEFRIED, K. MATIASEK, D. SEICHTER, I. RUSS, R. FRIES, AND I. MEDUGORAC. Confirmation of a non-synonymous SNP in *PNPLA8* as a candidate causal mutation for Weaver syndrome in Brown Swiss cattle. **Genetics Selection Evolution**, v. 48, n. 1, p. 21, 2016.

LANDER, E.S.; BOTSTEIN, D.. Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. **Science**, v. 236, n. 4808, p. 1567-1570, 1987.

LAWLOR, T.J., P.M. VANRADEN, D. NULL, J. LEVISEE, AND B. DORHORST. Using haplotypes to unravel the inheritance of Holstein coat color. In **Proceedings of the 1th world congress on genetics applied to livestock production**, p. 1-3, 2014.

LICHTEN, M., & GOLDMAN, A. S. H. Meiotic recombination hotspots. **Annual Reviews of Genetics**, *29*, 423–444, 1995

LIPKIN S. M., MOENS P. B., WANG V., LENZI M. & SHANMUGARAJAH D. Meiotic arrest and aneuploidy in MLH3-deficient mice. **Nature Genetics**, 31:385–390, 2002.

LU, S., ZONG, C., FAN, W., YANG, M., LI, J., CHAPMAN, A. R., BAI, F. Probing meiotic recombination and aneuploidy of single sperm cells by whole-genome sequencing. **Science**, v. 338, n. 6114, p. 1627-1630, 2012.

MA L., O'CONNELL J. R., VANRADEN P. M., SHEN B., PADHI A., SUN C. & DA Y. Cattle sex-specific recombination and genetic control from a large pedigree analysis. **PLoS Genetics**, 11:e1005387, 2015.

MCCLURE, M.C., D. BICKHART, D. NULL, P. VANRADEN, L. XU, G. WIGGANS, G. LIU, S. SCHROEDER, J. GLASSCOCK, J. ARMSTRONG, J.B. COLE, C.P. VAN TASSELL, AND T.S. SONSTEGARD. Bovine exome sequence analysis and targeted SNP genotyping of recessive fertility defects BH1, HH2, and HH3 reveal a putative causative mutation in SMC2 for HH3. **PloS one**, v. 9, n. 3, 2014.

MCCLURE, M., E. KIM, D. BICKHART, D. NULL, T. COOPER, J. COLE, G. WIGGANS, P. AJMONE-MARSAN, L. COLLI, E. SANTUS, G.E., LIU, S. SCHROEDER, L. MATUKUMALLI, C. VAN TASSELL, AND T. SONSTEGARD. Fine mapping for Weaver Syndrome in Brown Swiss cattle and the identification of 41 concordant mutations across NRCAM, PNPLA8 and CTTNBP2. **PLoS One**, v. 8, n. 3, 2013.

MEDUGORAC, I., D. SEICHTER, A. GRAF, I. RUSS, H. BLUM, K.H. GÖPEL, S. ROTHAMMER, M. FÖRSTER, AND S. KREBS. Bovine polledness – An autosomal dominant trait with allelic heterogeneity. **PloS one**, v. 7, n. 6, 2012.

MENZI, F., N. BESUCHET-SCHMUTZ, M. FRAGNIÈRE, S. HOFSTETTER, V. JAGANNATHAN, T. MOCK, A RAEMY, E. STUDER, K. MEHINAGIC, N. REGENSCHEIT, M. MEYLAN, F. SCHMITZ-HSU, AND C. DRÖGEMÜLLER. A transposable element insertion in APOB causes cholesterol deficiency in Holstein cattle. **Animal genetics**, v. 47, n. 2, p. 253-257, 2016.

MOURESAN, E. F., GONZÁLEZ-RODRÍGUEZ, A., CAÑAS-ÁLVAREZ, J. J., MUNILLA LEGUIZAMON, S. E. B. A. S. T. I. A. N., ALTARRIBA, J., DIAZ, C. & VARONA, L. Mapping recombination rate on the autosomal chromosomes based on the persistency of linkage disequilibrium phase among autochthonous beef cattle populations in Spain. **Frontiers in Genetics**, v. 10, p. 1170, 2019.

MYERS S., BOTTOLO L., FREEMAN C., MCVEAN G. & DONNELLY P. A fine-scale map of recombination rates and hotspots across the human genome. **Science**, 310:321–4, 2005.

NANI, J.P., PEÑAGARICANO, F. Whole-genome homozygosity mapping reveals candidate regions affecting bull fertility in US Holstein cattle. **BMC genomics**, v. 21, p. 1-9, 2020.

NEFF M. W., BROMAN K. W., MELLERSH C. S., RAY K., ACLAND G. M., AGUIRRE G. D., ZIEGLE J. S., OSTRANDER E. A. & RINE J. A second-generation genetic linkage map of the domestic dog, *Canis familiaris*. **Genetics**, 151:803–20, 1999.

NICHOLAS, F.W., AND M. HOBBS. Mutation discovery for M endelian traits in non-laboratory animals: a review of achievements up to 2012. **Animal genetics**, v. 45, n. 2, p. 157-170, 2014.

NULL, D.J., J.L. HUTCHISON, D.M. BICKHART, P.M. VANRADEN, J.B. COLE. Discovery of a haplotype affecting fertility in Ayrshire dairy cattle and identification of a putative causal variant. **Journal of dairy science**, v. 100, p. 199, 2017.

OLLIVIER L. Genetic differences in recombination frequency in the pig (Susscrofa). **Genome**, 38:1048–51, 1995.

PAN, J., SASAKI, M., KNIEWEL, R., MURAKAMI, H., BLITZBLAU, H. G., TISCHFIELD, S. E. & HOCHWAGEN, A. A hierarchical combination of factors shapes the genome-wide topography of yeast meiotic recombination initiation. **Cell**, v. 144, n. 5, p. 719-731, 2011.

PAIGEN, K., & PETKOV, P. (2010). Mammalian recombination hot spots: properties, control and evolution. **Nature Reviews Genetics**, v. 11, n. 3, p. 221–233.

PERIPOLLI, E., METZGER, J., DE LEMOS, M. V. A., STAFUZZA, N. B., KLUSKA, S., OLIVIERI, B. F. & LÔBO, R. B. Autozygosity islands and ROH patterns in Nellore lineages: Evidence of selection for functionally important traits. **BMC genomics**, v. 19, n. 1, p. 680, 2018.

ROSEN, B., BICKHART, D., SCHNABEL, R., KOREN, S., ELSIK, C., ZIMIN, A., DREISCHER, C., SCHULTHEISS, S., HALL, R., SCHROEDER, S.G. AND VAN TASSELL, C.P. February. Modernizing the bovine reference genome assembly. In **Proceedings of the 11th world congress on genetics applied to livestock production**, p. 11-16, 2018.

ROTHAMMER, S., A. CAPITAN, E. MULLAART, D. SEICHTER, I. RUSS, AND I. MEDUGORAC. The 80-kb DNA duplication on BTA1 is the only remaining candidate mutation for the polled phenotype of Friesian origin. **Genetics Selection Evolution**, v. 46, n. 1, p. 44, 2014.

SAHANA, G.; NIELSEN, U. S.; AAMAND, G. P.; LUND, M. S.; GULDBRANDTSEN, B. Novel harmful recessive haplotypes identified for fertility traits in Nordic Holstein cattle. **PLoS One**, v. 8, n. 12, p. e82909, 2013.

SCHWARZENBACHER, H., J. BURGSTALLER, F.R. SEEFRIED, C. WURMSER, M. HILBE, S. JUNG., C. FUERST, N. DINHOPL, H. WEISSENBÖCK, B. FUERST-WALTL, M. DOLEZAL, R. WINKLER, O. GRUETER, U. BLEUL, T. WITTEK, R. FRIES, AND H. PAUSCH. A missense mutation in *TUBD1* is associated with high juvenile mortality in Braunvieh and Fleckvieh cattle. **BMC genomics**, v. 17, n. 1, p. 400, 2016.

SCHÜTZ, E., C. WEHRHAHN, M. WANJEK, R. BORTFELD, W.E. WEMHEUER, J. BECK, AND B. BRENIG. The Holstein Friesian lethal haplotype 5 (HH5) results from a complete deletion of TBF1M and cholesterol deficiency (CDH) from an ERV-(LTR) insertion into the coding region of APOB. **PloS one**, v. 11, n. 4, 2016.

SHANKS, R.D., D.B. DOMBROWSKI, G.W. HARPESTAD, AND J.L. ROBINSON. Inheritance of UMP synthase in dairy cattle. **Journal of Heredity**, v. 75, n. 5, p. 337-340, 1984.

SHEN B., JIANG J., SEROUSSI E., LIU G. E. & MA L. Characterization of recombination features and the genetic basis in multiple cattle breeds. **BMC Genomics**, 19:304, 2018.

SHERMAN S. L., PETERSEN M. B., FREEMAN S. B., HERSEY J., PETTAY D., TAFT L. & HASSOLD, T. J. Non-disjunction of chromosome 21 in maternal meiosis I: evidence for a maternal age-dependent mechanism involving reduced recombination. **Human Molecular Genetics**, 3:1529-1535, 1994.

SHUSTER, D.E., M.E. KEHRLI JR., M.R. ACKERMANN, AND R.O. GILBERT. Identification and prevalence of a genetic defect that causes leukocyte adhesion deficiency in Holstein cattle. **Proceedings of the National Academy of Sciences**, v. 89, n. 19, p. 9225-9229, 1992.

SONSTEGARD, T. S.; COLE, J. B.; VANRADEN, P. M.; VAN TASSELL, C. P.; NULL, D. J.; SCHROEDER, S. G.; BICKHART, D.; MCCLURE, M. C. Identification of a nonsense mutation in CWC15 associated with decreased reproductive efficiency in Jersey cattle. **PLoS One**, v. 8, n. 1, p. e54872, 2013.

TAO S., WU J., YAO D., CHEN Y., YANG W. & TONG C. Identification of recombination events in outbred species with next-generation sequencing data. **BMC Genomics**, 19: 398, 2018.

TESELING CF, PARNELL P. How Angus breeders have reduced the frequency of deleterious recessive genetic conditions. **Association of Advancement Animal Breeding and Genetics. AAABG**, v. 20, p. 558-61, 2013.

THOMSEN, B., P.H. NISSEN, J.S. AGERHOLM, C. BENDIXEN. Congenital bovine spinal dysmyelination is caused by a missense mutation in the *SPAST* gene. **Neurogenetics**, v. 11, n. 2, p. 175-183, 2010.

TORTEREAU F., SERVIN B., FRANTZ L., MEGENS H. J., MILAN D., ROHRER G., WIEDMANN R., BEEVER J., ARCHIBALD A. L. & SCHOOK L. B. A high

density recombination map of the pig reveals a correlation between sex-specific recombination and GC content. **BMC Genomics**, 13:586, 2012.

VANRADEN, P., D. NULL, J. HUTCHISON, D. BICKHART, AND S. SCHROEDER. Jersey haplotype 2 (JH2). Changes to evaluation system (August 2014). Council on Dairy Cattle Breeding. 2014.

VANRADEN, P., D. NULL, J. HUTCHISON, AND T. COOPER. New fertility and stillbirth haplotypes and changes in haplotype status. Changes to evaluation system (August 2013). Council on Dairy Cattle Breeding. 2013.

VANRADEN, P.M., D.J. NULL, T.S. SONSTEGARD, H.A. ADAMS, C.P. VAN TASSELL, AND K.M. OLSON. Fine mapping and discovery of recessive mutations that cause abortions in dairy cattle. **Journal of dairy science**, v. 95, 2012.

VANRADEN, P. M.; OLSON, K. M.; NULL, D. J.; HUTCHISON, J. L. Harmful recessive effects on fertility detected by absence of homozygous haplotypes. **Journal of dairy science**, v. 94, n. 12, p. 6153-6161, 2011.

VENHORANTA, H., H. PAUSCH, K. FLISIKOWSKI, C. WURMSER, J. TAPONEN, H. RAUTALA, A. KIND, A. SCHNIEKE, R. FRIES, H. LOHI, AND M. ANDERSSON. In frame exon skipping in *UBE3B* is associated with developmental disorders and increased mortality in cattle. **BMC genomics**, v. 15, n. 1, p. 890, 2014.

WANG, J., FAN, H. C., BEHR, B., & QUAKE, S. R. Genome-wide single-cell analysis of recombination activity and de novo mutation rates in human sperm. **Cell**, v. 150, n. 2, p. 402-412, 2012.

WRIGHT, S. Coefficients of inbreeding and relationship. **The American Naturalist**,

v. 56, p. 330–338, 1922.

ZAVAREZ, L. B.; UTSUNOMIYA, Y. T.; CARMO, A. S.; NEVES, H. H. R.; CARVALHEIRO, R.; FERENCAKOVIC, M.; PÉREZ O'BRIEN, A. M.; CURIK, I.; COLE, J. B.; VAN TASSELL, C. P.; DA SILVA, M. V. G. B.; SONSTEGARD, T. S.; SÖLKNER, J.; GARCIA, J. F. Assessment of autozygosity in Nellore cows (Bos indicus) through high-density SNP genotypes. **Frontiers in Genetics**, v. 6, p. 1–8, 2015.

Chapter 2 – Identification of candidate lethal haplotypes in Nellore cattle

ABSTRACT - With genomic data, lethal recessives may be discovered from haplotypes that are common in the population but are never homozygous in live animals. The objective of the present study was to identify lethal haplotypes in a Nellore population, based on expected population frequencies. The method requires genotypes only from apparently normal individuals and not from affected embryos. The conflict.f90 software was used to correct Mendelian errors and fill missing SNP using parental genotypes. A total of 4,447 animals were genotyped with a high-density panel (777,962 SNP markers) and 4,041 with a panel containing 74,677 markers, which were imputed to the HD panel using findhap.f90 software v3. The genomic position of the markers were based on the ARS-UCD1.2 *Bos taurus* genome assembly. By default, the program first examined haplotypes constructed using 2,000 markers, then 632 markers, and finally identified haplotypes with ≤200 markers for further analysis. Expected numbers of homozygous individuals were calculated through two methods: Simple - assuming random mating and using the number of individuals genotyped divided by 4 and multiplied by the square of the carrier frequency; and Mating - using the actual mating pattern for calculating the number of carrier service sire × carrier maternal grandsire matings divided by 4. The probabilities of observing 0 homozygotes when n are expected were obtained by 2 analogous formulas that were used to obtain expectations. Crossover haplotypes were identified directly by findhap.f90 software v3 when any new haplotype of progeny had been formed from the original 2 haplotypes of a parent with just 1 recombination. For the functional analyses, we used SIFT scores to classify mutations as deleterious or tolerant. Twenty-six haplotypes had high expected frequency but no homozygotes observed. Of these, the haplotypes with the largest number of expected homozygotes were on chromosomes 8, located at 59,384,409-60,230,288 (simple method – 43; mating method – 9), and 23 located at 15,237,396-16,040,973 (simple method – 29; mating method - 4). For the simple and mating methods the probabilities ranged from 6.190E-02 to 2.523E-19 and 0.1001 to 2.38E-05, respectively. Two haplotypes on chromosome 1:56,408,787-56,947,331 and on 21:22,003,502-22,770,526 overlaps with previously known defects, Deficiency of Uridine Monophosphate Synthase and Brachyspina syndrome, respectively. Furthermore, the candidate lethal haplotypes on chromosome 7:52,418,587-53,136,816 and on chromosome 12:27,930,543-28,993,509 match with potential signatures of selection found previously in a similar population of Nellore cattle. We found 55 candidate genes responsible for harboring the deleterious mutations and 11 genes with tolerant mutations. The phenotypes underlying these putative regions are unknown for this population and should be the subject of future investigation. Tests to find lethal haplotypes carriers could help breeders to implement selection actions in order to eliminate these haplotypes from the population or manage matings in order to avoid homozygous.

Keywords: beef cattle, crossovers, expected haplotype frequency, homozygous individuals, sliding windows

2.1. INTRODUCTION

Nellore is the most important beef breed raised in Brazil and the use of artificial insemination has allowed great diffusion of superior genetic material; however, it has also increased the excessive use of certain paternal lineages (FARIA et al., 2009). The widespread use of a few Nellore bulls and artificial insemination generated long homozygous segments in the genomes of more than 70% of the animals in a Nellore population studied by Zavarez et al. (2015). These homozygous segments indicate the occurrence of inbreeding and increase the probability of expression of harmful recessive genes and may also reduce the animal's adaptive value (PERIPOLLI et al., 2018; BOSSE et al., 2019; MARTIKAINEN et al., 2020). The extent of the impact these alleles have on fertility and fitness in beef cattle is unknown yet, mainly because in general, embryo losses are not monitored (HOFF et al., 2017).

Pedigree analysis and strategies requiring genotypes from abnormal embryos were necessary to provide evidence about the effects of recessive alleles (Charlier et al,. 2008). These same authors stated that, the possibility of identifying lethal recessive variants through dense molecular markers, increased with the inclusion of large-scale genotyping technologies. However, the traditional approach is not able to identify deleterious genetic mutations in cases where the phenotype is not detectable in typical production systems, such as early embryonic deaths (WU et al., 2019).

VanRaden et al. (2011a) proposed a method to find lethal haplotypes that are common in the population without using any phenotypes. Haplotypes are constructed from SNP genotypes and those that appear in the homozygous state far less frequently than expected are identified for further study. This methodology has been used, for example, by Sonstegard et al. (2013), Fritz et al. (2018) and Hozé et al.

(2020) in dairy cattle breeds, Hoff et al. (2017) and Jenko et al. (2019) in beef cattle, and Häggman and Uimari (2016) in Yorkshire pigs.

The haplotypes found by this method are termed original haplotypes and often trace mainly to just one source ancestor; a descendant may receive the whole haplotype or just a part if a crossover occurs, these haplotypes are termed crossover haplotypes (VANRADEN et al., 2011b). Therefore, crossover haplotypes contain part of the original haplotype and part of some other haplotype. Further analyses can be done to detect crossover haplotypes, in order to refine the map location and include more carriers of the suspect region with the lethal mutation.

Identifying and monitoring lethal recessive haplotypes in a population can help breeders to anticipate problems caused by disadvantageous variants, through the implementation of selection and mating strategies and, potentially, gene editing (COLE, 2015; COLE, 2019). Studies aiming to identify lethal recessive haplotypes in Nellore cattle, using this method, were not found in the literature. The objectives of the present study were: 1) to identify lethal recessive haplotypes based on expected population frequencies; 2) to detect crossovers haplotypes that had at least 50% overlap with the original haplotypes; and, 3) to identify, through functional analyses, tolerant and deleterious mutations in candidate genes underlying lethal haplotypes.

2.2. MATERIAL AND METHODS

2.2.1. Pedigree and genotypic data information

Pedigree information recovered from historical records comprised 2,688,124 animals. These records were incomplete due to the use of multiple-sire matings; 57.4% and 22.4% of pedigrees included unknown sires and dams, respectively.

Unkown dams were part of the base population. Genotype inconsistences between parents and progeny were adjusted using conflict.f90 software, which corrects for Mendelian errors and fills missing SNP using parental genotypes (VANRADEN, 2015).

A total of 8,488 Nellore animals obtained from the Alliance Nellore dataset (www.gensys.com.br) were genotyped, including 2,025 females, 1,473 young males, and 949 bulls using the Illumina Bovine HD Bead chip (Illumina, San Diego, CA, USA) that contains 777,962 SNPs; and 716 females, 3,146 young males, and 179 bulls genotyped using the GeneSeek® (Genomic Profiler Indicus HD - GGP75Ki NEOGEN) which contains 74,677 SNPs (75K). Animals genotyped with lower density panel (75K) were imputed to the HD panel (42% missing genotypes) using findhap.f90 software v3 (VANRADEN, 2011).

Genotype quality control was performed using R software version 3.6.1 (R Core Team, 2019) with the SNPStats package version 1.38.0. (CLAYTON, 2020) and was implemented considering the SNPs markers with a GenCall score greater than 0.90 and located in autosomal chromosomes. In addition, only SNP with call rate greater than 0.90 and not in highly significant deviation from Hardy-Weinberg equilibrium (P>10-5) were considered. SNP at the same genomic position and monomorphic were removed. Samples with call rates below 0.90 were also removed. The SNP positions within a chromosome were based on the ARS-UCD1.2 *Bos taurus* genome assembly (ROSEN et al., 2020). After editing and merging the 2 markers datasets, HD and 75K, 596,834 SNPs and 8,470 animals remained to carry out the further analyses.

2.2.2. Identification of candidate lethal haplotypes

The haplotypes were constructed using the sliding windows method implemented in findhap.f90 software v3 (VANRADEN, 2011). By default, the program first examined haplotypes of 2,000 markers, then 632 markers, and finally identified haplotypes with up to 200 markers, used for further analyses. Haplotypes that were never homozygous state but had the highest frequencies were examined first because of their potential importance and to maximize power of detection. New haplotypes, created by recent crossovers, had no opportunity to become homozygous and also, those were not likely to be lethal if they were simply a mixture of 2 ancestral non-lethal haplotypes.

Following VanRaden et al. (2011a), two methods were used to calculate the expected number of homozygous individuals for each haplotype: Simple method – it was assumed that all members of the population were on random mating over time. Expected number of homozygous was given by the number of genotyped individuals divided by 4 and multiplied by the square of the carrier frequency; Mating method – the actual mating pattern generating the genotyped individuals in the population, was used. In this case, the number of carrier service sire × carrier maternal grandsire pairs divided by 4 was used to obtain the expected number of homozygous. Allele frequencies for maternal granddams and maternal grandsire were assumed to be equal.

The probabilities of observing 0 homozygotes when n are expected were obtained by 2 analogous formulas that were used to obtain expectations, following VanRaden et al. (2011a) and Jenko et al. (2019). The first test, for the simple method the probability is: 2DD = (1 □²/4)□, where the probability of non-observing the homozygous animals (□) depends on the carrier frequency of the heterozygous

animals (□) and the number of genotyped animals (□). For the mating method that used the actual mating pattern, probability follows a Bernoulli process and is equal to 0.75 raised to the power of the observed number of carrier service sire × carrier maternal grandsire pairs.

Due to thousands of haplotypes that could have 0 homozygotes by chance, to define a recessive lethal haplotype region, we set the following conditions, following Wu et al. (2019): (1) the haplotype carrier frequency had to be higher than 1%, (2) the number of expected homozygous individuals for the haplotype had to be higher than the observed homozygotes, (3) the probabilities of observing 0 homozygotes had to be small enough, close to 0. All the haplotypes satisfying these conditions were selected.

Further analyses were performed to identify crossovers haplotypes, directly from an output file from findhap.f90 software v3 (VANRADEN, 2011). Any other haplotype that had 50% or more overlap and identical SNPs allele to the lethal original haplotype was considered as a crossover haplotype. With these analyses, a larger number of carriers, with part of the defective haplotype, were detected (VANRADEN et al., 2011a). Details about the use of crossovers and accuracy of detection methods are reported by VanRaden et al. (2011b).

2.2.3. Identification of mutations and enrichment analyses

To identify the causal variants underlying the putative lethal regions the VEP (variant effect predictor) tool, from ENSEMBL software (https://www.ensembl.org/info/docs/tools/vep/index.html) was used, considering the ARS-UCD1.2 *Bos taurus* genome assembly (ROSEN et al., 2020). SIFT

(Sorting Intolerant From Tolerant) scores were used to predict whether an amino acid substitution affects or not the protein function. The SIFT provides a score and a qualitative prediction, that is, the mutation can be classified as "deleterious" or "tolerant". If the score is less than 0.05 it is likely to be deleterious, otherwise it is "tolerant" (NG & HENIKOFF, 2003).

The Cytoscape plug-in version 3.8.0 of ClueGO v2.5.7 software and CluePedia v1.5.7 software (SHANNON et al., 2003) were used to perform combined functional analysis for annotated genes that harbored tolerant and deleterious mutations with high impact, separately. GO terms with P_value less 0.05 were deemed as significant. Positional candidate genes related to significant GO groups and terms were considered as functional candidates. Developmental genes of all stages of life were described and marked as candidate genes.

2.3. RESULTS AND DISCUSSION

2.3.1. Candidate lethal haplotypes

We have used haplotypes in sliding windows with variable number of SNPs (5 to 186). This approach seems to provide a good balance between detection of recessive lethal alleles and haplotype diversity (VANRADEN et al., 2011a). The average length of each window was 0.836201 Mb, slightly smaller compared to Hoff et al. (2017) that found, on average, windows of 1.1428 (Mb) in length and used 20-marker fixed sliding windows. A total of 3,002 segments were built and, on average, 900 haplotypes on each one were found.

Using the thousands of haplotypes found, two tests were applied to search for haplotypes carrying recessive lethal alleles. The tests were based on the number of

expected recessive homozygous individuals either in the whole population or from matings between carriers. Twenty-six haplotypes, distributed on thirteen chromosomes and ranked by mean carrier's frequency, had a sufficiently high expected frequency of homozygotes in both methods, but none were observed (Table 1). The frequency of carrier animals with the original haplotype is usually about twice the lethal haplotype mean frequency, since a heterozygous animal counts only 0.50 to allele frequency due to one bad and one normal allele.

Table 1 – Potentially lethal haplotypes, sorted by carrier frequencies in genotyped Nellore animals

Potentially lethal defect[1]	Chr[2]	Haplotype				Number of carriers	Expected Homozygotes numbers	
		Start position[3]	End position[3]	Length (Mb)	Carrier frequency, %		Simple[4]	Mating[5]
1133.2	8	59,384,409	60,230,288	0.845879	7.10	1,203	43	9
2596.2	23	15,237,396	16,040,973	0.803577	5.84	989	29	4
1582.4	12	2,546,283	3,201,799	0.655516	5.33	902	24	2
158.1	1	137,838,183	138,582,556	0.744373	5.26	891	23	4
2542.4	22	32,650,017	33,168,208	0.518191	5.19	880	23	2
1583.4	12	3,212,875	4,146,558	0.933683	5.14	870	22	2
943.7	7	19,010,499	19,846,412	0.835913	5.08	860	22	4
63.1	1	56,408,787	56,947,331	0.538544	4.99	845	21	4
1910.1	15	27,939,003	28,605,653	0.66665	4.90	830	20	5
944.7	7	19,854,110	20,775,225	0.921115	4.45	753	17	4
1958.12	15	66,815,440	67,511,873	0.696433	4.36	738	16	3
1148.3	8	70,484,376	71,306,212	0.821836	4.14	702	15	3
984.4	7	52,418,587	53,136,816	0.718229	3.26	552	9	3

2444.1	21	22,003,502	22,770,526	0.767024	3.21	544	9	2
1773.4	13	79,152,234	80,162,937	1.010703	3.15	533	8	2
336.58	2	124,279,927	125,250,396	0.970469	3.03	514	8	2
338.55	2	126,166,386	127,051,942	0.885556	3.03	514	8	2
339.63	2	127,053,602	127,875,933	0.822331	3.01	510	8	2
214.60	2	22,551,147	23,432,529	0.881382	2.92	494	7	2
215.58	2	23,439,757	24,004,909	0.565152	2.91	493	7	3
217.74	2	24,769,013	25,749,934	0.980921	2.87	486	7	2
496.112	3	118,774,674	119,885,137	1.110463	2.72	461	6	2
1613.5	12	27,930,543	28,993,509	1.062966	1.98	335	3	2
2788.34	26	14,680,252	15,760,181	1.079929	1.98	335	3	2
2166.93	17	68,089,794	68,912,917	0.823123	1.97	334	3	3
416.100	3	51,380,123	52,274,782	0.894659	1.81	307	3	2

[1]Defects are identified by DNA segment number and haplotype within segment (e.g., 1133.2 indicates segment 1133, haplotype 2)
[2]Chromosome
[3]ARS-UCD1.2 *Bos taurus* genome assembly (ROSEN et al., 2020)
[4]Number of individuals genotyped divided by 4 and multiplied by square of carrier frequency
[5]Number of carrier service sire × carrier maternal grandsire matings divided by 4

The carrier frequencies of the potentially lethal haplotypes ranged from 1.81 to 7.10 %. Our results were similar to those reported by Sahana et al. (2013) in a Nordic Holstein cattle population. They found 17 haplotypes harboring possible recessive lethal alleles and carrier frequencies, ranging from 2.7 to 6.7%. On the other hand, we have found lethal haplotypes carrier frequencies lower than Jenko et al. (2019) in five beef cattle breeds, Aberdeen Angus, Charolais, Hereford, Limousin, and Simmental. The authors reported frequencies ranging from 1.2 to 33.5%. The high haplotype frequencies found by these authors suggest there is selection that maintains the high frequency on these populations.

The number of expected homozygous obtained through the method that considers random mating (simple) was higher (ranged from 2 to 42) than that from the mating method (ranged from 2 to 9 - Table 1). This result could be expected due to incomplete pedigree and non-genotyped ancestors in our dataset. With a complete pedigree and almost all ancestral bulls genotyped, as in US dairy cattle, the expected number of homozygotes with both methods would be similar (VANRADEN et al., 2011a).

Thousands of haplotypes showed zero homozygous. However, only the haplotypes with the highest expected number of homozygous and small probabilities of zero homozygous occurrences were selected. For the simple and mating methods the probabilities of observing zero homozygous, ranged from 2.523E-19 to 6.190E-02 and 2.38E-05 to 0.1001, respectively. Adjacent segments showing similar number of carriers might be a good sign indicating that several nearby segments may all be tracking the same lethal variant, such as on chromosome 2 (336.58; 338.55; 339.63

and 214.60; 215.58; 217.74 haplotypes), on chromosome 7 (943.7; 944.7 haplotypes) and on chromosome 12 (1582.4; 1583.4 haplotypes).

Most of the studies found in literature used the UMD v3.1 assembly, previous to the ARS-UCD1.2 (ROSEN et al., 2020) used in the present study. In order to find similarities between our sequences and those already reported by different authors, the suite-2sequences - BLAST tool (ZHANG et al. 2000, available at https://blast.ncbi.nlm.nih.gov/Blast.cgi), based on Percent Identity parameter, was used.

The haplotype 63.1 on chromosome 1:56,408,787-56,947,331 (Table 1) is an overlap region previously reported as a known defect in Holstein cattle (VANRADEN et al., 2011a). Comparing this region to that reported by VanRaden et al. (2011), between 56-57Mbp on chromosome 1, we have found 99.98% similarity. This haplotype region harbors recessives alleles responsible for the Deficiency of Uridine Monophosphate Synthase (DUMPS), which is known to contribute to the embryonic and fetal mortality in cattle (SHANKS & ROBINSON, 1989; GHANEM et al., 2006). The autosomal recessive form of the gene DUMPS, if present in homozygous condition, may lead to embryonic mortality before day 40 of gestation (SHANKS et al., 1984).

The region 22,003,502-22,770,526 on chromosome 21, which harbors the haplotype 2444.1 (Table 1), overlap with a region reported by Sahana et al. (2013) that showed significant effects on non-return rates and calving interval. The region we found showed 99.99% of similarity with that from Sahana et al. (2013). Furthermore, this haplotype is close to or overlaps with a known defect at the Brachyspina locus on BTA21. The observed carrier haplotype frequency in our

population (3.2%) was lower than those described by VanRaden et al. (2011a) and Charlier et al. (2012) of 6.0% and 7.4%, respectively, for this recessive defect in the US Holstein and Dutch Holstein-Friesian breeds, respectively. But, it was close to the observed carrier frequency (4%) in the Nordic Holstein population reported by Sahana et al. (2013). Brachyspina syndrome is a rare congenital defect affecting cattle which is characterized by severely reduced body weight, growth retardation, extensive vertebral malformations causing a significant shortening of the spine and long and slender limbs (CHARLIER et al., 2012). In addition, affected calves exhibit inferior brachygnatism (i.e. uneven alignment of the upper and lower teeth), as well as malformation of the inner organs, in particular, heart, kidneys and gonads (AGERHOLM et al., 2006).

The locus of the candidate lethal haplotype, 984.4 found on chromosome 7:52,418,587-53,136,816 and 1613.5 on chromosome 12:27,930,543-28,993,509 (Table 1), match with the potential signatures of selection found by Zavarez et al. (2015) in a similar population of Nellore cattle. These two haplotypes, 984.4 and 1613.5, showed 84.24% and 99.98%, respectively, of similarity with the regions reported by those authors. The region of the haplotype 984.4 is orthologous to the human chromosome segment 5q31-q33, contains the Th2 cytokine gene cluster, which is related to allergy control and resilience against infectious diseases such as malaria (GARCIA et al., 1998; RIHET et al., 1998; FLORI et al., 2003; HERNANDEZ-VALLADARES et al., 2004) and leishmaniasis (JERONIMO et al., 2007). This region was also reported as being involved in the control of parasitemia in cattle infected by *Trypanosoma congolense* (HANOTTE et al., 2003).

Considering the crossover haplotypes, the number of markers per window ranged from 9 to 183 SNPs (Table 2). From the 26 potentially lethal original haplotypes 12 had from 1 (haplotype 1958.12) to 25 (haplotype 1910.1) crossovers, including up to 447 carrier animals (haplotype 1613.5 – Table 2) from the suspected area with the lethal segment. The haplotypes 63.1 and 2444.1 (Table 1) that were previously reported for harboring know defects as DUMP and Brachyspina syndrome, respectively, had at least 2 crossovers (Table 2) and included together ten carriers of the suspected area with the lethal segment.

Crossover haplotypes were detected directly from genotyped animals within the pedigree and contain part of the original haplotype and part of some other haplotype. Depending on exact location, animals with crossovers haplotypes might or might not contain the lethal mutation, so the animal's status could be labeled as inconclusive or semi-lethal (VANRADEN et al., 2011b). These animals could be mated to animals carrying the original haplotype, to further narrow the suspect areas with the lethal segment.

Considering the increasing number of genotyped animals, a larger number of rare recessive alleles may be discovered in a near future. These lethal recessives might have a great impact on the population's fitness and genotyping large number of animals should allow their management. Testing animals for lethal haplotypes can help breeders to anticipate problems caused by disadvantageous variants through the implementation of selection actions, management of matings and, potentially, gene-editing, maintaining diversity within breed and avoiding inbreeding and with that, decreasing lethal alleles (COLE, 2015; COLE, 2019).

Table 1 - Crossovers haplotypes identified from the original lethal haplotypes

Original haplotype[1]	Chr[2]	Number of crossovers haplotypes	Number of carriers[3]	Suspect area with the lethal segment			Frequency of crossovers haplotypes, %
				Start position[4]	End position[4]	Number of SNP's	
1133.2	8	8	18	59,579,444	59,693,913	33	0.10625
2596.2	23	3	12	15,289,678	15,863,128	121	0.07083
158.1	1	5	11	137,838,183	138,148,961	59	0.06494
2542.4	22	10	38	33,069,600	33,111,433	9	0.2243
63.1	1	2	4	56,444,206	56,495,128	23	0.02361
1910.1	15	25	98	28,249,372	28,292,541	26	0.57848
1958.12	15	1	6	66,869,687	67,511,873	176	0.03542
984.4	7	5	22	52,418,587	52,888,439	124	0.12986
2444.1	21	3	6	22,003,502	22,391,774	114	0.03542
339.63	2	4	9	127,053,602	127,408,785	102	0.05313
1613.5	12	17	447	28,020,799	28,090,368	23	2.63902
2166.93	17	4	31	68,135,597	68,912,917	183	0.18301

[1]Defects are identified by DNA segment number and haplotype within segment (e.g., 1133.2 indicates segment 1133, haplotype 2)
[2]Chromosome
[3]Number of carriers of the crossover haplotypes
[4]ARS-UCD1.2 *Bos taurus* genome assembly (ROSEN et al., 2020)

2.3.2. Identification of mutations and enrichment analyses

The deleterious and tolerant mutations with high impact found through the SIFT scores, are reported on Tables S1 and S2, respectively. The SIFT score predicts whether an amino acid substitution in a protein will have a phenotypic effect; some DNA repair mechanisms allow the cell to "tolerate" the presence of DNA damage, even if temporarily (KUNZ et al., 2000). The mutations were classified as high impact based in a functional impact score (FIS) that analyses evolutionary information of the amino acids, assuming which are highly important for the structure and function of the protein (REVA e al., 2011).

We found 55 genes responsible for harboring deleterious mutations and 11 genes related with tolerant mutations, which were submitted to enrichment analysis. GO terms related to immune system and pathways were found using Cytoscape plug-in of ClueGO software. For deleterious genes (Figure 1), terms as positive regulation of NK T cell differentiation (GO:0051137), T cell proliferation (GO:0042111), T cell activation (GO:0042110), T cell lineage commitment (GO:0002360), B cell activation (GO:0042113), B cell receptor signaling pathway (0050853), positive regulation of lymphocyte activation (GO:0050870), lymphocyte proliferation (GO:0046651), were found.

Natural Killer (NK) Cells are similar to T and B cells, and also developed from common lymphoid progenitor cells (PAUL & LAL, 2017). However, as cells of the innate immune system, NK cells are classified as group I Innate Lymphocytes (ILCs) and respond quickly to a wide variety of pathological challenges (ABEL et al., 2018). The same authors reported that NK cells may play an important role in pregnancy,

being found in the placenta; furthermore, NK cells may help detecting and controlling early signs of cancer and is known for killing virally infected cells.

The defective lymphocytes level is associated with a wide variety of immune responses, such as viral infections, blood cancers, autoimmune diseases, stem cell transplantation, and pregnancy which indicate their potential diversity and a relationship between the deleterious mutations found with body's defense (SOJKA et al., 2014). The haplotypes where the deleterious genes are located could be related with the deficient defenses in bovines, but validation using phenotypes are necessary to confirm.

The tolerant genes found (Figure 2) are related to taste transduction pathway (Kegg pathway database – map04742), which was reported in some previous cattle studies (ABDALLA et al., 2016; OLIVIERI et al., 2016; CHEN et al., 2018) and is responsible for five basic taste systems recognized by animals (bitter, sweet, sour, salty and umami). The palatability is a sensorial response to feed intake and is related to the animal alimentary choices (MARGOLSKEE, 1993). Genes and haplotypes found, related to palatability, could indicate an adaptation to adverse feed conditions tolerance.

The serotonergic synapse pathway (Kegg pathway database map04726), responsible for Serotonin metabolism, plays important roles in physiological functions such as learning and memory, emotion, sleep, pain, motor function and endocrine secretion, as well as in pathological states including abnormal mood and cognition (ALBERT & TIBERI, 2001; MILAN et al., 2008). Serotonin was also related to gastrointestinal tract operation, bone metabolism, milk production, liver metabolism and cell division, important systems to organism survival (LEZOUALC'H & ROBERT,

2003; HERNANDEZ et al., 2008; COLLIER et al., 2012) and related to tolerant genes in this study.

Another important pathway related to tolerant genes is *PPAR* signaling pathway (Kegg pathway database - map03320). The *PPAR* metabolism plays an important role in the cellular lipid circulation and oxidation and was related with productive traits in bovines as lipid content in *Longissimus thoracis* muscle tissue (BERTON et al., 2016; SANTOS SILVA et al., 2019).

The *BRCA2* gene, found on functionally grouped terms network for tolerant mutations (Figure 2), was related with Fanconi anemia in humans (HOWLETT et al., 2002) and pathways in cancer, such as pancreatic and breast cancer (WOOSTER et al., 1995; IQBAL et al., 2012). Homologous recombination also was related with this gene which, from the perspective of evolution, is important to generate genetic diversity. Moreover, this process is essential to maintaining genetic information, guaranteeing its accurate transmission from parent to offspring. (VON NICOLAI et al., 2018).

Although candidate genes and GO terms were associated with body's defense, cancers and autoimmune diseases, the phenotypes underlying these supposed regions are unknown for this population. These defects and deficiencies should be subject of future investigation and could be confirmed by finding a few bulls and cows that are carriers of the lethal haplotypes and mating them.

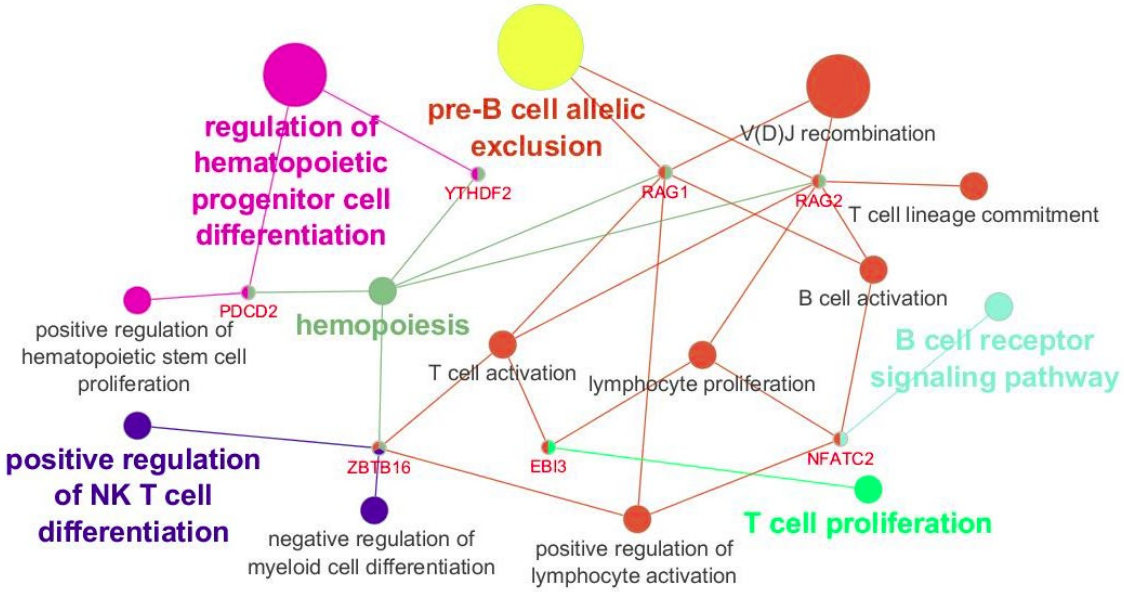

Figure 1 - Functionally grouped terms network for deleterious genes with high impact: each color represents a significant functional group with associated GO terms (P_value < 0,05) in which candidate genes were distributed

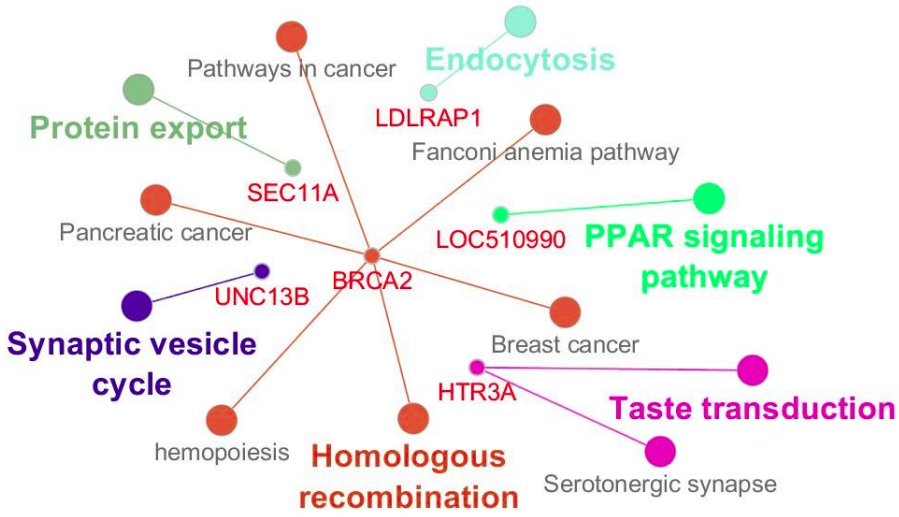

Figure 2 - Functionally grouped terms network for tolerant genes with high impact: each color represents a significant functional group with associated GO terms (P_value < 0,05) in which candidate genes were distributed

2.4. CONCLUSIONS

A total of twenty-six potentially lethal haplotypes were identified and candidate regions will be study in confirmatory analyzes (molecular and statistical) later. Two of them are in overlap with previously known defects such as Deficiency of Uridine Monophosphate Synthase and Brachyspina syndrome. We have detected 1 to 25 crossovers in 12 lethal original haplotypes, allowing the identification of up to 447 carrier animals from the suspected area with the lethal segment. We found 55 candidate genes that may harbor deleterious mutations and 11 genes related with tolerant mutations, both associated with body's defense, cancers and autoimmune diseases. Tests to find lethal haplotypes carriers could help breeders to implement selection actions in order to eliminate these haplotypes from the population or manage matings, in order to avoid lethal alleles dispersion.

2.5. REFERENCES

ABDALLA, E. A., PEÑAGARICANO, F., BYREM, T. M., WEIGEL, K. A., & ROSA, G. J. M. Genome-wide association mapping and pathway analysis of leukosis incidence in a US Holstein cattle population. **Animal genetics**, v. 47, n. 4, p. 395-407, 2016.

ABEL, A. M., YANG, C., THAKAR, M. S., & MALARKANNAN, S. Natural killer cells: development, maturation, and clinical utilization. **Frontiers in immunology**, v. 9, p. 1869, 2018.

AGERHOLM, J S.; MCEVOY, F; ARNBJERG, J. Brachyspina syndrome in a Holstein calf. **Journal of veterinary diagnostic investigation**, v. 18, n. 4, p. 418-422, 2006.

ALBERT, P.R. & TIBERI, M. Receptor signaling and structure: insights from serotonin-1 receptors. **Trends in Endocrinology & Metabolism**, v. 12, n. 10, p. 453-460, 2001.

BANSAL, V.; HALPERN, A. L.; AXELROD, N.; BAFNA, V. An MCMC algorithm for haplotype assembly from whole-genome sequence data. **Genome Research**, v. 18, p. 1336–1346, 2008.

BERTON, M. P., FONSECA, L. F., GIMENEZ, D. F., UTEMBERGUE, B. L., CESAR, A. S., COUTINHO, L. L. & STAFUZZA, N. B. Gene expression profile of intramuscular muscle in Nellore cattle with extreme values of fatty acid. **BMC genomics**, v. 17, n. 1, p. 972, 2016.

BOSSE, M., MEGENS, H. J., DERKS, M. F., DE CARA, Á. M., & GROENEN, M. A. Deleterious alleles in the context of domestication, inbreeding, and selection. **Evolutionary applications**, v. 12, n. 1, p. 6-17, 2019.

CHARLIER, C.; COPPIETERS, W.; ROLLIN, F.; DESMECHT, D.; AGERHOLM, J. S.; CAMBISANO, N.;CARTA, E.; DARDANO, S.; DIVE, M.; FASQUELLE, C.;

FRENNET, J. C; HANSET, R.; HUBIN, X.; JORGENSEN, C.; KARIM, L.; KENT, M.; HARVEY, K.; PEARCE, B. R.; SIMON, P.; TAMA, N.; NIE, H.; VANDEPUTTE, S.; LIEN, S.; LONGERI, M.; FREDHOLM, M.; HARVEY, R. J.; GEORGES, M. Highly effective SNP-based association mapping and management of recessive defects in livestock. **Nature genetics**, v. 40, n. 4, p. 449, 2008.

CHARLIER, C., AGERHOLM, J. S., COPPIETERS, W., KARLSKOV-MORTENSEN, P., LI, W., DE JONG, G. & AHARIZ, N. A deletion in the bovine FANCI gene compromises fertility by causing fetal death and brachyspina. **PloS one**, v. 7, n. 8, 2012.

CHEN, N., CAI, Y., CHEN, Q., LI, R., WANG, K., HUANG, Y. & SONG, W. Whole-genome resequencing reveals world-wide ancestry and adaptive introgression events of domesticated cattle in East Asia. **Nature communications**, v. 9, n. 1, p. 1-13, 2018.

CLAYTON, D. snpStats: SnpMatrix and XSnpMatrix classes and methods. **R package version, 1.38.0**, 2020.

COLE, J.B. A simple strategy for managing many recessive disorders in a dairy cattle breeding program. **Genetics selection evolution**. v. 47, p. 94-107, 2015.

COLE, J.B. Promotion of alleles by genome engineering. **CAB reviews: Perspectives in agriculture, veterinary science, nutrition and natural resources**. v. 14, p. 1-15, 2015.

COLLIER, R. J.; HERNANDEZ, L. L.; HORSEMAN, N. D. Serotonin as a homeostatic regulator of lactation. **Domestic animal endocrinology**, v. 43, n. 2, p. 161-170, 2012.

DAETWYLER, H. D.;CAPITAN, A.; PAUSCH, H.; STOTHARD, P.; BINSBERGEN, R.; BRONDUM, R. F.; LIAO, X.;DJARI, A.; RODRIGUEZ, S. C.;GROHS, C.;ESQUERRÉ, D.; BOUCHEZ, O.; ROSSIGNOL, M. N.;KLOPP, C.; ROCHA, D.; FRITZ, S.;EGGEN, A.; BOWMAN, P. J.; COOTE, D.; CHAMBERLAIN, A. J.; ANDERSON, C.; VANTASSELL, C. P.; HULSEGGE, I.; GODDARD, M. E.;

GULDBRANDTSEN, B.; LUND, M. S.; VEERKAMP, R. F.; BOICHARD, D. A.; FRIES, R.; HAYES, B. J. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. **Nature genetics**, v. 46, n. 8, p. 858, 2014.

FARIA, F. J. C.; FILHO, A. V.; MADALENA, F. E.; JOSAHKIAN, L. A. Pedigree analysis in the Brazilian Zebu breeds. **Journal of Animal Breeding and Genetics**, v. 126, n. 2, p. 148-153, 2009.

FLORI, L., KUMULUNGUI, B., AUCAN, C., ESNAULT, C., TRAORE, A. S., FUMOUX, F. Linkage and association between Plasmodium falciparum blood infection levels and chromosome 5q31–q33. **Genes & Immunity**, v. 4, n. 4, p. 265-268, 2003.

FRITZ, S., HOZE, C., REBOURS, E., BARBAT, A., BIZARD, M., CHAMBERLAIN, A., ESCOUFLAIRE, C., VANDER JAGT, C., BOUSSAHA, M., GROHS, C. AND ALLAIS-BONNET, A. An initiator codon mutation in SDE2 causes recessive embryonic lethality in Holstein cattle. **Journal of dairy science**, v. 101, n.7, p.6220-6231, 2018.

GARCIA, A., MARQUET, S., BUCHETON, B., HILLAIRE, D., COT, M., FIEVET, N. Linkage analysis of blood Plasmodium falciparum levels: interest of the 5q31-q33 chromosome region. **The American journal of tropical medicine and hygiene**, v. 58, n. 6, p. 705-709, 1998.

GHANEM, M. E.; NAKAO, T.; NISHIBORI, M. Deficiency of uridine monophosphate synthase (DUMPS) and X-chromosome deletion in fetal mummification in cattle. **Animal reproduction science**, v. 91, n. 1-2, p. 45-54, 2006.

HÄGGMAN, J.; UIMARI, P. Novel harmful recessive haplotypes for reproductive traits in pigs. **Journal of Animal Breeding and Genetics**, v. 134, n. 2, p. 129-135, 2017.

HANOTTE, O., RONIN, Y., AGABA, M., NILSSON, P., GELHAUS, A., HORSTMANN, R. Mapping of quantitative trait loci controlling trypanotolerance in a cross of tolerant West African N'Dama and susceptible

East African Boran cattle. **Proceedings of the National Academy of Sciences**, v. 100, n. 13, p. 7443-7448, 2003.

HERNANDEZ, L. L., STIENING, C. M., WHEELOCK, J. B., BAUMGARD, L. H., PARKHURST, A. M., & COLLIER, R. J. Evaluation of serotonin as a feedback inhibitor of lactation in the bovine. **Journal of Dairy Science**, v. 91, n. 5, p. 1834-1844, 2008.

HERNANDEZ-VALLADARES, M., RIHET, P., OLE-MOIYOI, O. K., AND IRAQI, F. A. Mapping of a new quantitative trait locus for resistance to malaria in mice by a comparative mapping approach with human Chromosome 5q31-q33. **Immunogenetics**, v. 56, n. 2, p. 115-117, 2004.

HOFF, J. L.; DECKER, J. E.; SCHNABEL, R. D.; TAYLOR, J. F. Candidate lethal haplotypes and causal mutations in Angus cattle. **BMC genomics**, v. 18, n. 1, p. 799, 2017.

HOWLETT, N. G., TANIGUCHI, T., OLSON, S., COX, B., WAISFISZ, Q., DE DIE-SMULDERS, C. & IKEDA, H. Biallelic inactivation of BRCA2 in Fanconi anemia. **Science**, v. 297, n. 5581, p. 606-609, 2002.

HOZÉ, C., ESCOUFLAIRE, C., MESBAH-UDDIN, M., BARBAT, A., BOUSSAHA, M., DELOCHE, M.C., BOICHARD, D., FRITZ, S. AND CAPITAN, A. A splice site mutation in CENPU is associated with recessive embryonic lethality in Holstein cattle. **Journal of dairy science**, v. 103, n.1, p.607-612, 2020.

IQBAL, J., RAGONE, A., LUBINSKI, J., LYNCH, H. T., MOLLER, P., GHADIRIAN, P., & SENTER, L. The incidence of pancreatic cancer in BRCA1 and BRCA2 mutation carriers. **British journal of cancer**, v. 107, n. 12, p. 2005-2009, 2012.

JENKO, J., MCCLURE, M.C., MATTHEWS, D., MCCLURE, J., JOHNSSON, M., GORJANC, G. AND HICKEY, J.M. Analysis of a large dataset reveals haplotypes carrying putatively recessive lethal and semi-lethal alleles with pleiotropic effects on economically important traits in beef cattle. **Genetics Selection Evolution**, v.51, n.1, p.9, 2019.

JERONIMO, S. M., HOLST, A. K., JAMIESON, S. E., FRANCIS, R., MARTINS, D. R., BEZERRA, F. L. Genes at human chromosome 5q31. 1 regulate delayed-type hypersensitivity responses associated with Leishmania chagasi infection. **Genes & Immunity**, v. 8, n. 7, p. 539-551, 2007.

KUNZ, B.A.; STRAFFON, A.F.L; VONARX, E.J. DNA damage-induced mutation: tolerance via translesion synthesis. **Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis**, v. 451, n. 1-2, p. 169-185, 2000.

LEZOUALC'H, F. & ROBERT, S.J. The serotonin 5-HT4 receptor and the amyloid precursor protein processing. **Experimental gerontology**, v. 38, n. 1-2, p. 159-166, 2003.

MAGNABOSCO CD, CORDEIRO CMT, TROVO JD, MARIANTE AD, LÔBO RB, JOSAHKIAN LA. Catálogo de linhagens do germoplasma zebuíno: raça Nelore. **Brasília: Embrapa Recursos Genéticos e Biotecnologia**; 1997.51

MARGOLSKEE, R. F. The biochemistry and molecular biology of taste transduction. **Current opinion in neurobiology**, v. 3, n. 4, p. 526-531, 1993.

MARTIKAINEN, K.; KOIVULA, M.; UIMARI, P. Identification of runs of homozygosity affecting female fertility and milk production traits in Finnish Ayrshire cattle. **Scientific Reports**, v. 10, n. 1, p. 1-9, 2020.

MILLAN, M. J., MARIN, P., BOCKAERT, J., & LA COUR, C. M. Signaling at G-protein-coupled serotonin receptors: recent advances and future research directions. **Trends in pharmacological sciences**, v. 29, n. 9, p. 454-464, 2008.

NG, P.C.; HENIKOFF, S. SIFT: Predicting amino acid changes that affect protein function. **Nucleic acids research**, v. 31, n. 13, p. 3812-3814, 2003.

OLIVIERI, B. F., MERCADANTE, M. E. Z., CYRILLO, J. N. D. S. G., BRANCO, R. H., BONILHA, S. F. M., DE ALBUQUERQUE, L. G. & BALDI, F. Genomic regions associated with feed efficiency indicator traits in an experimental Nellore cattle population. **PloS one**, v. 11, n. 10, 2016.

PAUL, S. & LAL, G. The molecular mechanism of natural killer cells function and its importance in cancer immunotherapy. **Frontiers in immunology**, v. 8, p. 1124, 2017.

PERIPOLLI, E., METZGER, J., DE LEMOS, M. V. A., STAFUZZA, N. B., KLUSKA, S., OLIVIERI, B. F. & LÔBO, R. B. Autozygosity islands and ROH patterns in Nellore lineages: Evidence of selection for functionally important traits. **BMC genomics**, v. 19, n. 1, p. 680, 2018.

R CORE TEAM. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2019. URL http://www.R-project.org/.

REVA, B.; ANTIPIN, Y.; SANDER, C. Predicting the functional impact of protein mutations: application to cancer genomics. **Nucleic acids research**, v. 39, n. 17, p. e118-e118, 2011.

RIHET, P., TRAOR, EÌ, Y., ABEL, L., AUCAN, C., TRAOREÌ-LEROUX, T., AND FUMOUX, F. Malaria in humans: Plasmodium falciparum blood infection levels are linked to chromosome 5q31-q33. **The American Journal of Human Genetics**, v. 63, n. 2, p. 498-505, 1998.

ROSEN, B.D., D.M. BICKHART, R.D. SCHNABEL, S. KOREN, C.G. ELSIK, E. TSENG, T.N. ROWAN, P.M. VANRADEN, D.J. NULL, W.Y. LOW, A. ZIMIN, C. COULDREY, R. HALL, W. LI, A. RHIE, J. GHURYE, S.D. MCKAY, F. THIBAUD-NISSEN, J. HOFFMAN, W.M. SNELLING, W. WARREN, T.G. MCDANELD, J.A. HAMMOND, J.C. SCHWARTZ, W. NANDOLO, S.G. SCHROEDER, A.M. PHILLIPPY, J.B. COLE, C.P. VAN TASSELL, G. LIU, T.P.L. SMITH, AND J.F. MEDRANO. De novo assembly of the cattle reference genome with single-molecule sequencing. **GigaScience**, v. 9, n. 3, p. giaa021, 2020.

SAHANA, G., NIELSEN, U. S., AAMAND, G. P., LUND, M. S., & GULDBRANDTSEN, B. Novel harmful recessive haplotypes identified for fertility traits in Nordic Holstein cattle. **PLoS One**, v. 8, n. 12, 2013.

SANTOS SILVA, D. B., FONSECA, L. F. S., PINHEIRO, D. G., MUNIZ, M. M. M., MAGALHÃES, A. F. B., BALDI, F. & DE ALBUQUERQUE, L. G. Prediction of hub genes associated with intramuscular fat content in Nelore cattle. **BMC genomics**, v. 20, n. 1, p. 520, 2019.

SHANKS, R. D., DOMBROWSKI, D. B., HARPESTAD, G. W., & ROBINSON, J. L. Inheritance of UMP synthase in dairy cattle. **Journal of Heredity**, v. 75, n. 5, p. 337-340, 1984.

SHANKS, R D.; ROBINSON, J L. Embryonic mortality attributed to inherited deficiency of uridine monophosphate synthase. **Journal of dairy science**, v. 72, n. 11, p. 3035-3039, 1989.

SHANNON P, MARKIEL A, OWEN OZIER, BALIGA, NS, WANG JT, RAMAGE D, AMIN N, SCHWIKOWSKI B, IDEKER, T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. **Genome research**, v. 13, n. 11, p. 2498-2504, 2003.

SOJKA, D.K.; TIAN, Z.; YOKOYAMA, W.M. Tissue-resident natural killer cells and their potential diversity. In: **Seminars in immunology**. Academic Press, 2014. p. 127-131.

SONSTEGARD, T. S.; COLE, J. B.; VANRADEN, P. M.; VAN TASSELL, C. P.; NULL, D. J.; SCHROEDER, S. G.; BICKHART, D.; MCCLURE, M. C. Identification of a nonsense mutation in CWC15 associated with decreased reproductive efficiency in Jersey cattle. **PLoS One**, v. 8, n. 1, p. e54872, 2013.

VANRADEN, P. M.; OLSON, K. M.; NULL, D. J.; HUTCHISON, J. L. Harmful recessive effects on fertility detected by absence of homozygous haplotypes. **Journal of dairy science**, v. 94, n. 12, p. 6153-6161, 2011a.

VANRADEN, P. M., NULL, D. J., OLSON, K. M., & HUTCHISON, J. L. Reporting of haplotypes with recessive effects on fertility. **Interbull Bulletin**, n. 44, 2011b.

VANRADEN, P. 2011. findhap.f90 v3. Accessed Feb. 14, 2020. http://aipl.arsusda.gov/software/findhap

VANRADEN, P. 2015. conflict.f90. Accessed Feb. 14, 2020. http://aipl.arsusda.gov/software/genosim

VON NICOLAI, C., EHLÉN, Å., MARTINEZ, J. S., & CARREIRA, A. Dissecting the Recombination Mediator Activity of BRCA2 Using Biochemical Methods. In: **Methods in enzymology**. Academic Press, 2018. p. 479-511.

WOOSTER, R., BIGNELL, G., LANCASTER, J., SWIFT, S., SEAL, S., MANGION, J., & BARFOOT, R. Identification of the breast cancer susceptibility gene BRCA2. **Nature**, v. 378, n. 6559, p. 789-792, 1995.

WU, X., MESBAH-UDDIN, M., GULDBRANDTSEN, B., LUND, M. S., & SAHANA, G. Haplotypes responsible for early embryonic lethality detected in Nordic Holsteins. **Journal of dairy science**, v. 102, n. 12, p. 11116-11123, 2019.

ZHANG, Z., SCHWARTZ, S., WAGNER, L., & MILLER, W. A greedy algorithm for aligning DNA sequences. **Journal of Computational biology**, v. 7, n. 1-2, p. 203-214, 2000.

ZAVAREZ, L. B.; UTSUNOMIYA, Y. T.; CARMO, A. S.; NEVES, H. H. R.; CARVALHEIRO, R.; FERENCAKOVIC, M.; PÉREZ O'BRIEN, A. M.; CURIK, I.; COLE, J. B.; VAN TASSELL, C. P.; DA SILVA, M. V. G. B.; SONSTEGARD, T. S.; SÖLKNER, J.; GARCIA, J. F. Assessment of autozygosity in Nellore cows (Bos indicus) through high-density SNP genotypes. **Frontiers in Genetics**, v. 6, p. 1–8, 2015.

2.6. SUPPLEMENTARY INFORMATION

Table S1. Deleterious mutations with high impact predicted using SIFT that provides a score and a qualitative prediction. If the score is less than 0.05 it is likely to be deleterious

Variation	Location	Allele	Consequence	Impact	Symbol	Biotype	SIFT
rs433337286	1:56,882,970	C	start_lost	HIGH	*C1H3orf52*	protein_coding	deleterious_low_confidence(0)
rs440037384	1:138,550,395	A	start_lost	HIGH	*NUDT16*	protein_coding	deleterious_low_confidence(0.01)
rs460147042	1:138,550,397	C	start_lost	HIGH	*NUDT16*	protein_coding	deleterious_low_confidence(0.01)
rs462457881	12:3,474,092	C	start_lost	HIGH	-	protein_coding	deleterious_low_confidence(0)
rs471372165	13:79,384,616	G	start_lost	HIGH	*NFATC2*	protein_coding	deleterious_low_confidence(0)
rs441853148	15:66,862,809	G	start_lost	HIGH	*PRR5L*	protein_coding	deleterious_low_confidence(0)
rs464412220	15:67,005,344	T	start_lost	HIGH	*RAG1*	protein_coding	deleterious_low_confidence(0)
rs434862466	15:67,022,335	A	start_lost	HIGH	*RAG2*	protein_coding	deleterious_low_confidence(0)
rs454549578	15:67,022,337	A	start_lost	HIGH	*RAG2*	protein_coding	deleterious_low_confidence(0)
rs450202654	17:68,513,552	C	start_lost	HIGH	*EMID1*	protein_coding	deleterious_low_confidence(0.03)
rs135776479	17:68,748,094	C	start_lost	HIGH	*NIPSNAP1*	protein_coding	deleterious_low_confidence(0)
rs440383635	2:24,945,645	C	start_lost	HIGH	*CYBRD1*	protein_coding	deleterious_low_confidence(0)
rs450160527	2:25,556,575	C	start_lost	HIGH	*GORASP2*	protein_coding	deleterious(0)
rs451629699	2:25,641,619	T	start_lost	HIGH	*GAD1*	protein_coding	deleterious_low_confidence(0)
rs459159210	2:25,731,208	G	start_lost	HIGH	*SP5*	protein_coding	deleterious_low_confidence(0)
rs471178728	2:25,731,210	C	start_lost	HIGH	*SP5*	protein_coding	deleterious_low_confidence(0)

rs436813354	2:124,325,550	A	start_lost	HIGH	*PTPRU*	protein_coding	deleterious_low_confidence(0)
rs456957523	2:124,325,551	T	start_lost	HIGH	*PTPRU*	protein_coding	deleterious_low_confidence(0)
rs473889790	2:124,325,552	G	start_lost	HIGH	*PTPRU*	protein_coding	deleterious_low_confidence(0)
rs476822808	2:124,332,600	A	start_lost	HIGH	*MECR*	protein_coding	deleterious_low_confidence(0)
rs465468213	2:124,375,182	C	start_lost	HIGH	*SRSF4*	protein_coding	deleterious_low_confidence(0)
rs444833644	2:124,739,515	C	start_lost	HIGH	*YTHDF2*	protein_coding	deleterious_low_confidence(0)
rs464810604	2:125,015,868	T	start_lost	HIGH	-	protein_coding	deleterious(0.03)
rs436757659	2:125,015,869	G	start_lost	HIGH	-	protein_coding	deleterious(0)
rs440702315	2:125,147,679	G	start_lost	HIGH	*ATP5IF1*	protein_coding	deleterious(0.01)
rs437652889	2:125,150,827	A	start_lost	HIGH	*DNAJC8*	protein_coding	deleterious_low_confidence(0)
rs482934068	2:126,212,986	C	start_lost	HIGH	*TENT5B*	protein_coding	deleterious_low_confidence(0)
rs469825802	2:126,266,914	A	start_lost	HIGH	*KDF1*	protein_coding	deleterious_low_confidence(0)
rs444512736	2:126,286,601	C	start_lost	HIGH	*NUDC*	protein_coding	deleterious_low_confidence(0)
rs468175013	2:126,321,911	C	start_lost	HIGH	*GPN2*	protein_coding	deleterious_low_confidence(0)
rs433769150	2:126,321,912	G	start_lost	HIGH	*GPN2*	protein_coding	deleterious_low_confidence(0)
rs449020982	2:126,930,420	G	start_lost	HIGH	*ZNF593*	protein_coding	deleterious(0)
rs439539939	2:127,619,281	C	start_lost	HIGH	-	protein_coding	deleterious(0)
rs450423315	21:22,190,574	G	start_lost	HIGH	*IQGAP1*	protein_coding	deleterious(0.01)
rs801293745	21:22,190,575	G	start_lost	HIGH	*IQGAP1*	protein_coding	deleterious(0)
rs465233226	23:15,384,043	G	start_lost	HIGH	*FOXP4*	protein_coding	deleterious_low_confidence(0)

rs444874148	23:15,506,056	G	start_lost	HIGH	*TFEB*	protein_coding	deleterious(0.01)
rs440606560	23:15,911,773	T	start_lost	HIGH	*GUCA1B*	protein_coding	deleterious(0)
rs480270364	26:14,825,421	G	start_lost	HIGH	*CEP55*	protein_coding	deleterious(0)
rs469713365	26:14,913,074	A	start_lost	HIGH	*PDE6C*	protein_coding	deleterious_low_confidence(0)
rs455672985	3:51,533,775	T	start_lost	HIGH	*TGFBR3*	protein_coding	deleterious_low_confidence(0)
rs443018086	3:52,006,451	G	start_lost	HIGH	-	protein_coding	deleterious_low_confidence(0)
rs432685642	3:118,923,928	C	start_lost	HIGH	*NDUFA10*	protein_coding	deleterious_low_confidence(0)
rs436229475	3:119,044,500	C	start_lost	HIGH	-	protein_coding	deleterious(0)
rs450032510	3:119,577,931	G	start_lost	HIGH	*AKAP17A*	protein_coding	deleterious(0)
rs478429685	7:19,175,928	C	start_lost	HIGH	*KDM4B*	protein_coding	deleterious_low_confidence(0)
rs481818602	7:19,470,141	G	start_lost	HIGH	*MYDGF*	protein_coding	deleterious_low_confidence(0)
rs449436638	7:19,486,477	G	start_lost	HIGH	*TNFAIP8L1*	protein_coding	deleterious(0)
rs461514633	7:19,486,478	T	start_lost	HIGH	*TNFAIP8L1*	protein_coding	deleterious(0)
rs447078341	7:19,596,357	A	start_lost	HIGH	*LRG1*	protein_coding	deleterious_low_confidence(0)
rs448683482	7:19,615,384	T	start_lost	HIGH	*PLIN4*	protein_coding	deleterious(0)
rs457689039	7:19,645,378	G	start_lost	HIGH	*HDGFL2*	protein_coding	deleterious(0.01)
rs448890978	7:19,773,296	A	start_lost	HIGH	*FSD1*	protein_coding	deleterious(0.01)
rs449131679	7:19,790,931	T	start_lost	HIGH	*SHD*	protein_coding	deleterious_low_confidence(0)
rs453881827	7:19,819,609	C	start_lost	HIGH	*EBI3*	protein_coding	deleterious_low_confidence(0)
rs472199688	7:19,819,610	A	start_lost	HIGH	*EBI3*	protein_coding	deleterious_low_confidence(0)

rs475696745	7:19,966,093	C	start_lost	HIGH	*ZBTB7A*	protein_coding	deleterious(0)
rs479116586	7:20,027,845	G	start_lost, splice_region_variant	HIGH	*EEF2*	protein_coding	deleterious(0)
rs440886483	7:20,069,101	G	start_lost	HIGH	*NMRK2*	protein_coding	deleterious_low_confidence(0)
rs481011291	7:20,358,237	A	start_lost	HIGH	*HMG20B*	protein_coding	deleterious_low_confidence(0)
rs448146438	7:20,358,238	C	start_lost	HIGH	*HMG20B*	protein_coding	deleterious_low_confidence(0)
rs470001999	7:20,358,239	G	start_lost	HIGH	*HMG20B*	protein_coding	deleterious_low_confidence(0)
rs440244635	7:20,382,520	C	start_lost	HIGH	*C7H19orf71*	protein_coding	deleterious(0)
rs479200726	7:20,433,112	A	start_lost	HIGH	*SMIM24*	protein_coding	deleterious(0)
rs449348342	7:20,433,113	C	start_lost	HIGH	*SMIM24*	protein_coding	deleterious(0)
rs443692192	7:52,501,562	C	start_lost	HIGH	-	protein_coding	deleterious_low_confidence(0)
rs468561788	7:52,714,253	A	start_lost	HIGH	-	protein_coding	deleterious_low_confidence(0.04)
rs433994295	7:52,714,254	T	start_lost	HIGH	-	protein_coding	deleterious_low_confidence(0)
rs440618000	7:52,730,046	T	start_lost	HIGH	*HDAC3*	protein_coding	deleterious_low_confidence(0)
rs435923040	7:52,744,831	C	start_lost	HIGH	*FCHSD1*	protein_coding	deleterious(0)
rs432901049	8:59,833,776	T	start_lost	HIGH	*CCDC107*	protein_coding	deleterious_low_confidence(0)
rs450375337	8:59,895,109	A	start_lost	HIGH	*TLN1*	protein_coding	deleterious_low_confidence(0)
rs383060400	8:60,031,706	T	start_lost	HIGH	-	protein_coding	deleterious_low_confidence(0.03)
rs482024336	8:60,072,941	T	start_lost	HIGH	-	protein_coding	deleterious_low_confidence(0.03)
rs472858978	8:60,152,259	C	start_lost	HIGH	-	protein_coding	deleterious(0)
rs210250623	8:60,181,901	G	start_lost	HIGH	-	protein_coding	deleterious(0)

rs456832685	8:70,495,253	T	start_lost	HIGH	*LOXL2*	protein_coding	deleterious_low_confidence(0)
rs457135410	8:70,576,903	C	start_lost	HIGH	*ENTPD4*	protein_coding	deleterious(0)
rs451460801	8:70,670,618	A	start_lost	HIGH	*SLC25A37*	protein_coding	deleterious_low_confidence(0.03)
rs456110553	8:70,791,518	C	start_lost	HIGH	*NKX3-1*	protein_coding	deleterious_low_confidence(0)
rs441946501	8:70,791,519	C	start_lost	HIGH	*NKX3-1*	protein_coding	deleterious_low_confidence(0)

Table S2. Tolerant mutations with high impact predicted using SIFT that provides a score and a qualitative prediction. If the score is higher than 0.05 it is likely to be tolerant

Variation	Location	Allele	Consequence	Impact	Symbol	Biotype	SIFT
rs453127518	1:138,091,994	C	start_lost	HIGH	-	protein_coding	tolerated_low_confidence(0.66)
rs443435063	12:28,655,955	G	start_lost	HIGH	*BRCA2*	protein_coding	tolerated(0.49)
rs449573204	13:79,533,513	C	start_lost, splice_region_variant	HIGH	*ATP9A*	protein_coding	tolerated_low_confidence(0.16)
rs438305750	17:68,513,551	G	start_lost	HIGH	*EMID1*	protein_coding	tolerated_low_confidence(0.12)
rs468758918	17:68,513,553	A	start_lost	HIGH	*EMID1*	protein_coding	tolerated_low_confidence(0.19)
rs480831588	2:126,928,736	C	start_lost	HIGH	-	protein_coding	tolerated_low_confidence(0.8)
rs451973172	2:127,508,675	A	start_lost	HIGH	*LDLRAP1*	protein_coding	tolerated(0.07)
rs209684412	21:22,241,342	C	start_lost	HIGH	*ZNF774*	protein_coding	tolerated_low_confidence(0.07)
rs473138603	21:22,361,952	A	start_lost	HIGH	*SEC11A*	protein_coding	tolerated_low_confidence(0.52)
rs448683482	7:19,615,384	G	start_lost	HIGH	*PLIN4*	protein_coding	tolerated(0.07)
rs460268702	7:20,313,628	G	start_lost	HIGH	*CACTIN*	protein_coding	tolerated_low_confidence(0.05)
rs432526231	7:20,346,464	C	start_lost	HIGH	-	protein_coding	tolerated_low_confidence(0.07)
rs516152548	7:20,346,465	C	start_lost	HIGH	-	protein_coding	tolerated_low_confidence(1)
rs459151927	7:20,408,373	A	start_lost	HIGH	-	protein_coding	tolerated(0.05)
rs449216624	7:20,645,303	G	start_lost	HIGH	*NCLN*	protein_coding	tolerated(0.09)
rs450405781	7:53,021,318	G	start_lost	HIGH	*PCDH12*	protein_coding	tolerated_low_confidence(0.33)
rs434402545	8:59,387,357	C	start_lost	HIGH	*UNC13B*	protein_coding	tolerated_low_confidence(0.11)

Chapter 3 – Identification of recombination events in Nellore males

ABSTRACT - The meiotic exchange of DNA between homologous chromosomes is known as recombination. Failures in recombination may lead to potentially deleterious outcome, such as aneuploidy. The objectives of the present study were to calculate recombination rate, to identify hotspot regions, to construct a recombination map for Nellore males and to identify, through functional analyses, candidate genes underlying hotspot regions. Pedigree information comprised of 2,688,124 animals. A total of 4,447 animals were genotyped with a high-density panel (777,962 SNP markers) and 4,041 with a panel containing 74,677 markers. Recombination rates were measured by an indirect method using pedigree. We extracted progeny-sire pairs from the pedigree and both, were genotyped and phased in order to infer about paternal recombination events during meiosis. Hotspot regions were defined as SNP intervals with recombination rate > 2.5 standard deviations above the mean. To construct recombination maps, we estimated recombination rate between consecutive SNPs. We extracted 21,391 crossover events and 659 paternal meiosis with an average number of crossovers per meiosis of 32.4 for Nellore males. There was a total of 520 hotspot regions, especially in chromosomes 1, 6 and 11, with the highest recombination rates. We have found 52 candidate genes underlying hotspot regions and associated GO terms related pathways. Some pathways as Lysine degradation, Pyruvate metabolism, Viral myocarditis, Neuroactive ligand-receptor interaction, Peroxisome, Axon guidance, Basal transcription factors and GO terms related to transcription and translation processes, were found. These results provide useful insights into the genetic mechanism and evolution of recombination within an important livestock breed raised in Brazil.

Keywords: beef cattle, crossover, hotspots, males, pedigree, phasing

3.1. INTRODUCTION

Preserving genetic diversity is important to allow the long-term sustainability of breeding programs in Brazil. For understanding genetic diversity throughout the genome, Weng et al., (2019) reported that is important characterizing patterns of recombination rates. In eukaryotes, meiotic recombination is a fundamental biological process in which parental genetic materials are exchanged during egg or sperm formation (Ma et al., 2015). This process produces new combinations of genetic variants, resulting in offspring with traits that differ from those in their parents and contributing to genetic diversity (BARTON & CHARLESWORTH, 1998; COOP & MYERS, 2007; MA et al., 2015; WANG et al., 2016; SHEN et al., 2018).

Recombination rates vary widely between and within chromosomes, species, sex and possibly age; that is, this variation is under both genetic and environmental controls (OLLIVIER, 1995; BROMAN et al., 1998; KOREN et al., 2002; MYERS et al., 2005). In addition, some studies have suggested that maternal age had also a significant effect on crossover frequency (SHERMAN et al., 1994; KONG et al., 2004). Studies in several mammals including mice, dogs, humans and swine performed by Dietrich et al. 1996, Neff et al., 1999, Kong et al., 2002 and Tortereau et al., 2012, respectively, showed higher recombination rates in females compared to males. Ma et al. (2015) evaluated a large sample of Holstein cattle from the USDA national database and reported for the first time that bulls have more recombination than cows.

In most species, recombination events are not totally placed randomly throughout the chromosomal DNA, but tend to occur in concentrated areas with elevated recombination activity, called hotspots regions (KAUPPI et al., 2004). Usually, hotspots are small regions in which recombination rate is significantly higher

than in surrounding regions (PAIGEN & PETKOV, 2010; WENG et al., 2014). As recombination is under genetic and environmental control, it is possible to manipulate it by performing environmental modifications or via gene editing through the evidence of favorable alleles (JENKO et al., 2015). These manipulations may increase response to selection in traits of economic interest, if used in breeding programs, releasing greater genetic variability to each generation (BATTAGIN et al., 2016). A strategy to explore and quantify the potential effect of shifting recombination hotspots on genetic gain in livestock breeding programs was tested by Gonen et al. (2017). These authors showed, through the manipulation of the locations of these recombination hotspots, which was possible to increase the overall response to selection and decrease loss of genetic variation.

Recombination events can be measured both directly and indirectly. The direct methods measure the recombination frequency by examining proteins related to some specific stages of meiosis, such as Single-Sperm Genome-Wide Genotyping (LU et al., 2012; WANG et al., 2012) or Molecular Assay (PAN et al., 2011). Indirect methods build recombination maps based on the identification of local patterns in Linkage Disequilibrium (MOURESAN et al., 2019) or, as in this study, using genomic information from pedigree data (MA et al., 2015; SHEN et al., 2018).

Most studies about recombination events and their importance are done on humans and mice (JENSEN-SEAMAN et al., 2004; MCVEAN et al., 2004; SMAGULOVA et al., 2011). Studies on other mammalian species, such as cattle (ZHOU et al., 2018; MOURESAN et al., 2019), can provide comparative information to better understand recombination events and the evolution of recombination in mammals. The objectives of the present study were to calculate recombination rate,

identify hotspots regions, to construct a recombination map for Nellore males and to identify, through functional analyses, candidate genes underlying hotspots regions.

3.2. MATERIAL AND METHODS

3.2.1. Pedigree and genotypic data information

Pedigree information recovered from historical records comprised 2,688,124 animals. These records were incomplete due to use of multiple-sire matings; 57.4% and 22.4% of pedigrees included unknown sires and dams, respectively. It is important to notice that unknown dams are those in the base population.

A total of 8,488 Nellore animals obtained from the Alliance Nellore dataset (www.gensys.com.br) were genotyped, including 2,025 females, 1,473 young males, and 949 bulls using the Illumina Bovine HD Bead chip (Illumina, San Diego, CA, USA) that contains 777,962 SNPs (HD) and 716 females, 3,146 young males, and 179 bulls genotyped using the GeneSeek® (Genomic Profiler Indicus HD - GGP75Ki NEOGEN) which contains 74,677 SNPs (75K). Genotype inconsistences between parents and progeny were corrected using conflict.f90 software, which corrects for Mendelian errors and fills missing SNP using parental genotypes (VANRADEN, 2015).

3.2.2. Quality control of genomic data

Genotype quality control was performed using R software version 3.6.1 (R Core Team, 2019) with SNPStats package version 1.38.0. (CLAYTON, 2020). Only SNPs markers with a GenCall score greater than 0.90 and located on the autosomal chromosomes were used. In addition, only SNP with call rate greater than 0.90 located on the autosomal chromosomes and not in highly significant deviation from

Hardy-Weinberg equilibrium (P>10-5) were considered. SNP at the same genomic position, with minor allele frequency (MAF) less than 0.05 and monomorphic were removed. Samples with call rates below 0.90 were also removed. The SNP positions within a chromosome were based on the ARS-UCD1.2 *Bos taurus* genome assembly (ROSEN et al., 2020). After editing the 2 markers datasets, 410,674 SNPs and 4,424 animals remained on HD panel; and 61,603 SNPs and 4,038 animals remained on 75K panel to carry out the further analyses.

3.2.3. Analysis and identification of recombination events

Pedigree method was used for detection of recombination events using a similar approach described in more details previously in dairy cattle breeds (MA et al., 2015; SHEN et al., 2018). We extracted progeny-sire pairs from the pedigree of Nellore cattle, both genotyped. An animal may appear in more than one pair based on the pedigree structure, since some bulls have hundreds of progeny. The genotype of an animal was phased based on the sire genotype, and crossovers locations were identified by comparing one haplotype of the offspring to the two haplotypes of the corresponding parent (sire). The whole process to find crossovers includes two steps: 1) phasing and 2) finding crossovers.

We only used sire for phasing. Phasing was focusing on heterozygote SNPs. For a heterozygote SNP (Aa) in the focal individual, the phasing was to know the parental origin of alleles "A" and "a", based on the sire genotype. For example, if sire was "AA", then we know "A" was from the sire, and "a" came from dam. If sire was "aa", we can also figure out the parental origin. However, if the sire was “Aa”, we were unable to phase this SNP. Once finished the phasing, we found crossovers by comparing the phased haplotype to the sire haplotypes. The number of meiosis is the

number of progeny-sire pairs and the number of crossovers is the total number of crossovers between two informative SNPs (MA et al., 2015).

For constructing recombination maps, recombination rate between consecutive SNPs was estimated. Each crossover was evenly divided to all SNP intervals between the starting and ending informative SNPs (MA et al., 2015). We went through whole dataset in order to calculate the number of crossovers between two SNPs. For quality control purposes, similar to Shen et al. (2018), we removed pairs that had more than 60 crossover events genome-wide, and also removed all crossovers SNP adjacent to each other, remaining on output file only SNPs with more than ten distance positions.

Following previous studies in cattle, the recombination rate was calculated as the number of crossover divided by the number of meiosis. (SANDOR et al., 2012; MA et al., 2015; SHEN et al., 2018), and SNP intervals with recombination rates of 2.5 standard deviations above the average were considered hotspot regions. We herein used the term “hotspot region” instead of “hotspot” because our SNP intervals were much larger than typical human or mouse recombination hotspots (a few kb or smaller) (SHEN et al., 2018).

3.2.4. Functional and enrichment analyses

To identify the genes in the hotspots regions the BioMart tool, from ENSEMBL software (https://www.ensembl.org/info/docs/tools/vep/index.html), was used, considering the ARS-UCD1.2 *Bos taurus* genome assembly (ROSEN et al., 2020). The Cytoscape plug-in version 3.8.0 of ClueGO v2.5.7 software and CluePedia v1.5.7 software (SHANNON et al., 2003) were used to perform combined functional

analysis for annotated genes. GO terms with P_value less 0.05 were deemed as significant. Positional candidate genes related to significant GO groups and terms were considered as functional candidates. Genes during the development of all stages of life were described and marked as candidate genes.

3.3. RESULTS AND DISCUSSION

3.3.1. Analysis and identification of recombination events

The principle of the method used here to detect recombination, is to track the transmission of alleles from one generation to the next. Initially, a total of 2,381 progeny-sire pairs genotyped by 75k and HD panels were extracted and inferred over 1 million crossover events and, after quality control, 21,391 crossover events and 659 paternal meiosis remained. The average number of crossovers per meiosis was 32.4 for Nellore males. This result is higher than previous studies in dairy cattle (MA et al., 2015; SHEN et al., 2018). These authors compared recombination rates between males and females, using a 50k panel and reported recombination rates for males ranging from 22.2 to 25.5 for Holstein, Jersey, Brown Swiss and Ayrshire breeds. Differences could be due to various factors such as the amount and structure of the data, breeds and also to the genotyping panels used.

Although in this study it was not possible to compare male and female recombination maps, it was reported by Ma et al. (2015) for the first time in cattle, a male recombination map 10% longer than the female map, in contrast to what was observed in many mammalian species (Dietrich et al. 1996, Neff et al., 1999, Kong et al., 2002 and Tortereau et al., 2012), that females have more recombination than males. Some theories of recombination rate evolution showed that the more intense

artificial selection directed to the males could be a reason, if the selection has a positive effect direct or indirect on the recombination (BLIRT & BELL, 1987; BUTTIN, 2005).

From the identification of the number of crossovers between two informative SNPs, the recombination rate between SNPs was estimated and the recombination map was constructed (Figure 1). The average recombination rate was 8.32e-05 and the highest recombination rate found across the whole genome was 0.0015 Morgan for 1 Mb distance. For a HD panel, as in this study, the distance between SNPs is smaller than previous studies, which used a lower density panel (50K) and showed an average value of 0.01 Morgan for 1Mb distance (WENG et al., 2014; MA et al., 2015; SHEN et al., 2018).

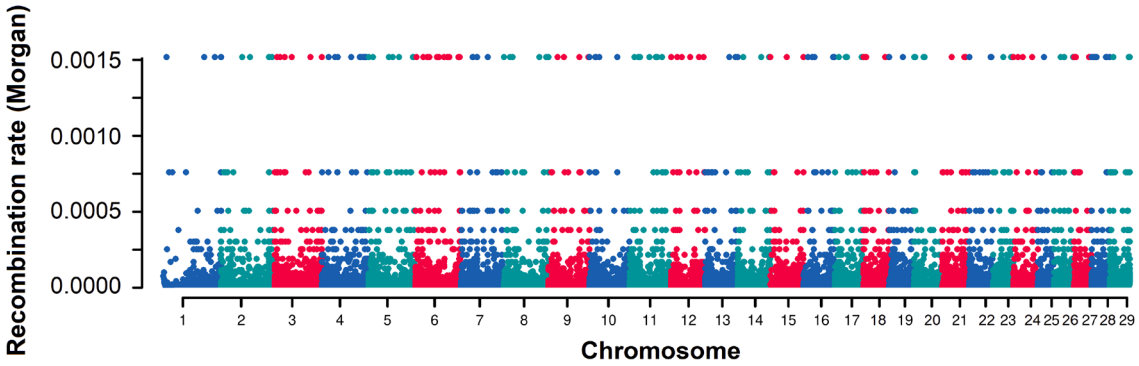

Figure 1 – Recombination map for Nellore males

The relative physical position for each SNP interval on a chromosome was calculated by standardizing the original physical position by the chromosome length: a value of zero corresponds to the beginning of a chromosome and a value of one corresponds to the end (MA et al 2015, Figure 2). Following these authors, a smooth-spline model was fitted across all 29 autosomes. Cattle autosomes are acrocentric, centromeres are located at the beginning and telomeres at the end of chromosomes

(BAND et al., 2000). In this study, males had a higher recombination rate at the beginning (centromeres in cattle) and in the middle of chromosomes and considerably higher recombination rate near the end of chromosomes (telomere). In fact, as cattle centromeres are located almost at the chromosome extremity, we observed a mixed effect from both centromere and telomere as already described by MA et al (2015). A higher recombination rate for males in the end of chromosome is consistent with observations in previous studies in mice (LYNN et al., 2005), human (KONG et al., 2010) and cattle (MA et al., 2015; SHEN et al., 2018).

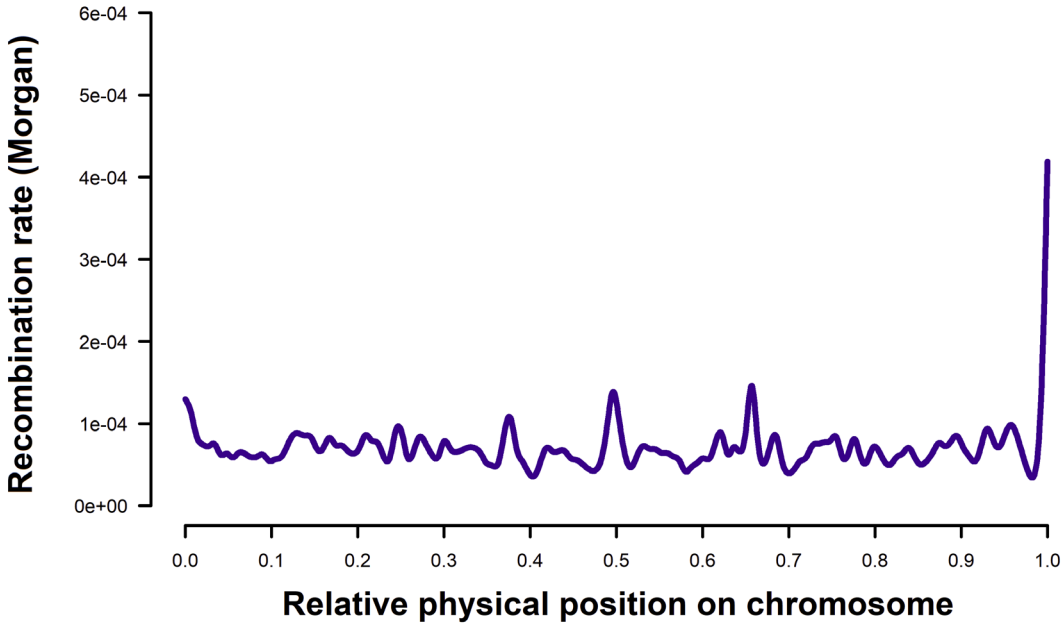

Figure 2 - Distribution of male recombination rates along a chromosome in Nellore cattle

Recombination events do not occur complete randomly throughout the genomic DNA, instead they are more likely to take place in some regions of the genome than in others (KAUPPI et al., 2004). So, the term hotspot has been used to indicate a locus or genomic region suffering a greater than average frequency of

meiotic recombination (LICHTEN & GOLDMAN, 1995). The pedigree approach has also been used to detect hotspots location in different cattle breeds (SHEN et al., 2018). By defining hotspots as SNP intervals with recombination rate >2.5 standard deviations greater than the mean (MA et al., 2015; SHEN et al., 2018), we identified 520 hotspots regions for males (Table S1) across all autosomal chromosomes distributed on 115 progeny-sire pair, and from those, each had from 2 to 6 hotspots regions. Some chromosomes had a greater number of hotspots (Figure 3), especially chromosomes 1, 6 and 11, possibly due to longer length. Shen et al. 2018 authors found variability in the number of hotspots regions among different breeds (Jersey, Ayrshire and Brown Swiss). Moreover, males displayed a total of 233 hotspots shared by all breeds.

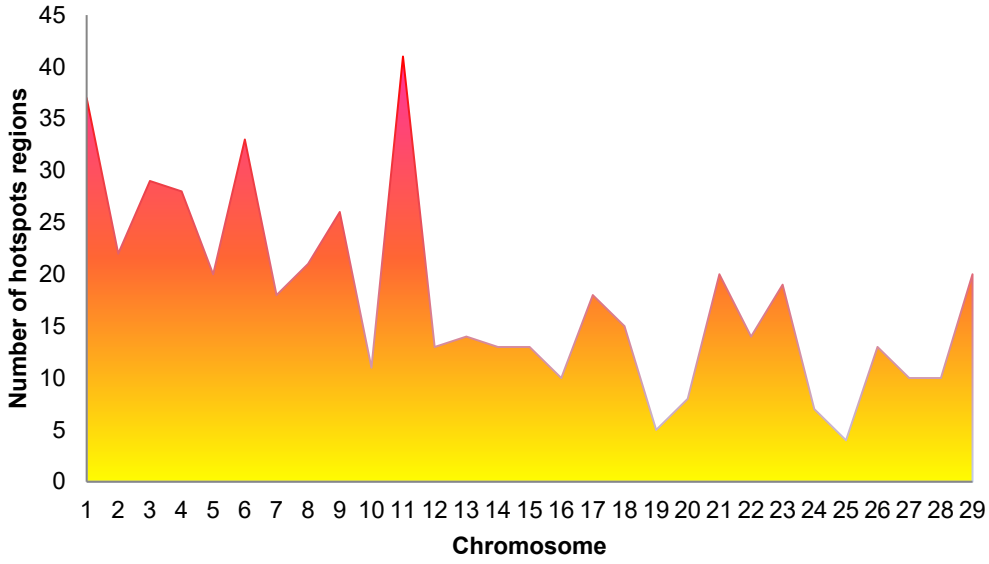

Figure 3 – Distribution of number of hotspot regions with recombination rate > 2.5 standard deviations above the mean per chromosome

3.3.2. Functional and enrichment analyses

There were found 52 candidate genes underlying in 520 hotspots regions, which were submitted to enrichment analysis (Tables S1 and S2). Metabolic

Pathways and Gene Ontology terms (GO) were found using Cytoscape plug-in of ClueGO software (Figures 4). Pathways as Lysine degradation (KEGG: 00310), Pyruvate metabolism (KEGG: 00620), Viral myocarditis (KEGG: 05416), Neuroactive ligand-receptor interaction (KEGG:04080), Peroxisome (KEGG: 04146), Axon guidance (KEGG: 04360), Hippo signaling pathway (KEGG: 04392), Glucagon signaling pathway (KEGG: 04922), Chemokine signaling pathway (KEGG: 04062), Basal transcription factors (KEGG: 03022), Thyroid hormone signaling pathway (KEGG: 04919), PPAR signaling pathway (KEGG: 03320), Rap1 signaling pathway (KEGG: 04015) and MAPK signaling pathway (KEGG: 04010), were found (Figure 4).

An important pathway was PPAR signaling pathway (Kegg pathway database - map03320) (Figure 4). The PPAR metabolism plays an important role in the cellular lipid circulation and oxidation and was reported in previous studies of transcriptomic with productive traits in Nellore cattle, as fatty acid and lipid content in *Longissimus thoracis* muscle tissue (BERTON et al., 2016; SANTOS SILVA et al., 2019).

Two candidate genes (*CCN2* and *FAT4*) (Figure 4) act in Hippo signaling pathway, an evolutionarily conserved signaling pathway that controls organ size in many species, from fungi to arthropods and mammals (ZHAO et al., 2010). Mutations in some genes present in this pathway lead to tissue overgrowth, also called hippopotamus-like phenotype (SAUCEDO & EDGAR, 2007). Some authors also related this pathway with development of human cancers (DONG et al., 2007; STEINHARDT et al., 2008; XU et al., 2009; ZHOU et al., 2009). Hippo signaling pathway is strongly related to Wtn pathway, also found in the present study, acting in tissue development as well. Unrolled activation of Wtn pathway is associated mainly with digestive tract cancer in humans (BARKER & CLEVERS, 2006).

The cellular senescence is an irreversible process to control the cell proliferation and can be initiated by telomere shortening, oncogene activation, irradiation, DNA damage and oxidative stress. This process is strictly related to aging. Senescence stops the proliferation of premalignant cells and thereby lowers the risk for cancer initiation (LECOT et al., 2016). According to Kuroda et al. (2015), fetal bovine colon epithelial cells immortalized in vitro i.e., with senescence inactivated, when infected with pathogenic bacteria, the expression of genes related to cessation of cell division is not detected, inactivating the cell's ability to convert into an immunological phenotype. These results indicate that recombination in regions determinants to cell senescence can affect their damage response ability.

The gene *CACNA1D* (Figure 4), participates in 32 metabolic pathways process (Figure 4). According results from Fonseca et al. (2020) with Nellore cattle muscle tissue transcriptome, some of them, as oxytocin signaling pathway, type II diabetes mellitus, insulin secretion, carbohydrate digestion and absorption, are strictly related to lipid metabolism. Additionally, the gene *CCND2* (Figure 4) plays a role in JaK-STAT signaling pathway, also related to metabolic syndromes as obesity and insulin resistance. Mutations in these genes are associated to autism spectrum disorders and epilepsy, and megalencephaly-polymicrogyria-polydactyly-hydrocephalus syndrome 3 (MPPH) a kind of megalencephaly (INABA et al., 1992; KAMIDE et al., 2009).

The *SLC16A10* and *LCP2* (Figure 4) genes are involved in thyroid hormone signaling pathway and T cell receptor signaling pathway, respectively. The *SLC16A10* encodes an integral membrane protein that functions as a transporter of thyroid hormone. In Nellore cattle muscle tissue, other solute carrier family members

were associated with high marbling score while Lymphocyte Cytosolic Protein family members were associated with low marbling score (FONSECA et al., 2020). In humans, mutations in *SLC16A10* gene are related to Allan-Herndon-Dudley syndrome, a rare disorder of brain development that causes intellectual disability and movement problems, exclusively in males (SCHWARTZ et al., 2005; MARANDUBA et al., 2005). Also in humans, the *LCP2* mutation is related with Wiskott-Aldrich syndrome with microtrombocitopenia, eczema and recurrent respiratory, gastrointestinal and skin infections due to severe combined immunodeficiency, also exclusively expressed in males (TRASHER et al., 2009).

The mitogen-activated protein kinase (MAPK) (Figure 4) cascade is a highly conserved module that is involved in various cellular functions, including cell proliferation, differentiation, migration and can be active in response to pro-inflammatory stimuli (CHEN et al., 2001). Rap1 is a small GTPase that controls diverse processes, such as cell adhesion, cell-cell junction formation, cell polarity and also regulates MAP kinase (MAPK) activity (Kooistra et al.; 2007). The GTPases gene expression were related to Myofibrillar fragmentation index (MFI) in Nellore cattle muscle (Muniz et al.; 2020).

Another important recombinant gene found in many studies with Nellore cattle is *PRDM9,* which is responsible to encode a zinc finger protein, with histone methyltransferase activity, that catalyzes histone H3 lysine 4 trimethylation during meiotic prophase, determining the location of recombination hotspots and facilitating the hotspots association with the chromosome axis, giving rise to genetic exchange between chromosomes (PAIGEN & PETKOV, 2018). In this research, only a part of the *PRDM9* gene sequence is overlapping a hotspot region. In cattle, according to

Zhou et al. (2014), a specific allele is the causal variant for the *PRDM9* association with cattle recombination. In humans, low levels of *PRDM9* mRNA are associated with high infertility risk. The *PRDM9* expression can be stimulated with curative hormonal therapy becoming an alternative to Cryptorchidism-Dependent Male Infertility treatment (HADZISELIMOVIC et al., 2018).

Information on recombination events contributes to the knowledge of the genetic mechanism, genetic diversity and how the maternal and paternal alleles provide new genetic variants important for the selection and evolution of a population. In addition, from them, it is possible to estimate the gametic variation in the population.

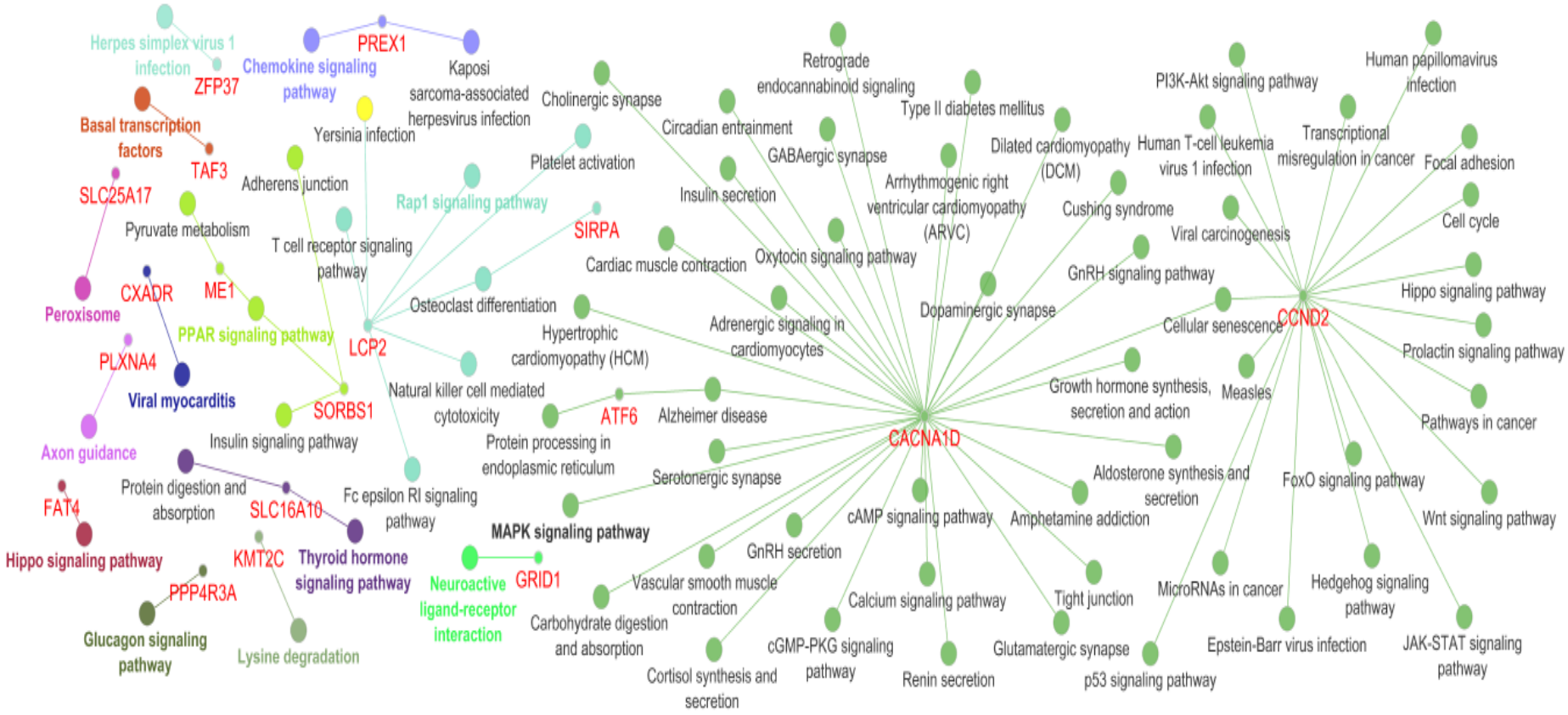

Figure 4 - Functionally grouped terms network for genes: each color represents a significant metabolic pathway group (P_value < 0.05) in which candidate genes were distributed

3.4. CONCLUSIONS

We characterized cattle meiotic recombination landscape in Nellore males that will be useful in genetic studies of different beef cattle breeds. We extracted 21,391 crossover events and 659 paternal meiosis and, the average number of crossovers per meiosis was 32.4 for Nellore males. There were found 520 hotspots regions, especially in chromosomes 1, 6 and 11 possibly due to longer length. From hotspots regions, we found the cattle PRDM9 gene on chromosome 1 in our population, although only a part of the PRDM9 gene sequence is overlapping a hotspot region. A total of 52 candidate genes underlying hotspots regions and associated to GO terms related pathways, were found. Some pathways as Lysine degradation, Pyruvate metabolism, Viral myocarditis, Neuroactive ligand-receptor interaction, Peroxisome, Axon guidance, Basal transcription factors and GO terms related to transcription and translation processes, were found. These results provide useful insights into the genetic mechanism and evolution of recombination within an important livestock breed raised in Brazil.

3.5. REFERENCES

AGUILERA, A.; GAILLARD, H. Transcription and recombination: when RNA meets DNA. **Cold Spring Harbor perspectives in biology**, v. 6, n. 8, p. a016543, 2014.

BAND, Mark R. et al. An ordered comparative map of the cattle and human genomes. **Genome research**, v. 10, n. 9, p. 1359-1368, 2000.

BARKER, N.; CLEVERS, H. Mining the Wnt pathway for cancer therapeutics. **Nature reviews Drug discovery**, v. 5, n. 12, p. 997-1014, 2006.

BARTON, N. H.; CHARLESWORTH, B. Why sex and recombination?. **Science**, v. 281, n. 5385, p. 1986-1990, 1998.

BATTAGIN M., GORJANC G., FAUX A. M., JOHNSTON S. E., & HICKEY J. M. Effect of manipulating recombination rates on response to selection in livestock breeding programs. **Genetics Selection Evolution**, v. 48, n. 1, p. 1-12, 2016.

BLIRT, A.; BELL, G. Mammalian chiasma frequencies as a test of two theories of recombination. **Nature**, v. 326, n. 6115, p. 803-805, 1987.

BROMAN K. W., MURRAY J. C., SHEFFIELD V. C., WHITE R. L. & WEBER J. L. Comprehensive human genetic maps: individual and sex-specific variation in recombination. **American Journal of Human Genetics**, 63:861–9, 1998.

BUTLIN, R. K. Recombination and speciation. **Molecular Ecology**, v. 14, n. 9, p. 2621-2635, 2005.

CHEN Z, GIBSON TB, ROBINSON F, SILVESTRO L, PEARSON G, XU B, WRIGHT A, VANDERBILT C, COBB MH. MAP kinases. **Chemical reviews**, v. 101, n. 8, p. 2449-2476, 2001.

CLAYTON, D. snpStats: SnpMatrix and XSnpMatrix classes and methods. **R package version, 1.38.0**, 2020.

COOP G. & MYERS S. R. Live hot, die young: transmission distortion in recombination hotspots. **PLoS Genetics**, 3:e35, 2007.

DIETRICH W. F., MILLER J., STEEN R., MERCHANT M. A., DAMRON-BOLES D., HUSAIN Z., DREDGE R., DALY M. J., INGALLS K. A. & O'CONNOR T.J. A comprehensive genetic map of the mouse genome. **Nature**, 380:149–52, 1996.

DONG, J., FELDMANN, G., HUANG, J., WU, S., ZHANG, N., COMERFORD, S. A., GAYYED, M. F., ANDERS, R. A., MAITRA, A. AND PAN, D. Elucidation of a universal size-control mechanism in Drosophila and mammals. **Cell**, v. 130, n. 6, p. 1120-1133, 2007.

KAMIDE, Kei et al. Genetic polymorphisms of L-type calcium channel α1C and α1D subunit genes are associated with sensitivity to the antihypertensive effects of L-type dihydropyridine calcium-channel blockers. **Circulation Journal**, v. 73, n. 4, p. 732-740, 2009.

INABA, Toshiya et al. Genomic organization, chromosomal localization, and independent expression of human cyclin D genes. **Genomics**, v. 13, n. 3, p. 565-574, 1992.

GONGADZE, G. M. 5S rRNA and ribosome. **Biochemistry (Moscow)**, v. 76, n. 13, p. 1450-1464, 2011.

HADANY, L.; FELDMAN, M. W. Evolutionary traction: the cost of adaptation and the evolution of sex. **Journal of evolutionary biology**, v. 18, n. 2, p. 309-314, 2005.

HADZISELIMOVIC, F., CATHOMAS, G., VERKAUSKAS, G., DASEVICIUS, D., & STADLER, M. B. PRDM Histone Methyltransferase mRNA Levels Increase in Response to Curative Hormone Treatment for Cryptorchidism-Dependent Male Infertility. **Genes**, v. 9, n. 8, p. 391, 2018.

HUANG W, SHERMAN BT, LEMPICK RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. **Nature protocols**, v. 4, n. 1, p. 44, 2009.

KOOISTRA, MATTHIJS RH; DUBÉ, NADIA; BOS, JOHANNES L. Rap1: a key regulator in cell-cell junction formation. **Journal of cell science**, v. 120, n. 1, p. 17-22, 2007.

JENKO J., GORJANC G., CLEVELAND M. A., VARSHNEY R. K., WHITELAW C. B. & WOOLLIAMS J. A. Potential of promotion of alleles by genome editing to improve quantitative traits in livestock breeding programs. **Genetics Selection Evolution**, 47:55, 2015.

JENSEN-SEAMAN, M. I., FUREY, T. S., PAYSEUR, B. A., LU, Y., ROSKIN, K. M., CHEN, C. F., & JACOB, H. J. Comparative recombination rates in the rat, mouse, and human genomes. **Genome research**, v. 14, n. 4, p. 528-538, 2004.

KHANSEFID, M., MILLEN, C. A., CHEN, Y., PRYCE, J. E., CHAMBERLAIN, A. J., VANDER JAGT, C. J., & GODDARD, M. E. Gene expression analysis of blood, liver, and muscle in cattle divergently selected for high and low residual feed intake. **Journal of animal science**, v. 95, n. 11, p. 4764-4775, 2017.

KIRKPATRICK, DAVID T.; FAN, QINGQING; PETES, THOMAS D. Maximal stimulation of meiotic recombination by a yeast transcription factor requires the

transcription activation domain and a DNA-binding domain. **Genetics**, v. 152, n. 1, p. 101-115, 1999.

KONG A., BARNARD J., GUDBJARTSSON D. F., THORLEIFSSON G., JONSDOTTIR G., SIGURDARDOTTIR S. & LAMB N. E. Recombination rate and reproductive success in humans. **Nature genetics**, v. 36, n. 11, p. 1203-1206, 2004.

KONG A., GUDBJARTSSON D. F., SAINZ J., JONSDOTTIR G. M., GUDJONSSON S. A. & RICHARDSSON B. A high-resolution recombination map of the human genome. **Nature Genetics**, 31:241–7, 2002.

KONG, Augustine et al. Fine-scale recombination rate differences between sexes, populations and individuals. **Nature**, v. 467, n. 7319, p. 1099-1103, 2010.

KOREN A., BEN-AROYA S. & KUPIEC M. Control of meiotic recombination initiation: a role for the environment? **Current Genetics**, 42:129–39, 2002.

KURODA, K., KIYONO, T., ISOGAI, E., MASUDA, M., NARITA, M., OKUNO, K., & FUKUDA, T. Immortalization of fetal bovine colon epithelial cells by expression of human cyclin D1, mutant cyclin dependent kinase 4, and telomerase reverse transcriptase: an in vitro model for bacterial infection. **PLoS One**, v. 10, n. 12, p. e0143473, 2015.

LECOT, PACOME et al. Context-dependent effects of cellular senescence in cancer development. **British journal of cancer**, v. 114, n. 11, p. 1180-1184, 2016.

LICHTEN, Michael; GOLDMAN, Alastair SH. Meiotic recombination hotspots. **Annual review of genetics**, v. 29, n. 1, p. 423-444, 1995.

LIPKIN S. M., MOENS P. B., WANG V., LENZI M. & SHANMUGARAJAH D. Meiotic arrest and aneuploidy in MLH3-deficient mice. **Nature genetics**, v. 31, n. 4, p. 385-390, 2002.

LYNN, Audrey et al. Sex, not genotype, determines recombination levels in mice. **The American Journal of Human Genetics**, v. 77, n. 4, p. 670-675, 2005.

MA L., O'CONNELL J. R., VANRADEN P. M., SHEN B., PADHI A., SUN C. & DA Y. Cattle sex-specific recombination and genetic control from a large pedigree analysis. **PLoS Genetics**, 11:e1005387, 2015.

MARANDUBA CM, FRIESEMA EC, KOK F, KESTER MH, JANSEN J, SERTIÉ AL, PASSOS-BUENO MR, VISSER TJ. Decreased cellular uptake and metabolism in Allan-Herndon-Dudley syndrome (AHDS) due to a novel mutation in the MCT8 thyroid hormone transporter. **Journal of Medical Genetics**, v. 43, n. 5, p. 457-460, 2006.

MCVEAN, G. A., MYERS, S. R., HUNT, S., DELOUKAS, P., BENTLEY, D. R., & DONNELLY, P. The fine-scale structure of recombination rate variation in the human genome. **Science**, v. 304, n. 5670, p. 581-584, 2004.

METZGER, D. E., GASPEROWICZ, M., OTTO, F., CROSS, J. C., GRADWOHL, G., & ZARET, K. S. The transcriptional co-repressor Grg3/Tle3 promotes pancreatic endocrine progenitor delamination and β-cell differentiation. **Development**, v. 139, n. 8, p. 1447-1456, 2012.

MUNIZ, M. M. M., FONSECA, L. F. S., MAGALHÃES, A. F. B., DOS SANTOS SILVA, D. B., CANOVAS, A., LAM, S., & DE ALBUQUERQUE, L. G. Use of

gene expression profile to identify potentially relevant transcripts to myofibrillar fragmentation index trait. **Functional & Integrative Genomics**, p. 1-11, 2020.

MYERS S., BOTTOLO L., FREEMAN C., MCVEAN G. & DONNELLY P. A fine-scale map of recombination rates and hotspots across the human genome. **Science**, 310:321–4, 2005.

NEFF M. W., BROMAN K. W., MELLERSH C. S., RAY K., ACLAND G. M., AGUIRRE G. D., ZIEGLE J. S., OSTRANDER E. A. & RINE J. A second-generation genetic linkage map of the domestic dog, *Canis familiaris*. **Genetics**, 151:803–20, 1999.

KAMAL, N. et al. Aging of the Cells: Insight into Cellular Senescence and Detection Methods. **European Journal of Cell Biology**, p. 151108, 2020.

OLLIVIER L. Genetic differences in recombination frequency in the pig (Susscrofa). **Genome**, 38:1048–51, 1995.

PAIGEN, Kenneth; PETKOV, Petko M. PRDM9 and its role in genetic recombination. **Trends in Genetics**, v. 34, n. 4, p. 291-300, 2018.

ROSEN, B., BICKHART, D., SCHNABEL, R., KOREN, S., ELSIK, C., ZIMIN, A., DREISCHER, C., SCHULTHEISS, S., HALL, R., SCHROEDER, S.G. AND VAN TASSELL, C.P. February. Modernizing the bovine reference genome assembly. In **Proceedings of the 11th world congress on genetics applied to livestock production**, p. 11-16, 2018.

SAUCEDO, Leslie J.; EDGAR, Bruce A. Filling out the Hippo pathway. **Nature reviews Molecular cell biology**, v. 8, n. 8, p. 613-621, 2007.

SCHWARTZ CE, MAY MM, CARPENTER NJ, ROGERS RC, MARTIN J, BIALER MG, WARD J, SANABRIA J, MARSA S, LEWIS JA, ECHEVERRI R, LUBS

HA, VOELLER K, SIMENSEN RJ, STEVENSON RE. Allan-Herndon-Dudley syndrome and the monocarboxylate transporter 8 (MCT8) gene. **The American Journal of Human Genetics**, v. 77, n. 1, p. 41-53, 2005.

SHEN B., JIANG J., SEROUSSI E., LIU G. E. & MA L. Characterization of recombination features and the genetic basis in multiple cattle breeds. **BMC Genomics**, 19:304, 2018.

SHERMAN S. L., PETERSEN M. B., FREEMAN S. B., HERSEY J., PETTAY D., TAFT L. & HASSOLD, T. J. Non-disjunction of chromosome 21 in maternal meiosis I: evidence for a maternal age-dependent mechanism involving reduced recombination. **Human Molecular Genetics**, 3:1529-1535, 1994.

SMAGULOVA, F., GREGORETTI, I. V., BRICK, K., KHIL, P., CAMERINI-OTERO, R. D., & PETUKHOVA, G. V. Genome-wide analysis reveals novel molecular features of mouse recombination hotspots. **Nature**, v. 472, n. 7343, p. 375-378, 2011.

STEINHARDT, A. A., GAYYED, M. F., KLEIN, A. P., DONG, J., MAITRA, A., PAN, D., MONTGOMERY, E. A. AND ANDERS, R. A. Expression of Yes-associated protein in common solid tumors. **Human pathology**, v. 39, n. 11, p. 1582-1589, 2008.

TAO S., WU J., YAO D., CHEN Y., YANG W. & TONG C. Identification of recombination events in outbred species with next-generation sequencing data. **BMC Genomics**, 19: 398, 2018.

TIZIOTO, P. C., COUTINHO, L. L., OLIVEIRA, P. S., CESAR, A. S., DINIZ, W. J., LIMA, A. O., & TULLIO, R. R. Gene expression differences in Longissimus

muscle of Nelore steers genetically divergent for residual feed intake. **Scientific reports**, v. 6, p. 39493, 2016.

TORTEREAU F., SERVIN B., FRANTZ L., MEGENS H. J., MILAN D., ROHRER G., WIEDMANN R., BEEVER J., ARCHIBALD A. L. & SCHOOK L. B. A high density recombination map of the pig reveals a correlation between sex-specific recombination and GC content. **BMC Genomics**, 13:586, 2012.

THRASHER, Adrian J. New insights into the biology of Wiskott-Aldrich syndrome (WAS). **ASH Education Program Book**, v. 2009, n. 1, p. 132-138, 2009.

TUROWSKI, Tomasz W.; TOLLERVEY, David. Cotranscriptional events in eukaryotic ribosome synthesis. **Wiley Interdisciplinary Reviews: RNA**, v. 6, n. 1, p. 129-139, 2015.

VANRADEN, P. 2015. conflict.f90. Accessed Feb. 14, 2020. http://aipl.arsusda.gov/software/genosim

WANG Z., SHEN B., JIANG J., LI J. & MA L. Effect of sex, age and genetics on crossover interference in cattle. **Scientific Reports**, 6: 37698, 2016.

WENG Z., WOLC A., SU H., FERNANDO R. L., DEKKERS J. C., ARANGO J. & GARRICK D. J. Identification of recombination hotspots and quantitative trait loci for recombination rate in layer chickens. **Journal of Animal Science and Biotechnology**, 10:20, 2019.

XU, D., YIN, C., WANG, S., & XIAO, Y. JAK-STAT in lipid metabolism of adipocytes. **Jak-stat**, v. 2, n. 4, p. e27203, 2013.

YOSHIONARI, S., KOIKE, T., YOKOGAWA, T., NISHIKAWA, K., UEDA, T., MIURA, K., & WATANABE, K. Existence of nuclear-encoded 5S-rRNA in bovine mitochondria. **FEBS letters**, v. 338, n. 2, p. 137-142, 1994.

YUSUPOVA, Gulnara; YUSUPOV, Marat. High-resolution structure of the eukaryotic 80S ribosome. **Annual review of biochemistry**, v. 83, p. 467-486, 2014.

ZHOU, D., CONRAD, C., XIA, F., PARK, J. S., PAYER, B., YIN, Y., LAUWERS, G. Y., THASLER, W., LEE, J. T., AVRUCH, J. Mst1 and Mst2 maintain hepatocyte quiescence and suppress hepatocellular carcinoma development through inactivation of the Yap1 oncogene. **Cancer cell**, v. 16, n. 5, p. 425-438, 2009.

ZHOU, Y., SHEN, B., JIANG, J., PADHI, A., PARK, K. E., OSWALT, A., & LIU, G. E. Construction of PRDM9 allele-specific recombination maps in cattle using large-scale pedigree analysis and genome-wide single sperm genomics. **DNA Research**, v. 25, n. 2, p. 183-194, 2018.

3.6. SUPPLEMENTARY INFORMATION

Table S1 - Hotspot regions, candidate genes and recombination rate corresponding to all regions identified on each chromosome

Chr	Hotspots regions	Gene name	Recombination rate, Morgan
1	157585388:157586519, 157683452:157698749, 34145088:34158319, 157797187:157802222, 157797187:157801621, 158126167:158132513, 157802222:157806548, 115346504:115352068, 158468861:158472415, 157585388:157588023, 157801621:157804612, 142405365:142410366, 157683452:157698749, 157585388:157586519, 157797187:157801621, 19466098:19474211, 158119988:158132513, 157683452:157698749, 158126167:158132513, 157683452:157698749, 157797187:157801621, 157797187:157802222, 157683452:157698749, 157585388:157586519, 157683452:157698749, 158126167:158132513, 158126167:158132513, 157830112:157836118, 158119988: 158132513, 68665933:68669127, 158119988:158132513, 157797187:157802222, 158119988:158132513, 157797187:157801621, 157683452:157698749, 25660987:25687581, 158126167:158134933	*CXADR* *MBNL1* *PRDM9**	0,0455
2	121754873:121771831, 52960199:52967645, 15986560:15990269, 121453656:121457537, 121754873:121759767, 121453656:121457537, 11595323:11648427, 121453656:121457537, 121381878:121387409, 121381878:121386429,	*TLE3* *ARMH4* *DNAH6*	0,0258

	121450368:121457537, 6662451:6671897, 121381878:121386429, 30463341:30479449, 128883483:128897943, 73005102:73007928, 121450368:121457537, 121381878:121386429, 121450368:121457537, 121754873:121759767, 121453656:121457537, 121453656:121461611		
3	91145906:91169646, 164057:186018, 119664785:119670899, 26966191:26969648, 13568637:13569008, 164057:186018, 164057:186018, 164057:186018, 164057:186018, 86675644:86681231, 164057:186018, 25093821:25146422, 25906297:25909664, 7612629:7616547, 14179595:14190553, 119664785:119670899, 164057:186018, 164057:186018, 4564463:4566759, 25299753:25301251, 164057:186018, 164057:186018, 164057:186018, 164057:186018, 5866987:5871582, 25503219:25510917, 78319914:78324851, 43343187:43344719, 23549457:23551968	*DYSF* *NEK5* *KLF12* *ARGLU1* *TAF3*	0,0273
4	259459:312770, 38620896:38631379, 76149388:76155958, 95886201:95886905, 109674081:109688536, 74159961:74175801, 74159961:74175801, 95886201:95886905, 86187083:86196399, 33162534:33163202, 74159961:74175801, 106161098:106166511, 40353988:40378150, 33162534:33163202, 114806483:114814577, 259459:312770, 76149388:76155958, 104360705:104361555, 76149388:76155958, 74159961:74175801, 16647766:16653619, 259459:312770,	*PREX1* *SIRPA*	0,0334

	16708640:16715416, 108532241:108536358, 112558092:112560366, 98852463:98862387, 39530043:39562903, 76149388:76155958,		
5	76928770:76955820, 55934:70070, 10584328:10588467, 94640187:94645353, 105783913:105792437, 10789597:10790381, 52297605:52306381, 83085284:83091036, 107997689:108009993, 55934:70070, 60539375:60551021, 112059920:112065881, 50884368:50885416, 70872655:70875199, 45811558:45835410, 12153175:12156021, 55934:70070, 6516409:6535073, 113698370:113705205, 21534623:21539198	*FAT4* *CDH11* *DRG2*	0,0220
6	19455356:19461213, 112276249:112290590, 112372627:112380794, 85606270:85611091, 62166581:62183567, 112366262:112375974, 35323822:35324677, 112273871:112289712, 19455356:19461213, 62146067:62166581, 112372627:112380794, 15495537:15504808, 112172440:112173077, 112372627:112380794, 72695832:72713559, 35323822:35324677, 2096690:2097147, 112273871:112289712, 38093634:38107020: 19455356:19461213, 49491476:49494665: 54524486:54528699, 80573537:80620310, 35323822:35324677, 73016696:73031707, 112169109:112173077, 112276249:112289712, 113875552:113899944, 59466469:59470927, 109203123:109206318, 112372627:112380794, 35323822:35324677, 29193954:29195775	*GID4*	0,0379

7	23306087:23312839, 41337364:41355713, 68559500:68571791, 41337364:41355713, 41337364:41355713, 102350636:102356861, 103840539:103856004, 92653838:92679332, 80446362:80449409, 41337364:41349352, 40427586:40439279, 40427586:40439279, 29467673:29468663, 30885328:30888884, 41337364:41349352, 88452518:88459418, 41337364:41349352, 41337364:41349352	*DRC3* *ASNSD1*	0,0205
8	101765294:101779139, 113096592:113097086, 51572977:51577763, 22254903:22262434, 113096592:113097086, 113228120:113248793, 28226040:28233824, 113096592:113097086, 113096592:113097086, 7719044:7729470, 102295909:102299719, 107497827:107503497, 113222022:113248793, 90234863:90240175, 46262708:46270870, 113096592:113097086, 47870392:47876015, 113222022:113248793, 15766677:15770929, 255535:286793, 113096592:113097086	*ARHGAP15*	0,0250
9	78704840:78711796, 104247340:104264220, 78704840:78711796, 78704840:78708021, 104573598:104584056, 78704840:78711796, 41655841:41659327, 104247340:104273427, 8997683:9019348, 81708299:81718669, 78704840:78711796, 23295206:23299146, 7593418:7601779, 78704840:78708021, 78704840:78711796, 78704840:78711796, 50497047:50517434, 78704840:78708021, 39201239:39206672, 78704840:78711796, 78704840:78708021, 78704840:78711796, 102124709:102126215, 104573598:104584056, 78704840:78708021, 104162871:104172188	*TFCP2L1* *LCP2*	0,0288

10	15381084:15382216, 66988:101366, 16964692:16989105, 61312:66988, 14749251:14755257, 24787530:24795109, 70954763:70984525, 16662791:16670545, 15381084:15382216, 70421117:70424666, 13960912:13965699	*PPP4R3A* *SERPINA6*	0,0137
11	70299253:70300762, 83043794:83045506, 26415232:26423476, 70227491:70230562, 92700381:92702088, 26415232:26423476, 26415232:26423476, 26415232:26423476, 12933413:12936553, 70143837:70146001, 70218600:70219183, 70212253:70213891, 70299253:70300762, 70299253:70300762, 70143837:70146001, 50447269:50457912, 85200355:85204252, 26592724:26600136, 70143837:70146001, 70212253:70213891, 70212253:70213891, 60772699:60832079, 70227491:70230562, 82856965:82864317, 70098903:70099572, 70212253:70213891, 70053919:70057177, 83043794:83045997, 70143837:70146001, 70138936:70146001, 66802506:66804587, 70299253:70300762, 83043794:83045997, 26415232:26423476, 70218600:70219183, 70367825:70369782, 70299253:70300762, 82856965:82864317, 90307530:90316731, 26415232:26423476, 22168408:22171523	*SYNE3*	0,0546
12	27282553:27292887, 82296607:82302471, 5340085:5352883, 64546377:64556327, 48598258:48599688, 113099:167702, 71020218:71041923, 21504853?21512267, 21504853:21512267, 15739554:15742096,	*CACNA1D* *MRPS24* *TAMM41*	0,0144

	84327563:84339094, 46026076:46037738, 15502303:15515789		
13	5151232:5157560, 460596:480468, 460596:480468, 61599112:61602070, 53234735:53253023, 76932120:76935047, 15739903:15770313, 82985637:82989996, 19310717:19312814, 25299196:25300377, 9497368:9505489, 15979372:15988961, 53234735:53253023, 53234735:53253023	*FAM83B*	0,0129
14	79635503:79651320, 82346440:82356992, 168379:171657, 168379:171657, 76480481:76499794, 82346440:82348667, 59618866:59621332, 168379:171657, 7784547:7793973, 25775165:25794361, 82356992:82357727, 13901418:13905019, 18964633:18974308	*CCBE1*	0,0137
15	84968321:84978286, 10995940:11010786, 173832:189216, 173832:189216, 9690073:9722923, 10995940:11010786, 42277063:42291092, 84968321:84978286, 84968321:84971729, 84968321:84978286, 69016162:69047614, 84968321:84978286, 10995940:11010786		0,0129
16	47718710:47721000, 22173768:22178185, 78158279:78160467, 74879871:74880754, 62488851:62501389, 26666858:26670241, 11126904:11128117, 31360463:31364117, 11527818:11528798, 27997792:28008613		0,0121

17	68089794:68098247, 67686:77144, 68089794:68098247, 68089794:68098247, 44205761:44209502, 30310316:30324194, 9921767:9925309, 68220932:68228521, 67686:77144, 68220932:68228521, 32527275:32536552, 68220932:68228521, 68220932:68228521, 68220932:68228521, 27320876:27322800, 53444697:53460483, 68089794:68098247, 67686:77144		0,0190
18	6737514:6744277, 261398:321344, 63178062:63182112, 32826230:32828541, 16888217:16921890, 65312387:65315292, 63036702:63039977, 63685080:63713399, 1360037:1376774, 5623041:5624880, 63178062:63182112, 63178062:63182112, 55206784:55214914, 25142836:25145458, 54702299:54706635	*MYOM1*	0,0190
19	69645:78647, 69645:78647, 14280647:14292397, 34489987:34538767, 69645:78647	*BCAR4* *FRA10AC1* *SORBS1*	0,0068
20	209803:280086, 2145037:2169730, 27113936:27115634, 263956:280086, 3301288:3302789, 263956:280086, 3291156:3302789, 24425516:24427310	*CSMD1*	0,0099
21	58975521:58986906, 497917:523477, 58982450:58986906, 51918826:51921606, 12441354:12444186, 58975521:58986906, 23586456:23662517, 497917:523477, 49799924:49812870, 49799924:49812870, 58975521:58986906, 58975521:58986906,	*GRID1* *BCAN*	0,0182

	22099281:22104557, 497917:523477, 68819945:68826917, 58975521:58986906, 58982450:58986906, 2425242:2448779, 60283958:60289056, 56223597:56237279		
22	220440:231768, 55839788:55843712, 220440:231768, 8976409:8988872, 220440:231768, 14498016:14508669, 41026863:41036177, 35021875:35025130, 26604210:26618658, 53982308:53983106, 220440:231768, 220440:231768, 47361554:47366781, 220440:231768	*COPS9* *ATF6*	0,0167
23	132041:153394, 25933580:26634448, 25933580:26634448, 29060809:29065047, 38639751:38644120, 25933580:26634448, 132041:153394, 50518740:50520053, 103505:148759, 5483877:5503102, 103505:148759, 25931945:26634448, 48788499:48795668, 45901742:45903775, 45901742:45903775, 42490382:42499049, 45901742:45903775, 103505:148759, 45901742:45903775	*FGGY* *PLXNA4* *KMT2C*	0,0212
24	159114:163270, 37294413:37306947, 23519904:23521719, 58525356:58529713, 8085365:8088507, 11833368:11836987, 47622688:47623979	*CCND2* *SLC25A17*	0,0083
25	35423349:35427557, 10519056:10534330, 15674595:15679671, 7842311:7846650	*PACSIN2*	0,0038

26	50551472:50559542, 25809196:25812988, 202644:260240, 25809196:25812988, 14944920:14986751, 14944920:14986751, 25270536:25288590, 16994096:16994955, 10576401:10584983, 202644:260240, 25809196:25812988, 25509679:25549986, 48455887:48475379	*RBM47* *SNX24* *5S_rRNA*	0,0144
27	4661291:4662153, 11742651:11766045, 168177:172066, 2800634:2808855, 10003070:10009199, 1569517:1574999, 2706332:2710325, 43608430:43611981, 38372239:38378986, 14602673:14638858	*ZFP37*	0,0114
28	12552487:12557982, 558599:592200, 45919071:45925555, 39555688:39560270, 7863709:7867773, 40396484:40407364, 5469703:5470800, 36690584:36706337, 37776092:37780777, 45919071:45925555	*AFG1L*	0,0137
29	51054077:51058936, 13709839:13719674 51054077:51058936, 51054077:51058936, 46609259:46615913, 426723:511905, 51054077:51058936, 48209831:48220436, 51054077:51058936, 51058936:51070239, 426723:511905, 9104654:9113947, 51030756:51032127, 6214562:6232619, 9102521:9113947, 4045383:4058837, 51054077:51058936, 426723:511905, 40666106:40671249, 51054077:51058936	*ME1* *SLC16A10*	0,0175

*****Only a part of the *PRDM9* gene sequence is overlapping a hotspot region

Table S2. Official candidate gene symbol and respectively gene name

Gene Symbol	Gene Name
5S_rRNA	5S ribosomal RNA
AFG1L	AFG1 Like ATPase
ARGLU1	arginine and glutamate rich 1
ARHGAP15	Rho GTPase activating protein 15
ARMH4	Armadillo Like Helical Domain Containing 4
ASNSD1	asparagine synthetase domain containing 1
ASXL3	additional sex combs like 3, transcriptional regulator
ATF6	activating transcription factor 6
BCAN	Brevican
BCAR4	breast cancer anti-estrogen resistance 4
CACNA1D	calcium voltage-gated channel subunit alpha1 D
CCBE1	collagen and calcium binding EGF domains 1
CCND2	cyclin D2
CDH11	cadherin 11
COPS9	COP9 signalosome subunit 9
CSMD1	CUB and Sushi multiple domains 1
CXADR	coxsackie virus and adenovirus receptor
DNAH6	Dynein Axonemal Heavy Chain 6
DRC3	dynein regulatory complex subunit 3
DRG2	developmentally regulated GTP binding protein 2
DYSF	dysferlin, limb girdle muscular dystrophy 2B (autosomal recessive)
FAM83B	family with sequence similarity 83 member B

FAT4	FAT atypical cadherin 4
FGGY	FGGY carbohydrate kinase domain containing
FRA10AC1	FRA10AC1 homolog (human)
GID4	GID complex subunit 4 homolog
GRID1	glutamate ionotropic receptor delta type subunit 1
KLF12	Kruppel like factor 12
KMT2C	lysine methyltransferase 2C
LCP2	lymphocyte cytosolic protein 2
MBNL1	muscleblind like splicing regulator 1
ME1	malic enzyme 1
MRPS24	mitochondrial ribosomal protein S24
MYOM1	myomesin 1, 185kDa
NEK5	NIMA related kinase 5
PACSIN2	protein kinase C and casein kinase substrate in neurons 2
PLXNA4	plexin A4
PPP4R3A	protein phosphatase 4 regulatory subunit 3A
PREX1	phosphatidylinositol-3,4,5-trisphosphate dependent Rac exchange factor 1
*PRDM9**	PR-Domain Containing Protein 9
RBM47	RNA binding motif protein 47
SERPINA6	serpin family A member 6
SIRPA	signal regulatory protein alpha
SLC16A10	solute carrier family 16 member 10
SLC25A17	solute carrier family 25 member 17

SNX24	sorting nexin 24
SORBS1	sorbin and SH3 domain containing 1
SYNE3	spectrin repeat containing nuclear envelope family member 3
TAF3	TATA-box binding protein associated factor 3
TAMM41	TAM41 mitochondrial translocator assembly and maintenance Homolog
TFCP2L1	transcription factor CP2 like 1
TLE3	transducin like enhancer of split 3
ZFP37	ZFP37 zinc finger protein

*****Only a part of the *PRDM9* gene sequence is overlapping a hotspot region