Copy Number Variations In South African Nguni Cattle: Prevalence, Characterization And Genetic Diversity

By

Magretha Diane Wang

Dissertation presented for the degree of
Doctor Of Philosophy (Agriculture)
in the Faculty of Agrisciences at
Stellenbosch University

Supervisor: Dr Farai Catherine Muchadeyi

Co-Supervisors: Prof. Kennedy Dzama and Dr Jasper Rees

December 2016

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

_____ _____

Magretha Diane Wang August 2016

Abstract

Copy Number Variations (CNVs) comprise of deletions, duplications and insertions larger than 1kb that occur within genomes. The identification of CNVs within regions of the bovine genome important for adaptation renders them a potential role in breed formation and adaptation. South African Nguni cattle are adapted and demonstrate an enhanced ability to endure the harsh environmental conditions of Southern Africa. This study investigated the prevalence of CNVs in the genome of South African Nguni cattle. CNV occurrence and distribution within Nguni subpopulations was assessed and comparisons with other South African cattle breeds were performed. The dynamics between CNVs and haplotype blocks (HPBs), correlations amongst CNVs and the genic locality of CNVs were investigated with the objective of determining CNV prevalence in adaptation. The Illumina BovineSNP50 beadchip was used in the first experiment to genotype 492 South African Nguni cattle sampled nationwide. *PennCNV* software identified 334 CNV regions (CNVRs) of between 30kb and 1Mb in length. Population structure analyses was performed and HPBs identified using *ADMIXTURE* and *PLINK* software respectively. Five subpopulations were evident with some degree of CNV segregation amongst populations. CNVRs covered or lay within 10Mb of 289 genes of which 149, 28, 44, 2 and 14 genes exclusive to the five sub-populations were identified. Some degree of overlap between CNVRs and the 541 HPBs was evident. In the second experiment, 59 Nguni genotypes were analyzed using the Bovine 50K Beadchip in conjunction with six other South African breeds. PennCNV software identified 356 unique CNVRs. One hundred and sixty three CNVRs identified in more than 1 animal were utilized as genetic markers to assess within and between breed genetic diversity (GD). Between breed group GD scores were 2.510, 6.115 and 4.233 for the Sanga, Taurine and composite breeds respectively. One hundred and two (Taurine) and seven (Sanga and composite) of the CNVRs demonstrated a significant ($p \leq 0.05$) association with one another. *PANTHER* overrepresentation analyses demonstrated significant representation of a number of processes, functions, components and proteins by correlated CNVR genes. CNVR based phylogenetic clustered animals of the same breed group together. In the third experiment 24 Nguni animals were sequenced at 7X coverage using illumina next generation sequencing technologies. Reads were mapped to the UMD3.1 reference genome and *RAPTR-SV* software was utilized to identify CNVs. CNVs identifed were filtered according to the number of reads that support the event with low (F10), medium (F45) and high stringencies (F75). Adjacent and overlapping CNVs were merged to form 399, 55 and 23 unique CNVRs that covered or lay within 1Mb of 358, 51 and 23 genes at F10, F45 and F75 stringencies respectively. NGS tools identified smaller CNVs compared to those reported from SNP data. Despite discrepancies between array and NGS methods, CNVR genes represented the same specific ontologies. The study demonstrated CNVRs to be prevalent in South African Nguni cattle, with potential role in breed formation and adaptation. CNVR GD scores, population structure, distribution and incidence dynamics were thus ascertained for the South African Nguni.

Opsomming

Kopie Getal Variasies (KGV) bestaan uit genomiese delesies, duplikasies of invoegings groter as 1kb in die genoom. Die identifisering van KGVs binne streke van die bees genoom, belangrik vir aanpassing, maak dat hulle 'n potensiële rol in ras vorming en aanpassing kon speel. Suid-Afrikaanse Nguni beeste is aangepas en bestand teen die harde klimaat toestande wat ervaar word in Suidelike Afrika. Hierdie studie het die teenwoordigheid van KGV's in die genoom van Suid-Afrikaanse Nguni beeste bestudeer. Die voorkoms en verspreiding van KGV's binne die Nguni sub-populasies is ge-assesseer en vergelyk met ander Suid-Afrikaanse bees rasse. Die dinamika tussen KGV's en haplotipe blokke (HPB), die korrelasie tussen verskillende KGV's en die posisie op die genoom is bestudeer met die doel om KGV voorkoms in aanpassing te bepaal. In die eerste eksperiment is die Illumina BovineSNP50 beadchip gebruik om die genotipes van 492 Nguni beeste, wat landwye ingesamel is, te bepaal. Drie honderd vier en dertig KGV Areas (KGVA) met lengtes tussen 30kb en 1Mb is met *PennCNV* sagteware geidentifiseer. Populasie struktuur analise sowel as HPB evaluasie is uitgevoer met onderskeidelik die *ADMIXTURE* en *PLINK* sagteware. Vyf sub-populasies is duidelik onderskeidbaar met n sekere graad van KGV segregasie. Die KGVA is waargeneem oor 10Mb van 298 gene; en onderskeidelik 149, 28, 44, 2 en 14 gene kon toegeskryf word aan elk van die vyf sub-populasies. 'n Sekere graad van oorvleuling kon waargeneem word tussen die KGVA's en die 541 HPB. In die tweede eksperiment is genotipes van 59 Nguni beeste ge-analiseer met die Bovine 50K Beadchip, saam met ses ander Suid-Afrikaanse bees rasse. *PennCNV* het 356 unieke KGVA's ge-identifiseer. Genetiese diversiteit (GD) is bepaal op graad van 163 KGVA's, wat versprei was oor meer as een bees. Die GD tellings tussen verskillende bees ras groepe, was 2.510, 6.115 en 4.233 vir die Sanga, Taurine en saamgestelde rasse respektiewelik. 'n Totaal van 102 (Taurine) en sewe (Sanga en saamgestelde ras) KGVA's het n beduidende assosiasie ($p \leq 0.05$) getoon met mekaar. Oor-representasie analise met die sagteware *PANTHER*, demonstreer n oorweldige verteenwoordiging van prosesse, funksies, komponente en proteïene wat korreleer met die KGVA gene. KGVA filogenetiese bome het diere van dieselfde rastipe saam groepeer. Dat spesifieke KGV's kan onderskei tussen verskillende rasse was ook opvallend. Die derde eksperiment het die genome van 24 Nguni beeste bepaal (teen 7x dekking) deur die Illumina "Next Generation Sequencing (NGS)" tegnologie. Genomiese fragmente is toegevoeg aan die oorspronlike UMD3.1 verwysings genoom, en die *RAPTR-SV* sagteware is gebruik om KGV's te identifiseer. Die KGV's is gefilter op die hoeveelheid fragmente wat die DNA basis volgorde ondersteun met lae (F10), gemiddeld (F45) en hoë (F75) strenghede. Aangrensende en oorvleulende KGV's was saamgesmelt om 399, 55 en 23 unieke KGVA's te vorm wat verpreid is oor 1Mb. Ongeveer 358, 51 en 23 gene kon ge-identifiseer word by F10, F45 en F75 onderskeidelik. NGS tegnologie kon kleiner KGV's identifiseer, wanneer vergelyk word met data vanaf SNPs. Ten spyte van teenstrydighede tussen die twee metodes, was dieselfde spesifieke ontologieë verteenwoordig deur die KGVA gene. In die geheel, demonstreer hierdie studie dat KGVA's algemeen voorkom in Suid-Afrikaanse Nguni beeste, met potensiële rolle in ras formasie en adaptasie. KGVA GD tellings, bevolkingstruktuur, verspreiding en voorkoms dinamika is toe vasgestel vir die Suid-Afrikaanse Nguni.

Acknowledgements

The study was funded by the National Research Fund DST internship and the Agriculture Research Council's Biotechnology Platform and Personal Development Program. Without their generous funds, this study would not have been possible.

My colleagues and friends at the Biotechnology Platform at the Agriculture Research Council - special mention of Minique, Khanyi, Khulekani, Erika, Stephanie and Pitunia with whom I could bounce ideas and discuss software, methodologies and results.

Kennedy Dzama and Jasper Rees - I am immensely grateful for their time, patience and care.

Farai Muchadeyi - my primary supervisor, mentor, word of reason, encouragement, direction and understanding - someone who I greatly respect and appreciate. Her careful guidance and continued support played a vital role in the production of this thesis.

My siblings and their "other halves" - Karl Wang, Melis Wang, Sven Wang, Lou Wang, Dave Pringle and Ta Pringle - people who have inspired me, supported me, helped me keep perspective and always believed in me.

My incredible parents - Ron and Biddy Wang who have never stopped believing in me, encouraging me and supporting me. Who always trust me to make decisions, giving me the space and freedom to learn, grow and pursue my dreams.

My Lord and Saviour, Jesus Christ who fights for me, strengthens me, equips me and guides me.

Contents

Addendum C

Addendum D

List Of Abbreviations

AFR - Afrikaner

AMOVA - Analysis Of Molecular Variance

ANG - Angus

ANML – Number of Animals

ATP - Adenosine Triphosphate

ATT - African Tsetse Fly-Transmitted Trypanosomosis

AvL – Average Length

BON - Bonsmara

BP – Biological Process

BRD - Breed

Camp - Cyclic Adenosine Monophosphate

CC – Cellular Component

CHR - Chromosome

CN – Copy Number

CNV - Copy Number Variations

CNVRs - Copy Number Variation Regions

DAG - Diacylglycerol

DAPC - Discriminant Analysis Of Principle Components

DLRS - Derivative Log Ratio Spread

DNA - Deoxyribonucleic Acid

DRK - Drakensberger

ECF - East Coast Fever

EIV - Eigenvalue

EUR - Europe

EXP – Expected number of genes

F10 - Low Filtering Stringency

F45 - Medium Filtering Stringency

F75 - High Filtering Stringency

GCWF - GC Wave Factor

GEN – Number of Genes

GLB - Global

GO - Gene Ontology

HER - Hereford

HOL – Holstein

HPB - Haplotype Block

IND - India

IP3 - Inositol 1,4,5-Triphosphate

IQR - Inter-Quartile Range

ISR - Israel

Kb - Kilobase

LD - Linkage Disequilibrium

MAF - Minor Allele Frequency

MaxL – Maximum Length

Mb – Megabase

MF – Molecular Function

MinL – Minimum Length

NAHR - Nonallelic Homologous Recombination

NEL - Nellore

NGS – Next Generation Sequencing

NGU – Nguni

NGxAN – Nguni Angus

PC - Principle Component

PrC – Protein Class

PCA - Principle Component Analyses

PERCN – percentage of chromosome length

Phipt - Pairwise Population

QC - Quality Control

REF - Reference

SD - Standard Deviation

SHRD – Number of CNVRs shared with other studies

SNP - Single Nucleotide Polymorphism

TP – Type of representation

USA – United States of America

List Of Genes

ABL1 - Protein Kinase Abl1

ADCY1 - Adenylate Cyclase 1

ADRA1B - Alpha-1b Adrenergic Receptor

AHSP - Alpha Hemoglobin Stabilizing Protein

AIF1L - Allograft Inflammatory Factor 1-Like

ANAPC10 - Anaphase-Promoting Complex Subunit 10

ANKRD50 - Ankyrin Repeat Domain 50

ARL6 – Adenosine Diphosphated-Ribosylation Factor-Like 6

ATXN7L3B - Ataxin 7-Like 3B

BDKRB1 - Bradykinin Receptor B1

CD79A - B-Cell Antigen Receptor Complex-Associated Protein Alpha Chain

CDKN1C - Cyclin-Dependent Kinase Inhibitor 1

CFTR - Cystic Fibrosis Transmembrane Conductance Regulator

CHIC2 - Cysteine-Rich Hydrophobic Domain 2

COL13A1 - Collagen Type XIII Alpha 1

DUSP18 - Dual Specificity Phosphatase 18

FBXW7 - F-Box And WD Repeat Domain Containing 7

FBXW9 - F-Box/WD Repeat-Containing Protein 9

FOXP1 - Forkhead Box Protein P1

GABRA2 - Gamma-Aminobutyric Acid Type A Receptor Alpha2 Subunit

GABRB1 - Gamma-Aminobutyric Acid Receptor Subunit Beta-1

GSTT1 - Glutathione S-Transferase Theta-1

GSTT3 - Glutathione S-Transferase Theta-3

HHIP - Hedgehog Interacting Protein

HSF1 - Heat Shock Transcription Factor 1

HSF4 - Heat Shock Transcription Factor 4

HSP1 - Heat Shock Protein 1

HSP90AA1 - Heat Shock Protein 90 Alpha Family Class A Member 1

HSPA12B - Heat Shock Protein Family A (Hsp70) Member 12B

HSPA5 - Heat Shock Protein Family A (Hsp70) Member 5

HSPA6 - Heat Shock Protein Family A Member 6

HSPB1 - Heat Shock Protein Family B Member 1

HSPB8 - Heat Shock Protein Family B (Small) Member 8

HSPBP1 - Heat Shock Binding Protein 1

IFGBP3 - Insulin Like Growth Factor Binding Protein 3

IFN-Γ - Interferon-Γ

IGF-II - Insulin-Like Growth Factor 2

IGFBP3 - Insulin-Like Growth Factor 1 Binding Protein 3

IGLL1 - Immunoglobulin Lambda-Like Polypeptide 1

IL-12 - Interleukin-12

IL12B - Interleukin-12 Subunit Beta

IL15 - Interleukin-15

IL27RA - Interleukin 27 Receptor Subunit Alpha

KLHL2 - Kelch-Like Family Member 2

LFNG - O-Fucosylpeptide 3-Beta-N-Acetylglucosaminyltransferase

LOC527441 - Low Affinity Sodium-Glucose Cotransporter-Like

LRFN5 - Leucine Rich Repeat And Fibronectin Type III Domain Containing 5

LSP1 - Lymphocyte-Specific Protein 1

LYAR - Ly1 Antibody Reactive Homolog

MFGE8 - Milk Fat Globule-EGF Factor 8 Protein

MGC157405 - Bos Taurus Pregnancy-Associated Glycoprotein

MIMT1 - MER1 Repeat Containing Imprinted Transcript 1

MR1 - Major Histocompatibility Complex, Class I-Related

NLRP5 - Nacht, Lrr And Pyd Domains-Containing Protein 5

NSG1 - Neuron-Specific Protein Family Member 1

NTNG2- Netrin G2

OTOP1 - Otopetrin 1

P2RX7 - Purinergic Receptor P2X 7

PCDH10 - Protocadherin 10

PCDH7 - Protocadherin 7

PECAM1 - Platelet/Endothelial Cell Adhesion Molecule 1

PRDX2 - Peroxiredoxin-2

PTGS2 - Prostaglandin-Endoperoxide Synthase 2

RPS19 - 40S Ribosomal Protein S19

SERPINB6 - Serpin Family B Member 6

SLC - Solute Carrier

SLC5A1 - Solute Carrier Family 5 Member 1

SMARCB1 - SWI/SNF-Related Matrix-Associated Actin-Dependent Regulator Of Chromatin Subfamily B Member 1

SMTN - Smoothelin

SPRY1 - Protein Sprouty Homolog 1

STX18 - Syntaxin 18

TAOK3 - TAO Kinase 3

THBS1 - Thrombospondin-1

TMEM128 - Transmembrane Protein 128

TNFAIP8 - Tumor Necrosis Factor, Alpha-Induced Protein 8

TNNT3 - Troponin T, Fast Skeletal Muscle

TSPAN32 - Tetraspanin-32

WBSCR17 - Williams-Beuren Syndrome Chromosome Region 17

WDR1 - WD Repeat Domain 1

ZBTB49 - Zinc Finger And BTB Domain Containg 49

List Of Tables

List Of Figure

Chapter 1: General Introduction

1.1 Background

Genetic diversity is a multifaceted value that forms surety in the instance of changes in production circumstances, disease threat or market demands (Reist-marti *et al.*, 2003). The degree of genetic variability within a population, and hence the population genetic diversity is an essential source of biodiversity, providing the basic material for adaptation (Fisher 1930; Hughes *et al.,* 2008). Structural variation ranging from single nucleotide polymorphisms (SNPs) to large chromosomal events is primarily responsible for the genetic diversity evident between breeds and individuals (Freeman *et al.,* 2006). Within this broad range of variants, copy number variations (CNVs) comprise duplications, deletions and insertions within the genome that are larger than 1 kb (Freeman *et al.,* 2006). While SNPs were considered the predominant form of genomic variation, accounting for much normal phenotypic variation, CNVs were deemed mutational occurances only demonstrating deleterious effects like neurocognitive diseases and cancers in humans (The International Human Hapmap Consortium 2005; Freeman *et al.,* 2006). Two separate studies performed in 2004 (Sebat *et al.,* 2004; Iafrate *et al.,* 2004), however reported the widespread presence of copy number variation in normal individuals. Subsequent studies in humans (Zarrei *et al.,* 2015), dogs (Molin *et al.,* 2014), horses (Doan *et al.,* 2012; Ghosh *et al.,* 2014; Wang *et al.,* 2014), chickens (Crooijmans *et al.,* 2013; Völker *et al.,* 2010) and cattle (Liu *et al.,* 2010) among other species exhibited a similar pattern placing CNVs as a vital component of genetic diversity. Cattle breed differences in CNV prevalence have been alluded to and it has been alleged that cattle CNVs occur independently in breeds, thus playing a role in breed disparities and hence breed formation and adaptation (Bickhart *et al.,* 2016; Xu *et al.,* 2016; Liu *et al.,* 2010). Taurine and Indicine cattle breeds exhibit reasonable levels of functional genetic diversity with Indicine cattle displaying an enhanced ability to adapt to tropical climates (Hanotte *et al.,* 2003; Hansen 2004). African Taurine, Indicine and composite cattle breeds have more CNV loci than the European Taurine breeds (Liu *et al.,* 2010; Matukumalli *et al.,* 2009). CNVs have also been identified in regions covering a number of genes that are recognized to play a role in cattle environmental responses and adaptation (Matukumalli *et al.,* 2009; Kijas *et al.,* 2011), with CNV incidence tending to parallel breed history and breed formation patterns (Bickhart *et al.,* 2016; Xu *et al.,* 2016; Hou *et al.,* 2011). Several predicted CNV loci fall within gene boundaries and may play a vital role in ascertaining subspecies divergence (Matukumalli *et al.,* 2009).

Africa demonstrates a primarily tropical environment with harsh climatic conditions, feed and water scarcity and widespread pathogens and epidemic diseases (Hoffmann 2010; Mirkena *et al.,* 2010). Despite a general reduction in the burden of livestock diseases through the accessibility and efficacy of vaccines and drugs coupled with superior diagnostic technologies, new diseases have emerged and disease prevalence remains a primary concern in tropical agriculture (Lamy *et al.,* 2012). With 150 different breeds/populations, locally adapted cattle of Africa demonstrate far greater levels of resistance to such conditions due to the development of unique adaptive traits that enable them to survive and produce despite the harsh conditions (Mirkena *et al.,* 2010). Environmentally adapted breeds facilitate sustainable food production in lower-input

farming systems that exhibit a reduced impact on the environment while also delivering animal draught power, fuel from dung and clothing (Reist-marti *et al.,* 2003).

The Nguni breed of South Africa is increasingly attracting international interest, mainly due to its resilience to tick-borne diseases, low internal parasite load, high reproductive performance, good walking and foraging ability and low maintenance requirements, acquired through centuries of adapting to the abrasive environmental conditions of South Africa (Muchenje *et al.,* 2008; Mapiye *et al.,* 2009; Rechav and Kostrzewski 1991). This is a distinct, conserved, Sanga type cattle breed that has undergone little synthetic breeding and is considered to comprise of multiple subpopulations (Bester *et al.,* 2001; Mapiye *et al.,* 2009).

1.2 Problem Statement

Adaptation is vital for species to survive novel and transforming environments. The underlying genomic mechanisms involved in adaptation are however not fully understood. Africa comprises a harsh climate with abrasive environmental conditions including disease, extreme climates and food and water scarcity posing notable challenges in the livestock production sector. While exotic cattle breeds have undergone years of intensive selection for production traits, indigenous breeds have primarily been subject to natural selection with adaptation encompassing the primary driving force of survival. Having undergone little synthetic breeding, local breeds for the most part lack genetic characterisation. With an enhanced ability to endure under harsh environmental conditions, South African Nguni cattle comprise one such breed demonstrating notable adaptation to harsh conditions while comprising a primarily uncharacterized and untapped genetic resource. Considered genetically diverse, Nguni cattle hold promise in low input production systems and in crossbreeding schemes aimed at combining the production traits of exotic breeds with the adpatation of the indigenous breeds.

Genomic CNVs are modifications in DNA structure comprising of deletions, duplications and insertions greater than 1kb in size. Thought to be primary role-players in breed formation and adaptation, CNVs potentially arise independently in breeds, thus contributing to between breed discrepancies (Bickhart *et al.,* 2016; Lingyang Xu *et al.,* 2016). The publication of two alternative reference genomes (Btau4.6.1 and UMD3.1) has enabled new avenues of bovine genomics. Recent studies primarily in Taurine (Holstein, Hereford, Angus) and Indicine (Nelore) breeds render CNVs embodying 20% of the autosomal bovine genome with 0.91-4.7% encompassing CNV regions (Bae *et al.,* 2010; Fadista *et al.,* 2010; Liu *et al.,* 2010; Cicconardi *et al.,* 2013). CNV associations with gut nematodes were also reported in Angus cattle (Hou *et al.,* 2012). The prevalence of CNVs in South African Nguni cattle is however unknown. With enhanced adaptation and disease resistance, CNVs may comprise a significant source of Nguni phenotypic diversity. Characterization of the existence and distribution of CNVs is therefore a vital step towards dissecting the molecular mechanisms underlying phenotypic variation in this breed. Genetic characterization of Nguni cattle is essential for the efficient management and conservation of the Nguni breed. In addition, Nguni cattle comprise an apposite model for studying the underlying genetic components of adaptation.

1.3 Justification

The genetic improvement of a number of domesticated cattle breeds worldwide has been achieved through the development and focus on intense selection programs. The role that CNVs play within breeds to ensure diversity and adaptation has, however not yet been fully investigated. Understanding the multiple components of functional breed diverstiy has important implications for breed management and genetic improvement practices, especially in breeds that are locally adapted and have not undergone intense artificial selection. With CNVs demonstrating a possible correspondence with breed diversity and adaptation, Nguni cattle present a valuable breed in which to investigate CNV prevalence and distribution. CNV loci have been found within gene boundaries, with the incidence of some coinciding with breed histories and breed formation patterns (Hou *et al.,* 2011; Matukumalli *et al.,* 2009). The availability of two cattle reference genomes (Btau 4.6.1 and UMD3.1) (The Bovine Genome sequencing and analysis consortium, 2009) and the develeopment of genomewide single nucleotide polymorphism (SNP) genotying arrays has enabled new avenues of research in bovine genomics. Initial CNV analyses were performed utilizing array comparative hybridization methodologies. The development of CNV discovery tools utilizising next generation sequencing and SNP array data hold opportunity for the in depth investigation into the prevalence of CNVs (Bickhart *et al.,* 2015; Wang *et al.,* 2007; Zhao *et al.,* 2013).

1.4 Objectives

This study aimed to identify, characterize and validate CNVs within the genome of South African Nguni cattle. The prevalence of CNVs in breed formation, segregation and adaptation was investigated. The distribution of CNVs within Nguni populations in comparison to other South African cattle breeds was interrogated together with the possibility of CNVs occurring at multiple genomic locations involved in specific biological processes, cellular components, molecular functions and proteins. The utility and complementarity of the Bovine 50K Beadchip and next generation sequencing technologies for CNV identification was investigated.

1.5 Thesis Overview And Layout

This study therefore assessed CNVs within the genome of South African Nguni cattle using the Bovine 50K Beadchip and next generation sequencing technologies. The thesis is structured in the form of a general introduction chapter, a literature review and three stand alone experimental chapters (chapters 3-5). A general discussion and conclusion is presented at the end. Current knowledge pertaining to bovine CNVs and the potential role of CNVs in tropical adaptation of African cattle breeds was discussed in the literature review (chapter 2) that was published in *Animal Genetics*. In chapter 3 the Bovine 50K Beadchip was utilized to assess genomic population structure and CNV prevalence in 492 South African Nguni cattle. After filtering data for quality using *PLINK* software, *PennCNV* software identified CNVs in each of 492 animals. *ADMIXTURE, R, gPLINK* and *Haploview* was utilized to perform population structure analyses and to determine Haplotype blocks. The distribution of CNVs among populations identified and across haplotype

blocks was investigated. The *PANTHER* database was used to assess biological processes, molecular functions and cellular component's of genes covered or lying within 10Mb of CNVs identified. This chapter was published in *BMC Genomics*. chapter 4 then assessed CNVs as a measure of genetic diversity. In order to better understand the prevalence of CNVs in genetic diversity, 2 South African Taurine, 2 South African composite, 2 South African Sanga (including the Nguni) and 1 Sanga Taurine cross breed breeds were assessed for CNV prevalence and genetic diversity. Correlations between CNVs identified were investigated within and across the different breed groups. Geno ontology analyses were executed. The possible simultaneous occurrence of multiple CNV loci involved in similar mechanisms and the representation of CNV genes in biological processes, cellular components and molecular functions was assessed. CNVs identified in more than 1 animal were utilized as genetic markers to assess within and between breed genetic diversity. This chapter is prepared for submission in a international peer reviewed journal. Chapter 5 comprised a validation chapter assessing CNVs in Nguni cattle using whole genome sequencing technologies. Twenty-four animal were selected and sequenced at 10X coverage using illumina next generation sequencing technologies. *FastQC* was used to determine read quality and reads were subsequently trimmed using Trimmomatic. Trimmed reads were mapped to the UMD3.1 reference genome using Burrows Wheeler Alignment and samtools. *RAPTR-SV* software was then used to cluster CNVs which were filtered according to the number of reads that supported the CNV events at F10, F45 and F75. In house scripts amalgamated adjacent and overlapping CNVs into CNVRs and the *PennCNV* scan_CNV.pl script was used to identify genes lying within 10Mb of CNVRs. Gene ontologies for identified CNVRs were obtained on the *PANTHER* databases and comparisons were drawn with the results from chapter 3 and 4. Chapter 6 (general discussion) presented a critical discussion of the overall study and describes implications of the study findings.

1.6 References

Bae, J.S., H.S. Cheong, L.H. Kim, S. NamGung, T.J. Park, J. Chun, J.Y. Kim, C.F.A. Pasaje, J.S. Lee, and H.D. Shin. 2010. "Identification of Copy Number Variations and Common Deletion Polymorphisms in Cattle." *BMC Genomics* 11: 232. doi:10.1186/1471-2164-11-232.

Bester, J., I.E. Matjuda, J.M. Rust, H.J. Fourie. 2001. The Nguni: case study. In: *FAO Community-based management of animal genetic resources.* Rome: UNDP, GTZ, CTA, FAO. pp. 45-68.

Bickhart, D. M., J.L. Hutchison, L. Xu, R.D. Schnabel, J.F. Taylor, J.M. Reecy, S. Schroeder, C.P. Van Tassell, T.S. Sonstegard, and G.E. Liu. 2015. "RAPTR-SV : A Hybrid Method for the Detection of Structural Variants." *Bioinformatics*, 1–7.

Bickhart, D.M., L. Xu, J.L. Hutchison, J.B. Cole, D.J. Null, S.G. Schroeder, J. Song, et al. 2016. "Diversity and Population-Genetic Properties of Copy Number Variations and Multicopy Genes in Cattle," 1–10. doi:10.1093/dnares/dsw013.

Cicconardi, F., G. Chillemi, A. Tramontano, C. Marchitelli, A. Valentini, P. Ajmone-Marsan, and A. Nardone. 2013. "Massive Screening of Copy Number Population-Scale Variation in Bos Taurus Genome." *BMC Genomics* 14 (1): 124. doi:10.1186/1471-2164-14-124.

Crooijmans, R.P.M., M.S. Fife, T.W. Fitzgerald, S. Strickland, H.H. Cheng, P. Kaiser, R. Redon, and M.A.M. Groenen. 2013. "Large Scale Variation in DNA Copy Number in Chicken Breeds." *BMC Genomics* 14 (1): 398. doi:10.1186/1471-2164-14-398.

Doan, R., N.D. Cohen, J. Sawyer, N. Ghaffari, C.D. Johnson, and S.V. Dindot. 2012. "Whole-Genome Sequencing and Genetic Variant Analysis of a Quarter Horse Mare." *BMC Genomics* 13 (1). BioMed Central Ltd: 78. doi:10.1186/1471-2164-13-78.

Fadista, J., B. Thomsen, L. Holm, and C. Bendixen. 2010. "Copy Number Variation in the Bovine Genome." *BMC Genomics* 11: 284. doi:10.1186/1471-2164-11-284.

Fisher, R.A. 1930. "The Genetical Theory of Natural Selection." *Genetics* 154: 272. doi:10.1038/158453a0.

Freeman, J.L., G.H. Perry, L. Feuk, R. Redon, S.A. McCarroll, D.M. Altshuler, H. Aburatani, et al. 2006. "Copy Number Variation: New Insights in Genome Diversity." *Genome Research* 16 (617): 949–61. doi:10.1101/gr.3677206.

Ghosh, S., Z. Qu, P.J. Das, E. Fang, R. Juras, E.G. Cothran, S. McDonell, et al. 2014. "Copy Number Variation in the Horse Genome." *PLoS Genetics* 10 (10). doi:10.1371/journal.pgen.1004712.

Hanotte, O., Y. Ronin, M. Agaba, P. Nilsson, A. Gelhaus, R. Horstmann, Y. Sugimoto, et al. 2003. "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 of the United States of America* 100: 7443–48. doi:10.1073/pnas.1232392100.

Hansen, P. J. 2004. "Physiological and Cellular Adaptations of Zebu Cattle to Thermal Stress." *Animal Reproduction Science* 82–83: 349–60. doi:10.1016/j.anireprosci.2004.04.011.

Hoffmann, I. 2010. "Climate Change and the Characterization, Breeding and Conservation of Animal Genetic Resources." *Animal Genetics* 41: 32–46. doi:10.1111/j.1365-2052.2010.02043.x.

Hou, Y., G.E. Liu, D.M. Bickhart, M.F. Cardone, K. Wang, E. Kim, L.K. Matukumalli, et al. 2011. "Genomic Characteristics of Cattle Copy Number Variations." *BMC Genomics* 12: 127. doi:10.1186/1471-2164-12-127.

Hou, Y., G.E. Liu, D.M. Bickhart, L.K. Matukumalli, C. Li, J. Song, L.C. Gasbarre, C.P. Van Tassell, and T.S. Sonstegard. 2012. "Genomic Regions Showing Copy Number Variations Associate with Resistance or Susceptibility to Gastrointestinal Nematodes in Angus Cattle." *Functional and Integrative Genomics* 12: 81–92. doi:10.1007/s10142-011-0252-1.

Hughes, A.R., B.D. Inouye, M.T.J. Johnson, N. Underwood, and M. Vellend. 2008. "Ecological Consequences of Genetic Diversity." *Ecology Letters*. doi:10.1111/j.1461-0248.2008.01179.x.

Iafrate, A.J., L. Feuk, M.N. Rivera, M.L. Listewnik, P.K. Donahoe, Y. Qi, S.W. Scherer, and C. Lee. 2004. "Detection of Large-Scale Variation in the Human Genome." *Nature Genetics* 36 (9): 949–51. doi:10.1038/ng1416.

Kijas, J.W., W. Barendse, W. Barris, B. Harrison, R. McCulloch, S. McWilliam, and V. Whan. 2011. "Analysis of Copy Number Variants in the Cattle Genome." *Gene* 482. Elsevier B.V.: 73–77. doi:10.1016/j.gene.2011.04.011.

Lamy, E., S. van Harten, E. Sales-baptista, and M. Manuela. 2012. "Factors Influencing Livestock

Productivity .” In *Environmental Stress and Amelioration in Livestock Production*, edited by V. Sejian, S.M.K. Naqvi, T. Ezeji, J. Lakritz, and R. Lal, 19–52. Sprinter-Verlag Berlin Heidelberg. doi:10.1007/978-3-642-29205-7.

Liu, G.E., Y. Hou, B. Zhu, M.F. Cardone, L. Jiang, A. Cellamare, A. Mitra, et al. 2010. “Analysis of Copy Number Variations among Diverse Cattle Breeds.” *Genome Research* 20: 693–703. doi:10.1101/gr.105403.110.

Mapiye, C., M. Chimonyo, K. Dzama, J. G. Raats, and M. Mapekula. 2009. “Opportunities for Improving Nguni Cattle Production in the Smallholder Farming Systems of South Africa.” *Livestock Science* 124 (1–3). Elsevier B.V.: 196–204. doi:10.1016/j.livsci.2009.01.013.

Matukumalli, L.K., C.T. Lawley, R.D. Schnabel, J.F. Taylor, M.F. Allan, M.P. Heaton, J. O’Connell, et al. 2009. “Development and Characterization of a High Density SNP Genotyping Assay for Cattle.” *PLoS ONE* 4. doi:10.1371/journal.pone.0005350.

Mirkena, T., G. Duguma, A. Haile, M. Tibbo, a. M. Okeyo, M. Wurzinger, and J. Sölkner. 2010. “Genetics of Adaptation in Domestic Farm Animals: A Review.” *Livestock Science* 132. Elsevier B.V.: 1–12. doi:10.1016/j.livsci.2010.05.003.

Molin, A., J. Berglund, M.T. Webster, and K. Lindblad-Toh. 2014. “Genome-Wide Copy Number Variant Discovery in Dogs Using the CanineHD Genotyping Array.” *BMC Genomics* 15 (1): 210. doi:10.1186/1471-2164-15-210.

Muchenje, V., K. Dzama, M. Chimonyo, J. G. Raats, and P. E. Strydom. 2008. “Meat Quality of Nguni, Bonsmara and Aberdeen Angus Steers Raised on Natural Pasture in the Eastern Cape, South Africa.” *Meat Science* 79 (1): 20–28. doi:10.1016/j.meatsci.2007.07.026.

Rechav, Y., and M.W. Kostrzewski. 1991. “Relative Resistance of Six Cattle Breeds To the Tick Boophilus Decoloratusin South Africa.” *Onderstepoort J. Vet. Res* 58: 181–86.

Reist-marti, S.B., H. Simianer, J. Gibson, O. Hanotte, and J.E.O. Rege. 2003. “Weitzman’s Approach and Conservation of Breed Diversity : An Application to African Cattle Breeds” 17 (5): 1299–1311. doi:10.1046/j.1523-1739.2003.01587.x.

Sebat, J., B. Lakshmi, J. Troge, J. Alexander, J. Young, P. Lundin, S. Maner, et al. 2004. “Large-Scale Copy Number Polymorphism in the Human Genome” 305 (July): 525–28. doi:10.1126/science.1098918.

The Bovine Genome sequencing and analysis consortium, C.G. Elsik, R.L. Tellam, and K.C. Worley. 2009. “The Genome Sequence of Taurine Cattle: A Window to Ruminant Biology and Evolution.” *Science* 324: 522–28. doi:10.1126/science.1169588.

The International Human Hapmap Consortium. 2005. “A Haplotype Map of the Human Genome.” *Nature* 437: 1299–1320. doi:10.1038/nature04226.

Völker, M., N. Backström, B.M. Skinner, E.J. Langley, S.K. Bunzey, H. Ellegren, and D.K. Griffin. 2010. “Copy Number Variation, Chromosome Rearrangement, and Their Association with Recombination during Avian Evolution.” *Genome Research* 20: 503–11. doi:10.1101/gr.103663.109.

Wang, K., M. Li, D. Hadley, R. Liu, J. Glessner, S.F.A. Grant, H. Hakonarson, and M. Bucan. 2007. “PennCNV: An Integrated Hidden Markov Model Designed for High-Resolution Copy Number

Variation Detection in Whole-Genome SNP Genotyping Data." *Genome Research* 17: 1665–74. doi:10.1101/gr.6861907.

Wang, W., S. Wang, C. Hou, Y. Xing, J. Cao, K. Wu, C. Liu, et al. 2014. "Genome-Wide Detection of Copy Number Variations among Diverse Horse Breeds by Array CGH." *PLoS ONE* 9 (1). doi:10.1371/journal.pone.0086860.

Xu, L., J.B. Cole, D.M. Bickhart, Y. Hou, J. Song, P.M. VanRaden, T.S. Sonstegard, C.P. Van Tassell, and G.E. Liu. 2014. "Genome Wide CNV Analysis Reveals Additional Variants Associated with Milk Production Traits in Holsteins." *BMC Genomics* 15: 683. doi:10.1186/1471-2164-15-683.

Xu, L., Y. Hou, D.M. Bickhart, Y. Zhou, E. Hamidi, J. Song, T.S. Sonstegard, C.P. Van Tassell, and G.E. Liu. 2016. "Population-Genetic Properties of Differentiated Copy Number Variations in Cattle." *Nature Publishing Group*. Nature Publishing Group, 1–8. doi:10.1038/srep23161.

Zarrei, M., J.R. Macdonald, D. Merico, and S.W. Scherer. 2015. "A Copy Number Variation Map of the Human Genome." *Nature Publishing Group* 16 (February): 172–83. doi:10.1038/nrg3871.

Zhao, M., Q.Q. Wang, P. Jia, and Z. Zhao. 2013. "Computational Tools for Copy Number Variation (CNV) Detection Using next-Generation Sequencing Data: Features and Perspectives - Springer." *BMC Bioinformatics* 14 (Suppl 11). BioMed Central Ltd: S1. doi:10.1186/1471-2105-14-S11-S1.

Genomic Population Structure And Prevalence Of Copy Number Variations In South African Nguni Cattle

Magretha D Wang[1,2], Kennedy Dzama [1], Charles A Hefer[2], Farai C Muchadeyi[2]

[1]Department of Animal Sciences, University of Stellenbosch, Private Bag X1, Matieland 7602, South Africa

[2]Biotechnology Platform, Agricultural Research Council, Private Bag X5, Onderstepoort 0110, South Africa

Published in *BMC Genomics* (2015) 16:894. DOI: 10.1186/s12864-015-2122-z

Chapter 3: Genomic Population Structure And Prevalence Of Copy Number Variations In South African Nguni Cattle

3.1 Abstract

Copy number variations (CNVs) are modifications in DNA structure comprising of deletions, duplications, insertions and complex multi-site variants. Although CNVs are proven to be involved in a variety of phenotypic discrepancies, the full extent and consequence of CNVs is yet to be understood. To date, no such genomic characterization has been performed in indigenous South African Nguni cattle. Nguni cattle are recognized for their ability to sustain harsh environmental conditions while exhibiting enhanced resistance to disease and parasites and are thought to comprise of up to nine different ecotypes. Illumina BovineSNP50 beadchip data was utilized to investigate genomic population structure and the prevalence of CNVs in 492 South African Nguni cattle. *PLINK, ADMIXTURE, R, gPLINK* and *Haploview* software was utilized for quality control, population structure and haplotype block determination. *PennCNV* hidden markov model identified CNVs and genes contained within and 10Mb downstream from reported CNVs. *PANTHER* and *Ensembl* databases were subsequently utilized for gene annotation analyses. Population structure analyses on Nguni cattle revealed 5 sub-populations with a possible sub-structure evident at K equal to 8. Four hundred and thirty three CNVs that formed 334 CNVRs ranging from 30 kb to 1Mb in size are reported. Only 231 of the 492 animals demonstrated CNVRs. Two hundred and eighty nine genes were observed within CNVRs identified. Of these 149, 28, 44, 2 and 14 genes were unique to sub-populations A, B, C, D and E respectively. Gene ontology analyses demonstrated a number of pathways to be represented by respective genes, including immune response, response to abiotic stress and biological regulation processess. CNVs may explain part of the phenotypic diversity and the enhanced adaptation evident in Nguni cattle. Genes involved in a number of cellular components, biological processes and molecular functions are reported within CNVRs identified. The significance of such CNVRs and the possible effect thereof needs to be ascertained and may hold interesting insight into the functional and adaptive consequence of CNVs in cattle.

Keywords

Breed diversity, Nguni cattle, genetic variation, adaptation

3.2 Background

Copy number variants (CNVs) are segments of DNA that are 1kb or larger in size and display a variable copy number relative to a reference genome, hence comprising deletions, duplications and insertions(Tuzun *et al.,* 2005). A number of recent studies demonstrated CNVs to be prevalent in bovine genomes (Liu *et al.,* 2010; Liu and Bickhart 2012). CNVs are reported to affect a greater percentage of genomic sequences and have been identified in regions covering a number of genes that are recognized to play a role in cattle environmental responses and adaptation (Kijas *et al.,* 2011). CNV region (CNVR) incidence also

demonstrates some tendancy to parallel breed history and breed formation patterns (Hou *et al.,* 2011; Kijas *et al.,* 2011).

The development and focus on intense selection programs have greatly enhanced the genetic improvement of a number of domesticated cattle breeds worldwide. Understanding the multiple components of functional breed diversity have important implications for breed management and genetic improvement practices, especially in breeds that are locally adapted and have not undergone intense artificial selection. South African Nguni cattle represent such a distinct, conserved, Sanga type cattle breed that has undergone little synthetic breeding (Bester et al. 2001; Mapiye et al. 2009; Rechav and Kostrzewski 1991). Having endured natural selection pressures from a variety of disease agents and harsh climatic conditions, Nguni cattle have proven to prevail in suboptimal environmental circumstances (Marufu *et al.,* 2011). These indigenous South African cattle are also recognized for their small frame size and diversely patterned and multi-coloured hides.

The availability of two cattle reference genomes (Btau4.6.1 and UMD3.1) (The Bovine Genome sequencing and analysis consortium, 2009) and the develeopment of genomewide single nucleotide polymorphism (SNP) genotying arrays has enabled new avenues of research in bovine genomics. Although SNPs have been the primary focus of variant screening and association analyses, the recent development of CNV discovery tools utilizising both sequencing and SNP data hold opportunity for the in depth investigation into the prevalence of additional types of genomic variation (Bickhart *et al.,* 2015; Wang *et al.,* 2007; Zhao *et al.,* 2013). The role that CNVs play within breeds to ensure diversity and adaptation has not yet been investigated. Nguni cattle have undergone scant synthetic breeding and are well adapted to their primary environment. With CNVs demonstrating a possible correspondence with breed diversity and adaptation, Nguni cattle present a valuable breed in which to investigate CNV prevalence and distribution.

This study investigated the population structure, haplotype block structure and the occurance and distribution of CNVs in Nguni cattle of South Africa using genotype data from the Ilumina Bovine SNP50K panel. Extensive linkage disequilibrium studies have been performed in cattle (Gautier *et al.,* 2007; McKay *et al.,* 2007). Haplotype block (HPB) structure studies are however not as widespread (Mokry *et al.,* 2014). The characterization of HPB structure at the population level contribute towards understanding the nature of non-linear association between phenotypes and genes (Mokry *et al.,* 2014). This study determined the prevalence of CNVs within Nguni cattle followed by an analysis of their distribution within the different ecotypes inferred by population structure analysis. The prevalence of HPB structures in CNV formation was also investigated.

3.3 Materials And Methods

3.3.1 Sample Collection And *Data Generation*

Blood samples collected in 10ml EDTA VACUETTE® tubes by means of venal puncture of the caudal vein and hair root samples were collected from 492 Nguni animals distibuted across South Africa (Figure 3.1). Genomic DNA was extracted by means of the Qiagen DNeasy Blood and Tissue Kit from the blood samples. Proteinase-K digestions followed by phenol, chloroform, isoamyl alcohol extraction ad ethanol precipitation were utilised for the extraction of genomic DNA from hair root samples (Green and Sambrook 2012). The quantity and quality of extracted DNA was assessed by means of the Qubit and those samples exhibiting a minimum concentration of 50µl were subsequently genotyped with the Illumina BovineSNP50 (Illumina Inc., San Diego, CA) containing 54,001 highly informative markers that uniformly span the bovine genome. Illumina BovineSNP50 BeadChip SNP markers were designed based on the Btau4.6.1 reference genome. Markers were clustered and genotyped by means of Illumina *GenomeStudio* v2.0 software. Fifty four of the genotyped samples were derived from a previous study (Makina *et al.,* 2014). Ethics approval was obtained for the study (Ref. Nr.:2014/CAES/101).

Figure 3.1 Geographic origin of the 492 Nguni cattle sampled in the current study (http://www.google.co.za/maps).

3.3.2 SNP Quality Assessment

SNP quality control and sample pruning was performed by means of *PLINK* (version 1.9) (Purcell *et al.,* 2007) SNPs with a minor allele frequency of greater than 0.02 and/or genotype rate of less than 0.95 were filtered from the dataset.

3.3.3 Determination Of Population Structure

One of the SNPs was removed for each pair of SNPs demonstrating an LD of greater than 0.1 on a sliding window of 30 SNPs. Relationship-based pruning was performed and one member of each pair of animals with an observed genomic relatedness of greater than 0.25 was removed from further analyses to correct for population stratification (Thornton *et al.,* 2014). *ADMIXTURE* (Alexander *et al.,* 2009) was subsequently used to determine population structure of unrelated animals. *ADMIXTURE* was run from K = 2 to K = 10 and a cross-validation procedure was used to ascertain the best k (Figure 3.2). That k-value that generated the lowest cross-validation standard error was determined as being the most probable population sub-structuring. Q estimate matrices barplots were generated with *R* (http://cran.r-project.org) for each value of k, and animals were sorted according into ecotypes based on this population structure.

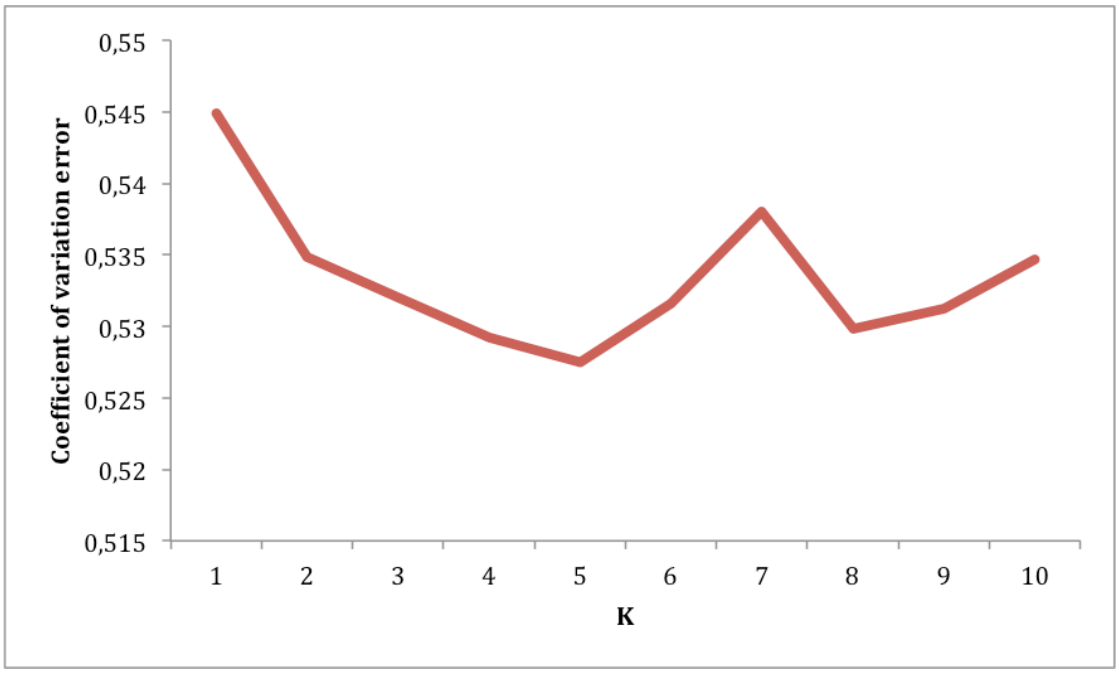

Figure 3.2 Cross-validation plot demonstrating the effect of different K-values on the cross-validation error.

A discriminant analysis of principle components (DAPC) was performed using *ADEGENET* 2.0.0 in *R* (Thornton *et al.,* 2014). In the absence of group priors, DAPC infers genetic clusters from sequential K-means and model selections. The find.clusters script was utilized to determine clusters with a maximum of 9 groups. The cumulative variance against the number of retained principle components (PCs) (Figure 3.3), demonstrated the greatest amount of variance being explained by 100 PCs which were therefore utilized in conjunction with 2 discriminant functions (Figure 3.4) to determine group clustering. A scatterplot of the DPCA was subsequently generated.

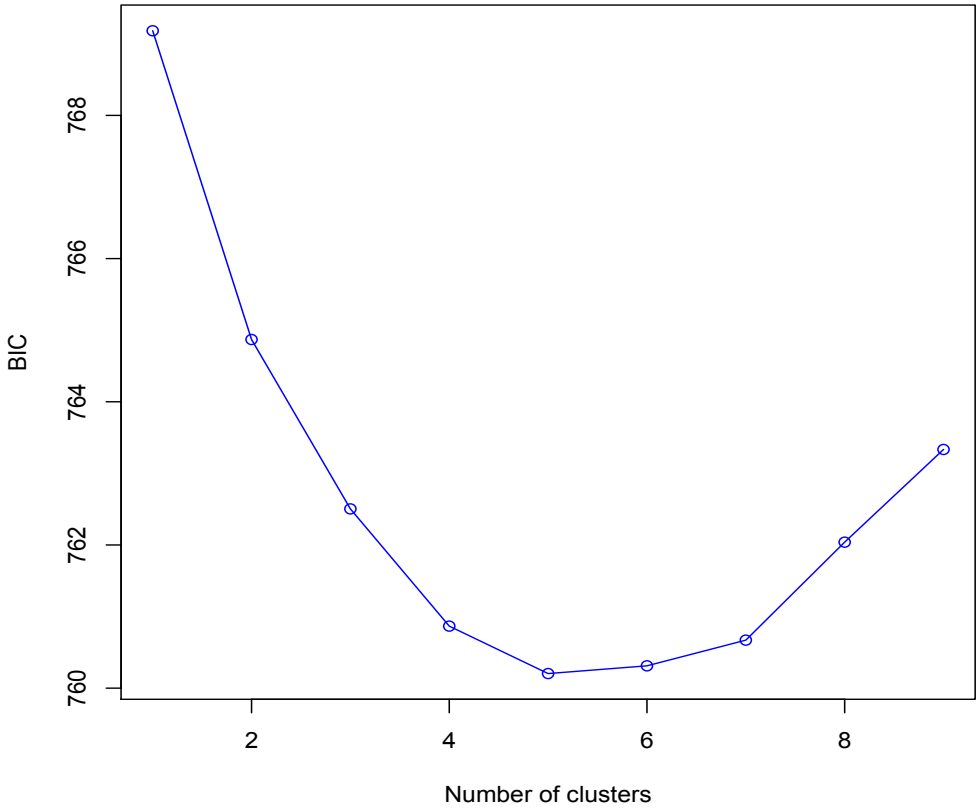

Figure 3.3 A linear graph demonstrating the bayesian information criterion against the number of clusters**.**

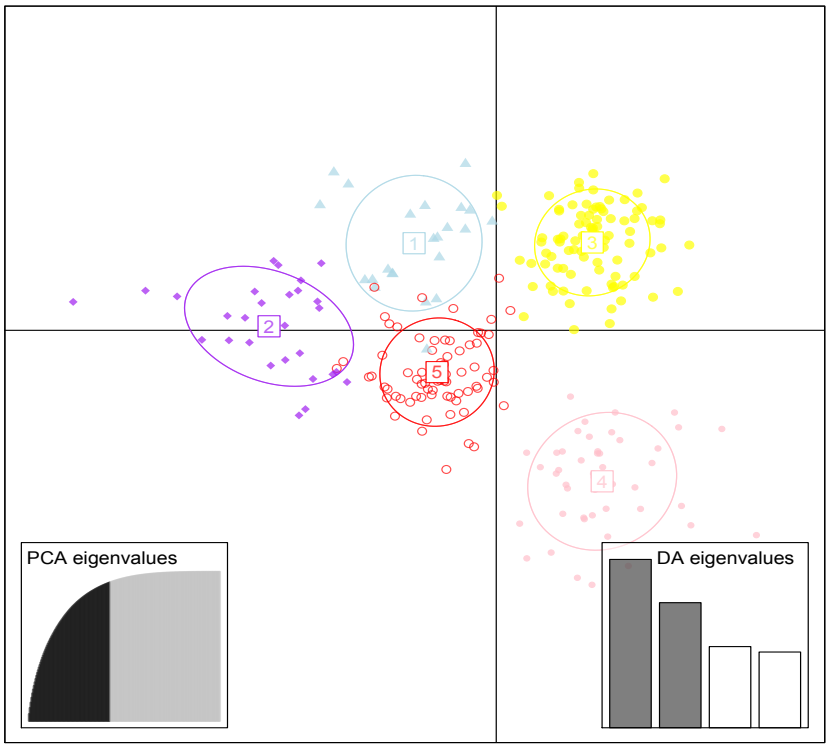

Figure 3.4 A DPCA plot demonstrating the group clustering with the subfigure 1 and 2 exhibiting discriminant eigenvalues and PCA eigenvalues.

3.3.4 Analysis Of HPB

PLINK software (http://pngu.mgh.harvard.edu/purcell/*PLINK*, (Purcell *et al.,* 2007) was utilized to impute haplotypes based on single SNP tests for each of the 29 bovine autosomes of 492 Nguni animals. Variants were pruned for LD using an independent pairwise parameters of window size 30, step size 5 and a r^2 threshold of 0.1. Haplotype blocks were estimated using *Haploviews* interpretation of Gabriel *et al.* (2002) for each of the 29 bovine autosomes under *PLINK*'s default block settings. Gene ontology analyses of HPB regions was performed against the *Bos taurus* reference gene list by means of the *PANTHER* databases (Mi *et al.,* 2013).

3.3.5 Generation Of CNV Calls And *CNV Filtering*

The Log Rratio, B allele frequency, G type, chromosome and position were exported from *GenomeStudio* for each animal for analyses using *PennCNV* (Wang *et al.,* 2007). *PennCNV* has outperformed a number of CNV detection packages on multiple occasions demonstrating a greater specificity and sensitivity for CNV calling and reasonably little bias (Zhang *et al.,* 2014; Castellani *et al.,* 2014). *PennCNV* utilizes a first order hidden markov model, which assumes that the hidden copy number state at each SNP is subject to the copy number state of the most preceding SNP for high resolution CNV discovery with whole genome SNP genotyping data (Wang *et al.,* 2007). The Viterbi algorithm is subsequently utilized to determine the most probable sequence of hidden states chromosome by chromosome (Wang *et al.,* 2007). A dynamic programming algorithm, the Viterbi algorithm was applied to predict the Viterbi Path which generates the most probable sequence of hidden states representing discrete copy numbers along the chromosomes (Xu *et al.,* 2011).

The *PennCNV* compile_pfb script (Wang *et al.,* 2007) was utilized to create a pfb file from the data. The detect_cnv.pl was run to detect CNVs on 29 autosomes. A number of animals (125) exhibited an absolute genomic waviness factor of greater than 0.04. GC content within 1Mb region (500K per side) surrounding each marker was calculated and utilized to create the bovine gcmodel. A second analyses including the –gcmodel option was also run for comparative purposes.

In order to minimize the rate of false positives, extensive quality control was applied by means of the filter_cnv.pl script (Wang *et al.,* 2007). Two separate filtering criterions were utilized. By means of *Golden Helix SVS* software, the median DLRS and GCWF values, were utilized to determine the upper outlier threshold set at 1.5 inter-quartile range (IQRs) from the third quartile of all DLRS and GCWF values respectively. Upper outlier thresholds of 0.318 and 0.072 for DLRS and GCWF were thus determined The second filtering was also performed utilizing more stringent standards where only those CNVs that demonstrated a standard deviation (SD) less than 0.3, B allele drift of less than 0.01 and waviness factor 65 of less than 0.04 were kept.

3.3.6 Statistical Analyses

Bioinformatic tools together with Microsoft Excel software were utilized to organize and analyse the data. A python script developed in house merged overlapping and adjacent CNVs to form CNVRs (Additional file 3.1). Pivot tables summarized data statistics.

3.3.7 Gene Ontology Analyeses

RefGene and RefLink annotations (*USCS*, downloaded on http://genome.ucsc.edu/goldenpath/gbdDescriptionsOld.html) were used to identify genes located within a 10Mb window surrounding a CNV. Norris and Whan (2008) have shown that CNVs have a demonstrated effect on surrounding genes in a number of species. The coincidence of CNVs and corresponding genes identified by the different models was visualized by means of the bioinformatics and evolutionary genomics *VENN* diagram web tool (http://bioinformatics.psb.ugent.be/webtools/Venn/) The hypothesis that genes were over or under represented in *PANTHER* pathways, biological processes, cellular components and molecular pathways was tested by means of the bonferoni correction on the *pantherdb.org*. *Bos taurus* gene ontologies were ascertained by means of the *Ensembl* and *PANTHER*.

3.4 Results And Discussion

3.4.1 SNP Quality Control

The Illumina Bovine SNP50 beadchip v2 comprising of 54,609 markers was utilized in the study (Illumina Inc., San Diego, CA). Of these 54,609, 54,060 SNP probes map to the most current UMD3.1 bovine reference genome. After genotyping, a total of 1,340 variants were removed due to missing genotypes, and a further 11,232 variants were removed due to having a minor allele frequency of less than 0.02 and an additional 1,724 variants with a call rate of less than 95% in the sampled population. In summary, 40,313 SNPs remained after applying extensive quality control (QC) pruning.

3.4.2 Population Structure Analysis

3.4.2.1 Population Structure QC

The 40,313 SNPs that remained after QC were further pruned for linkage disequilibrium (LD) using a threshold of r^2 = 0.1. LD trimming resulted in another 29,836 SNPs pruned from the dataset, resulting in a final set of 10,477 SNPs used in the downstream analysis. Of the 492 animals sampled, 230 demonstrated an identy by descend (IBD) value of greater than 0.25 with animals within the dataset and were subsequently removed. Two hundred and sixty two unrelated animals thus remained for population structure analyses. Previous research suggests Nguni cattle populations to comprise of up to 9 different eco-types (Bester *et al.,* 2001). This estimation was then used to perform for a cross validation for 10 different K values. Standard error estimates for K ranged from 0.545 for K=1 to 0.527 for K=5 (Figure 3.5).

3.4.2.2 Population Structure Statistics And *Classification*

Organization of the data according to ancestry percentages, demonstrated 5 distinct sub-population clusters (Figure 3.5). Instead of exhibiting the typical "v" shape graph which congests at the optimal K, the K graph demonstrated a "w" type of formation, with K equal to 8 (K8) following closely behind the optimal of K5. Admixture between sub-populations was evident. Sub-populations were assigned alphabetical tags. Nguni cattle have only recently been incorporated into synthetic breeding schemes, and for many years subsisted under natural selection pressures (Horsburgh *et al.,* 2013). It can thus be expected that crossing between eco-types would be evident. The observed clustering may therefore be subsequent to such crossing between ecotypes or an indication of subpopulations that diverged more recently from one another. It is however, important to note that the ecotype structure of the studied animals was unknown upon sampling of animals used in the analyses. Discriminant principle component analyses (DPCA) also demonstrated 5 clusters within the 262 Nguni animals and is presented in Figure 3.3 and Figure 3.4.

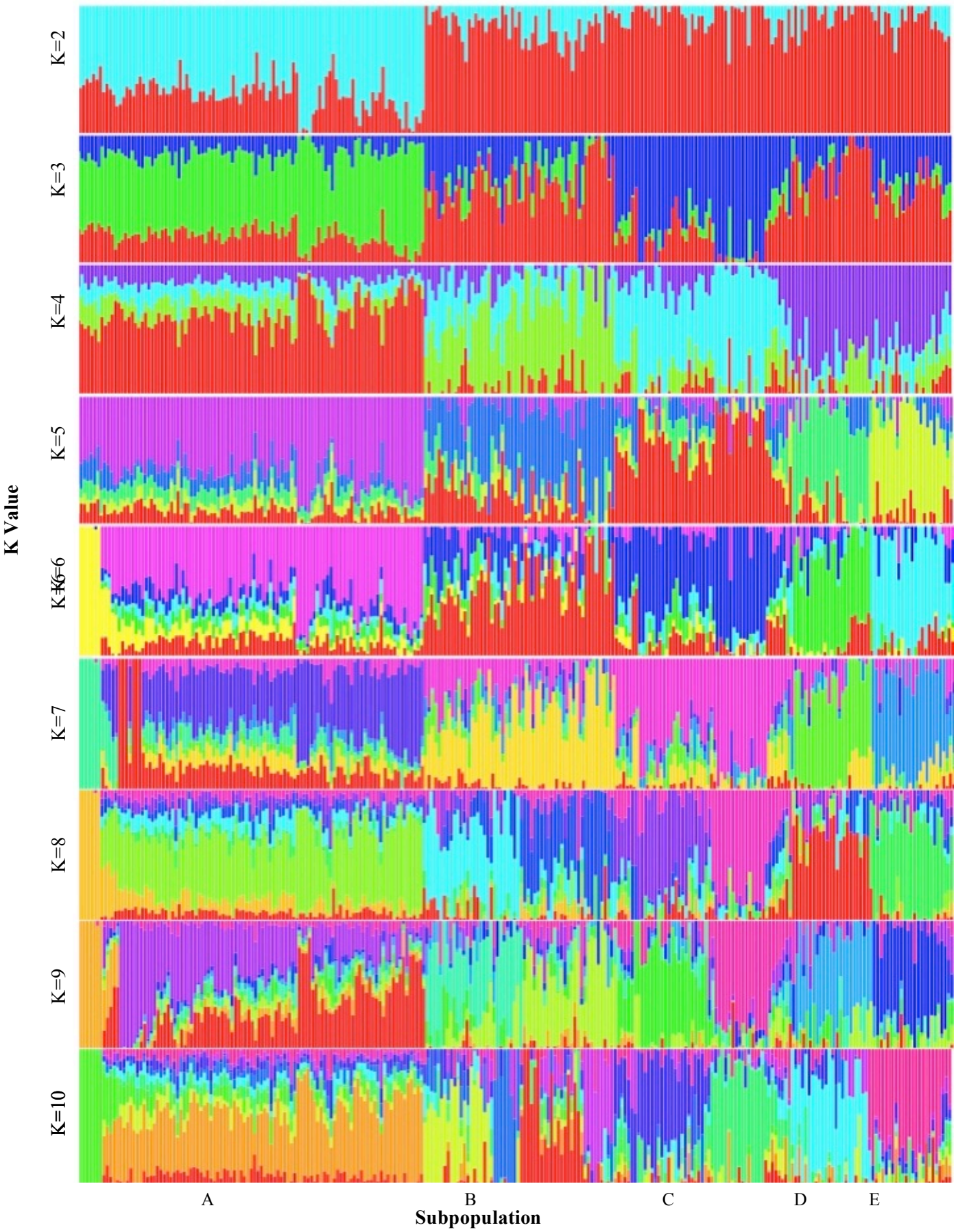

Figure 3.5 Boxplot demonstrating the population structure of the Nguni cattle for K = 2 to K =10.

3.4.3 Haploblock Analysis

3.4.3.1 Haploblock Statistics

A haplotype block is a combination of allelles that are linked on a common chromosome and inherited concurrently from a single ancestor (The International Human Hapmap Consortium 2005). Five hundred and fourty one haplotype blocks were identified across all 492 animals. Of these, 297 covered 3 or more SNPs. HPBs ranged in length from 84 base pairs on chromosome 8 to 199 730 base pairs on chromosome 1 (Table 3.1). The average length of the haplotype blocks was 79 686.68 (SD ± 67 651.42) base pairs across chromosomes with a total HPB length of 41.5Mbs. Large amounts of variation in haplotype structure and size between chromosomes were observed. Chromosome 1, 2, 3 and 8 exhibited the most haplotype blocks at 43, 33, 37 and 30 respectively (Table 3.1). Althought the largest HPB was found on chromosome 1, chromosome 10 contained the highest average HPB length of 123kbps and the also second highest percentage of its genome comprising of HPBs (Table 3.1). Previously, a negative correlation was reported between the average HPB length and recombination rate (Greenwood *et al.,* 2004), and there also exists evidence of differences in recombination rates between cattle breeds (Thomsen *et al.,* 2001).

The smallest number of haplotype blocks were identified on chromosomes 22, 27 and 28, with chromosome 22 exhibiting the smallest percentage of its length consisting of HPBs. The exact boundaries of HPBs are not resilient to variations in SNP density as the average size of HPBs may decrease with the greater sequence coverage of the HPB that results from elevated marker density (Ke *et al.,* 2004). Khatkar *et al.* (2007) reported 727 haplotype blocks covering more than 3 SNPs in 1,000 Holstein-Friesian bulls using 9 195 SNPs in Hardy-Weinberg equilibrium mapped to the Btau 3.1 bovine assembly. Haploblocks reported in this study were on average 1kb larger than those reported by Khatkar *et al.* (2007).

Table 3.1 Haplotype block chromosomal distribution and characteristics. Chromosome number (CHR), chromosome length (CHRLN), number of SNPs (SNP) and HPBs (HPB), minimum length (MinL), maximum length (MaxL), average length (AvL) and total length (HPBLN) of HPBs and percentage of chromosome covered by HPBs (PCN).

CHR	CHRLN	SNP	HPB	MinL	MaxL	AvL	HPBLN	PCN
1	161 428 367	2 637	43	2 809	199 730	10 5617.15	433 0346	2.68
2	141 965 563	1 691	33	1 641	190 889	77 656.88	2 485 053	1.75
3	126 844 711	1 716	38	108	192 734	84 265.08	3 117 846	2.46
4	123 809 850	1 358	22	3 406	198 460	99 861.67	2 097 117	1.69
5	125 249 322	1 349	21	148	191 441	80 559.43	1 691 769	1.35
6	122 519 025	1 438	21	2 660	197 610	82 146.24	1 725 092	1.41
7	113 029 157	1 328	31	1 969	199 428	88 447.73	2 653 463	2.35
8	116 846 264	1 306	30	84	197 993	95 895.72	2 781 006	2.38
9	108 503 706	1 253	23	449	173 897	61 200.77	1 346 440	1.24
10	105 982 576	1 059	22	10 923	194 738	123 553.18	2 718 192	2.56
11	109 987 751	1 071	15	9 816	191 201	96 847.67	1 452 730	1.32
12	85 119 472	2 140	25	620	194 123	75 741.70	1 817 826	2.14
13	84 213 851	1 220	16	382	196 561	57 314.19	917 043	1.09
14	81 216 349	1 121	25	108	160 297	48 024.25	1 152 607	1.42
15	84 472 747	1 015	10	6 788	176 328	73 205.13	585 651	0.69
16	77 710 258	852	22	178	188 386	74 494.59	1638 903	2.11
17	76 280 064	987	12	2 032	194 454	650 94.33	781 144	1.02
18	65 811 054	769	13	5 603	193 175	63 212.5	758 563	1.15
19	64 845 320	864	14	1 014	197 085	82 993.29	1 161 920	1.79
20	75 686 341	756	12	1 527	195 895	72 733.91	800 085	1.06
21	69 078 422	755	10	12 797	173 099	86 163.67	775 483	1.12
22	61 598 339	819	9	9 271	176 207	49 994.00	399 961	0.65
23	52 334 015	1 980	11	2 414	142 956	48 732.73	536 071	1.02
24	64 508 398	1 950	12	95	195 261	65 562.08	786 757	1.22
25	44 081 797	1 650	11	1 343	180 067	56 038.46	616 434	1.40
26	51 826 547	2 017	11	281	187 010	85 746.20	857 473	1.65
27	48 460 478	1 784	8	151	191 285	70 888.75	567 118	1.17
28	45 964 680	1 890	10	675	140 455	41 588.63	332 719	0.72
29	51 812 796	1 538	11	3 683	172 838	64 737.73	712 126	1.37
Tot		40 313	541	84	199 730	79 686.58	41 596 938	

3.4.3.2 Haploblock Gene Ontology

Haplotype blocks have discrete boundaries that are defined by recombination hotspots (Villa-Angulo *et al.,* 2009). In the past HPB analyses were primarily used to identify tag SNPs (Zhang *et al.,* 2005). In this study 232 genes were present within the 541 HPB identified (Additional file 3.2). Five genes, including *Bos taurus fat mass* and *obesity aassociated* (*FTO), family with sequence similarity 155* (*FAM155A), Glypican* (*GPC5), Na+/K+ transporting ATPase interacting 2* (*NKAIN2), UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase-like 6* (*GALNTL6)* and *cysteine conjugate-beta lyase 2* (*CCBL2*) are covered by two separate haplotype blocks lying in close proximity to each other (Additional file 3.2). We used gene ontology (GO) terms to classify these genes into a number of biological process, molecular functions and cellular components. Furthermore, we used the *PANTHER* database to identify protein features associated with GO terms (Additional file 3.3). A total of 122 genes involved in metabolic processes and

143, 226 and 188 genes involved in biological regulation and biological process and cellular processes respectively were positioned within HPB regions ascertained. Of interest were genes involved in immune system process (18), immune response (7), immune system development (9) and positive regulation of response to stimulus (17) (Additional file 3.3). Gibson *et al.* (2013), utilised exome-chip data to demonstrate patterns of linkage disequilibrium and subsequent haplotype structure to be informative of gene function and possible relationships between genes and specific phenotype clusters. Nguni cattle are suited to survive in harsh environmental conditions with enhanced disease and parasite resistance as well as heat tolerance (Mapiye *et al.,* 2009). It is therefore not surprising that genes involved in processes like immunity and stimulus responses lie within the HPBs identified.

3.4.4 CNV Identification

3.4.4.1 CNV Model Quality Control

As with all current CNV detection methodologies deducing variations in copy number from SNP data encompasses a number of areas at which error can be introduced and ascertainment biases presented (Castellani *et al.,* 2014; Redon *et al.,* 2006). The Bovine 50K Beadchip is limited to detected variations in the copy numbers of sequences present in the reference population that was used to design the probes, while it does no provide details regarding the location of duplicated copies (Alkan *et al.,* 2011). A number of factors influence the accuracy of CNV breakpoint detection, including batch effects, population stratification, experimental differences and the robustness of the statistical model (Dellinger *et al.,* 2010). SNPs utilized are also selected to have a minimum minor allele frequency and tend to be those that segregrate within multiple breeds (Clark *et al.,* 2005). The tendancy of SNP arrays to demonstrate greater sensitivity to deletions than duplications is particularly note worthy in areas with insufficient probe density to use B allele frequency measurements which may result in the majority of the smaller CNV events being deletion events partially owing to an ascertainment bias (Alkan *et al.,* 2011). With this in mind, four models utilizing different filtering stringencies were used to identify CNVs in Nguni cattle (see Materials and Methods) and are presented in Table 3.2. Four hundred and thirty three CNVs were identified by all four filtering techniques in 231 animals (Table 3.2). Discrepancies in the number of CNVs identified by each of the models was evident. Model 1 identified 353 CNVRs in the 379 animals that had a average length of 259 kb (Table 3.2). Inclusion of the gcmodel enabled additional animals to pass QC filtering and subseqeuntly corresponded with an elevated number of CNVs being identified. Great variation in the size and number of CNVRs has been reported in cattle (Hou *et al.,* 2012; Jiang *et al.,* 2013). CNVs in this study ranged from 30kb to 1Mb in size (Table 3.2). All models demonstrated a similar pattern of CNV numbers across animals, although models 3 and 4 determined a number of novel CNVs. All CNVs identified by models 1 and 2 were identified by either model 3 or 4 or by both (Figure 3.6).

Table 3.2 Summary statistics of four CNV detection filtering models. The stringencies (GCWF and DLRS), the number of animals (ANMLs), animals that passed (QCPS) and animals with CNVs present in their genome (ANMLsCNVs), the number of CNVRs and the average length (AvL) of the CNVRs identified within Nguni cattle.

MDL	GCWF	DLRS	GCMDL	ANMLs	QCPS	ANMLsCNVs	CNVRs	AvL
1	0.040	0.300	Yes	492	379	281	353	259 283.62
2	0.040	0.300	No	492	326	231	334	270 939.14
3	0.070	0.318	Yes	492	453	361	501	237 869.23
4	0.070	0.318	No	492	462	352	486	240 572.18

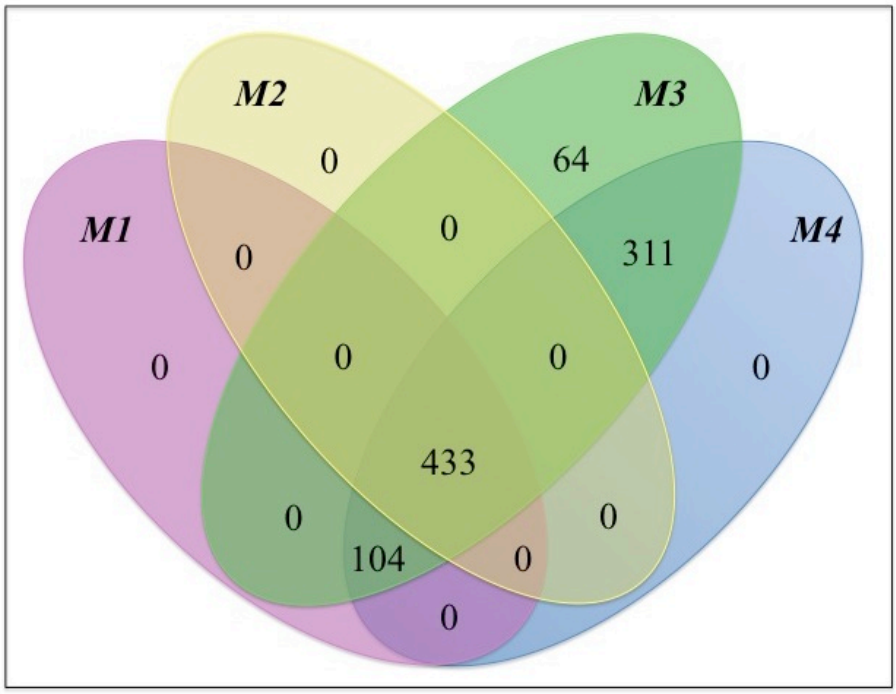

Figure 3.6 A Venn diagram showing the overlap of CNVs identified by each of the four models in 492 Nguni cattle.

3.4.4.2 CNV Statistics

Only those CNVs identified by all models were utilized for further analyses, to ensure validity of variable regions. Only 326 animals passed the *PennCNV* filtering. A total of 334 CNVRs were identified across models in 231 of these animals (Table 3.3). CNVR were between 30 kb and 1.2 Mb in length (Table 3.3). We identified 90 animals that contain a single copy number variation in their entire genome. One animal contained 22 CNVs in its genome. The average number of CNVs per animals was 2.61 (SD ± 2.63) which is similar to the 3.2 CNVs per animal reported by (Bae et al. 2010) in Korean cattle. Those animals with multiple CNVs detected in their genome, demonstrated a seemingly random spread of CNVs across chromosomes. Overall, 334 CNVRs were identified in 231 animals which was notably less than the 281 and 3088 CNVs identified by (Hou et al. 2011) and (Hou, Bickhart, et al. 2012)) respectively in a 39 and 47

animals from a variety of African breeds. The smallest CNV was 30kbs in length and demonstrated a single copy duplication (Table 3.4). Single copy deletions were identified in most of the animals while only 1 animal had a double copy duplication. This discrepancy in copy number of CNV may be an artifact of the *PennCNV* algorithm which has been seen to identify many more deletions than duplications (Eckel-Passow *et al.,* 2011). SNP array platforms tend to also demonstrate reduced precision in detecting single copy gains relative to deletions, of which this may be an artifact (Alkan *et al.,* 2011). (Jiang et al. 2013)identified 367 CNVRs comprising of 232 deletions, 111 duplications and 15 CNVRs of both gain and loss events by means of *PennCNV* analyses of high-density SNP genotyping data from 96 Chinese Holsteins. Hou *et al.* (2011) on the other hand, reported 682 CNVRs encompassing 370 loss, 216 gain and 96 loss and gain events in the same region in 521 animals representing 21 different breeds, also based on SNP genotyping arrays. Although Jiang *et al.* (2013) highlighted the differences in size and structure of populations, a difference in platforms and algorithms used and CNV discovery and filtering techniques also contributed to such incongruities. When CNVs from this study were compared to CNVs published in four other studies, very little overlap in the exact CNV breakpoints existed between studies. A number of CNVs identified in this study were however positioned in close proximity (<1Mb) to those CNVs identified by Bae *et al.* (2010), Bickhart *et al.* (2012), Fadista *et al.* (2010) and Hou *et al.* (2011) in other cattle breeds. This clustering of CNV regions demonstrated the potential for certain regions of the genome to be more susceptible to CNVs within cattle breeds. The form and exact locality of these CNVs may be what contributes to the nature and degree of variation exhibited by gene expression of adjacent genes. (Fadista et al. 2010)) reported CNV distribution in cattle to reflect chromosomal size with the most CNVs being identified on the largest chromosomes. Our data, however does not follow this pattern entirely. Chromosome 6 had the greatest number (18) of CNVs while chromosome 18 contained no CNVs (Table 3.3). This reflects findings of (Guryev *et al.,* 2008), who reported chromosome 18 to be a "cold spot for CNVs" in rats. Chromosome 18 together with chromosomes 5, 27 and 29 are reported to demonstrate a preponderance of segmental duplications in the bovine genome (Liu *et al.,* 2010). A noticeable feature of CNVs, particularly larger CNVs, is their prevalence in regions with known segmental duplications (Bickhart *et al.,* 2012). Also known as low copy repeats (LCRs), these segmental duplications are duplicated fragments of DNA that are more than 1 kb in size and can be found either on the homologous chromosome or on a separate, non-homologous chromosome with a minimum of 90% sequence identity (Sharp *et al.,* 2005). In this study we identified 15, 3, 7 and 9 CNVRs on chromosomes 5, 18, 27 and 29 respectively (Table 3.4). SNPs were reported as being sparse in regions of segmental duplications and may explain the comparatively lower numbers of CNVs on these chromosomes (Liu *et al.,* 2009). Segmentally duplicated domains are known to encode protein products that play a prominent role in species adaptation (Duda and Palumbi 1999), which makes identification of CNVs in these regions crucial. Techniques such as next generation sequencing may be more suitable for the detection of CNVs, particularly on chromosomes previously reported to harbour low number of CNVs.

Table 3.3 Summary statistics of CNV deletions and duplications. The copy number (CN), number of animals (ANMLs), number of CNVs (CNVs), minimum length (MinL), maximum length (MaxL) and average length (AvL) of CNVs.

CN	ANMLs	CNVs	MinL	MaxL	AvL
0	16	7	44 415	76 444	53 931.94
1	406	308	36 419	1 053 438	143 300.88
3	179	142	30 336	953 806	164 468.69
4	1	1	102 466	102 466	102 466.00

* double deletion (CN = 0), single deletion (CN = 1), single duplication (CN = 3) and double duplication (CN = 4)

Table 3.4 Autosomal distribution of CNVs identified in 492 Nguni cattle. CNVR count (CNVRs), total length (CNV,LN), percentage of chromosome length (PERCN) and minimum (MinL) maximum (MaxL) and average (AvL) lengths of CNVRs identified.

CHR	CNVRs	CNVLN	PERCN	MinL	MaxL	AvL
1	34	4 533 994	2.81	36 419	680 994	133 352.76
2	16	1 884 357	1.33	44 214	260 334	117 772.31
3	19	4 020 748	3.17	53 857	949 810	211 618.32
4	23	3 218 422	2.60	48 441	397 435	139 931.39
5	15	1 655 058	1.32	47 847	257 875	110 337.20
6	25	4303 075	3.51	31 128	953 806	172 123.00
7	11	1 792 440	1.59	52 476	306 135	162 949.09
8	6	794 463	0.68	76 217	237 689	132 410.50
9	11	1 230 570	1.13	30 336	289 059	111 870.00
10	7	822 052	0.78	44 415	184 185	117 436.00
11	13	1 265 163	1.15	52 654	199 903	97 320.23
12	19	2 775 332	3.26	48 596	392 714	146 070.11
13	6	1 295 356	1.54	86 589	522 669	215 892.67
14	12	2 133 059	2.63	48 512	741 197	177 754.92
15	11	1 539 814	1.82	51 632	390 973	139 983.09
16	10	1 379 434	1.78	40 032	242 142	137 943.40
17	10	2 570 441	3.37	74 327	1 285 287	257 044.10
18	3	298 969	0.45	63 682	161 641	99 656.33
19	3	415 596	0.64	106 928	182 010	138 532.00
20	11	1 615 406	2.13	49 902	378 113	146 855.09
21	9	964 270	1.40	42 434	156 070	107 141.11
22	6	1 942 282	3.15	73 778	1 171 794	323 713.67
23	4	506 937	0.97	42 345	211 284	126 734.25
24	12	1 744 861	2.70	38 738	343 135	145 405.08
25	4	1 369 746	3.11	66 262	1 041 448	342 436.50
26	11	1 958 085	3.78	73 168	518 655	178 007.73
27	7	784 830	1.62	50 958	261 955	112 118.57
28	7	1 354 237	2.95	117 087	414 660	193 462.43
29	9	1 179 028	2.28	54 840	367 944	131 003.11

3.4.4.3 Gene Ontology

Four hundred and fifty eight genes located within 10Mb of CNVRs were identified. A number of genes including *milk fat globule-EGF factor 8 protein* (*MFGE8*), *collagen type XIII alpha 1* (*COL13A1*), *cystic fibrosis transmembrane conductance regulator* (*CFTR*), *bradykinin receptor B1* (*BDKRB1*), *prostaglandin-*

endoperoxide synthase 2 (*PTGS2*), *major histocompatibility complex, class I-related* (*MR1*), *platelet/endothelial cell adhesion molecule 1* (*PECAM1*) and *leucine rich repeat* and *fibronectin type III domain containing 5* (*LRFN5*) involved in immune system response or B-cell mediated immunity were overrepresented within identified CNVs (Additional file 3.3). Copy number variations in immune related genes have previously been linked to disease (Fadista *et al.,* 2010). Variation in the genes comprising the major histocompatibility complex have been reported to play a pivotal role in the predisposition of cattle to diseases such as dermatophilosis, mastitis and tick born infections (Ibeagha-Awemu *et al.,* 2008). Stothard *et al*. (2011) reported CNVs that are closely associated with immune and lactation genes. Bickhart *et al.* (2012) reported that 15 of the 25 most variable copy number genes they identified, had functions associated with immune response and host defense, such as defensin, interferon and GIMAP (GTPase and IMAP) families. Anhidrotic ectodermal dysplasia in cattle is associated with a deletion that may range between 2–160kb of the genome and includes the third exon of the EDA gene (Drögemüller *et al.,* 2001). Flisikowski *et al.* (2010) demonstrated a 110kb microdeletions in the *MER1 repeat containing imprinted transcript 1* (*MIMT1*) gene region to be linked to the incidence of abortions and stillbirths in cattle. A 2.8 kb deletion in the *SLC4A2* gene was reported by (Meyers *et al.,* 2010) to cause osteopetrosis in Red Angus cattle. Two causal deletions in the *CLDN-16* gene were linked to renal tubular dysplasia in Japanese black cattle (Hirano *et al.,* 2000).

Sixteen CNVRs were detected in 8 or more animals in this study (Table 3.5 and Additional file 3.4). These CNVRs contained a number of genes involved in immune system processes, cell communication, response to toxic substances and cell communication. The CNVR on chromosome 1 located between base pair 104 798 012 and 105 264 358 observed in multiple animals contained the *sucrase-isomaltase* (*SI*), intestine-specific gene (Additional file 3.4). Nguni cattle are reported to exhibit a superior feed conversion rate when compared to other indigenous breeds (Schoeman 1988).

CNVs have potential to not only change gene dosage and structure, but may modify gene regulation as well as expose recessive alleles (Zhang *et al.,* 2009). A total of 458 genes were located adjacent to (within 10Mb), or within an identified CNV. Comparison of those genes contained within CNVRs identified within this study with those identified within other breeds (Bae *et al.,* 2010; Bickhart *et al.,* 2012; Hou *et al.,* 2011) revealed 402 (87%) genes that were unique to the Nguni (Table 3.6). The only gene identified close to a CNVR in all four studies was *immunoglobulin lambda-like polypeptide 1* (*IGLL1*). *IGLL1* is one of the polypeptides of the immunoglobulin light chain gene pool in domestic cattle that play a role in B cell production (Ekman *et al.,* 2009). This gene lies adjacent to its associated solute carrier (SLC) polypeptide (Ekman *et al.,* 2009). Immunoglobulins are the molecular mediators of the adapative humoral response of jawed vertebrates (Gnathostomata). The evident variation in copy number at this gene in a number of bovine breeds may explain the variation in the adaptive immunity evident between breeds, but further investigations into the role of this CNV needs to be ascertained. The *Bos taurus pregnancy-associated glycoprotein* (MGC157405) gene is the only gene represented across CNVRs of Hou *et al.* (2011), Bickhart *et al.* (2012)

and this study and forms part of the cellular defense response. Ten genes are shared between this study and that of Hou *et al.* (2011) and Bae *et al.* (2010), including *o-fucosylpeptide 3-beta-N-acetylglucosaminyltransferase* (*LFNG*) and *ADP-ribosylation factor-like 6* (*ARL6*) that are both involved in metabolic and cellular processes. B cell mediated immunity, mesoderm development and cell communication pathways also demonstrate representation by genes shared (Additional file 3.3). Twenty nine genes located within close proximity of the Nguni CNVRs were also reported to be associated with CNVRs in Korean cattle (Bae *et al.,* 2010) (Additional file 3.5). Overlapping genes were associated with a number of biological processes including positive regulation of cell proliferation, cell communication, detection of stimulus, cellular process, metabolic process and susceptibility to natural killer cell mediated cytotxicity (Additional file 3.5). Thirteen of the genes associated with CNVRs in this study overlap with genes covered by CNVRs reported by Hou *et al.* (2011) in variety of cattle breeds, including African Breeds. The funtional annotation of these 13 genes were associated with immune system processes, cell communication and lipid metabolic processes (Additional file 3.5).

Table 3.5 Copy numbers (CN) and genes (GEN) of CNVRs present in 5 or more Nguni animals.

CNVR	CN*				GEN
	0	1	3	4	
chr17:73713062-74998349		8	13		*CHCHD10 IGLL1 LOC527441 SLC5A1 VPREB3 ZNF280A ZNF280B ZNF70 DERL3 GSTT1 GSTT3 GSTT4 MIF SLC2A11 SMARCB1 DDT GGT1 GGT5 SUSD2 C17H22orf13 LOC531152 MIR2323 RTDR1 SNRPD3 SPECC1L UPB1*
chr1:104798012-105264358		16	1		*SI*
chr24:28154039-28497174		13	3		*CDH2*
chr7:75305297-75370366	1	8	5		*GABRG2*
chr5:3260057-3434356		7	4	2	*ATXN7L3B*
chr6:43037439-43089739	12	1			*GBA3*
chr19:49657396-49784054		10	2		*LYAR NSG1 OTOP1 STX18 TMEM128 WDR1 ZBTB49*
chr6:108998175-109951981		5	7		-
chr9:3651455-4439872		10	1	1	*PECAM1 POLG2*
chr1:32509969-32781614	1	7	3		-
chr6:71910076-72118486			11		*CHIC2*
chr28:21101833-21762976		5	5		*CTNNA3*
chr22:59487979-60960603		8	1		-
chr6:53514737-53692295		9			*ACAD9 C22H3orf37 CNBP COPG1 EFCC1 GATA2 ISY1 MIR2374 RAB7A RPN1 EFCC1 IQSEC1 ISY1 CHCHD4 HDAC11 NUP210 TMEM43 WNT7A XPC*
chr14:54875898-55141942			8		*ANGPT1*
chr25:41191025-42687812		5	3		*BRAT1 CARD11 GNA12 GRIFIN LFNG MIR2390 MIR2890*

* double deletion (CN = 0), single deletion (CN = 1), single duplication (CN = 3) and double duplication (CN = 4)

Table 3.6 CNVRs identified in this study and by Bae *et al.* (2010), Hou *et al.* (2011) and Bickhart *et al.* (2012), the number of animals in the study (ANMLs), the breed (BRD), the number of genes (GEN Num), CNVRs identified by other studies (SHRD) and number of genes that are unique to the study (GEN).

REF	ANMLs	BRD	CNVR	GEN Num	SHRD	GEN
Wang	326	NG	217	458	0	402
Bae	265	KOR, TAU, COMP	570	704	15	533
Hou	521	IND, AFR, OUTGR	667	491	15	291
Bickhart	6	NEL, HOL, HER,ANG	1 344	388	0	315

Five of the genes identified within CNVs in this study were also identified by Bae *et al.* (2010) in 265 Korean cattle (Additional file 3.5) while another 5 corresponded to findings of (Hou et al. 2011) in multiple different Indicine, Taurine, composite and African breeds. Bickhart *et al.* (2012) speculated that the distinctions in selected breeds for specific traits could be linked to specific CNVs and that discrepancies in CNVs and subsequent CNVRs between different breeds could thus be expected. The greatest amount of gene overlap was between this study and that by Hou *et al.* (2011). This corresponds with the proposition of CNVs segregating within breeds as they analysed the greatest variety of cattle breeds (366 Taurine, 46 composite, 70 Indicine and 39 african cattle) within their study.

Additional file 3.6 demonstrates biological processes and cellular components that were represented by genes covered within CNVRs or lying within close proximity of CNVRs identified by all four models. No molecular functions demonstrated significant over-representation by CNVR genes. The biological processes with the greatest number of genes represented included biological process, primary metabolic process, cellular metabolic process, primary to stimulus and cellular process. Nervous system development ($p = 0.008$) and single-organizm behaviour biological pathways ($p=0.003$) and dendrite cellular component ($p = 0.05$) demonstrate significant ($p \leq 0.05$) overrepresentation. Genes involved in these processors were evident in CNVRs identified in all ecotypes. Hansen (2009) denoted metabolic regulatory ability that results in a reduction in body temperature to be one of the factors that contribute to superior thermotolerance within cattle species. Whether the presence of CNVs at these genes may relate to the enhanced ability of Nguni cattle to handle harsh environmental conditions needs further investigation. Non-significant overrepresentation by CNV genes in 3055 biological processes, 593 molecular functions and 391 cellular components was evident. These systems included cellular response to transforming growth factor beta stimulus, regulation of B cell proliferation, positive regulation of viral release from host cell functions.

Previous findings have demonstrated CNVRs to be located in areas containing genes associated with environmental responses like sensory, defense and immunological functions and regulatory processors (Hou *et al.,* 2012; Seroussi *et al.,* 2010). Similar patterns are evident within Nguni cattle and suggest CNVs to potentially play an important role in the adaptative traits evident in Nguni cattle populations.

3.4.4.4 CNVs And *Population Structure*

CNV characteristics for each subpopulation are presented in Table 3.7. Sub-population A had the highest average number of CNVs per animal while sub-population D had the smallest average CNV length. Sub-population A had the greatest number of animals in the study (n=103) and also presented with the most CNVRs (n=121) (Table 3.7). A number of CNVRs were shared between populations. The most widespread CNVR was identified on chromosome 6, covering the *protocadherin 7* (*PCDH7*) and *cysteine-rich hydrophobic domain 2* (*CHIC2*) genes and present in sub-populations A, B, C and E (Table 3.8). Increasing evidence has suggested that CNVs play a primary role in interindividual diversity (Sebat *et al.,* 2004), attributing to both normal phenotypic variation and major variations in complex traits such as susceptibility to disease (Feuk *et al.,* 2006; Freeman *et al.,* 2006). Within Nguni cattle sub-populations a broad array of phenotypes are evident with great variations in coat colour, behaviour and immune response being evident (Bester *et al.,* 2001). As little research into the genotypic makeup of the Nguni ecotypes has been performed, little is known about what differentiates these ecotypes on a genomic scale. Eighteen CNVRs were identified in multiple animals and are reported in Additional file 3.4. On closer inspection of these CNVRs, some noteworthy association can be seen. The CNVR located on chromosome 1 (chr1:104798012-105264358) was identified in 7 animals. Four of the animals belong to sub-population A while 10 of the 11 animal genomes containing the CNVR on chromosome 4 (chr4:108834886-109130345) belonged to sub-population A. CNVR chr6:71910076-72118486 was present in 13 animals with 6 and 5 animals from sub-populations A and C respectively.

Table 3.7 Summary statistics of CNVs identified in five Nguni cattle subpopulations. The number of animals (ANMLs), animals with CNVs (ANMLsCNVs), CNVRs (CNVRs), the average number of CNVRs per animal (Av/An) the minimum (MinL), average (AvL) and maximum (MaxL) lengths of CNVs and the number of genes (GEN).

Pop	ANMLS	ANMLs CNVs	CNVRs	Av/An	MinL	MaxL	AvL	GEN
A	103	62	121	1.71	42 164	1 066 850	171 789.26	39
B	57	27	39	0.98	62 327	741 252	186 667.09	5
C	53	26	39	1.26	50 170	518 655	167 637.18	65
D	23	6	8	0.39	82 202	180 684	146 892.13	50
E	25	12	20	1.44	42 164	1066 850	223 319.41	195
Total	261	133	268	1.32	42 164	1 066 850	178 994.23	339

Two hundred and eighty eight genes were identified to be associated with CNVRs in sub-populations A, B, C, D and E (Table 3.8). A number of genes only identified within specific sub-populations were present (Table 3.8). Sub-population A has the most (149) unique genes that are not recorded in the other sub-population groups. The *ataxin 7-like 3B* (*ATXN7L3B*) and *tumor necrosis factor, alpha-induced protein 8* (*TNFAIP8*) genes were present in CNVRs in sub-populations B, C and E and A, C and E respectively and play a role in the immune system process, and the response to stress.

Table 3.8 The number of CNVR genes (GEN Num) and the CNVR genes (GEN) of Nguni cattle subpopulations (Pop).

Pop	GEN Num	GEN
A B C E	2	*PCDH7 CHIC2*
B C E	1	*ATXN7L3B*
A C E	1	*NXNL2*
A B C	4	*TNFAIP8 CTNNA3 SI LOC780933*
C E	2	*KCND3 ATP5G3*
B E	1	*ARL6*
A E	17	*RAB40C KLHL1 CISD1 IPMK PWWP2B MRPL28 VPREB3 DECR2 TRNT1 PCDH10 ARL4C ZNF70 NME4 CHCHD10 IGLL1 TMEM8A OTOP1*
A D	2	*CLRN1 LRFN5*
B C	2	*HPS3 LOC514194*
A C	8	*BICD2 CENPP ATG2B CDH12 BDKRB1 BDKRB2 ZWINT MR1*
A B	11	*GABRG2 PDLIM1 LOC509513 DCTD NDST4 CDH2 C28H10orf35 COL13A1 PROM1 ADCY1 TMPRSS15*
E	14	*GRAP2 SERPINB8 CADPS2 HERC4 ENTHD1 KCND2 PPP1R14C FKBP5 MSX1 CTSD FARS2 HTATSF1 NUP210 SORBS2*
D	2	*ASPH FSTL5*
C	44	*NUP35 URB2 HCK INSL6 PDPN PLGRKT PECAM1 ZC3H7B GDA MMS22L C6H4orf32 RHAG CPS1 TM9SF4 POFUT1 GLYATL3 SERINC1 GBE1 TM4SF18 IL1R2 C23H6orf141 CYB5R1 WBSCR17 CDH10 PHYHIPL ATF2 CNTNAP3 ADCY8 ANKRD50 CRISP2 FAM204A MRPS31 CD274 SPAM1 CELF4 KCMF1 CRISP3 HMGXB4 CDC73 KIF3B CELF2 RAB21 LACTB2 RANGAP1*
B	28	*KATNBL1 MPPED2 C15H11orf70 FAM5C SH3BP4 HLTF C21H14orf49 TYW3 PAQR3 CHRM5 MIR1256 GJA1 RPL37A GPC5 CLN5 UBE2U OXR1 FAM98A COX7C SMAD4 ACSL1 LPHN2 TNNI3K CRYZ EMC7 PET112 DHX29 CADM1*
A	149	*TBC1D19 PTGER3 SEC62 LOC527441 NR3C2 CA8 PFKP DDT STUB1 GGT1 AMPH FBXL16 WDR24 C15H11orf96 PRKAR2B TMEM128 RPUSD1 FAF1 NPRL3 LARGE GRB10 AXIN1 LUC7L C11H2orf28 PDIA2 PROP1 MSLN PLEKHA3 NOL4 PDGFD LYAR SPECC1L RNF185 AMY2B SUSD2 QRFPR POLR3K RFC3 ARL4A ACSL6 WFIKKN1 CLN8 ACYP2 SLC22A18 GBA3 MIR2390 FUBP3 SLC5A1 SNRPD3 C25H16orf13 SELM FGGY OTX2 KCTD16 PTGS2 CARD11 C1QTNF7 ARHGDIG DDI1 HAGHL MIF NAP1L4 MTRR H2AFY2 ALX1 ERICH1 CHTF18 FGF9 WDR1 PLEKHA1 GNG13 SRSF6 RRAGC ADIG SEMA3A UPB1 FZD1 SORCS3 NARFL LUZP2 SMARCB1 C15H11orf58 HBA SELPLG BCHE ZNF703 TMEM119 HBQ1 RGS11 MGAT4C LIN7C ITFG3 LMF1 OSTN TMEM225 GSTT4 ASS1 NRG3 ALKBH3 STAB2 CTXN3 RHBDF1 PATZ1 C21H14orf2 SNRNP25 INO80D PRR5L DRG1 ZBTB49 C17H22orf13 SLC25A21 METRN FAM173A ZNF280A KCNJ3 RHOT2 ST6GAL2 PPAP2B INPP5J GSTT3 GSTT1 QTRTD1 GGT5 HTRA1 CARS SEMA3C LOC615200 SOX2 CFTR ZNF280B PHLDA2 LPHN3 LYPLAL1 HBM LSAMP NXPH2 KCNQ1 LIMK2 SLC2A11 FAM195A GRIN3A CDKN1C DRD1 AGPAT9 PIK3IP1 DERL3 SMTN LOC516108 XRCC2*

3.4.4.5 CNVs And *Haplotype Blocks*

Thirty four HPBs lay either within, across or adjacent to CNVRs identified within the Nguni cattle population (Additional file 3.7). Half of these occurances were at CNVR sites that were present in multiple individuals, with one such CNVR on chromosome 1 that was present in 17 animals (Additional file 3.7). Another HPB overlaped a CNVR associated with genes *Ly1 antibody reactive homolog* (LYAR), *neuron-specific protein family member 1* (NSG1), *otopetrin 1* (OTOP1), *syntaxin 18* (STX18), *transmembrane*

protein 128 (TMEM128), *WD repeat domain 1* (WDR1) and *zinc finger* and *BTB domain containg 49* (*ZBTB49*) was present in 12 animals. Genes present in CNVRs that overlap or share cut-off points with HPBs contributed to a number of biological, cellular and molecular pathways (Figure 3.7). Of the biological pathways, metabolic processes demonstrated the greatest gene representation. Other interesting biological pathways represented by genes covered by both HPB and CNVR were the immune system processes, biological regulation and cellular processes. Four cellular component pathways demonstrated representation. Of the molecular pathways represented, protein binding transcription factor had the greatest number of genes denoted within CNVR-HPB overlap regions. Other molecular functions of interest included receptor activity, enzyme regulator activity and catalytic activity.

CNVs have been reported to be in LD with surrounding SNPs, demonstrating conserved long-range haplotypes (de Smith *et al.,* 2008). Meiotic crossing over hotspots flanked by recombinationally inert DNA is thought to be a major contributing factor in the presence of haplotype block structures (Kauppi *et al.,* 2004). Whether the mechanisms involved in meiotic recombination crossing-over may play a role in the variations in copy number is something that could be looked into as the exact mechanisms of CNV formation is yet to be fully understood.

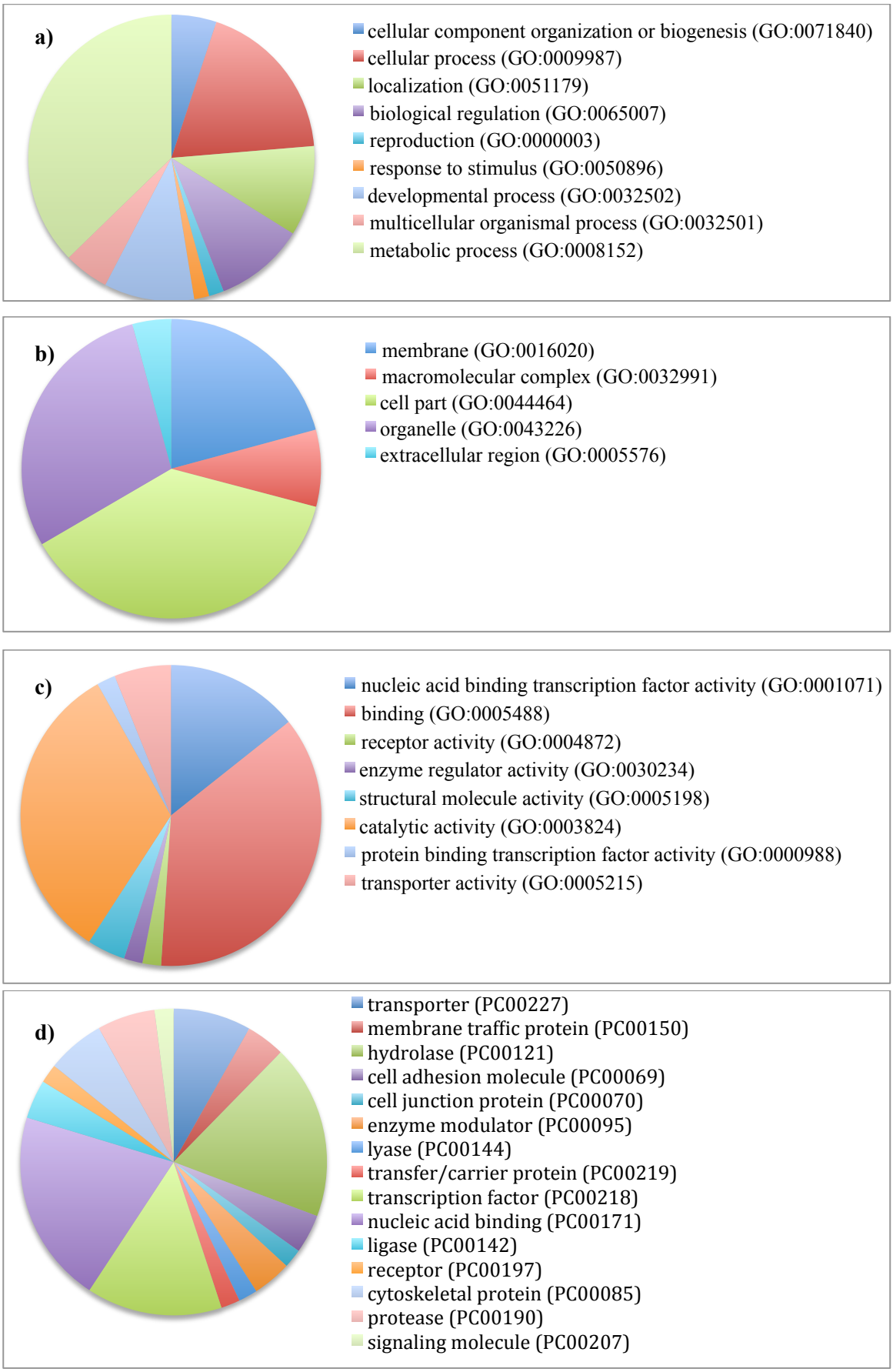

Figure 3.7 *PANTHER* pie chart of the **a)** biological pathways, **b)** cellular components **c)** molecular functions and **b)** protein classes represented within genes of CNVRs that overlap or share breakpoints with HPBs identified in 492 South African Nguni cattle.

3.5 Conclusions

Population structure analyses revealed the presence of 5 subpopulations with some degree of admixture occuring between groups. A total of 334 CNVRs were identified and characterized within the genome of 492 Nguni cattle. Different filtering techniques were modelled. The inclusion of the gcmodel with the higher waviness stringency proved to demonstrate the greatest repeatability with CNVs identifed across models.

Eighteen CNVRs were identified in multiple animals. Among these regions, segregation within as well as across sub-popuplation groups was evident. Specific CNVRs may play a role in the variation exhibited among Nguni ecotypes. Some of these CNVRs may also be distinct to Nguni cattle, contributing towards some of the distinctive phenotypic traits for which they are recognized. Until the twentieth century, Nguni cattle were primarily exposed to natural selection pressures and subsequently exhibit enhanced adaptive traits together with broad phenotypic diversity. Genes within CNVs demonstrated overrepresentation in a number of biological, molecular and cellular pathways and may therefore be potential contributors to the phenotypic diversity evident in Nguni cattle populations.

3.6 References

Alexander, D.H., J. Novembre, and K. Lange. 2009. “Fast Model-Based Estimation of Ancestry in Unrelated Individuals.” *Genome Research* 19: 1655–64. doi:10.1101/gr.094052.109.

Alkan, C., B.P. Coe, and E.E. Eichler. 2011. “Genome Structural Variation Discovery and Genotyping.” *Nature Reviews. Genetics* 12: 363–76. doi:10.1038/nrg2958.

Bae, J.S., H.S. Cheong, L.H. Kim, S. NamGung, T.J. Park, J. Chun, J.Y. Kim, C.F.A. Pasaje, J.S. Lee, and H.D. Shin. 2010. “Identification of Copy Number Variations and Common Deletion Polymorphisms in Cattle.” *BMC Genomics* 11: 232. doi:10.1186/1471-2164-11-232.

Bester, J., I.E. Matjuda, J.M. Rust, H.J. Fourie. 2001. The Nguni: case study. In: *FAO Community-based management of animal genetic resources.* Rome: UNDP, GTZ, CTA, FAO. pp. 45-68.

Bickhart, D. M., J.L. Hutchison, L. Xu, R.D. Schnabel, J.F. Taylor, J.M. Reecy, S. Schroeder, C.P. Van Tassell, T.S. Sonstegard, and G.E. Liu. 2015. “RAPTR-SV : A Hybrid Method for the Detection of Structural Variants.” *Bioinformatics*, 1–7.

Bickhart, D.M., Y. Hou, S.G. Schroeder, C. Alkan, M.F. Cardone, L.K. Matukumalli, J. Song, et al. 2012. “Copy Number Variation of Individual Cattle Genomes Using next-Generation Sequencing.” *Genome Research* 22: 778–90. doi:10.1101/gr.133967.111.

Castellani, C. A., M.G. Melka, A. E. Wishart, M. E. O. Locke, Z. Awamleh, R. L. O’Reilly, and S. M. Singh. 2014. “Biological Relevance of CNV Calling Methods Using Familial Relatedness Including Monozygotic Twins.” *BMC Bioinformatics* 15 (1). BMC Bioinformatics: 114. doi:10.1186/1471-2105-15-114.

Clark, A.G., M.J. Hubisz, C.D. Bustamante, S.H. Williamson, and R. Nielsen. 2005. “Ascertainment Bias in Studies of Human Genome-Wide Polymorphism.” *Genome Research* 15 (11): 1496–1502. doi:10.1101/gr.4107905.

de Smith, A.J., R.G. Walters, L.J.M. Coin, I. Steinfeld, Z. Yakhini, R. Sladek, P. Froguel, and A.I.F. Blakemore. 2008. "Small Deletion Variants Have Stable Breakpoints Commonly Associated with Alu Elements." *PLoS ONE* 3 (8). doi:10.1371/journal.pone.0003104.

Dellinger, A.E., S.M. Saw, L.K. Goh, M. Seielstad, T.L. Young, and Y.J. Li. 2010. "Comparative Analyses of Seven Algorithms for Copy Number Variant Identification from Single Nucleotide Polymorphism Arrays." *Nucleic Acids Research* 38 (9). doi:10.1093/nar/gkq040.

Drögemüller, C., O. Distl, and T. Leeb. 2001. "Partial Deletion of the Bovine ED1 Gene Causes Anhidrotic Ectodermal Dysplasia in Cattle." *Genome Research* 11: 1699–1705. doi:10.1101/gr.182501.

Duda, T F, and S R Palumbi. 1999. "Molecular Genetics of Ecological Diversification: Duplication and Rapid Evolution of Toxin Genes of the Venomous Gastropod Conus." *Proceedings of the National Academy of Sciences of the United States of America* 96: 6820–23. doi:10.1073/pnas.96.12.6820.

Eckel-Passow, J.E., E.J. Atkinson, S. Maharjan, S.L.R. Kardia, and M. de Andrade. 2011. "Software Comparison for Evaluating Genomic Copy Number Variation for Affymetrix 6.0 SNP Array Platform." *BMC Bioinformatics* 12 (1). BioMed Central Ltd: 220. doi:10.1186/1471-2105-12-220.

Ekman, A., M. Niku, J. Liljavirta, and A. Iivanainen. 2009. "Bos Taurus Genome Sequence Reveals the Assortment of Immunoglobulin and Surrogate Light Chain Genes in Domestic Cattle." *BMC Immunology* 10 (1): 22. doi:10.1186/1471-2172-10-22.

Fadista, J., B. Thomsen, L. Holm, and C. Bendixen. 2010. "Copy Number Variation in the Bovine Genome." *BMC Genomics* 11: 284. doi:10.1186/1471-2164-11-284.

Feuk, L., A.R. Carson, and S.W. Scherer. 2006. "Structural Variation in the Human Genome." *Nature Reviews. Genetics* 7: 85–97. doi:10.1038/nrg1767.

Flisikowski, K., H. Venhoranta, J. Nowacka-Woszuk, S.D. Mckay, A. Flyckt, J. Taponen, R. Schnabel, et al. 2010. "A Novel Mutation in the Maternally Imprinted PEG3 Domain Results in a Loss of MIMT1 Expression and Causes Abortions and Stillbirths in Cattle (Bos Taurus)." *PLoS ONE* 5 (11): 1–9. doi:10.1371/journal.pone.0015116.

Freeman, J.L., G.H. Perry, L. Feuk, R. Redon, S.A. McCarroll, D.M. Altshuler, H. Aburatani, et al. 2006. "Copy Number Variation: New Insights in Genome Diversity." *Genome Research* 16 (617): 949–61. doi:10.1101/gr.3677206.

Gabriel, S.B., S.F. Schaffner, H. Nguyen, J.M. Moore, J. Roy, B. Blumenstiel, J. Higgins, et al. 2002. "The Structure of Haplotype Blocks in the Human Genome." *Science (New York, N.Y.)* 296 (5576): 2225–29. doi:10.1126/science.1069424.

Gautier, M., T. Faraut, K. Moazami-Goudarzi, V. Navratil, M. Foglio, C. Grohs, A. Boland, et al. 2007. "Genetic and Haplotypic Structure in 14 European and African Cattle Breeds." *Genetics* 177 (2): 1059–70. doi:10.1534/genetics.107.075804.

Gibson, J., W. Tapper, S. Ennis, and A. Collins. 2013. "Exome-Based Linkage Disequilibrium Maps of Individual Genes: Functional Clustering and Relationship to Disease." *Human Genetics* 132: 233–43. doi:10.1007/s00439-012-1243-6.

Green, M.R., and J. Sambrook. 2012. "Isolation and Quantification of DNA." In *Molecular Clonging: A*

Laboratory Manual, edited by J. Inglis, A. Boyle, and A. Gann. Vol. 33. Cold Spring Harbor Laboratory Press. doi:10.3724/SP.J.1141.2012.01075.

Greenwood, T.A., B.K. Rana, and N.J. Schork. 2004. "Human Haplotype Block Sizes Are Negatively Correlated with Recombination Rates." *Genome Research* 14: 1358–61. doi:10.1101/gr.1540404.

Guryev, V., K. Saar, T. Adamovic, M. Verheul, S.A.A.C. van Heesch, S. Cook, M. Pravenec, et al. 2008. "Distribution and Functional Impact of DNA Copy Number Variation in the Rat." *Nature Genetics* 40 (5): 538–45. doi:10.1038/ng.141.

Hansen, P.J. 2004. "Physiological and Cellular Adaptations of Zebu Cattle to Thermal Stress." *Animal Reproduction Science* 82–83: 349–60. doi:10.1016/j.anireprosci.2004.04.011.

Hirano, T., N. Kobayashi, T. Itoh, A. Takasuga, T. Nakamaru, S. Hirotsune, and Y. Sugimoto. 2000. "Null Mutation of PCLN-1/Claudin-16 Results in Bovine Chronic Interstitial Nephritis." *Genome Research* 10: 659–63. doi:10.1101/gr.10.5.659.

Horsburgh, K.A., S. Prost, A. Gosling, J. Stanton, C. Rand, and E.A. Matisoo-Smith. 2013. "The Genetic Diversity of the Nguni Breed of African Cattle (Bos Spp.): Complete Mitochondrial Genomes of Haplogroup T1." *PloS One* 8 (8): e71956. doi:10.1371/journal.pone.0071956.

Hou, Y., D.M. Bickhart, M.L. Hvinden, C. Li, J. Song, D.A. Boichard, S. Fritz, et al. 2012. "Fine Mapping of Copy Number Variations on Two Cattle Genome Assemblies Using High Density SNP Array." *BMC Genomics* 13: 376. doi:10.1186/1471-2164-13-376.

Hou, Y., G.E. Liu, D.M. Bickhart, M.F. Cardone, K. Wang, E. Kim, L.K. Matukumalli, et al. 2011. "Genomic Characteristics of Cattle Copy Number Variations." *BMC Genomics* 12: 127. doi:10.1186/1471-2164-12-127.

Hou, Y., G.E. Liu, D.M. Bickhart, L.K. Matukumalli, C. Li, J. Song, L.C. Gasbarre, C.P. Van Tassell, and T.S. Sonstegard. 2012. "Genomic Regions Showing Copy Number Variations Associate with Resistance or Susceptibility to Gastrointestinal Nematodes in Angus Cattle." *Functional and Integrative Genomics* 12: 81–92. doi:10.1007/s10142-011-0252-1.

Ibeagha-Awemu, Eveline M., Patrick Kgwatalala, Aloysius E. Ibeagha, and Xin Zhao. 2008. "A Critical Analysis of Disease-Associated DNA Polymorphisms in the Genes of Cattle, Goat, Sheep, and Pig." *Mammalian Genome* 19: 226–45. doi:10.1007/s00335-008-9101-5.

Jiang, L., J. Jiang, J. Yang, X. Liu, J. Wang, H. Wang, X. Ding, J. Liu, and Q. Zhang. 2013. "Genome-Wide Detection of Copy Number Variations Using High-Density SNP Genotyping Platforms in Holsteins." *BMC Genomics* 14 (1). BMC Genomics: 131. doi:10.1186/1471-2164-14-131.

Kauppi, L., A.J. Jeffreys, and S. Keeney. 2004. "Where the Crossovers Are: Recombination Distributions in Mammals." *Nature Reviews. Genetics* 5 (June): 413–24. doi:10.1038/nrg1346.

Ke, X., S. Hunt, W. Tapper, R. Lawrence, G. Stavrides, J. Ghori, P. Whittaker, et al. 2004. "The Impact of SNP Density on Fine-Scale Patterns of Linkage Disequilibrium." *Human Molecular Genetics* 13 (6): 577–88. doi:10.1093/hmg/ddh060.

Khatkar, M.S., K.R. Zenger, M. Hobbs, R.J. Hawken, J.A.L. Cavanagh, W. Barris, A.E. McClintock, et al. 2007. "A Primary Assembly of a Bovine Haplotype Block Map Based on a 15,036-Single-Nucleotide

Polymorphism Panel Genotyped in Holstein-Friesian Cattle." *Genetics* 176: 763–72. doi:10.1534/genetics.106.069369.

Kijas, J.W., W. Barendse, W. Barris, B. Harrison, R. McCulloch, S. McWilliam, and V. Whan. 2011. "Analysis of Copy Number Variants in the Cattle Genome." *Gene* 482. Elsevier B.V.: 73–77. doi:10.1016/j.gene.2011.04.011.

Liu, G.E., and D.M. Bickhart. 2012. "Copy Number Variation in the Cattle Genome." *Functional and Integrative Genomics* 12: 609–24. doi:10.1007/s10142-012-0289-9.

Liu, G.E., Y. Hou, B. Zhu, M.F. Cardone, L. Jiang, A. Cellamare, A. Mitra, et al. 2010. "Analysis of Copy Number Variations among Diverse Cattle Breeds." *Genome Research* 20: 693–703. doi:10.1101/gr.105403.110.

Liu, G.E., M. Ventura, A. Cellamare, L. Chen, Z. Cheng, B. Zhu, C. Li, J. Song, and E.E. Eichler. 2009. "Analysis of Recent Segmental Duplications in the Bovine Genome." *BMC Genomics* 10: 571. doi:10.1186/1471-2164-10-571.

Makina, S.O., F.C. Muchadeyi, E. van Marle-KÃster, M.D. MacNeil, and A. Maiwashe. 2014. "Genetic Diversity and Population Structure among Six Cattle Breeds in South Africa Using a Whole Genome SNP Panel." *Frontiers in Genetics* 5: 1–7. doi:10.3389/fgene.2014.00333.

Mapiye, C., M. Chimonyo, K. Dzama, J. G. Raats, and M. Mapekula. 2009. "Opportunities for Improving Nguni Cattle Production in the Smallholder Farming Systems of South Africa." *Livestock Science* 124 (1–3). Elsevier B.V.: 196–204. doi:10.1016/j.livsci.2009.01.013.

Marufu, M.C., L. Qokweni, M. Chimonyo, and K. Dzama. 2011. "Relationships between Tick Counts and Coat Characteristics in Nguni and Bonsmara Cattle Reared on Semiarid Rangelands in South Africa." *Ticks and Tick-Borne Diseases* 2 (3). Elsevier GmbH.: 172–77. doi:10.1016/j.ttbdis.2011.07.001.

McKay, S.D., R.D. Schnabel, B.M. Murdoch, L.K. Matukumalli, J. Aerts, W. Coppieters, D. Crews, et al. 2007. "Whole Genome Linkage Disequilibrium Maps in Cattle." *BMC Genetics* 8: 74. doi:10.1186/1471-2156-8-74.

Meyers, S.N., T.G. McDaneld, S.L. Swist, B.M. Marron, D.J. Steffen, D. O'Toole, J.R. O'Connell, J.E. Beever, T.S. Sonstegard, and T.P.L. Smith. 2010. "A Deletion Mutation in Bovine SLC4A2 Is Associated with Osteopetrosis in Red Angus Cattle." *BMC Genomics* 11: 337. doi:10.1186/1471-2164-11-337.

Mokry, F., M. Buzanskas, M. de Alvarenga Mudadu, D. do Amaral Grossi, R. Higa, R. Ventura, A. de Lima, et al. 2014. "Linkage Disequilibrium and Haplotype Block Structure in a Composite Beef Cattle Breed." *BMC Genomics* 15 (Suppl 7). BioMed Central Ltd: S6. doi:10.1186/1471-2164-15-S7-S6.

Norris, B.J., and V.A. Whan. 2008. "A Gene Duplication Affecting Expression of the Ovine ASIP Gene Is Responsible for White and Black Sheep." *Genome Research* 18: 1282–93. doi:10.1101/gr.072090.107.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M.A.R. Ferreira, D. Bender, J. Maller, et al. 2007. "PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses." *American Journal of Human Genetics* 81 (3): 559–75. doi:10.1086/519795.

Rechav, Y., and M.W. Kostrzewski. 1991. "Relative Resistance of Six Cattle Breeds To the Tick Boophilus

Decoloratusin South Africa." *Onderstepoort J. Vet. Res* 58: 181–86.

Redon, R., S. Ishikawa, K.R. Fitch, L. Feuk, H. George, T.D. Andrews, H. Fiegler, et al. 2006. "Global Variation in Copy Number in the Human Genome." *Nature* 444 (7118): 444–54. doi:10.1038/nature05329.Global.

Schoeman, S.J. 1988. "Recent Research into the Production Potential of Indigenous Cattle with Special Reference to the Sanga." *South African Journal of Animal Science* 19 (2): 55–61.

Sebat, J., B. Lakshmi, J. Troge, J. Alexander, J. Young, P. Lundin, S. Maner, et al. 2004. "Large-Scale Copy Number Polymorphism in the Human Genome" 305 (July): 525–28. doi:10.1126/science.1098918.

Seroussi, E., G. Glick, A. Shirak, E. Yakobson, J.I. Weller, E. Ezra, and Y. Zeron. 2010. "Analysis of Copy Loss and Gain Variations in Holstein Cattle Autosomes Using BeadChip SNPs." *BMC Genomics* 11: 673. doi:10.1186/1471-2164-11-673.

Sharp, A. J., D. P. Locke, S. D. McGrath, Z. Cheng, J. A. Bailey, R. U. Vallente, L. M. Pertz, et al. 2005. "Segmental Duplications and Copy-Number Variation in the Human Genome." *American Journal of Human Genetics* 77: 78–88. doi:10.1086/431652.

Stothard, P., J. Choi, U. Basu, J.M. Sumner-Thomson, Y. Meng, X. Liao, and Stephen S Moore. 2011. "Whole Genome Resequencing of Black Angus and Holstein Cattle for SNP and CNV Discovery." *BMC Genomics* 12: 559. doi:10.1186/1471-2164-12-559.

The Bovine Genome sequencing and analysis consortium, C.G. Elsik, R.L. Tellam, and K.C. Worley. 2009. "The Genome Sequence of Taurine Cattle: A Window to Ruminant Biology and Evolution." *Science* 324: 522–28. doi:10.1126/science.1169588.

The International Human Hapmap Consortium. 2005. "A Haplotype Map of the Human Genome." *Nature* 437: 1299–1320. doi:10.1038/nature04226.

Thomsen, H., N. Reinsch, N. Xu, J. Bennewitz, C. Looft, S. Grupe, C. Kühn, et al. 2001. "A Whole Genome Scan for Differences in Recombination Rates among Three Bos Taurus Breeds." *Mammalian Genome* 12: 724–28. doi:10.1007/s00335-001-2068-0.

Thornton, T., M.P. Conomos, S. Sverdlov, E.M. Blue, C.Y. Cheung, C.G. Glazner, S.M. Lewis, and E.M. Wijsman. 2014. "Estimating and Adjusting for Ancestry Admixture in Statistical Methods for Relatedness Inference, Heritability Estimation, and Association Testing." *BMC Proceedings* 8 (Suppl 1). BioMed Central Ltd: S5. doi:10.1186/1753-6561-8-S1-S5.

Tuzun, E., A.J. Sharp, J.A. Bailey, R. Kaul, V.A. Morrison, L.M. Pertz, E. Haugen, et al. 2005. "Fine-Scale Structural Variation of the Human Genome." *Nature Genetics* 37 (7): 727–32. doi:10.1038/ng1562.

Villa-Angulo, R., L.K. Matukumalli, C.A. Gill, J. Choi, C.P. Van Tassell, and J.J. Grefenstette. 2009. "High-Resolution Haplotype Block Structure in the Cattle Genome." *BMC Genetics* 10: 19. doi:10.1186/1471-2156-10-19.

Wang, K., M. Li, D. Hadley, R. Liu, J. Glessner, S.F.A. Grant, H. Hakonarson, and M. Bucan. 2007. "PennCNV: An Integrated Hidden Markov Model Designed for High-Resolution Copy Number Variation Detection in Whole-Genome SNP Genotyping Data." *Genome Research* 17: 1665–74. doi:10.1101/gr.6861907.

Xu, Y., B. Peng, Y. Fu, and C. I. Amos. 2011. "Genome-Wide Algorithm for Detecting CNV Associations with Diseases." *BMC Bioinformatics* 12: 331. doi:10.1186/1471-2105-12-331.

Zhang, F., W. Gu, M.E. Hurles, and J.R. Lupski. 2009. "Copy Number Variation in Human Health, Disease, and Evolution." *Annual Review of Genomics and Human Genetics* 10: 451–81. doi:10.1146/annurev.genom.9.081307.164217.

Zhang, K., Z. Qin, T. Chen, J.S. Liu, M.S. Waterman, and F. Sun. 2005. "HapBlock: Haplotype Block Partitioning and Tag SNP Selection Software Using a Set of Dynamic Programming Algorithms." *Bioinformatics* 21 (1): 131–34. doi:10.1093/bioinformatics/bth482.

Zhang, X., R. Du, S. Li, F. Zhang, L. Jin, and H. Wang. 2014. "Evaluation of Copy Number Variation Detection for a SNP Array Platform." *BMC Bioinformatics* 15 (1). BMC Bioinformatics: 50. doi:10.1186/1471-2105-15-50.

Zhao, M., Q.Q. Wang, P. Jia, and Z. Zhao. 2013. "Computational Tools for Copy Number Variation (CNV) Detection Using next-Generation Sequencing Data: Features and Perspectives - Springer." *BMC Bioinformatics* 14 (Suppl 11). BioMed Central Ltd: S1. doi:10.1186/1471-2105-14-S11-S1.

Genetic Diversity Of South African Cattle Inferred Using Copy Number Variations

Magretha Diane Wang[1,2], Kennedy Dzama [1], Farai Catherine Muchadeyi[2]

[1]Department of Animal Sciences, University of Stellenbosch, Matieland, South Africa

[2]Biotechnology Platform, Agricultural Research Council, Private Bag X5, Onderstepoort 0110, South Africa

Prepared for publication in peer reviewed journal

Chapter 4: Genetic Diversity Of South African Cattle Inferred Using Copy Number Variations

4.1 Abstract

Copy number variations (CNVs) comprise structural variants that alter the structure of the DNA in the form of deletions, duplications and insertions larger than 1kb in size. CNVs are thought to be primary role-players in breed formation and adaptation. South Africa boasts a diverse ecology with harsh environmental conditions and a broad spectrum of parasites and diseases that pose challenges to livestock production. Composite cattle breeds have also been developed to utilize the hardiness of indigenous Sanga populations and the production potential of the exotic Taurine breeds. The prevalence of CNVs within these breeds of cattle is however not understood. Illumina Bovine SNP50 data and *PennCNV* were utilized to identify CNVRs within the genome of 287 animals from 7 South African cattle breeds representing Sanga, Taurine, composite and cross breeds. 356 Unique CNV regions (CNVRs) of between 36kb to 4.1Mb in size were identified. The null hypothesis that one CNVR loci is independent of another was tested using the *GENEPOP* software. 102 and 7 of the CNVRs in the exotic Taurine and indigenous Sanga and composite cattle breeds demonstrated a significant ($p \leq 0.05$) association. *PANTHER* overrepresentation analyses of correlated CNVRs demonstrated significant enrichment of a number of biological processes, molecular functions, cellular components and protein classes. CNVR genetic variation between and within breed group was measured using phiPT which allows intra-individual variation to be suppressed and hence proved suitable for measuring binary CNV presence/absence data. PhiPT within breed group values were 2.510, 6.115 and 4.233 for the Sanga, Taurine and Composite breeds respectively. Among breed group genetic variation was lower for the pure breeds at 0.085 (Sanga) and 0.113 (Taurine) than for the composite breeds (3.897). Phylogenetic trees were drawn. CNVRs primarily clustered animals of the same breed type together. This study successfully identified, characterized and analyzed 356 CNVRs within 7 South African cattle breeds. CNVR correlations were evident, with many more correlations being present among the exotic Taurine breeds. CNVR genetic diversity of Sanga, Taurine and composite South African cattle breeds was ascertained with breed types exposed to similar selection pressures demonstrating analogous incidences of CNVRs.

Keywords

Genetic diversity, CNVs, population structure, South African cattle.

4.2 Background

Copy number variations are deletions, duplications and insertions greater than 1kb in size that modify the DNA structure and play a significant role in the genomic variability and hence diversity evident within and among breeds (Liu *et al.,* 2010). They have been observed to affect a greater percentage of genomic sequences relative to other forms of genomic variations like single nucleotide polymorphisms (SNPs) (Zhang *et al.,* 2009; Liu and Bickhart 2012; Hou *et al.,* 2012). SNPs and microsatellite analyses have been used to assess population structures and genetic diversity in order to gain insight into origin, history and adaptation of cattle. CNV loci have however been found within gene boundaries, with the incidence of some coinciding with breed histories and breed formation patterns (Hou *et al.,* 2011; Matukumalli *et al.,* 2009). Covering a greater number of sequences than SNPs, CNVs have been demonstrated to alter gene dosage, disturb coding sequences or sway gene regulation (Stranger *et al.,* 2007). CNVs have been proposed to play a role in genetic adaptation (Liu *et al.,* 2010). Stranger *et al.* (2007) demonstrated SNPs and CNVs to capture 83.6% and 17.7% of the observed genetic variation with very little overlap in the variation captured by the two variant types. It was thus hypothesized that ascertaining the genetic variations captured by CNVs will generate supplementary information regarding the genetic variation which may add to that already obtained from SNPs. CNVs may hence be a suitable genomic marker for ascertaining cattle origins and history as well as divergence amongst breeds.

A number of Taurine, Sanga and composite breeds are found in South Africa. While exotic Taurine breeds demonstrate improved production subsequent to the development and elevated focus of intense selection programs, indigenous Sanga breeds of South Africa are recognized for their innate ability to handle the range of harsh climatic conditions, feed and water scarcity together with a widespread array of diseases and pathogens customary to South Africa (Hoffmann, 2010; Mirkena *et al.,* 2010). Composite breeds, like the Bonsmara have been developed to merge the adaptive ability of indigenous cattle with the productive ability of the exotic breeds (Bonsma 1980). Makina *et al.* (2014) assessed the genetic variation of composite, Sanga and Taurine cattle breeds of South Africa, using genome wide SNP data. Considering the evidenced adaptation of indigenous Sanga South African breeds that has also been introgressed into composite breeds, the determination of genetic variation of CNVs in these breeds may hold further insight into understanding the multiple components of functional breed diversity and the subsequent implications thereof. This may have important inference on current breed management and genetic improvement practices.

This study therefore comprised an investigation into the diversity of six South African cattle breeds (Angus, Drakensberger, Afrikaner, Holstein, Nguni and Bonsmara) from each of 3 breed groups (Taurine, Sanga and composite) and one cross breed (Nguni X Angus) utilizing CNVRs. It was hypothesized that CNVR genetic diversity would parallel the breed history and adaptation, with greater variation evidenced among breeds

more distantly related or under different selection pressures. Illumina BovineSNP50 genotyping methodology was used in conjunction with *PennCNV* to identify CNVs and subsequent genes enriched. CNVRs were used to ascertain levels of genetic diversity and to determine the measure of pairwise correlation in CNVR presence within and among breeds.

4.3 Materials And Methods

4.3.1 Sample Collection And *Genotyping.*

287 animals comprising of two exotic *Bos taurus* (45 Holstein and 32 Angus), two South African Sanga (59 Nguni and 48 Afrikaner), two composite (46 Bonsmara and 48 Drakensberger) and one crossbreed (10 Nguni Angus) breeds were sampled from throughout South Africa. Ethics approval was obtained for the study (Ref. Nr.:2014/CAES/101). The protocol utilized for the collection of samples, DNA extraction and genotyping has been published (Makina *et al.,* 2014). Genomic DNA was extracted from blood, hair and semen utilizing methods discussed in the previous chapter. Samples were selected against full-sib and half sib animals using pedigree data such that genetic diversity represented within the dataset was maximized. Blood samples were obtained from all animals with the exception of the 45 Holstein animals for which semen samples were obtained with permission from an artificial insemination company. Genomic DNA was extracted from blood samples using the Qiagen DNeasy extraction kit as per the manufacturer's protocol. Dithiothreitol (DTT) with Proteinase K was added in the first step of Qiagen DNeasy DNA extraction protocol for the extraction of genomic DNA from the semen samples. The Qubit® 2.0 Fluorometer and the Nanodrop Spectrophotometer (Nanodrop ND-1000) quantified the DNA.

4.3.2 SNP Quality Control.

SNP quality control was performed for all animals using *PLINK* v.1.07. Those SNPs with a MAF of less than 0.02, call rate of less than 95% and missing genotype frequency of more than 10% were excluded from further analyses. 45 925 SNPs thus remained for further analyses. A *PennCNV* input file containing LogR ratio and B allele frequency data of 45 925 resultant SNPs across 287 animals was generated in *GenomeStudio* Software 2011.1 and exported for further analyze.

4.3.4 CNVs

4.3.4.1 Identification

PennCNV software utilizes a first order hidden markov model, which assumes that the hidden copy number state at each SNP is subject to the copy number state of the most preceding SNP for high resolution CNV discovery with whole genome SNP genotyping data (Wang *et al.,* 2007). *PennCNV* has outperformed a number of CNV detection packages especially with regard to specificity and sensitivity of CNV calling (Castellani *et al.,* 2014; Zhang *et al.,* 2014). Despite the normalization of raw intensities prior to LogR calculations, false call rates are still common in datasets, especially in the presence of potential batch effects (Diskin *et al.,* 2008; Zhang *et al.,* 2014). Batch effects influence individual probes differently, thus effecting the LogR ratio and subsequent CNV detection (Scharpf *et al.,* 2012). *PennCNV* was thus utilized to identify

CNVs that were simultaneously screened for waviness and the derivative log ratio using the filter *PennCNV* script. The Viterbi algorithm then determined the most probable sequence of hidden states chromosome by chromosome (Wang *et al.,* 2007). A dynamic programming algorithm, the Viterbi algorithm is applied to predict the Viterbi Path which generates the most probable sequence of hidden states representing discrete copy numbers along the chromosomes (Xu *et al.,* 2011).

The bioinformatics and evolutionary genomics *VENN* diagram web tool (http://bioinformatics.psb.ugent.be/webtools/Venn/) was utilized to create a Venn diagram demonstrating the overlap between CNVs identified in different breeds. Overlapping CNVs across and within breeds were aggregated to delineate a set of CNVRs utilizing bioinformatics approaches (Additional file 3.1) (Redon *et al.,* 2006) .

4.3.4.2 Gene Ontology And *Representation*

Genomic regions of CNVRs identified were uploaded into *UCSC* and details of the regions together with the reflink and refGene genes covered were obtained. *VENN* (http://bioinformatics.psb.ugent.be/webtools/Venn/) was utilized to construct a Venn diagram demonstrating the overlap of those genes enriched within CNVs identified across breeds. Gene ontologies were determined by means of the *PANTHER* databases (Helleday 2003). The hypothesis that genes were over or under represented in *PANTHER* pathways, biological processes, cellular components and molecular pathways was tested using the bonferoni correction at a significance level of 0.05.

4.3.4.3 CNVR Correlations And *Representation*

A pairwise association testing the null hypothesis that genotypes at one locus were independent of genotypes at the other locus was performed using *GENEPOP* (Raymond and Rousset 1995). Only those CNVR identified in 2 or more individuals were used. Contingency tables, demonstrating the relationship between all pairs of loci within and between breeds was created. A markov chain algorithm described by Raymond and Rousset (1995) computed a G-test and probability test for each table. CNVRs demonstrating a significant correlation with a p-value of less than 0.05 were uploaded onto *UCSC* to ascertain genomic region information. A *PANTHER* overrepresentation analyses using the bonferoni correction for multiple testing was performed on genes covered by correlated CNVRs to ascertain whether any molecular functions, biological processes or cellular components were significantly ($p<0.05$) overrepresented by correlated CNVRs,

4.3.5 CNVR Genetic Diversity

A CNVR dataset was created from CNVRs identified in 197 animals from 7 cattle breeds. CNVR were each treated as individual loci and only those CNVRs identified in more than 2 animals were utilized so as to reduce the rate of false positives within the dataset (Jakobsson *et al.,* 2008). Three input files were generated. The first contained individual animals with binomial presence/absence data for each of the 163 CNVR loci

that remained post pruning. The second dataset contained presence/absence data of the 163 CNVR loci for each of the 7 cattle breeds, while the third dataset contained information on the CNVR loci frequencies for each of the 7 cattle breeds.

4.3.5.1 Analysis Of Molecular Variance And *Principle Component Analysis Plot*

An analysis of molecular variance (AMOVA) was subsequently performed on the pruned data on 163 CNVR loci in 197 samples. Animals were grouped according to their breed types. The Holstein and Angus animals were grouped together in the Taurine group. The Nguni and Afrikaner animals made up the Sanga group while the Bonsmara, Drakensberger and Nguni Angus cross animals made up the composite/cross group. *GenAlex* software was utilized to perform an analysis of molecular variance on the dataset (Peakall and Smouse 2012) together with a principle component analyses. A tri-matrix of squared Euclidean distances was used to calculate the pairwise population values (PhiPT) by means of an AMOVA using 9999 permutations. PhiPT values, which are analogous to Wrights' F_{ST} indices, measure population genetic differentiation from binary data and were used to measure the genetic variation of CNVRs within and among cattle breeds for each of the breed categories. This measures allows intra-individual variation to be suppressed and hence proved suitable for measuring binary CNV presence/absence data (Nabais *et al.,* 2014). A genetic distance trimatrix was utilized to determine standardized eigenvectors for principle components 1 to 100. Eigenvalues present the amount of genetic variation contained by each respective principle component (PC). In order to determine how many PCs to contain within the model, each eigenvalue was divided by the total sum of eigenvalues in order to establish the fraction of total variance retained versus the number of eigenvalues. Kaiser's stopping rule states that only PCs demonstrating eigenvalues over 1.00 should be considered in the analysis. This comprises the most utilized method for determining the number of PCs to retain in the analyses (Peres-Neto *et al.,* 2005). Principle component 15 demonstrated an eigenvector of 1.159 and explained 96.3% variance and was thus chosen as the cutoff component (Additional file 4.1). PCA plots were drawn for PC1 to PC5.

4.3.5.2 Population Structure

STRUCTURE v2.3 was utilized to perform a model based clustering analyses of population structure as reported by Pritchard *et al.* (2000) and Falush *et al.* (2007). The model used did not assume any specific mutation process and considering the exact mutation and inheritance patterns of CNVs is not as yet fully understood (Zhang *et al.,* 2014), it was deemed suitable for CNV analyses. Multiple analyses were performed for K = 2 to K=8. The membership coefficient Q estimate matrix was plotted as a barplot using the *R* barplot function.

4.3.5.3 Genetic Distance Cluster Analyses

The *R* package *hclust* was used to compute a distance matrix from binomial CNVR present/absence data for each animal that was then used to perform a hierarchal dissimilarity cluster analysis on regions with variable copy numbers. This was performed for each of the three datasets and plotted to demonstrate clusters.

4.4 Results And Discussion

4.4.1 SNP Quality Control

The Illumina Bovine SNP50 beadchip v2 was utilized for this study (Illumina Inc., San Diego, CA). Of the 54,609 markers on the beadchip, 45,924 SNPs had a call rate and MAF of greater than 0.95 and 0.02 respectively and thus remained for further analyses.

4.4.2 CNVs

4.4.2.1 Quality Control

The *PennCNV* CNV detection models determined to be most repeatable were those which utilized a greater waviness stringency of 0.04 as suggested in chapter 3 (M. D. Wang *et al*. 2015). All CNVs identified with this waviness stringency in the absence of the gcmodel, were identified by the other models. It was thus decided for this study to utilize model 2 with a GCWF of 0.04, DLRS of 0.3 and no gcmodel for CNV identification.

4.4.2.2 Statistics And *Distribution*

One thousand and fifty five unique CNVs were identified in 197 of the 287 cattle. CNVs ranged from 31kb to 2.9Mb in size, with an average length of 301kb (Table 4.1). The majority of the CNVs were single copy deletions (625). Four hundred and five single copy duplications together with 5 double copy duplication and 20 double copy deletions were reported. The smallest CNV was a single copy duplication, while the largest was a single copy deletion. The greater number of deletions identified reflect findings of Jiang *et al.* (2012) who report 81 of the 99 CNVRs identified in Chinese Holstein cattle to be deletion events. Discrepancies in deletion vs. duplication events were however apparent. Hou *et al.* (2011) reported 281 CNVs in 39 African cattle of which only 68 were deletion events. Zhang *et al.* (2014) identified and characterized CNVs in the genome of Qinchuan cattle using the BovineHD beadchip and *PennCNV*. They reported 367 unique CNVs in 6 Qinchuan cattle of which 132 were loss events.

Table 4.1 CNV summary statistics of Copy number (CN), Number of CNVs (CNVs) and maximum (MaxL), minimum (MinL) and average (AL) CNV lengths.

CN	CNVs	MinL (bp)	MaxL (bp)	AL (bp)
0	20	44 415	227 892	109 759.2
1	625	36 419	2 933 073	361 997.179
3	405	31 397	1 297 541	217 608.642
4	5	93 420	572 953	218 348.800
Total	**1055**	**31 397**	**2 933 073**	**301 105.844**

Adjacent and overlapping CNVs were joined to form 356 unique CNVRs (Table 4.2). CNVRs ranged from 36kb to 4.1Mb in length with an average length of 287kb across breeds. (Jiang *et al.,* 2013) report 358 autosomal CNVRs in 96 Chinese Holstein cattle ranging from 10.76kb to 2.8Mb in size using the bovine

high-density beadchip. The size of CNVRs reported in cattle using the bovine 54k beadchip demonstrate notable discrepancies (Bae *et al.,* 2010; Hou *et al.,* 2011; Hou *et al.,* 2012; Jiang *et al.,* 2013). Hou *et al.* (2011) report 682 CNVRs in 539 cattle ranging from 32.57 kb to 5.57Mb in size, similar to those reported in this study. Bae *et al.* (2010) on the other hand report 855 CNVRs in 265 Hanwoo cattle (*Bos taurus coreanae*).

The most CNVRs were identified on chromosomes 4 and 6, while chromosomes 22 and 28 had the least CNVRs. Jiang *et al.* (2012) also reported chromosome 6 to exhibit the most CNVRs in Chinese Holstein. Chromosome 25 presented the greatest portion of its length to be covered by CNVRs. The largest CNVR was present on chromosome 11, while the smallest occurred on chromosome 1. The percentage of chromosomes covered by variations in copy number ranged from 1.15% of chromosome 28 to 14.14% of chromosome 25. The observed 4.03% of the total genome that demonstrated CNVRs is similar to findings of Hou *et al.* (2012) who found 4.7% of the genome of Angus cattle to be regions variable in copy number. Segmental duplications, with the greatest enrichment on chromosomes 5, 18, 27 and 29 in bovine, are considered to be associated with CNV prevalence (Liu *et al.,* 2009; Conrad and Antonarakis 2007). These chromosomes did not however, demonstrate the greatest CNV enrichment in this study.

Table 4.2 Table depicting the count, minimum (MinL), maximum (MaxL) and average (AvL) lengths and total length (LN) of unique CNVRs identified on each of the 29 Btau chromosomes of 287 cattle from 7 different breeds.

Chr	CNVR	CNVRs Ln (bp)	MinL (bp)	MaxL (bp)	AvL (bp)	% TotLn
1	23	3 488 143	36 419	607 020	151 658.39	2.16
2	11	2 205 853	50 633	648 246	200 532.09	1.55
3	21	4 753 565	39 373	1 281 217	226 360.24	3.75
4	24	6 586 652	60 330	2 723 817	274 443.83	5.32
5	15	5 223 799	51 928	1 438 360	348 253.27	4.17
6	24	6 751 891	38 235	2 273 588	281 328.79	5.51
7	20	4 205 711	52 472	1 366 647	210 285.55	3.72
8	10	1 374 518	63 621	308 120	137 451.80	1.18
9	13	4 765 050	53 174	2 079 181	366 542.31	4.39
10	12	1 880 912	44 415	582 405	156 742.67	1.77
11	11	6 609 094	71 882	4 181 753	600 826.73	6.01
12	17	4 667 872	60 967	2 010 326	274 580.71	5.48
13	11	1 843 708	73 286	346 832	167 609.82	2.19
14	11	3 176 676	47 051	1 039 469	288 788.73	3.91
15	8	1 988 273	116 374	483 531	248 534.13	2.35
16	12	2 983 492	70 038	555 338	248 624.33	3.84
17	6	2 514 470	54 358	1 880 338	419 078.33	3.30
18	10	3 827 829	159 263	1 352 214	382 782.90	5.82
19	14	4 108 113	72 145	839 290	293 436.64	6.34
20	14	3 959 577	58 641	982 995	282 826.93	5.23
21	10	3 051 928	42 434	1 047 092	305 192.80	4.42
22	4	3 228 078	77 923	2 409 975	807 019.50	5.24
23	5	3 431 423	60 886	2 997 091	686 284.60	6.56
24	7	1 720 967	42 164	864 422	245 852.43	2.67
25	13	6 233 821	66 465	1 827 519	479 524.69	14.14
26	7	1 897 312	80 466	862 302	271 044.57	3.66
27	8	1 289 928	62 602	368 664	161 241.00	2.66
28	4	527 460	63 417	292 126	131 865.00	1.15
29	11	4 146 255	43 671	2 554 531	376 932.27	8.00
Total	**356**	**102 442 370**	**36 419**	**4 181 753**	**287 759.47**	4.03

4.4.2.3 CNVR GO Over-Representation

A *PANTHER* overrepresentation test using a Bonferroni correction for multiple testing was performed for genes covered by CNVR identified. Five GO biological processes, one molecular function and 25 cellular components demonstrated a significant ($p<0.05$) over representation by CNVR genes and are presented in Tables 4.3, 4.4 and 4.5. Protein kinase binding is one such molecular function that is overrepresented by CNVR genes. Protein kinases catalyze the transfer of phosphates from high-energy phosphate donating molecules to specific substrates, playing a vital regulatory role in cell function and constituting one of the largest and most functionally diverse gene families. Kinases and hence kinase binding activity is essential in metabolism, protein regulation, cell signaling, cellular transport, secretory processes and many other cellular pathways. Twenty-four cellular components including extracellular membrane-bounded organelle, nucleus, intracellular organelle and endomembrane system also demonstrated overrepresentation by CNVR genes

(Table 4.3). Metabolic processes play an important role in not only generating energy for basic functions but also in regulating body temperature especially during periods of heat stress (Thornton *et al.,* 2009). Hou *et al.* (2012) also reported CNVR genes in 27 breeds from around the globe over representing metabolic processes, while Bickhart *et al.* (2012) and Seroussi *et al.* (2010) reported overrepresentation of cellular metabolism by CNVR genes identified in a variety of different cattle breeds. Cellular metabolic process, primary metabolic process, organic substance metabolic process and metabolic process biological processes were all overrepresented by CNVR genes identified in this study.

Table 4.3 Complete GO molecular functions (MF) with significant ($p<0.05$) enrichment by genes covered by CNVRs in 7 South African cattle breeds.

MF	REF	GN	EXP	TP	FOLD	P-VAL
Kinase binding	431	35	15.79	+	2.22	3.29E-02
Unclassified	5 446	175	199.51	-	0.88	0.00E00
Molecular transducer activity	2 055	38	75.28	-	0.50	1.11E-03
Signal transducer activity	1 899	30	69.57	-	0.43	4.69E-05
Receptor activity	1 862	25	68.21	-	0.37	8.49E-07
Signaling receptor activity	1 705	17	62.46	-	0.27	4.28E-09
Transmembrane signaling receptor activity	1 621	14	59.38	-	0.24	7.41E-10
G-protein coupled receptor activity	1 357	5	49.71	-	< 0.2	3.50E-13

*REF – the number of genes in the reference genome involved in the molecular functions, GN – the number of CNVR genes involved in molecular functions, EXP – expected number of genes for significant overrepresentation of molecular functions, TP – type of representation: either over (+) or under (-), FOLD – the number of CNVR genes divided by the number of genes expected for a significant overrepresentation of the molecular functions, P-VAL – p-value

Table 4.4 Complete GO cellular components (CC) with significant ($p<0.05$) enrichment by genes covered by CNVRs in 7 South African cattle breeds.

CC	REF	GN	EXP	TP	FOLD	P-VAL
Bounding membrane of organelle	1 276	77	46.75	+	1.65	1.63E-02
Membrane-bounded vesicle	2 690	149	98.55	+	1.51	1.72E-04
Vesicle	2 784	153	101.99	+	1.50	1.80E-04
Extracellular organelle	2 219	118	81.29	+	1.45	2.77E-02
Extracellular vesicle	2 219	118	81.29	+	1.45	2.77E-02
Extracellular exosome	2 219	118	81.29	+	1.45	2.77E-02
Extracellular membrane-bounded organelle	2 219	118	81.29	+	1.45	2.77E-02
Endomembrane system	2 402	127	88.00	+	1.44	1.67E-02
Nuclear part	2 987	154	109.43	+	1.41	6.58E-03
Cytoplasm	7 819	398	286.44	+	1.39	5.09E-14
Nucleus	5 074	251	185.88	+	1.35	5.35E-05
Membrane-bounded organelle	9 193	452	336.78	+	1.34	7.55E-15
Cytoplasmic part	5 289	260	193.76	+	1.34	4.54E-05
Intracellular membrane-bounded organelle	8 098	394	296.67	+	1.33	2.57E-10
Intracellular organelle part	5 781	281	211.78	+	1.33	2.44E-05
Organelle part	5 935	285	217.43	+	1.31	5.97E-05
Macromolecular complex	4 136	197	151.52	+	1.30	3.55E-02
Intracellular part	10 600	501	388.33	+	1.29	1.37E-14
Organelle	10 132	475	371.18	+	1.28	5.76E-12
Intracellular organelle	9 125	425	334.29	+	1.27	1.10E-08
Intracellular	11 162	512	408.91	+	1.25	2.98E-12
Cell part	13 285	572	486.69	+	1.18	2.23E-09
Cell	13 368	574	489.73	+	1.17	3.39E-09
Cellular_component	15 579	620	570.73	+	1.09	2.46E-03
Unclassified	4 293	112	157.27	-	0.71	0.00E00

*REF – the number of genes in the reference genome involved in the cellular components, GN – the number of CNVR genes involved in cellular components, EXP – expected number of genes for significant overrepresentation of cellular components, TP – type of representation: either over (+) or under (-), FOLD – the number of CNVR genes divided by the number of genes expected for a significant overrepresentation of the cellular components, P-VAL – p-value

Table 4.5 Complete GO biological processes (BP) with significant ($p<0.05$) enrichment by genes covered by CNVRs in 7 South African cattle breeds.

BP	REF	GN	EXP	TP	FOLD	P-VAL
Vesicle-mediated transport	780	55	28.57	+	1.92	2.91E-02
Cellular metabolic process	6 030	284	220.91	+	1.29	2.59E-03
Primary metabolic process	6 346	291	232.48	+	1.25	1.93E-02
Organic substance metabolic process	6 609	303	242.12	+	1.25	9.80E-03
Metabolic process	7 727	353	283.07	+	1.25	5.83E-04
Unclassified	4 618	129	169.18	-	0.76	0.00E00
Neurological system process	1 544	27	56.56	-	0.48	3.27E-02
G-protein coupled receptor signaling pathway	1 565	16	57.33	-	0.28	2.09E-07
Sensory perception	1 281	10	46.93	-	0.21	1.82E-07
Detection of stimulus involved in sensory perception	1 007	3	36.89	-	< 0.2	2.49E-09
Sensory perception of chemical stimulus	1 038	3	38.03	-	< 0.2	8.25E-10
Detection of stimulus	1 076	3	39.42	-	< 0.2	2.12E-10
Detection of chemical stimulus involved in sensory perception	963	1	35.28	-	< 0.2	4.90E-11
Detection of chemical stimulus	983	1	36.01	-	< 0.2	2.32E-11

*REF – the number of genes in the reference genome involved in the biological processes, GN – the number of CNVR genes involved in biological processes, EXP – expected number of genes for significant overrepresentation of respective biological processes, TP – type of representation: either over (+) or under (-), FOLD – the number of CNVR genes divided by the number of genes expected for a significant overrepresentation of the biological processes, P-VAL – p-value

4.4.2.4 CNVR Correlations, Gene Ontology And *Representation*

Of the 163 CNVR evident in more than 1 individual, 22 loci demonstrated a significant pairwise association ($p\leq0.05$) with at least one other loci across all 7 breeds, 11 of which demonstrated highly significant correlations ($p\leq0.002$). These loci culminated to form 74 significant correlations with a p-value of less than 0.05 (Additional file 4.2). Zhang *et al.* (2014) report a significant reduction in the CNVR associations with increase in CNVR prevalence. Associated CNVRs in this study, however were present in 2 to 78 animals (Additional file 4.3). On analyzing data independently for each of the indigenous (Nguni, Sanga, Bonsmara, Afrikaner, Drakensberger) and exotic (Holstein, Angus) breeds, only 7 loci were significantly correlated within indigenous breeds representing 6 significant correlations, while 102 loci within the exotic Taurine breeds presented 904 significant ($p\leq0.05$) correlations (Additional file 4.4). Deletions and duplications at the same loci were treated as independent CNVRs. Only one of the correlated loci pairs of all breeds demonstrated a deletion corresponding with duplication. The rest exhibited correlations occurring between CNVRs of the same copy number. Within the 6 CNVR correlations of the indigenous Sanga and composite breeds, 4 were between CNVR duplications and 2 were between a deletion and duplication (Additional file 4.5). The significant Taurine breed CNVR associations exhibited 866 deletion associations, 38 duplication associations and 2 deletion and duplication associations. Deletions interrupt genes while also causing a loss of biological function and are therefore currently seen as the most common CNV effecting phenotype (Liu and Bickhart 2012). Increased copy number may have a positive (McCarroll 2008) or negative (Lee and Lupski 2006) association with gene expression levels. Such a discrepancy between the occurrence of CNVR

correlations in the commercial Taurine and Sanga/composite breeds may be subsequent to a number of factors. Distinctions in CNVRs correlations specific to breeds and breed subpopulations, augments the notion that selection pressures play an important role in CNV formation (Hou *et al.,* 2011; Porto-Neto *et al.,* 2014). Jimenez (2014) proposes recombination, selection and mutations to potentially be the primary factories driving the genomic structure of variations within breeds and populations. The introduction of exotic Taurine breeds to a new environment may have placed specific pressures on the genome, resulting in the formation of CNVRs at specific loci involved in processes, functions or components vital for adaptation. Frequently encoding protein products that play a prominent role in species adaptation (Duda and Palumbi 1999), segmental duplications are an important cause of genomic instability that results in nonallelic homologous recombination (NAHR) during meiosis and genomic innovations and are currently recognized as one of the major catalysts and hotspots for CNV formation (She *et al.*, 2008; Nicholas *et al.*, 2009; Liu and Bickhart 2012; Alkan *et al*., 2009).

The 906 correlations evident among CNVRs of Taurine breeds encompass 849 genes. The 7 CNVR correlations evident among the indigenous animals, on the other hand covered 76 genes. Genes represented within correlated CNVRs were involved in a number of biological, molecular and cellular pathways and are presented in Table 4.6. Statistically significant CNVR correlations may indicate selection to cause the formation of CNVRs on different genomic regions that are involved in the same process. CNVs may alter gene structure, dosage or gene functioning by disrupting coding sequences, long range regulation or by exposing recessive alleles (Zhang *et al.,* 2009; Liu and Bickhart 2012; Stankiewicz and Lupski 2010). The phenotypic impact of CNVs is, however too a large extent related to the locations of the variant in relation to the genes (Buchanan and Scherer 2008). Gene copy number is conventionally positively correlated with gene expression (Stranger *et al.,* 2007), although cases of negative correlations have been reported (Lee and Lupski 2006). A Duplicated CNVR on chromosome 11 covering *AIF1L* (*allograft inflammatory factor 1-like*) and *ABL1* (*protein kinase abl1*) genes was correlated with a second duplication on chromosome 18 covering the *NLRP5* (*nacht, lrr* and *pyd domains-containing protein 5)* gene. The *AIF1L* is an important component of anti-inflammatory response (Kadoya *et al.,* 2014) and response to stress while *NLRP5* comprises part of the cellular defense response (Hutcheon *et al*., 2016). ABL1 gene mutations causes resistance to tyrosine kinase inhibitors which have been found to improve the management of chronic myeloid leukemia in humans (O'Hare *et al.,* 2007; Shah *et al.,* 2002). Of the 6 correlations present among CNVRs of the indigenous breeds, all except two were between duplicated regions. The only exceptions were correlations between a deletion on chromosome 6 and duplication on chromosome 29 and 26 respectively. Although no genes were covered by the deleted CNVR, the correlated duplication on chromosome 29 covered 24 genes including *TSPAN32* (*tetraspanin-32*), CDKN1C (*cyclin-dependent kinase inhibitor 1*) and *TNNT3* (*troponin T, fast skeletal muscle*) involved in a variety of biological processes, molecular functions and cellular components.

The representation of CNVR genes involved in processes, pathways and components that are involved in adaptation have implicated CNVRs to play a role in adaptation. The significant overrepresentation of such ontologies represented in Table 4.6 by correlated CNVRs further supports this proposal. These findings correspond to previous findings where environmental function genes and genes encoding secreted proteins demonstrate noticeable coincidence with CNVRs (Feuk *et al.,* 2006; Nguyen *et al.,* 2006; Sharp *et al.,* 2005).

Table 4.6 Ontologies (GO) with significant ($p<0.05$) enrichment by genes covered by correlated CNVRs in 7 South African cattle breeds.

GO	REF	GEN	EXP	TP	FOLD	P-VAL
CC						
Troponin complex	8	4	.13	+	> 5	9.80E-03
Intracellular organelle part	5 633	133	89.57	+	1.48	1.21E-04
Organelle part	5 796	134	92.17	+	1.45	3.80E-04
Membrane-bounded organelle	9 165	190	145.74	+	1.30	4.06E-04
Cytoplasm	7 752	160	123.27	+	1.30	1.85E-02
Intracellular membrane-bounded organelle	8 047	166	127.96	+	1.30	1.04E-02
Organelle	10 107	205	160.72	+	1.28	3.47E-04
Intracellular organelle	9 084	183	144.45	+	1.27	9.11E-03
Intracellular part	10 523	209	167.33	+	1.25	1.39E-03
Intracellular	11 092	211	176.38	+	1.20	4.64E-02
PrC						
Translation elongation factor	50	6	.80	+	> 5	3.49E-02
BP						
Cellular biosynthetic process	2 369	69	37.67	+	1.83	3.09E-03
Organic substance biosynthetic process	2 450	70	38.96	+	1.80	5.05E-03
Biosynthetic process	2 527	72	40.18	+	1.79	3.71E-03
G-protein coupled receptor signaling pathway	1 539	6	24.47	-	0.25	3.27E-02
Sensory perception	1 281	3	20.37	-	< 0.2	9.00E-03
Detection of stimulus	1 076	1	17.11	-	< 0.2	2.80E-03

*CC – cellular component, PrC – protein class and BP – biological process

4.4.3 Genetic Diversity Based On CNVR

4.4.3.1 Breed CNVRs

The most CNVRs were identified in the Nguni Angus breed (n =114), followed by the Holstein (n =102) and Angus (n=101) breeds. The Nguni Angus crossbreed also demonstrated the highest average CNVRs per animal at 16.29, considerably higher than the 1.81 average across breeds. Great variation in the size and number of CNVRs has been reported in cattle (Jiang *et al.,* 2012; Hou *et al.,* 2012). Bickhart *et al.* (2012)speculated that the distinctions in selected breeds for specific traits could be linked to specific CNVs. A number of studies utilizing SNP genotyping platforms to identify CNVs have been done on a variety of different cattle breeds of late. Jiang *et al.* (2013) identified 367 CNVRs by means of *PennCNV* analyses of high-density SNP genotyping data from 96 Chinese Holsteins. Hou *et al.* (2011) on the other hand, reports 682 CNVRs identified in 521 animals representing 21 different breeds also identified using Bovine50K SNP genotyping array. Discrepancies in CNVs and subsequent CNVRs between different breeds and even individuals could thus be expected. Although Jiang *et al.* (2013) highlight the differences in size and

structure of populations, could also contribute to such incongruities. Despite the Nguni Angus cross having noticeably fewer animals in the study, the most CNVRs (114) were identified in these 10 animals. 102 and 101 CNVRs were identified in 45 and 32 Holstein and Angus animals respectively. The least CNVRs were identified in the 46 and 48 animals of the two composite breeds (Bonsmara and Drakensberger) (Table 4.7). The Nguni Angus demonstrated the most CNVRs, with an average of 16 CNVRs per animal. The greater number of CNVRs evident in the exotic Taurine breeds reflects findings of Choi *et al.* (2013) who compared the genome of a Hanwoo bull to that of Holstein and Black Angus respectively using whole genome sequencing methodologies. The results from this study supports the proposition of Choi *et al.* (2013) that breeds that have undergone more intensive selection for production traits may demonstrate greater number of copies at loci involved in specific traits. Choi *et al.* (2013) suggested CNVs to be affected by recent intensive artificial selection schemes aimed at improving economically important production traits. Narang *et al.* (2014) proposed that the migration and adaptation of a population or breed to a completely different environment to which they have typically been accustomed to may require considerable changes on a genomic level that may be achieved via events like CNVs which may hence contribute towards adaptation. This may explain the lower than expected number of CNVRs present in the sanga breeds which has undergone intensive selection for production traits in recent years, while comprising an indigenous breed that is adapted to the environment. Composite breeds have been developed with the mind to take advantage of the adaptability of the indigenous breeds with the productive capability of the taurine breeds (Bonsma 1980). The lower number of CNVR evidence in the composite breeds may therefore be subsequent to the innate adaptive ability that has been bred in from the indigenous components of the breeds. Although the exact origin of Drakensberger cattle has been under dispute, the parallel of Drakensberger CNVs with that of the composite Bonsmara cattle aligns with the recent findings published by Decker *et al*. (2016) who demonstrate these two breeds to comprise 3 ancestries. The elevated CNVs present in the Holstein and Angus animals may be consequent to genomic variations incurred by environmental pressures of a “new” environment. Matukumalli *et al.* (2009) and Hou *et al.* (2011) however report Taurine breeds to have fewer CNVs than composite, Indicine and African breeds. The African and composite breeds in the study of Hou *et al.* (2011) were represented by fewer animals (39 and 46 respectively) and demonstrated an average of 7.21 and 7.17 CNVs per animal. This is not much more than the 6.23 average of 366 Taurine animals, but noticeably less than the 11.41 average of the 70 Indicine animals.

Table 4.7 CNV summary statistics for each of 7 South African cattle breeds (Afrikaner – ANG, Angus – ANG, Bonsmara –BON, Drakensberger – DRK, Holstein – HOL, Nguni – NGU and Nguni Angus cross – NGxAN).

BRD	ANML	AN CNV	CNVR	Av	MinL (bp)	MaxL (bp)	AL(bp)	GEN
AFR	48	31	76	2,45	36 419	4 181 753	498 498.79	96
ANG	32	25	101	4,04	42 946	4 181 753	581 476.86	430
BON	46	35	60	1,71	52 472	4 181 753	668 772.47	96
DRK	48	24	63	2,63	38 235	4 181 753	353 594.71	29
HOL	45	28	102	3,64	42 164	4 181 753	558 378.40	207
NGU	59	47	95	2,02	44 415	4 181 753	467 388.03	142
NGxAN	10	7	114	16,29	54 147	4 181 753	584 980.73	616
	287	197	356	1,81	36 419	4 181 753	535 289.93	809

The chromosomal distribution of CNVRs across breeds demonstrates great variation in the size and number of CNVRs identified per autosome (Figure 4.1). Chromosome 4 and 6 possessed the most (25) CNVRs. The largest CNVR found on chromosome 11 (CNVR11) was 4.1Mb in length. This CNVR was present in the 76 animals from all 7 breeds. The smallest CNVR of 36 kb was identified in the Afrikaner cattle breed while the Bonsmara, despite demonstrating the least CNVRs, had the longest average CNVR.

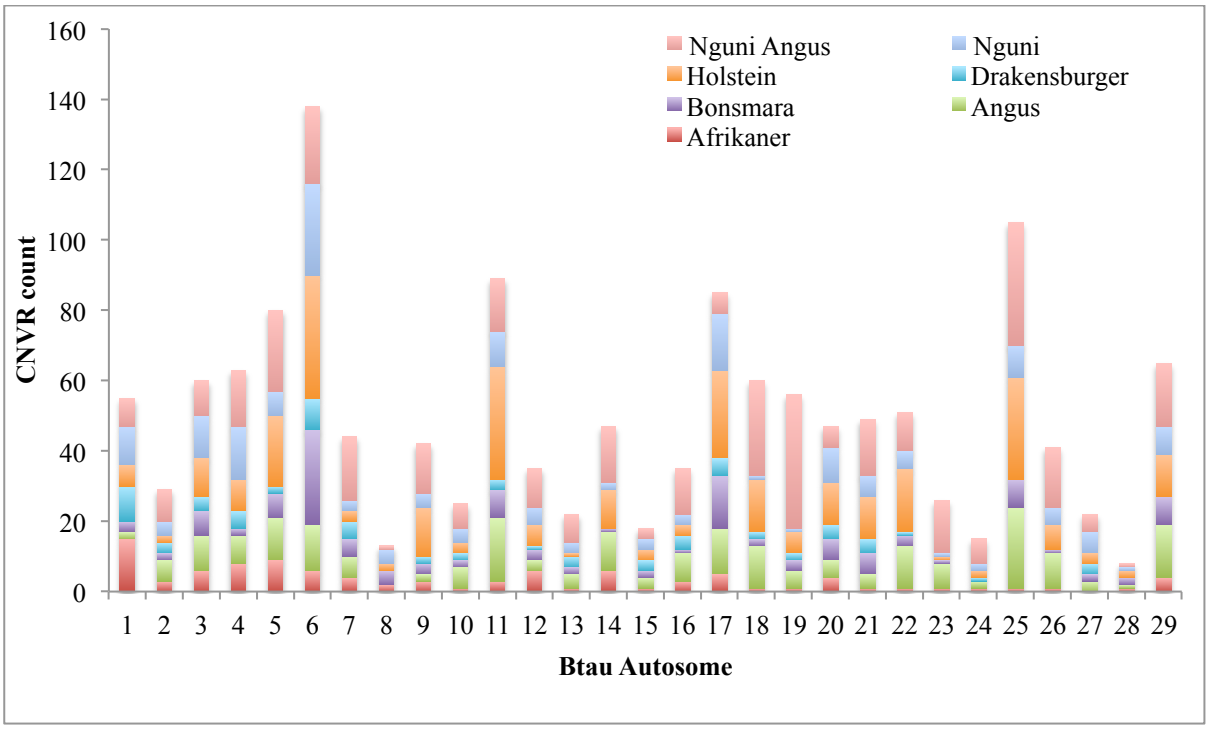

Figure 4.1 Chromosomal distribution of CNVRs for each of 7 South African cattle breeds.

Only 4 CNVRs were identified in all 7 cattle breeds with chromosome 17 and chromosome 11 presenting the 2 most common CNVR. Figure 4.2 demonstrates the spatial distribution of CNVs within each breed for the 4 mutual CNVRs that were identified in 53 to 78 animals. In all 4 instances Angus, Holstein and Nguni x Angus CNVs represented the largest portion of the CNVR while Drakensberger CNVs denoted the least. The consequence of such discrepancies in specific CNV regionality between breeds should be investigated.

Similar to the findings of Molin *et al.* (2014), the majority of the CNVs identified were shared between fewer breeds with the most CNVs (30) being shared between Angus and Nguni Angus cattle (Additional file 4.6). Choi *et al.* (2013) found great discrepancies in the prevalence of CNVRs when comparing the genomes of Hanwoo cattle to those of Holstein and Black Angus. Cicconardi *et al.* (2013) reported little variation in CNV distribution on chromosomes across five Italian cattle breeds, proposing CNV region (CNVR) variation to be greater between individuals than between breeds. Molin *et al.* (2014) identified 15 of the 72 CNVs identified in 351 dogs representing 30 different breeds to be breed specific CNVs. CNVRs identified in a single breed may pose interest for the investigation into breed specific traits (Molin *et al.,* 2014).

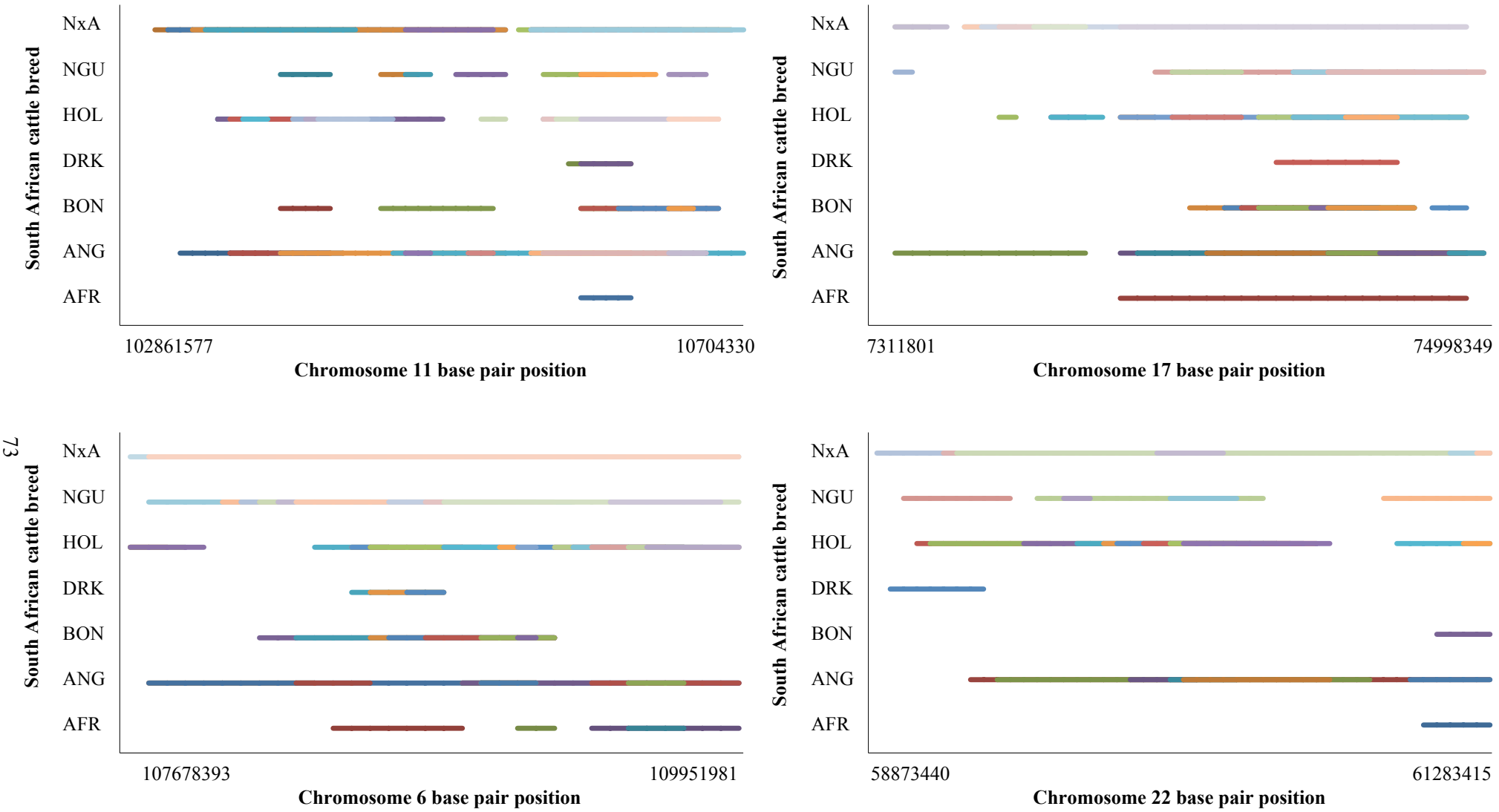

Figure 4.2 CNV chromosomal distribution in 7 South African cattle breeds at four difference chromosomal location namely **a)** chr11:102861577-10704330, **b)** chr17:7311801-74998349, **c)** chr6:107678393-109951981 and **d)** chr22:58873440-61283415.

4.4.3.2 Breed CNVRs Gene Ontology

Eight hundred and nine genes were covered by the 356 CNVRs identified across 7 South African cattle breeds (Table 4.7). Drakensberger cattle had the least CNVR genes, while Angus had the most of the purebreeds and Nguni Angus had the most overall. Of the 809 genes, 6 genes (*low affinity sodium-glucose cotransporter-like* (*LOC527441*), *netrin G2* (*NTNG2*), *otopetrin 1* (*OTOP1*), *solute carrier family 5 member 1* (*SLC5A1*), *transmembrane protein 128* (*TMEM128*) and *WD repeat domain 1* (*WDR1*)) were common to all breeds. Three hundred and eighty nine CNVR genes were breed specific (Additional file 4.7). The most CNVR genes were shared between Angus and Nguni Angus animals. Afrikaner, Angus, Bonsmara, Drakensberger, Holstein, Nguni and Nguni Angus animals had 17, 57, 26, 13, 19, 26 and 231 breeds specific CNVRs. Heat shock proteins *HSPBP1* (*heat shock binding protein 1*), *HSPB1* (*heat shock protein family B member 1*), *HSPA5* (*heat shock protein family A* (*Hsp70) member 5*) and *HSP90AA1* (*heat shock protein 90 alpha family class A member 1*) considered to play a vital role in balancing immunity and survival during times of stress (Morange 2006), were covered by CNVRs in Nguni, Angus, Holstein and/or Nguni Angus breeds/crossbreed. Severe reductions in *WDR1* (*WD40 repeat protein 1*), identified in 42 animals from breeds in this study were reported to disturb megakaryocyte maturation and platelet shedding, aggravate neutrophilic auto inflammatory disease and trigger embryonic lethality in mice (Kile *et al.,* 2007). *LSP1* (*Lymphocyte-specific protein 1*) and *IGF-II* (*insulin-like growth factor 2),* covered by CNVRs identified in Angus and Nguni Angus animals and *IGLL1* (*immunoglobulin lambdalike polypeptide 1)* overlapped by CNVRs in 44 animals from all breeds except Drakensberger were differentially expressed in cattle selected for resistance or susceptibility to intestinal nematodes (Araujo *et al.,* 2009). Other genes involved in immune response included *GSTT3* (*glutathione s-transferase theta-3*), *GSTT1* (*glutathione s-transferase theta-1*) and *SMARCB1* (*SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily B member 1*) that were present in 35, 33 and 40 animals respectively from all breeds except the Drakensberger.

4.4.3.3 Analysis Of Molecular Variance

No studies assessing partitioning of CNVR genetic variance in cattle have been performed. In this study, CNVRs were treated much the same as AFLP markers, using binomial present/absence data after which molecular variance was assessed. Breeds were grouped according to Sanga, taurus and composite/cross breeds. For all breed groups, the degree of variation within populations was considerably greater than that between populations. The majority of CNVR variation was therefore explained as interindividual variation as opposed to between breed variation. Inadequately distinguishing between CNVRs that are breed specific and those that are bovine specific may be the cause of the significantly higher degree of variation being evident within populations. We postulate that a large proportion of CNVRs are animal specific events, while only a few explicit CNVRs events that are specific to breeds. Pienaar *et al.* (2014) found high levels of within breed diversity for Afrikaner cattle using microsatellite data. Makina *et al.* (2014) found the Afrikaner breed to have the greatest number of alleles per locus when compared to the 5 other purebreeds in this study, while the Nguni had the least. Drakensberger cattle have the greatest genetic diversity of the 4 indigenous Sanga and composite breeds, while the two Taurine breeds were reported to have had the greatest gene

diversity (Makina *et al.,* 2014). Table 4.8 demonstrates pairwise population PhiPT values for CNVRs of 3 groups of South African cattle breeds. The greatest CNVR genetic variation was evident in composite/cross breed group, while the Sanga cattle breed demonstrated the least (Table 4.8). The Holstein and Angus breeds of the taurus cattle group have a longer history of artificial selection that has led to enhanced production (Choi *et al.,* 2013). The observed discrepancies evident between some breeds could very well be caused by genetic drift due to bottlenecks, natural selection and selective breeding (Hou *et al.,* 2011). Itsara *et al.* (2010) determined different mutation processes to contribute disproportionately to CNVs dependent on the size of the *de novo* event. The mutation rate of CNVs has been established to be considerably higher than that of SNPs, with great variation in mutation rates occurring between loci (Campbell *et al.,* 2011). CNVs have been suggested to be a mechanism by which the genome responds to selection pressures subsequent to genomic instability induced by such pressures (Wang *et al.,* 2015). The difference in between-breed occurrence of CNVRs may be more an artifact of commercial versus indigenous/non-commercial breeds, with selection pressures playing a pivotal role. Itsara *et al.* (2010) report lineage specific CNVRs, proposing CNVs in the Chinese cattle populations to be partly consequent to selective breeding during domestication but also subsequent to hybridization and introgression. The elevated CNVR genetic variation found within the composite breed group of this study, augment the idea of hybridization playing a role in CNV prevalence.

Table 4.8 Summary results of AMOVA within (WTHN), among (AMG) and total (TOT VAR) CNVR genetic variation for 3 breed groups (BRD GRP) of South African cattle breeds.

BRD GRP*	WTHN	AMG	TOT VAR
Sanga	2.510	0.085	2.596
Taurine	6.115	0.113	6.227
Composite	4.233	3.897	8.129
All	4.936	0.073	5.009

*Sanga – Nguni and Afrikaner, Taurine – Holstein and Angus, Composite – Bonsmara and Drakensberger

4.4.3.4 Principle Component Analyses

A genetic distance matrix was generated from the binomial presence/absence of CNVRs for each animal using *GenAlex*. This was used to compute principle component analyses of the dataset where a plot aligning the first and second eigenvectors displayed the spatial distribution of animals according to their CNVR distribution. Multiple analyses were performed using different stringencies to assess the data**.** The greatest amount of variation was captured in PC1 with an eigenvalue of 300.58, explaining 80.47% of the total variation captured among individuals. The Nguni Angus cross animals were the most differentiated from the rest of the animals at PC1 against PC2. With the exception of the Nguni Angus cross animals, all breeds clustered together (Figure 4.3). The Holstein animals clustered in the same region but with a larger spread (Figure 4.3). The Holstein animals pulled towards the top of the cluster, while the Angus and Afrikaner animals cluster more to the left. The Nguni, Drakensberger and Bonsmara animals had the most compact clustering, pulling more to the right of the x-axis. Considering the composite breeds are comprised of various crosses between the Sanga and Taurus breeds, one would expect them to lie between the two breed groups

that were however, not the case in this study. The exact mutation and inheritance patterns of CNVs are not fully understood (L. Zhang *et al.,* 2015). It has been proposed that forces such as recombination, selection and mutations are the primary factors driving the genomic architecture of large variations (Jimenez *et al.,* 2014). The distinction between the Sanga and Taurine breeds maybe evidence of such selection driven mutational CNVs. Zhang *et al.* (2015) report lineage specific CNVRs to align with Taurine and Indicine descent respectively. The distinction of the Nguni Angus crossbreed demonstrates possible genomic instability caused by crossbreeding that results in the formation of new CNVs distinct to the crossbred animals.

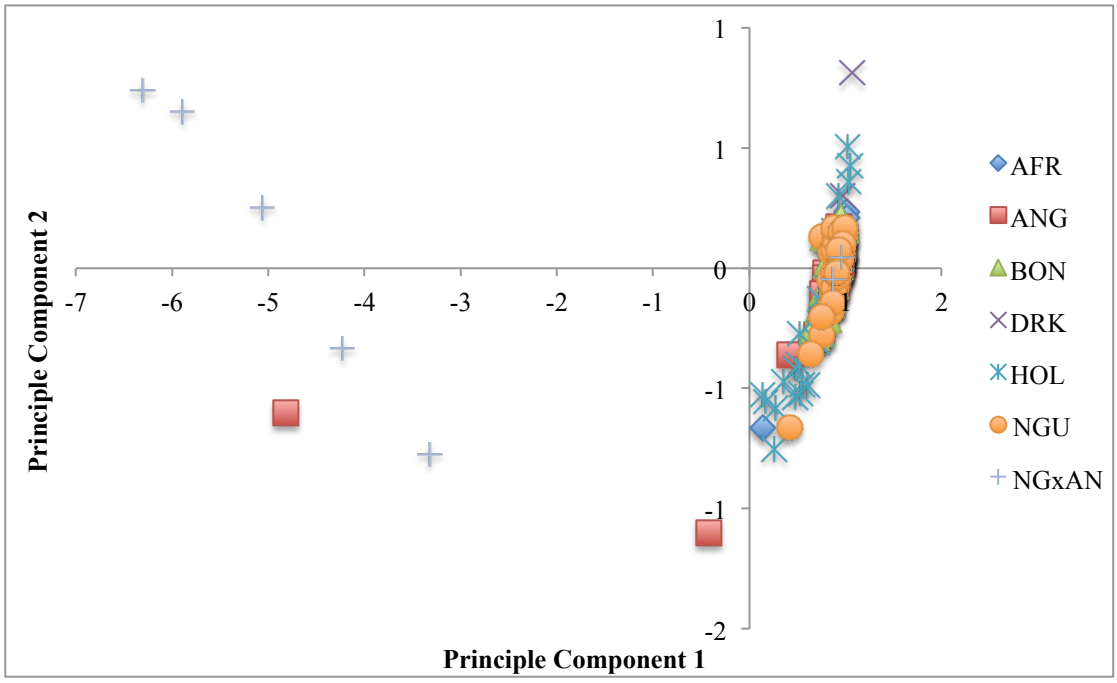

Figure 4.3 Principle components analyses for components 1 and 2 for CNVRs of animals from 6 different South African cattle breeds.

4.4.3.5 Population Structure

STRUCTURE was utilized in *R* to depict the population structure of breeds CNV frequencies. Figure 4.4 demonstrates the evolution of the population structure as K increased from 3 to 7. At K = 3, genomic signatures distinct to the Nguni Angus crossbred animals were evident while genomic signatures distinct to the Sanga breeds of cattle (Afrikaner, Drakensberger and Nguni) were picked during progression to K=7. Sanga cattle breeds comprise a crossbreed between indigenous Taurine and zebu cattle breed that are unique to Africa (Rege 1999). The presence of CNVs distinct to breed types corresponds with previous studies demonstrating CNV distribution within and among species to be shaped by mutation, selection and demographic history (Conrad and Hurles 2007). Bickhart *et al.* (2016) further demonstrate CNV distribution to parallel breed type, distinguishing taurine from indicine cattle breeds. In this study, a similar pattern is evident with sanga and taurine breeds demonstrating discrete genomic signatures. Some Nguni animals

demonstrate notable taurine specific signatures that may be subsequent to the Nguni breeders' society only being developed in more recent years. One could expect the composite breeds to possibly present with intermediary CNVs being captured from both taurine and sanga breed types. Structure analyses, however exhibits the Bonsmara composites to have more animals sharing signatures with the taurine animals. This may be as a result of the greater number of CNVs present in taurine breeds and hence having greater weight and subsequent carry over effect in the composite breeds. Levels of admixture were evident in the structure based clustering. This is in accordance with Decker *et al*. (2016) who demonstrated notable admixture present in South African cattle sanga breeds when using SNP data from breeds across the globe.

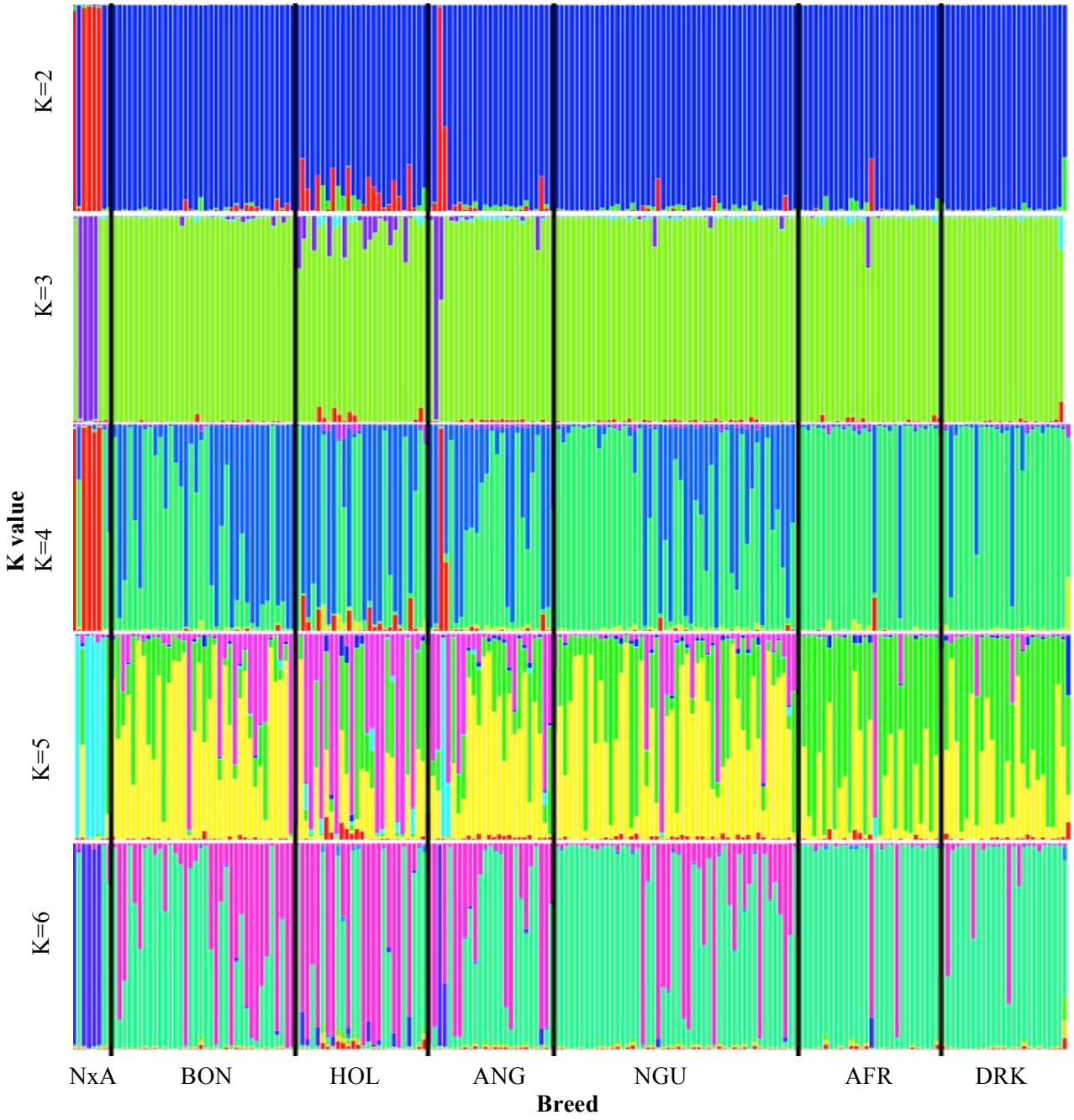

Figure 4.4 CNVR structure analyses for 287 animals from 7 different breeds of cattle for K = 3 to K = 7.

4.4.3.6 Phylogenetic Analyses

CNVs are considered to play a role in breed formation and adaptation, with copy number differences occurring between breeds (Liu *et al.,* 2010). Increasing evidence also suggests CNVs to play a primary role in interindividual diversity (Stankiewicz and Lupski 2002; Sebat *et al.,* 2004) attributed to both normal phenotypic variation and major variations in complex traits (Fellermann *et al.,* 2006; Feuk *et al.,* 2006). A cluster dendogram was generated from CNVRs identified in animals by means of *R hclust* (Figure 4.5). CNVRs for the most part clustered animals of the same breed together. Five of the 7 Nguni X Angus cross animals clustered together with 1 Angus animal in a clade distinct from the rest of the animals. A second clade was evident with a seemingly random mix of animals from different breeds with some animals clustering together within breeds, but others were seemingly random. The structure of the dendogram suggest a disparity with some CNVs being breed specific variations, while others may possibly be *Bos taurus*/*Bos indicus* CNVs or possibly indicators of interindividual variation. This corroborates findings of Bickhart *et al.* (2016) and Xu *et al.* (2016) who found CNVs to differentiate cattle groups. These authors demonstrated Taurine, Indicine and African breeds to be clearly distinguishable at K = 3 in CNV admixture analyses. Increasing K separated the beef from the dairy Taurine breeds (Bickhart *et al.,* 2016)

Hierarchal clustering analyses on CNVR frequency within breeds were performed. A cluster dendogram of breeds is depicted in Figure 4.6 Binomial clustering of CNVR presence generated two distinct clades separating the indigenous pure breeds from the two Taurine breeds and the Nguni Angus crossbreed. CNVR presence within the Nguni Angus animals placed them right next to the Angus animals and completely separated from the Nguni. The two frequency plots, however generated distinctly different distributions. CNVR frequency articulated as a percentage caused the Holstein and Nguni Angus animals to segregate away from the other animals while the Angus breed moved to between the Bonsmara/Nguni and Afrikaner/Drakensberger clades. On using the number of animals presenting the CNVR the Nguni Angus breed was completely isolated while the two Taurine breeds clustered together and the indigenous breeds clustered in a stepwise fashion. CNVR identified may therefore represent breed specific CNVRs as well as more random CNVR events. Those CNVRs occurring at a greater frequency within a breed may be indicators of breed specific CNVRs. Such breed specific CNVRs may, however also be evidence of selection driven CNVRs, with animals of the same breed being exposed to similar selection forces. The occurrence of some of these breed specific CNVRs in animals from other breeds, may be indicative of such a pattern.

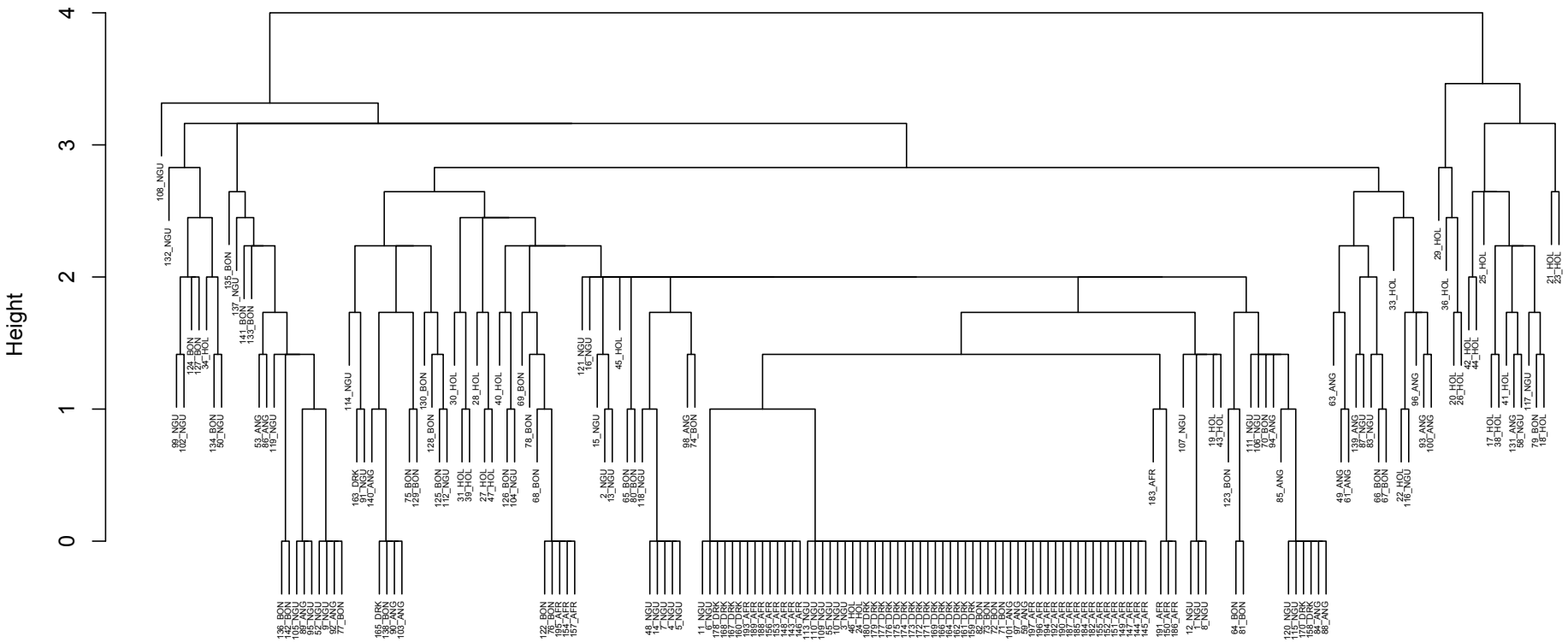

Figure 4.5 Hierarchal cluster analyses for CNVR presence of 287 South African cattle.

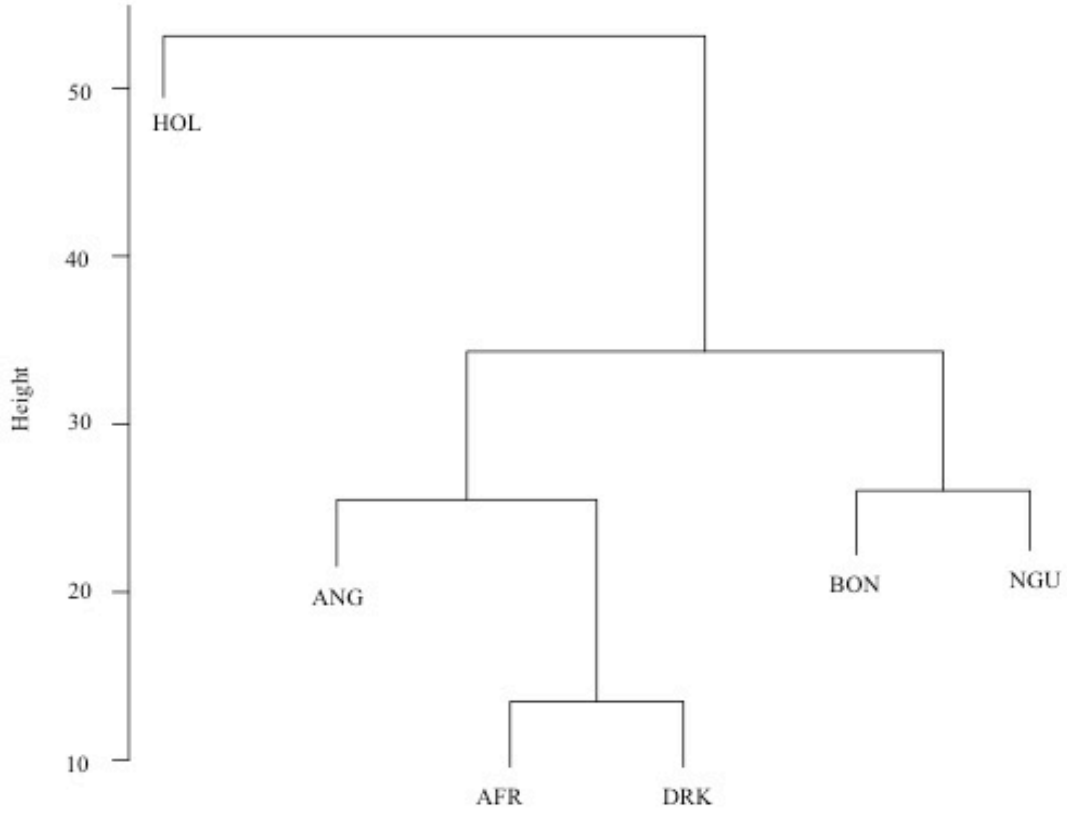

Figure 4.6 Hierarchal cluster analyses based on presence of CNVR in 6 South African cattle breeds.
*HOL = Holstein, ANG = Angus, AFR = Afrikaner, DRK = Drakensberger, BON = Bonsmara and NGU = Nguni

4.5 Conclusion

356 Unique CNVRs were identified in 287 animals from 2 Taurine, 2 composite, 2 Sanga and 1 Sanga Taurine cross South African cattle breeds using the Bovine 50K Beadchip. A number of cellular components, molecular functions and biological processes demonstrated overrepresentation by genes covered or lying within 10Mb of CNVRs identified. Correlations between CNVR presence were evident, with considerably more CNVR correlations occurring among the commercially bred Taurine breeds. Such correlations suggest selection pressures being exerted on different genomic regions involved in specific processes and functions. CNVs may be a means by which the genomes respond to selection pressures and subsequently adapts. Variations in CNVR presence between breeds were present with more CNVRs being present in the Nguni Angus cross and the two Taurine breeds. Composite and crossbred animals demonstrated the most within breed CNVR variation, while Sanga cattle demonstrated the least. The Nguni Angus cross demonstrated unique CNV genetic signatures, while some CNVs segregated in both the Taurine and Sanga breeds to some degree. This study indicates CNVRs to play a role in both interindividual and between breed variations. With Sanga and Taurine breeds having undergone different selection pressures, the variation in CNV incidence between these groups combined with the CNV correlations designate CNVRs to be genomic features prevalent in selection and adaptation. The distinct properties of CNVRs in the Nguni Angus cross animals need also be explored with possible implications in events like hybrid vigor.

4.6 References

Alkan, C., J.M. Kidd, T. Marques-bonet, G. Aksay, F. Hormozdiari, J.O. Kitzman, C. Baker, et al. 2009. "Personalized Copy-Number and Segmental Duplication Maps Using Next-Generation Sequencing." *Nature Genetics* 41 (10): 1061–67. doi:10.1038/ng.437.

Araujo, R.N., T. Padilha, D. Zarlenga, T. Sonstegard, E.E. Connor, C. Van Tassel, W.S. Lima, E. Nascimento, and L.C. Gasbarre. 2009. "Use of a Candidate Gene Array to Delineate Gene Expression Patterns in Cattle Selected for Resistance or Susceptibility to Intestinal Nematodes." *Veterinary Parasitology* 162: 106–15. doi:10.1016/j.vetpar.2008.12.017.

Bae, J.S., H.S. Cheong, L.H. Kim, S. NamGung, T.J. Park, J. Chun, J.Y. Kim, C.F.A. Pasaje, J.S. Lee, and H.D. Shin. 2010. "Identification of Copy Number Variations and Common Deletion Polymorphisms in Cattle." *BMC Genomics* 11: 232. doi:10.1186/1471-2164-11-232.

Bickhart, D.M., Y. Hou, S.G. Schroeder, C. Alkan, M.F. Cardone, L.K. Matukumalli, J. Song, et al. 2012. "Copy Number Variation of Individual Cattle Genomes Using next-Generation Sequencing." *Genome Research* 22: 778–90. doi:10.1101/gr.133967.111.

Bickhart, D.M., L. Xu, J.L. Hutchison, J.B. Cole, D.J. Null, S.G. Schroeder, J. Song, et al. 2016. "Diversity and Population-Genetic Properties of Copy Number Variations and Multicopy Genes in Cattle," 1–10. doi:10.1093/dnares/dsw013.

Bonsma, J. 1980. *Jan Bonsma and the Bonsmara Beef Cattle Breed*. Bonsmara Cattle Breeders Society's 21st Anniversary Publication.

Buchanan, J.A., and S.W. Scherer. 2008. "Contemplating Effects of Genomic Structural Variation." *Genetics in Medicine : Official Journal of the American College of Medical Genetics* 10 (9): 639–47. doi:10.1097/GIM.0b013e318183f848.

Campbell, C.D., N. Sampas, A. Tsalenko, P.H. Sudmant, J.M. Kidd, M. Malig, T.H. Vu, et al. 2011. "Population-Genetic Properties of Differentiated Human Copy-Number Polymorphisms." *American Journal of Human Genetics* 88 (3). The American Society of Human Genetics: 317–32. doi:10.1016/j.ajhg.2011.02.004.

Castellani, C. A., M.G. Melka, A. E. Wishart, M. E. O. Locke, Z. Awamleh, R. L. O'Reilly, and S. M. Singh. 2014. "Biological Relevance of CNV Calling Methods Using Familial Relatedness Including Monozygotic Twins." *BMC Bioinformatics* 15 (1). BMC Bioinformatics: 114. doi:10.1186/1471-2105-15-114.

Choi, J.W., K.T. Lee, X. Liao, P. Stothard, H.S. An, S. Ahn, S. Lee, S.Y. Lee, S.S. Moore, and T.H. Kim. 2013. "Genome-Wide Copy Number Variation in Hanwoo, Black Angus, and Holstein Cattle." *Mammalian Genome* 24 (3–4): 151–63. doi:10.1007/s00335-013-9449-z.

Cicconardi, F., G. Chillemi, A. Tramontano, C. Marchitelli, A. Valentini, P. Ajmone-Marsan, and A. Nardone. 2013. "Massive Screening of Copy Number Population-Scale Variation in Bos Taurus Genome." *BMC Genomics* 14 (1): 124. doi:10.1186/1471-2164-14-124.

Conrad, B., and S. E. Antonarakis. 2007. "Gene Duplication: A Drive for Phenotypic Diversity and Cause of

Human Disease." *Annual Review of Genomics and Human Genetics* 8: 17–35. doi:10.1146/annurev.genom.8.021307.110233.

Conrad, D.F., and M.E. Hurles. 2007. "The Population Genetics of Structural Variation." *Nature Gene* 39 (7 (supplement)): S30-36. doi:10.1038/ng2042.The.

Decker, J.E., S.O. Makina, L.K. Whitacre, J.E. Decker, J.F. Taylor, and M.D. Macneil. 2016. "Insight into the Genetic Composition of South African Sanga Cattle Using SNP Data from Cattle Breeds Worldwide." *Genetics Selection Evolution*, no. December. BioMed Central. doi:10.1186/s12711-016-0266-1.

Diskin, S.J., M. Li, C. Hou, S. Yang, J. Glessner, H. Hakonarson, M. Bucan, J.M. Maris, and K. Wang. 2008. "Adjustment of Genomic Waves in Signal Intensities from Whole-Genome SNP Genotyping Platforms." *Nucleic Acids Research* 36 (19): 1–12. doi:10.1093/nar/gkn556.

Duda, T F, and S R Palumbi. 1999. "Molecular Genetics of Ecological Diversification: Duplication and Rapid Evolution of Toxin Genes of the Venomous Gastropod Conus." *Proceedings of the National Academy of Sciences of the United States of America* 96: 6820–23. doi:10.1073/pnas.96.12.6820.

Falush, D., M. Stephens, and J. K. Pritchard. 2007. "Inference of Population Structure Using Multilocus Genotype Data: Dominant Markers and Null Alleles." *Molecular Ecology Notes* 7 (4): 574–78. doi:10.1111/j.1471-8286.2007.01758.x.

Fellermann, K., D. E. Stange, E. Schaeffeler, H. Schmalzl, J. Wehkamp, C. L. Bevins, W. Reinisch, et al. 2006. "A Chromosome 8 Gene-Cluster Polymorphism with Low Human Beta-Defensin 2 Gene Copy Number Predisposes to Crohn Disease of the Colon." *American Journal of Human Genetics* 79 (September): 439–48. doi:10.1086/505915.

Feuk, L., A.R. Carson, and S.W. Scherer. 2006. "Structural Variation in the Human Genome." *Nature Reviews. Genetics* 7: 85–97. doi:10.1038/nrg1767.

Helleday, T. 2003. "Pathways for Mitotic Homologous Recombination in Mammalian Cells." *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis* 532: 103–15. doi:10.1016/j.mrfmmm.2003.08.013.

Hoffmann, I. 2010. "Climate Change and the Characterization, Breeding and Conservation of Animal Genetic Resources." *Animal Genetics* 41: 32–46. doi:10.1111/j.1365-2052.2010.02043.x.

Hou, Y., D.M. Bickhart, M.L. Hvinden, C. Li, J. Song, D.A. Boichard, S. Fritz, et al. 2012. "Fine Mapping of Copy Number Variations on Two Cattle Genome Assemblies Using High Density SNP Array." *BMC Genomics* 13: 376. doi:10.1186/1471-2164-13-376.

Hou, Y., G.E. Liu, D.M. Bickhart, M.F. Cardone, K. Wang, E. Kim, L.K. Matukumalli, et al. 2011. "Genomic Characteristics of Cattle Copy Number Variations." *BMC Genomics* 12: 127. doi:10.1186/1471-2164-12-127.

Hou, Y., G.E. Liu, D.M. Bickhart, L.K. Matukumalli, C. Li, J. Song, L.C. Gasbarre, C.P. Van Tassell, and T.S. Sonstegard. 2012. "Genomic Regions Showing Copy Number Variations Associate with Resistance or Susceptibility to Gastrointestinal Nematodes in Angus Cattle." *Functional and Integrative Genomics* 12: 81–92. doi:10.1007/s10142-011-0252-1.

Hutcheon, C., P. Paulvannan, and N. Subramanian. 2016. “Cytoplasmic Sensing in Innate Immunity BT.” In *Encyclopedia of Cell Biology*, 710–26. Waltham: Academic Press. doi:http://dx.doi.org/10.1016/B978-0-12-394447-4.30108-0.

Itsara, A., H. Wu, J. D. Smith, D. A. Nickerson, I. Romieu, S. J. London, and E. E. Eichler. 2010. “De Novo Rates and Selection of Large Copy Number Variation.” *Genome Research* 20: 1469–81. doi:10.1101/gr.107680.110.

Jakobsson, M., S.W. Scholz, P. Scheet, J.R. Gibbs, J.M. VanLiere, H. Fung, Z.A. Szpiech, et al. 2008. “Genotype, Haplotype and Copy-Number Variation in Worldwide Human Populations.” *Nature* 451 (7181): 998–1003. doi:10.1038/nature06742.

Jiang, L., J. Jiang, J. Wang, X. Ding, J. Liu, and Q. Zhang. 2012. “Genome-Wide Identification of Copy Number Variations in Chinese Holstein.” *PLoS ONE* 7 (11). doi:10.1371/journal.pone.0048732.

Jiang, L., J. Jiang, J. Yang, X. Liu, J. Wang, H. Wang, X. Ding, J. Liu, and Q. Zhang. 2013. “Genome-Wide Detection of Copy Number Variations Using High-Density SNP Genotyping Platforms in Holsteins.” *BMC Genomics* 14 (1). BMC Genomics: 131. doi:10.1186/1471-2164-14-131.

Jimenez, D. E., T.C. de Lima Lins, P.O. Taviera, and R.W. Pereira. 2014. “A Prospective Screening of Gene Copy Number Variation in Brazilian Admixed Population Sample.” *Hereditary Genetics* 3 (1): 1–5. doi:10.4172/2161-1041.1000125.

Kadoya, M., A. Yamamoto, M. Hamaguchi, H. Obayashi, and K. Mizushima. 2014. “Biochemical and Biophysical Research Communications Allograft Inflammatory Factor-1 Stimulates Chemokine Production and Induces Chemotaxis in Human Peripheral Blood Mononuclear Cells.” *Biochemical and Biophysical Research Communications* 448 (3). Elsevier Inc.: 287–91. doi:10.1016/j.bbrc.2014.04.106.

Kile, B. T., A. D. Panopoulos, R. A. Stirzaker, D. F. Hacking, L. H. Tahtamouni, T. A. Willson, L. A. Mielke, et al. 2007. “Mutations in the Cofilin Partner Aip1/Wdr1 Cause Autoinflammatory Disease and Macrothrombocytopenia.” *Blood* 110 (7): 2371–80. doi:10.1182/blood-2006-10-055087.

Lee, J. A., and J. R. Lupski. 2006. “Genomic Rearrangements and Gene Copy-Number Alterations as a Cause of Nervous System Disorders.” *Neuron*. doi:10.1016/j.neuron.2006.09.027.

Liu, G.E., and D.M. Bickhart. 2012. “Copy Number Variation in the Cattle Genome.” *Functional and Integrative Genomics* 12: 609–24. doi:10.1007/s10142-012-0289-9.

Liu, G.E., Y. Hou, B. Zhu, M.F. Cardone, L. Jiang, A. Cellamare, A. Mitra, et al. 2010. “Analysis of Copy Number Variations among Diverse Cattle Breeds.” *Genome Research* 20: 693–703. doi:10.1101/gr.105403.110.

Liu, G.E., M. Ventura, A. Cellamare, L. Chen, Z. Cheng, B. Zhu, C. Li, J. Song, and E.E. Eichler. 2009. “Analysis of Recent Segmental Duplications in the Bovine Genome.” *BMC Genomics* 10: 571. doi:10.1186/1471-2164-10-571.

Makina, S.O., F.C. Muchadeyi, E. van Marle-KÃster, M.D. MacNeil, and A. Maiwashe. 2014. “Genetic Diversity and Population Structure among Six Cattle Breeds in South Africa Using a Whole Genome SNP Panel.” *Frontiers in Genetics* 5: 1–7. doi:10.3389/fgene.2014.00333.

Matukumalli, L.K., C.T. Lawley, R.D. Schnabel, J.F. Taylor, M.F. Allan, M.P. Heaton, J. O’Connell, et al.

2009. “Development and Characterization of a High Density SNP Genotyping Assay for Cattle.” *PLoS ONE* 4. doi:10.1371/journal.pone.0005350.

McCarroll, Steven A. 2008. “Extending Genome-Wide Association Studies to Copy-Number Variation.” *Human Molecular Genetics* 17 (R2): 135–42. doi:10.1093/hmg/ddn282.

Mirkena, T., G. Duguma, A. Haile, M. Tibbo, a. M. Okeyo, M. Wurzinger, and J. Sölkner. 2010. “Genetics of Adaptation in Domestic Farm Animals: A Review.” *Livestock Science* 132. Elsevier B.V.: 1–12. doi:10.1016/j.livsci.2010.05.003.

Molin, A., J. Berglund, M.T. Webster, and K. Lindblad-Toh. 2014. “Genome-Wide Copy Number Variant Discovery in Dogs Using the CanineHD Genotyping Array.” *BMC Genomics* 15 (1): 210. doi:10.1186/1471-2164-15-210.

Morange, M. 2006. “HSFs in Development.” *Handbook of Experimental Pharmacology* 172: 153–69. doi:10.1007/3-540-29717-0-7.

Nabais, C., H. Teixeira, and S. Rodrı. 2014. “Genetic Diversity and Differentiation of Juniperus Thurifera in Spain and Morocco as Determined by SSR.” *PLoS ONE* 9 (2). doi:10.1371/journal.pone.0088996.

Narang, A., P. Jha, D. Kumar, R. Kutum, A. K. Mondal, D. Dash, and M. Mukerji. 2014. “Extensive Copy Number Variations in Admixed Indian Population of African Ancestry: Potential Involvement in Adaptation.” *Genome Biology and Evolution* 6 (12): 3171–81. doi:10.1093/gbe/evu250.

Nguyen, D. Q., C. Webber, and C. P. Ponting. 2006. “Bias of Selection on Human Copy-Number Variants.” *PLoS Genetics* 2 (2): 198–207. doi:10.1016/j.mssp.2006.01.060.

Nicholas, T.J., Z. Cheng, M. Ventura, K. Mealey, E.E. Eichler, and J.M. Akey. 2009. “The Genomic Architecture of Segmental Duplications and Associated Copy Number Variants in Dogs.” *Genome Research* 19 (3): 491–99. doi:10.1101/gr.084715.108.

O’Hare, T., C. A. Eide, and M. W. N. Deininger. 2007. “Bcr-Abl Kinase Domain Mutations, Drug Resistance, and the Road to a Cure for Chronic Myeloid Leukemia.” *Blood*. doi:10.1182/blood-2007-03-066936.

Peakall, R., and P. E. Smouse. 2012. “GenALEx 6.5: Genetic Analysis in Excel. Population Genetic Software for Teaching and Research-an Update.” *Bioinformatics* 28 (19): 2537–39. doi:10.1093/bioinformatics/bts460.

Peres-Neto, P.R., D.A. Jackson, and K.M. Somers. 2005. “How Many Principal Components? Stopping Rules for Determining the Number of Non-Trivial Axes Revisited.” *Computational Statistics and Data Analysis* 49 (4): 974–97. doi:10.1016/j.csda.2004.06.015.

Pienaar, L. 2014. “Genetic Diversity of the Afrikaner.” University of the Freestate.

Porto-Neto, L.R., A. Reverter, K.C. Prayaga, E.K.F. Chan, D.J. Johnston, Rachel J. Hawken, Geoffry Fordyce, et al. 2014. “The Genetic Architecture of Climatic Adaptation of Tropical Cattle.” *PLoS ONE* 9 (11): e113284. doi:10.1371/journal.pone.0113284.

Pritchard, J K, M Stephens, and P Donnelly. 2000. “Inference of Population Structure Using Multilocus Genotype Data.” *Genetics* 155 (2): 945–59. doi:10.1111/j.1471-8286.2007.01758.x.

Raymond, M, and F Rousset. 1995. “Genepop (Version-1.2) - Population-Genetics Software for Exact Tests

and Ecumenicism." *Journal of Heredity* 86: 248–49.

Redon, R., S. Ishikawa, K.R. Fitch, L. Feuk, H. George, T.D. Andrews, H. Fiegler, et al. 2006. "Global Variation in Copy Number in the Human Genome." *Nature* 444 (7118): 444–54. doi:10.1038/nature05329.Global.

Rege, J.E.O. 1999. "The State of African Cattle Genetic Resourcs I. Classification Framework and Identification of Threatened and Extinct Breeds." In *Animal Genetic Resources/Ressources Génétiques animales/Recursos Genéticos Animales*, edited by S. Galal and J. Boyazoglu, 25thed., 1–26. Roma, Italy: Food and Agriculture Organization of the United Nations (FAO) and the United Nations Environment Programme (UNEP).

Scharpf, R. B., T. H. Beaty, H. Schwender, S. G. Younkin, A. F. Scott, and I. Ruczinski. 2012. "Fast Detection of de Novo Copy Number Variants from SNP Arrays for Case-Parent Trios." *BMC Bioinformatics* 13: 330. doi:10.1186/1471-2105-13-330.

Sebat, J., B. Lakshmi, J. Troge, J. Alexander, J. Young, P. Lundin, S. Maner, et al. 2004. "Large-Scale Copy Number Polymorphism in the Human Genome." *Science* 305: 525–28. doi:10.1126/science.1098918.

Seroussi, E., G. Glick, A. Shirak, E. Yakobson, J.I. Weller, E. Ezra, and Y. Zeron. 2010. "Analysis of Copy Loss and Gain Variations in Holstein Cattle Autosomes Using BeadChip SNPs." *BMC Genomics* 11: 673. doi:10.1186/1471-2164-11-673.

Shah, N. P., J. M. Nicoll, B. Nagar, M. E. Gorre, R. L. Paquette, J. Kuriyan, and C. L. Sawyers. 2002. "Multiple BCR-ABL Kinase Domain Mutations Confer Polyclonal Resistance to the Tyrosine Kinase Inhibitor Imatinib (STI571) in Chronic Phase and Blast Crisis Chronic Myeloid Leukemia." *Cancer Cell* 2 (2): 117–25. doi:10.1016/S1535-6108(02)00096-X.

Sharp, A. J., D. P. Locke, S. D. McGrath, Z. Cheng, J. A. Bailey, R. U. Vallente, L. M. Pertz, et al. 2005. "Segmental Duplications and Copy-Number Variation in the Human Genome." *American Journal of Human Genetics* 77: 78–88. doi:10.1086/431652.

She, X., Z. Cheng, S. Zöllner, D. M. Church, and E. E. Eichler. 2008. "Mouse Segmental Duplication and Copy Number Variation." *Nature Genetics* 40 (7): 909–14. doi:10.1038/ng.172.

Stankiewicz, P., and J. R. Lupski. 2002. "Genome Architecture, Rearrangements and Genomic Disorders." *Trends in Genetics* 18 (2): 74–82. doi:10.1016/S0168-9525(02)02592-1.

Stankiewicz, P, and J R Lupski. 2010. "Structural Variation in the Human Genome and Its Role in Disease." *Annual Review of Medicine* 61: 437–55. doi:10.1146/annurev-med-100708-204735.

Stranger, B.E., M.S. Forrest, M. Dunning, C.E. Ingle, C. Beazley, N. Thorne, R. Redon, et al. 2007. "Relative Impact of Nucleotide and Copy Number Variation on Gene Expression Phenotypes." *Science* 315 (5813): 848–53. doi:10.1126/science.1136678.Relative.

Thornton, P. K., J. van de Steeg, A. Notenbaert, and M. Herrero. 2009. "The Impacts of Climate Change on Livestock and Livestock Systems in Developing Countries: A Review of What We Know and What We Need to Know." *Agricultural Systems* 101 (3). Elsevier Ltd: 113–27. doi:10.1016/j.agsy.2009.05.002.

Wang, K., M. Li, D. Hadley, R. Liu, J. Glessner, S.F.A. Grant, H. Hakonarson, and M. Bucan. 2007.

"PennCNV: An Integrated Hidden Markov Model Designed for High-Resolution Copy Number Variation Detection in Whole-Genome SNP Genotyping Data." *Genome Research* 17: 1665–74. doi:10.1101/gr.6861907.

Wang, M. D., K. Dzama, C. A. Hefer, and F. C. Muchadeyi. 2015. "Genomic Population Structure and Prevalence of Copy Number Variations in South African Nguni Cattle." *BMC Genomics*. BMC Genomics, 1–16. doi:10.1186/s12864-015-2122-z.

Xu, L., Y. Hou, D.M. Bickhart, Y. Zhou, E. Hamidi, J. Song, T.S. Sonstegard, C.P. Van Tassell, and G.E. Liu. 2016. "Population-Genetic Properties of Differentiated Copy Number Variations in Cattle." *Nature Publishing Group*. Nature Publishing Group, 1–8. doi:10.1038/srep23161.

Xu, Y., B. Peng, Y. Fu, and C. I. Amos. 2011. "Genome-Wide Algorithm for Detecting CNV Associations with Diseases." *BMC Bioinformatics* 12: 331. doi:10.1186/1471-2105-12-331.

Zhang, F., W. Gu, M.E. Hurles, and J.R. Lupski. 2009. "Copy Number Variation in Human Health, Disease, and Evolution." *Annual Review of Genomics and Human Genetics* 10: 451–81. doi:10.1146/annurev.genom.9.081307.164217.

Zhang, L., S. Jia, M. Plath, Y. Huang, C. Li, C. Lei, X. Zhao, and H. Chen. 2015. "Impact of Parental Bos Taurus and Bos Indicus Origins on Copy Number Variation in Traditional Chinese Cattle Breeds." *Genome Biology and Evolution* 7 (8): 2352–61. doi:10.1093/gbe/evv151.

Zhang, Q., Y. Ma, X. Wang, Y. Zhang, and X. Zhao. 2014. "Identification of Copy Number Variations in Qinchuan Cattle Using BovineHD Genotyping Beadchip Array." *Molecular Genetics and Genomics* 290 (1): 319–27. doi:10.1007/s00438-014-0923-4.

Whole Genome Sequencing Of South African Nguni Cattle: Copy Number Variation Prevalence And Genetic Diversity

Magretha D Wang[1,2], Kennedy Dzama [1], Charles A Hefer[2], Farai C Muchadeyi[2]

[1]Department of Animal Sciences, University of Stellenbosch, Private Bag X1, Matieland 7602, South Africa

[2]Biotechnology Platform, Agricultural Research Council, Private Bag X5, Onderstepoort 0110, South Africa

Chapter 5: Whole Genome Sequencing Of 24 South African Nguni Cattle: Copy Number Variation Prevalence And Genetic Diversity

5.1 Abstract

Copy number variations (CNVs) are modifications in DNA structure comprising of deletions, duplications and insertions. Prevalent in bovine genomes, CNVs have been designated as playing a role in adaptation and interindividual and between breed variations in cattle. South African Nguni cattle have undergone years of natural selection in harsh environmental conditions that has resulted in a breed that is well adapted to the abrasive conditions of Southern Africa. To date no next generation sequence data of any cattle breed from Southern Africa has been published. Next generation sequencing data has been deemed a suitable means of supplementing array based CNV studies, as breakpoints can be more accurately determined while analyses are not limited to predefined marker regions. CNVs may be prevalent within the genome as a reflection of the genetic diversity and adaptation of Nguni cattle. In this study, twenty four South African Nguni cattle were sequenced on the Illumina Nextera HiSeq 2500 at low to medium coverage. Paired end reads were trimmed and mapped against the UMD3.1 and Btau4.6.1 reference genomes using *TRIMMOMATIC* v0.33, *Burrows Wheeler Alignment* and *SAMTOOLS*. The average mapping percentage was 97,05 and 97,29 for UMD3.1 and Btau4.6.1 references. The recently developed *RAPTR-SV* was utilized to identify regions of variable copy number by means of hybrid split-read and paired end method. CNVs were filtered according to the number of reads that support the event with low stringency (F10), medium stringency (F45) and high stringency (F75). Adjacent and overlapping CNVs were merged to form 399, 55 and 23 unique CNVRs of between 1 kb and 1.59 Mb in length at F10, F45 and F75 respectively. All CNVRs identified by higher stringencies were picked up in lower stringencies. Comparisons with chapter 3 Bovine 50K Beadchip data from the same breed demonstrated notable discrepancies with considerably more CNVs of smaller size being reported by sequencing data which had greater precision of identifying breakpoints. CNVRs at F10, F45 and F75 covered or lay within 1Mb of 358, 51 and 23 genes respectively that represented a number of biological processes, cellular components and molecular functions. The F75 CNVRs shared a single gene with CNVRs identified using array data, while F10 shared 9 CNVRs genes with those reported from array data. The occurrence of CNVRs within regions of the Nguni genome involved in processes such as biological regulation, metabolic process and response to stimulus designate a possible correspondence of CNVR prevalence with adaptation traits.

Keywords

Breed diversity, Nguni cattle, genetic variation, adaptation

5.2 Background

Copy number variations (CNVs) are genomic segments of DNA that display a variable copy number exhibiting deletions, duplications and insertions larger than 1kb relative to a reference genome (Tuzun *et al.,*

2005). A number of recent studies demonstrated CNVs to be prevalent in bovine genomes and indicated a possible association of CNVs in adaptation (Liu and Bickhart, 2012; Liu *et al.,* 2010). South African Nguni cattle represent a distinct, conserved, Sanga type cattle breed that has undergone little synthetic breeding (Bester *et al.,* 2001; Makina *et al.,* 2014). Having endured natural selection pressures from a variety of disease agents and harsh climatic conditions, Nguni cattle have proven to prevail in suboptimal environmental circumstances (Marufu *et al.,* 2011).

Wang *et al.* (2015) (Chapter 3) performed the first CNV discovery analyses in South African Nguni cattle using the Illumina Bovine 50K Beadchip v2. Three hundred and thirty four CNVRs ranging from 30kb to 1Mb in size were detected in 231 of the 492 animals. Despite the bovine 50K Beadchip being a common method for CNV detection in cattle (Hou *et al.,* 2012; Liu and Bickhart 2012; Hou *et al.,* 2012), next generation sequencing (NGS) tools are able to complement SNP detection methods with increased coverage and resolution, better estimation of copy number and CNV breakpoints and with increased capacity to identify novel CNVs (Zhao *et al.,* 2013). Next generation sequencing has advanced into a method of choice for screening CNVs and incorporates thorough characterization of CNVs (Zhao *et al.,* 2013). Unlike array-based approaches, NGS platforms are not limited to predefined genomic regions, but rather sample at random from the entire genome while also eliminating ascertainment biases associated with SNP array methodologies (Medvedev *et al.,* 2009). The few genomic studies that have been performed in Nguni cattle have been limited to microsatellite and Bovine 50K Beadchip studies with no whole genome sequencing analyses (Horsburgh *et al.,* 2013; Makina *et al.,* 2014). Advantages of NGS methodologies, however include higher resolution and coverage, greater accuracy in estimating copy numbers, greater ability to detect novel CNV and superior detection of breakpoints (Alkan *et al.,* 2011; Cantsilieris *et al.,* 2013). The high occurrence of novel CNVs identified in chapter 3 suggested Nguni cattle to possibly have a great number of CNVs not yet characterized. Chapter 4 also indicated breeds to have breed specific regions of variation within a larger CNVR identified across breeds. The greater resolution and coverage of NGS technologies combined with an enhanced ability to identify novel CNVs made the supplementation of SNP based CNV identification within Nguni cattle logical.

With a variety of features that can be extracted from NGS data, a diverse set of tools have been developed for CNV detection and characterization from NGS data and have been reviewed by Zhao *et al.,* (2013). Briefly these can be classified into five different approaches, namely: paired-end mapping, split read, read depth, *de novo* assembly of a genome and a combinations of the above approaches (Zhao *et al.,* 2013). The majority of these methods do however exhibit a trade off between precision of variant detection and structural variant resolution/breakpoint detection (Bickhart *et al.*, 2015). Of the multiple software's developed, *PEMER* (Korbel *et al.*, 2009) and *PINDEL* (Ye *et al.*, 2009) contributed the best quality structural variant predictions to the human 1000 genome project (Mills *et al.*, 2011). These methods are however inclined to false positives. Bickhart *et al.* (2015), therefore recently developed a hybrid method of structural variant detection called *RAPTR-SV* that combines the split read and paired end algorithms of

PINDEL and *PEMER* to achieve greater accuracy with fewer false positives. This method was therefore used to identify CNVs in the genomes of 24 Nguni animals sequenced at an average of between 1.97 and 14.4X coverage and mapped to the UMD3.1 reference genome.

5.3 Materials And Methods

5.3.1 Sample Collection

Twenty animals were selected from chapter 3's dataset to be sequenced. Animals were selected according to their CNV prevalence based on SNP50K array data and are presented in Figure 5.1. Animals were chosen to represent a spread of those animals with high numbers of CNVs, intermediate numbers and no CNVs identified by the four different stringency models that were utilised in chapter 3. Consent was obtained to include data from an additional four animals that had been previously selected to represent Nguni cattle populations for sequencing purposes.

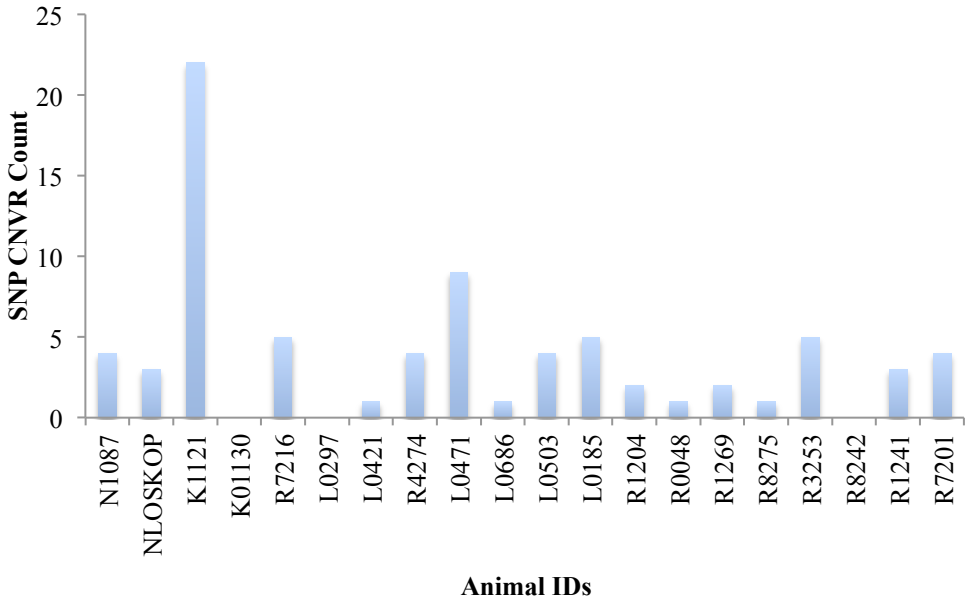

Figure 5.1 CNVR count for the 20 animals selected for sequencing from those analysed in chapter 3 using the Bovine 50K Beadchip.

5.3.2 Generation Of Sequence Data

DNA was extracted from hair samples collected from the 20 animals selected from the previous chapters' data set using the Macherey and Nagel NucleoSpin ® blood kit as per manufacturers instructions. Genomic DNA was extracted by means of the Qiagen DNeasy Blood and Tissue Kit from the blood samples of the four animals that were added to the study. DNA concentration for all samples was normalized to 250ng/µl and the paired-end Nextera DNA sample preparation kit was utilized to prepare paired-end libraries (Illumina inc., San Diego, CA, USA). Libraries were sequenced on the Illumina HiSeq 2500 platform at the Agriculture Research Council Biotechnology Platform. *FastQC* software was utilised to visualize read quality and *TRIMMOMATIC* v0.33 software was subsequently utilized to trim reads according to read

quality and length (Andrews 2010; Bolger *et al.,* 2014). Nextera Transposase adapter sequences were clipped. Three bases were removed from the 3' and 5' ends of each read and only reads of 50 bp or larger were kept. The Burrows Wheeler Alignment tool was utilized to align reads to the UMD 3.1 reference genome using default parameters in accordance with previous CNV sequencing publications (Li and Durbin 2010). *BWA* uses the Burrow-Wheeler Transform to efficiently and accurately align short reads against large reference genomes while allowing mismatches and gaps and supporting paired end reads (Li and Durbin, 2009). *SAMtools* v0.1.19-44428cd converted sam files into bam files and generated mapping statistics.

5.3.3 CNV

5.3.3.1 Identification

RAPTR-SV, developed by Bickhart *et al.* (2015) utilizes a hybrid split read and paired-end method for structural variation detection. The software utilizes an expansion of the paired-end discordancy algorithm of Hormozdiari *et al.*, (2010). This algorithm utilizes the mapping coordinates and orientations of both paired reads to identify insertions, deletions and duplications within the genome. In conjunction with the discordant read pair analyses, *RAPTR-SV* performs a split-read analyses adopted from Karakoc *et al.* (2011) to identify CNV breakpoints (Ye *et al.,* 2009). This method preselects targeted read pairs by selecting only one-end anchored read pairs, thereby reducing the typically high computation time of split read methodology. Bickhart *et al.* (2015) expanded this method to select target reads only from those that have more than 25% of their bases soft-clipped near the 5' or 3' terminal ends. Unmapped reads are cut in two, generating two separate half reads from which one half is aligned to the reference genome using *mrsFAST* alignment (Hach et al. 2010). When variant breakpoints exist at the exact division of the split read, both half reads will align to the reference genome and are termed balanced split reads. Unbalanced split reads on the other hand occur when the split read division does not cut exactly at the variant break point and hence has only one of the split reads aligning to the reference genome. Balanced and unbalanced reads covering the same breakpoint regions are grouped to form set of split reads that are then combined with prior discordant read sets such that breakpoint coordinates of parallel CNV calls can be determined.

RAPTR-SV involves a pre-processing step where *RepeatMasker* (http://www.repeatmasker.org/) software is utilized to mask repeats in the reference genome while *mrsFAST* short read alignment tool v2.0.5.4 (Hach *et al.,* 2010) is utilized to identify all read alignment positions in the reference genome in a "cache-oblivious" fashion (Bickhart *et al.,* 2015). Anchor reads, unmapped and soft clipped reads and discordant reads are categorized and stored in three metadata files. The cluster algorithm is then utilized to process the metadata, reading the data and assigning it to sets that are then filtered and sorted into deletions, insertions and tandem duplications (Bickhart *et al.,* 2015). Bioinformatics tools together with Microsoft Excel were utilized to organize and analyse the data. As CNVs only comprise variants larger than 1kb, all variants smaller than this were removed from the dataset. Next generation sequencing technologies have been demonstrated to generate more false positive CNV events and less false negative events. The *RAPTR-SV* indicates the number of discordant read pairs, balanced split read, unbalanced split reads and total weighted support of each CNV

identified. The elimination of CNVRs only supported by single reads is proposed to be suitable for CNVR detection (Bickhart *et al.,* 2015). CNVs were thus filtered at 3 different stringencies, namely low stringency F10, medium stringency F45 and high stringency F75.

A python script developed in house merged overlapping and adjacent CNVs to form CNVRs (Additional file 3.1). Pivot tables summarized data statistics. CNVs identified in animals were compared to those previously identified using the Bovine 50K Beadchip and reported in chapter 3. In order to identify genomic regions of overlap with CNVs reported by other authors in different cattle breeds, CNVRs of this study were run through the CNVR script together with CNVs using next generation sequencing technologies and reported by Choi *et al*. (2013), Bickhart *et al*. (2012), Liu *et al*. (2008) and Stothard *et al*. (2011). Regions of overlap across studies were thus determined and a Venn diagram was generated by means the online tool VENN (http://*Bioinformatics*.psb.ugent.be/cgi-bin/liste/Venn).

5.3.3.2 Gene Ontology Analyses

RefGene and RefLink annotations (USCS, downloaded on http://genome.ucsc.edu/goldenpath/gbdDescriptionsOld.html) were used to identify genes located within a 10Mb window surrounding a CNVR. Norris and Whan (2008) have shown that CNVs have a demonstrated effect on surrounding genes in a number of species. The hypothesis that genes were over or under represented in *PANTHER* pathways, biological processes, cellular components and molecular pathways was tested by means of the Bonferoni correction on the pantherdb.org. *Bos taurus* gene ontologies were ascertained by means of the *Ensembl* and *PANTHER* databases.

5.3.3.3 Correlation With Bovine 50K Beadchip Data

CNVRs identified were compared to CNVRs identified using the Bovine 50K Beadchip in chapter 3. CNVRs were tagged according to method used and the CNVR script developed in house was used on all CNVRs so as to delineated those CNVRs identified by both methods that overlapped or lay within close proximity of each other. Venn diagrams comparing CNVR genes and the subsequent biological processes, molecular functions and cellular components identified by next generation sequencing and the Bovine 50K Beadchip were generated.

5.4 Results And Discussion

5.4.1 Read Quality And *Its Variation Per Animal*

The *FastQC* report for the trimmed reads 1 and 2 demonstrated 21 of the animals to comprise good quality reads. Animals R03253, K01130 and R04274 demonstrated questionable per base sequence content, per base GC content, sequence duplication levels and overrepresented sequences. Low quality data of these three animals was further reflected in low mapping percentages (Table 5.1). All animals were sequenced at the same sequencing platform although sequencing was performed on different days according to the schedule of

the sequencing platform used. Discrepancies in read quality could thus be expected with a variety of factors, including DNA storage duration and quality and possible contamination.

5.4.2 Whole Genome Sequencing

Twenty-four Nguni animals were aligned to the UMD3.1 reference genomes using *Burrows Wheeler Alignment* and *SAMtools* after adapters were removed and reads were trimmed for length and base quality using *FastQC* and *TRIMMOMATIC* v0.33. Whole genome mapping percentages ranged from 36.28 to 99.79 percent. Although animals were sequenced with target coverage of 10x coverage ranged from 2.05x to 14.40x, with an average of 7.08x.

The Nguni comprise a Sanga type breed that was a result of crosses between South Asian *Bos indicus* bulls and African Taurine cows (Gifford-Gonzalez and Hanotte 2011). The UMD3.1 represents one of the most well annotated bovine *Bos taurus* assemblies that is currently available (The Bovine Genome sequencing and analysis consortium *et al.,* 2009; Zimin *et al.,* 2009). It would thus be expected that some degree of discrepancies would exist between mapping percentages to the Sanga and Taurine genomes. Despite this, fairly high mapping percentages were achieved across animals with the odd exception of a few animals where a number of factors such as poor quality DNA may have played a greater role (Table 5.1). Mapping percentages reflect those reported in CNV detection studies in cattle (Choi et al. 2013; Jansen et al. 2013). The three animals that exhibited low quality data and subsequent lower mapping percentages were excluded from further CNV analyses.

Table 5.1 Mapping statistics for 24 Nguni animals mapped to the UMD3.1 reference genome with animal ID (ANML ID), coverage and its standard deviation (Av. COV), mapping percentage against the UMD3.1 reference genome (UMD MAP%), properly paired read percentage (PPR PRR %) and singleton percent (SNGLTN %).

ANML ID	Av. COV	UMD MAP %	PPR PRP %	SNGLTN %
1130	4.29	78.28	75.82	0.44
7216	12.15	93.68	91.21	0.08
1121	7.29	98.73	95.98	0.08
0185	7.83	99.52	96.94	0.07
0297	7.55	98.98	96.53	0.08
0421	11.19	99.67	98.18	0.04
0471	6.42	99.60	97.33	0.05
0503	13.09	99.37	96.26	0.09
0686	6.28	98.94	46.92	0.65
1087	4.66	96.43	96.43	0.07
NLOSKOP	2.05	99.06	97.84	0.02
0048	5.45	99.79	96.52	0.08
1204	11.27	99.57	95.77	0.12
1241	11.17	99.71	95.17	0.09
1269	4.33	99.61	96.27	0.06
3253	2.30	55.94	53.36	0.14
4274	1.75	36.28	35.51	0.14
7201	4.45	99.53	95.01	0.10
8242	9.90	99.25	95.62	0.09
8275	14.40	94.80	92.84	0.10
NG6	7.17	99.64	98.13	0.10
KZN	4.25	92.82	91.10	0.07
5990	6.42	98.45	97.11	0.02
9363	4.33	93.34	91.44	0.08

5.4.3 CNVs

5.4.3.1 Identification

RAPTR-SV was utilized to identify CNVs in 21 Nguni animals mapped to the UMD3.1 reference with a mapping percentage of greater then 90%. Prior to filtering for length and applying filtering stringencies, 32 667 structural variants were identified of which a great number were filtered out leaving 543, 109 and 69 CNVs greater than 1kb in size at stringencies F10, F45 and F75 respectively across autosomes. More deletion events were identified across stringencies (Table 5.2) with the largest CNV events were deletion events. CNVs were between 1kb and 0.92Mb in size. This is a broader range than those reported by Stothard *et al.* (2011) in 2 animals (1 Holstein and 1 Black Angus) that ranged between 1.84 and 28 kb in size. Considering the greater number of animals in the present study, this can be expected. Only 13 of the 17

animals demonstrated CNVs at the lowest stringency, which was further reduced to 11 animals at F45 and only 6 animals at F75.

Table 5.2 Summary statistics of unique CNV deletions and duplication events (CN). The number of animals with the CNV (CNVs), the minimum length (MinL), maximum length (MaxL) and average length (AvL) of CNVs at 3 different filtering stringencies (Filter).

Filter	CN	CNVs	ANML	MinL	MaxL	AvL
F10	Deletion	387	13	1 033	866 481	87 901.48
	Duplication	156	13	1 005	921 664	93 456.80
F45	Deletion	69	11	1 091	708 995	81 776.16
	Duplication	40	8	1 005	691 751	62 442.35
F75	Deletion	42	6	1 091	708 995	53 974.90
	Duplication	27	6	1 005	691 751	67 039.26

Adjacent and overlapping CNVs were joined to form 185, 21 and 10 CNVRs at F10, F45 and F75 (Table 5.3). Across filtering stringencies CNVRs ranged between 1kb and 1.6Mb in length. The smallest CNVRs were tandem duplication events. The smallest deletion was identified within the lowest stringencies group. The greater filter stringency corresponded with larger CNVR events with deletion and duplication average lengths being 242 and 412 kb for F75 deletion and duplication events while only 69 and 141 kb for F10 deletion and duplication events (Table 5.3) respectively. Average CNVR lengths reported in cattle breeds by means of sequencing analyses ranged from 10.03 and 7.18 kb reported by Choi *et al.* (2013) for Black Angus Hanwoo and Holstein Hanwoo respective genome comparisons. Bickhart *et al.* (2012) reported 1265 CNVRs with an average length of 49.1kbp in 3 Angus, 1 Holstein, 1 Hereford and 1 Nellore cattle of which 476 comprised of novel CNVR events not previously reported. Shin *et al.* (2014) reported 6 811 CNV loss events in the genome of 10 Holstein and 22 Hanwoo beef cattle sequenced at a coverage of between 13.58 to 20 fold. Of these deletion events, 4 407 events were found in both Hanwoo and Holstein animals (Shin *et al.,* 2014). Deletion and duplication CNVs were treated separately such that CNV loss regions and CNV gain regions were identified independently. Although none of the deletion CNVRs and duplication CNVRs shared exact breakpoints, some overlap in duplication and deletion regions was evident, demonstrating possible complex CNVRs (Additional file 5.1). CNVR deletion and duplication events make up to 28.87 and 14.23 Mb of the bovine genome respectively. This reflects the 28.1 Mb reported in the first CNV sequencing analyses performed in bovine (Liu *et al.,* 2010). The greater preponderance of deletion events reported here is in accordance with literature that has demonstrated duplications to be more difficult to detect by means of NGS technologies (Teo *et al.,* 2012).

Table 5.3 Summary statistics of unique CNVR events for F10, F45 and F75 stringencies. The number (CNVRs), minimum length (MinL), maximum length (MaxL) and average length (AvL) of CNVRs together with the number of genes (GEN) within 10Mb of the CNVR.

Filter	CN	CNVRs		MinL	MaxL	AvL	SumL	GEN	
F10	Deletion	232	327	1 123	1 595 205	28 867 168	124 427.45	179	250
	Duplication	95		1 207	1 250 141	14 232 713	149 818.03	71	
F45	Deletion	28	43	1 183	1 570 699	4 619 188	164 971.00	19	33
	Duplication	15		4 180	1 213 929	2 112 695	140 846.33	14	
F75	Deletion	14	23	1 091	1 564 609	2 093 376	149 526.85	12	17
	Duplication	9		10 900	1 213 929	2 522 356	280 261.78	5	

The depth and breadth of sequence coverage is directly related to the sensitivity and specificity of variant detection (Koboldt *et al.,* 2010). Bickhart *et al.* (2015) recommend 10x coverage to however be sufficient for CNV identification using *RAPTR-SV* which utilizes a hybrid split read and paired-end mapping technique. Coverage range between 4 and 8X for this dataset. No noticeable correlations between coverage and CNV presence is however recognized in the data. Animal 7201, with the greatest number of CNVs at F10, had an average coverage of 4.45 that is close to the overall average for the dataset (Table 5.4). GC content of sequence reads ranged from between 42 to 48%, while whole genome coverage was between 2.05x and 14.40x (Table 5.1). Although a 10x coverage was targeted for sequencing, some animals achieved a considerably lower coverage. This may be as a result of contamination or a high number poor quality reads that failed QC. GC content has been observed to have a unimodel relationship with the depth of coverage, with regions of high or low coverage corresponding to reduced coverage (Abyzov *et al.,* 2011; Yoon *et al.,* 2009). Regions of low depth of coverage may have insufficient reads to discern copy number variants using split read and paired end mapping detection methodologies, however the depth of coverage method for CNV detection is the most affected by GC content bias (Teo *et al.,* 2012). A four fold plus depth of coverage is, however sufficient for read depth approaches (Alkan *et al.,* 2009; Mills *et al.*, 2011) and low to medium coverage of animals explains a major fraction of genomic variants (Jansen *et al.,* 2013). Coverages obtained in the 23 animals used for CNV detection reflect prior CNV detection studies. Bickhart *et al.* (2012) used a read depth approach to identify CNVs in cattle from different breeds at sequence coverages ranging from between 4 and 19x. (Stothard et al. 2011) report sequence coverage depth of 19 and 22 fold in the two bulls resequenced for SNP and CNV detection. Choi *et al.* (2013) reported CNVs identified in cattle sequenced at 10x, 17x and 57x. Jansen *et al.* (2013) reported variants in 43 key animals sequenced at coverages ranging from between 4.17 and 24.98 fold with an average of 7.46 x.

Of the 21 animals assessed for CNVs, only 13 animals exhibited CNVs on the 29 autosomes at stringency F10 (Table 5.4). Up to 143 unique CNVRs were detected in a single animal at lowest CNVR stringency. Animals 5990 and 68 contained the least CNVs that were between 200 780 and 293 949 bps in length, only detected at the lowest stringency. The two CNVs identified in animal 5990 were also identified in animal 68.

Table 5.4 Summary statistics of unique CNV deletions and duplication events for each of 13 Nguni cattle. The animal identity (ANML), number of CNVRs at F10 (F10), F45 (F45) and F75 (F75) and the minimum length (MinL), maximum length (MaxL) and average length (AvL) of CNVRs.

ANML	Cov	F10	F45	F75	MinL	MaxL	AvL
0048	5.45	12	1	0	1 272	294 751	187 901.00
1269	4.33	31	2	0	1 207	380 474	76 203.08
5990	6.42	2	0	0	200 780	208 534	204 657.00
7201	4.45	143	15	5	1 207	1 595 205	483 037.88
9363	4.33	25	4	3	4 098	1 151 793	136 833.77
l121	7.29	4	2	2	13 933	200 780	54 493.26
KZN	4.25	13	4	5	1 183	1 151 793	95 639.10
0471	6.42	11	1	0	1 420	294 751	173 709.45
68	6.28	3	0	0	200 780	293 949	226 010.75
85	7.83	14	2	0	1 537	294 751	160 746.13
97	7.55	24	2		1 272	294 751	131 603.75
1087	4.66	10	4	3	1 091	1 151 793	113 120.46
NG6	7.17	35	6	5	1 123	1 151 793	123 727.80

An analysis of chromosomal distribution of CNVRs demonstrated noticeable variation (Figure 5.2). The uncharacterized chromosomes, that comprise of sequences that cannot be uniquely mapped to the genome (Liu *et al.*, 2009), contained a large number of variation events. To eliminate confusion, only autosomes were thus assessed. Little correlation between CNVR distribution and chromosome length was evident in accordance with prior findings (Fadista *et al.*, 2010). At the highest stringency, deletion events were identified on chromosomes 6, 11, 17, and 25 while duplication events were only present on chromosomes 6 and 17. The largest CNVR deletion and duplication events were identified in the same individual 7201 on chromosome 6 (Additional file 5.1). A single genomic region on chromosome 17 was identified in 5 animals at F75 stringency as comprising both deletion and duplication events in all 5 animals. Chromosome 17 demonstrated the greatest number of CNVRs occurring in multiple individuals. This chromosome also demonstrated the greatest number of animals (93) exhibiting a single CNVR in chapter 3 (Additional file 3.3) and the most common CNVR across the 7 South African breeds (Figure 4.2) exhibiting a variation in copy number between bases 73 118 011 and 74 998 349. The smallest CNVR comprising of a bp tandem duplication was found on chromosome 17 which also demonstrated the greatest number of CNVRs across filtering with 41, 5 and 4 deletions and 8, 3 and 2 duplication events at stringencies F10, F45 and F75 respectively.

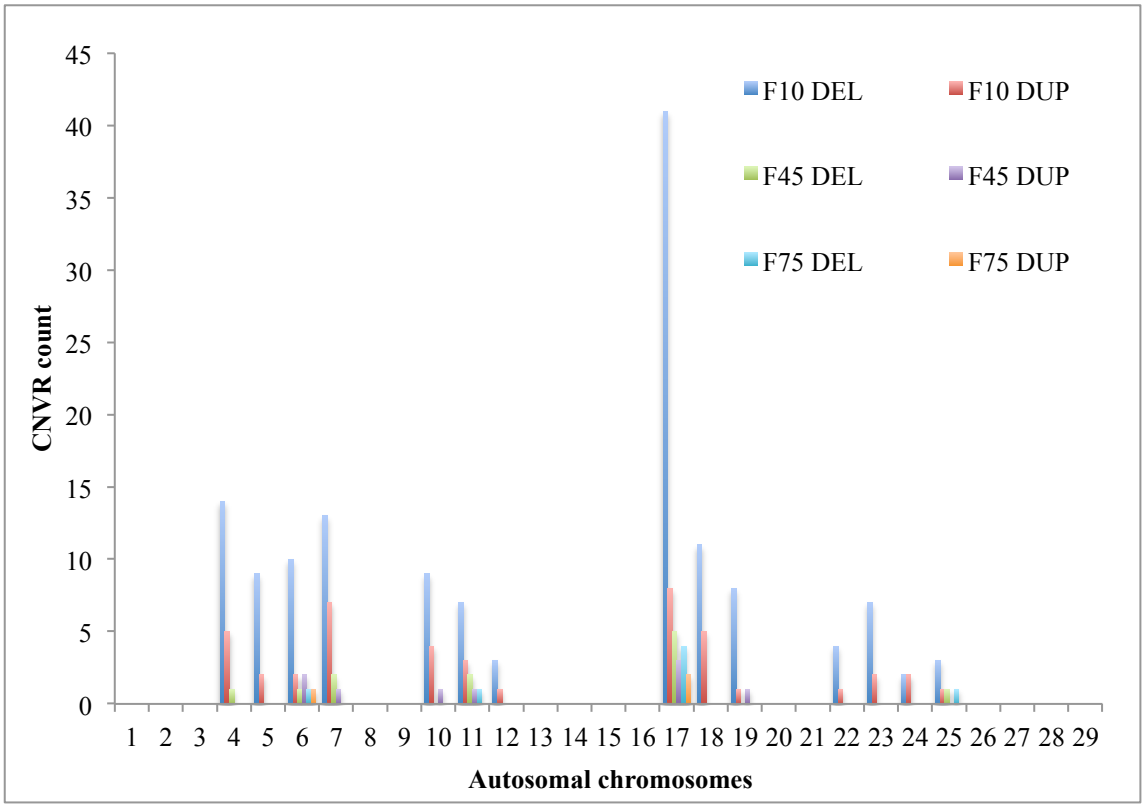

Figure 5.2 Autosomal chromosome distribution of deletion and duplication CNV events identified in 22 Nguni cattle.

5.4.3.2 CNVRs Common In SA And *Global Breeds From Literature*

Regions varying in copy number across breeds reported in various NGS studies by Choi *et al.* (2013), Bickhart *et al.* (2012), Liu *et al.* (2010) and Stothard *et al.* (2011). were investigated. This study demonstrated the greatest amount of overlap with that of Choi *et al.* (2013). Fourteen variable regions identified in this study were reported exclusively by Choi *et al.* (2013) in Chinese Holstein cattle (Figure 5.3). Although Chinese Holstein cattle were introduced into China, introgression of native Asian cattle breeds into the breed was evident (Ferreri *et al*., 2011). The presence of CNVRs common exclusively between the native South African Nguni and this breed could indicate possible regions of adaptive response. CNVR genes common between these two breeds include such as the *b-cell antigen receptor complex-associated protein alpha chain* (*CD79A*) involved in the *B-cell receptor complex*, *alpha-1b adrenergic receptor* (*ADRA1B*) involved in positive regulation of the force of heart contraction by *epinephrine-norepinephrine biological process* and *interleukin-12 subunit beta* (*IL12B*), *forkhead box protein P1* (*FOXP1*) and *interleukin-15* (*IL15*) involved in T cell differentiation. A single region lying on chromosome 23 between basepairs 28455706 and 28601986 demonstrated variation in this study as well as in that of Choi *et al.* (2013), Bickhart *et al.* (2012) and Liu *et al.* (2010). A CNVR on chromosome 6 and another on chromosome 12 reported in this study overlapped with CNVs reported by Bickhart *et al.* (2012), Stothard *et al.* (2011) and Liu *et al.* (2010).

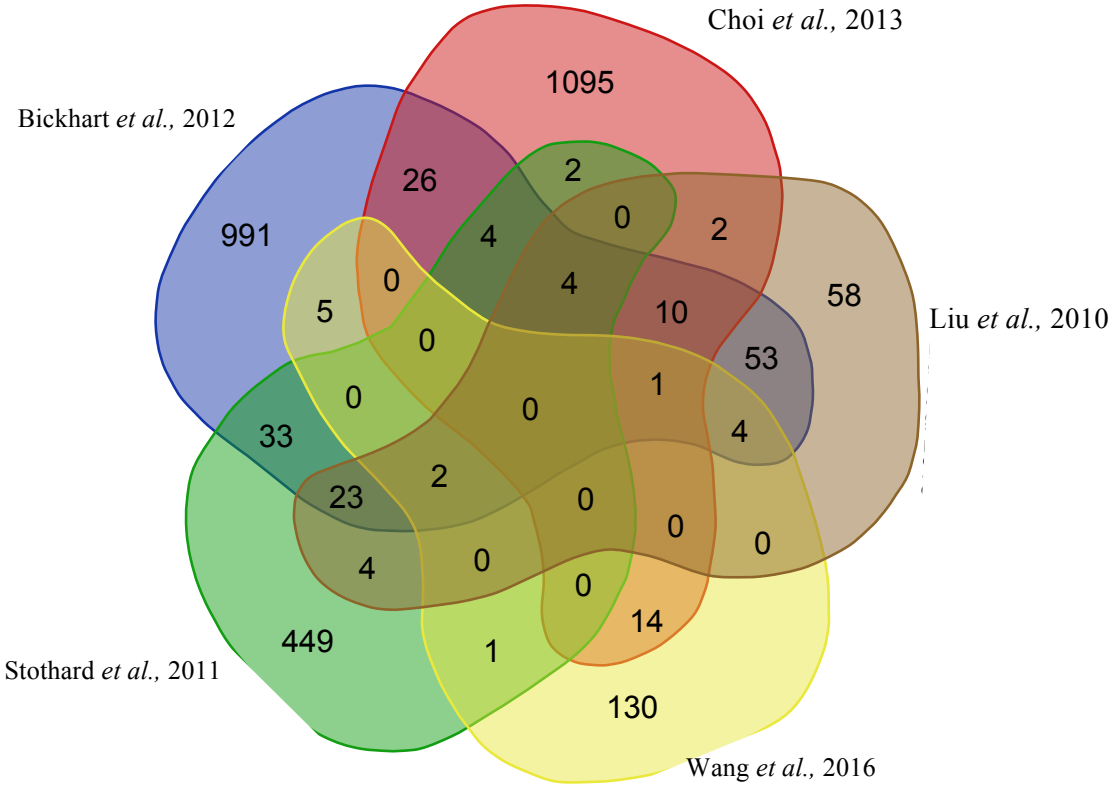

Figure 5.3 Venn diagram demonstrating the overlap of CNVRs identified in this study and that of Choi *et al.* (2013), Bickhart *et al.* (2012), Stothard *et al.* (2011) and Liu *et al.* (2010).

5.4.3.3 CNVR Gene Ontology

CNVRs overlapped or lay within 10Mb proximity of 218, 21 and 12 genes at F10, F45 and F75 respectively. This is considerably less than the 1 228 genes covered by deletion events reported by Shin *et al.* (2014). A number of genes involved in adaptive phenotypes like immune response and metabolic process were represented within CNVRs (Figure 5.4). Nguni cattle are recognized for their innate ability to survive in harsh environmental conditions (Mapiye *et al.,* 2009; Tada *et al.,* 2013). Parasite and disease resistance together with heat tolerance traits are among some of the phenotypes evident that enable this breed of cattle endure extreme climatic conditions and elevated levels of parasite and disease threats. *PANTHER* overrepresentation analyses demonstrated no specific biological processes, cellular components and molecular functions to be significantly ($p<0.05$) overrepresented by CNVR genes. A number of biological processes, cellular components, molecular functions and proteins of interest were, however represented by CNVR genes (Figure 5.4, 5.5, 5.6 and 5.7). A number of the processes, functions, components and proteins playing a role in adaptation have been reported in other CNV studies in cattle. Both Bickhart *et al.,* 2012)and Hou *et al.*, (2012) report immune system process, response to stimulus and metabolic process to be represented by CNVR genes in multiple cattle breeds. A number of proteins involved in adaptive responses are also represented within the CNVR genes reported here (Figure 5.7). These include defence/immunity proteins, signalling molecule proteins, kinases, receptors and phosphatase. This is not the first instance of

CNVRs covering defence/immunity proteins in cattle (Bickhart *et al*., 2012; Hou *et al*., 2012; Hou *et al*., 2012). Molecular functions represented by CNVR genes include catalytic activity, receptor activity, enzyme regulator activity and translation regulation activity.

A number of instances exhibiting a gene covered or within a 10Mb range of both a deletion and duplication event are evident (Additional file 5.2). *Anaphase-promoting complex subunit 10* (*ANAPC10*), *hedgehog interacting protein* (*HHIP*) and *kelch-like family member 2* (*KLHL2*) demonstrated the highest prevalence, occurring in 13 and 10 of the 17 animals in proximity of both deletion and duplication events. Both *ANAPC10* and *HHIP* genes were covered or lay within close of CNVRs reported by Choi *et al.* (2013) in Chinese Holstein cattle. *ANAPC10* gene plays a role in a number of biological processes that include mitotic nuclear division, cell division and protein ubiquitination (Additional file 5.3). *HHIP*, on the other hand plays a role in carbohydrate metabolic process, regulation of fibroblast growth factor receptor signalling pathway, signal transduction and oxidation-reduction process among other biological processes (Additional file 5.3). In humans, genome wide association analyses have implicated *HHIP* in chronic obstructive pulmonary disease, a complex disease with a strong influence of genetic predisposition and cigarette smoking (Zhou *et al.*, 2012). Molecular functions of HHIP include Zinc ion binding, quinon binding catalytic activity and hedgehog family protein binding. Next generation sequencing analyses in Hanwoo and Nguni cattle demonstrated breeds to both have variable copy numbers that cover the *carbonic anhydrase 10* (*CA10*) gene. *CA10* is one of the 16 carbonic anhydrase isoforms of the mammalian carbonic anhydrases (Aspatwar *et al.,* 2010). Human and mouse studies demonstrate carbonic anhydrase related proteins to play a significant role in neural functions and/or brain development. *CARP* genes comprise highly conserved genes across species, with *CA10* being universal across the animal kingdom (Aspatwar *et al.,* 2010).

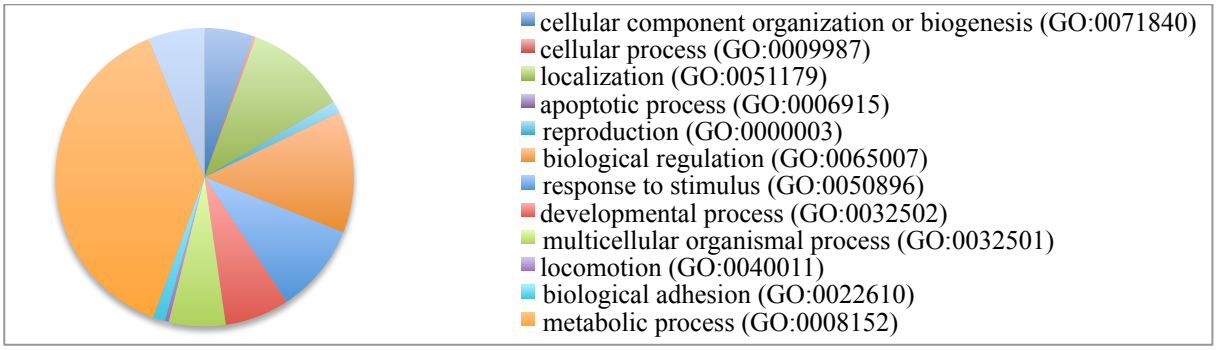

Figure 5.4 The distribution of CNVR genes across biological processes identified in 22 Nguni animals.

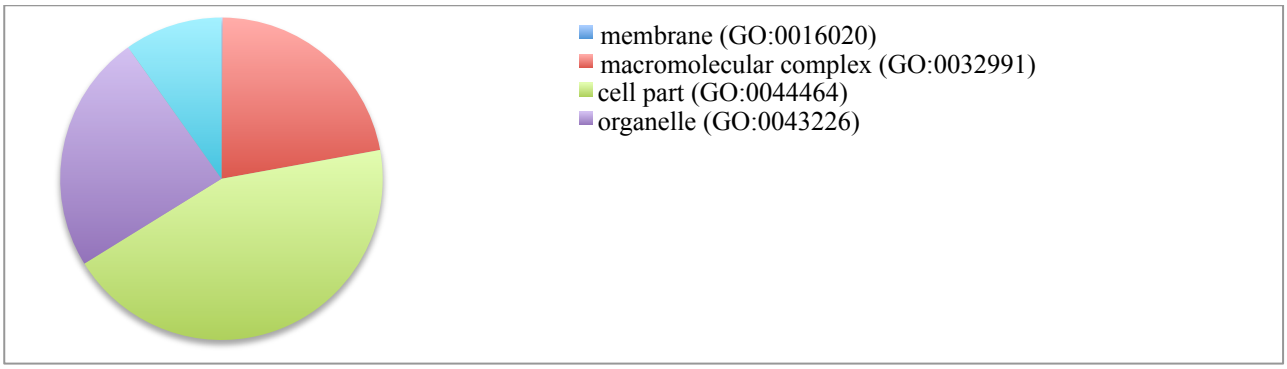

Figure 5.5 The distribution of CNVR genes across cellular components identified in 21 Nguni animals.

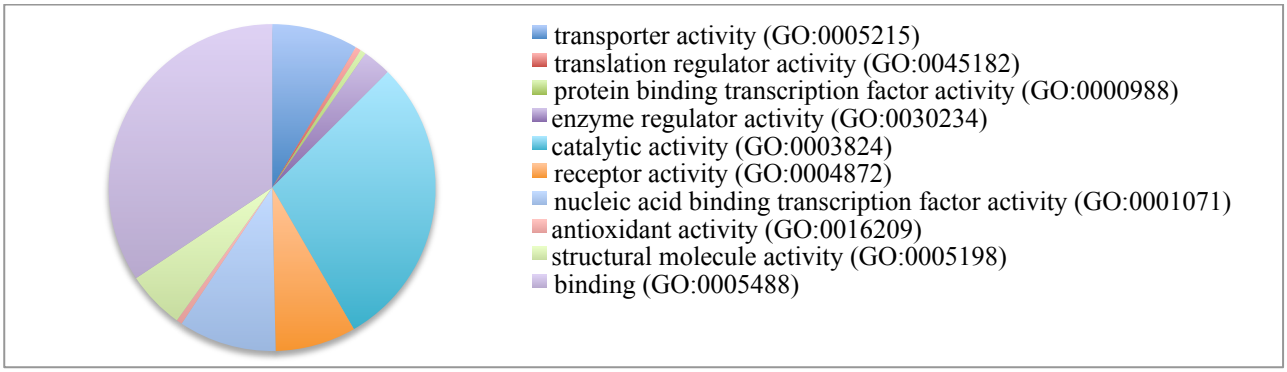

Figure 5.6 The distribution of CNVR genes across molecular functions identified in 21 Nguni animals.

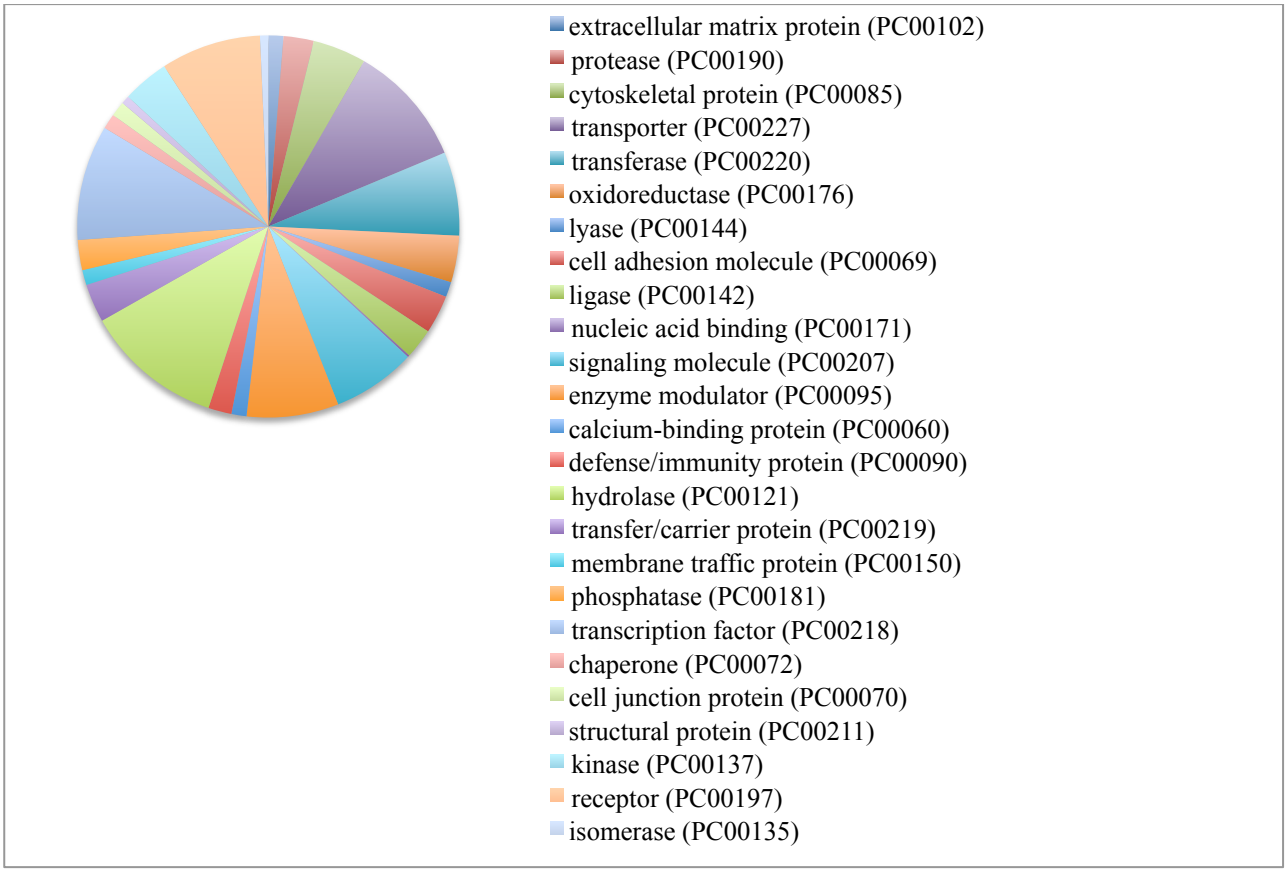

Figure 5.7 The distribution of CNVR genes across proteins identified in 21 Nguni animals.

The *insulin-like growth factor 1 binding protein 3* (*IGFBP3*) gene which has been associated with some growth and development traits in multiple Chinese beef cattle breeds (Gao *et al*., 2009), lay within close

proximity of a CNVR identified in Nguni's (Additional file 5.4). *IGFBP3* is more highly expressed in the mesenteric lymph node of cattle resistant to intestinal nematodes (Araujo *et al*., 2009). Hou *et al.* (2010) report an association between CNV prevalence nematode resistance in Angus cattle with nematode resistant animals demonstrating CNV associations with *gamma-aminobutyric acid type a receptor alpha 2 subunit* (*GABRA2*), *gamma-aminobutyric acid receptor subunit beta-1* (*GABRB1*) and *peroxiredoxin-2* (*PRDX2*) among other genes. In this study, variations in copy numbers in genomic regions of *GABRB2* and *PRDX2* genes were observed in the Nguni cattle. Mutations in the FBXW7 gene that encodes a number of the F-box protein family, have been detected in ovarian and breast cancer cell lines in humans, and have been implicated in the pathogenesis of human cancers (Uddin *et al*. 2016; Heo *et al.,* 2016). CNVRs covering both *f-box* and *WD repeat domain containing 7* (*FBXW7*) and *f-box/WD repeat-containing protein 9* (*FBXW9*) genes were identified in this study in Nguni cattle. Three animals demonstrated deletion events that covered or lay within close proximity of the *CD79A* gene found on chromosome 18 which plays a role in adaptive immune response and B cell activation, differentiation, proliferation and receptor signal pathways. (Gautier *et al*., 2009) reported *CD79A* to be under strong positive selection in West African cattle where infectious and parasitical parameters are considered to have been the primary pressures driving selection. Other CNVR genes involved in immune response processes include *thrombospondin-1* (*THBS1*), *interleukin-12 subunit beta* (*IL12B*), *interleukin-15* (*Il15*) and *IL27RA protein* (IL27RA) (Additional file 5.3) Studies in humans demonstrated variations in the copy number of *IL2B* together with that of T-beta genes to be associated with risk of developing systemic lupus erythematous, a systemic autoimmune disease (Yu *et al.,* 2013). Chickens bred for both high and low antibody response, when exposed to cold stress demonstrated equally enhanced expression of *IL12B* (Biscarini *et al.*, 2010). *IL12B* is one of two genes encoding for interleukin 12, a heterodimic cytokine that is generated in response to antigenic stimulation, playing an integral role in immunity (Trinchieri and Gerosa 1996). Early studies in young calves, report *interleukin-12* (*IL-12*) together with *interferon-γ* (*IFN-γ*) and inducible nitric oxide synthase mRNA expression to be involved in the immunity of calves to the tick borne, haemoparasitic disease babesiosis (Goff *et al*., 2001). This was further validated by Aguilar-Delfin *et al.,* (2003) in a study done in genetically modified mice, where early production of *IL-12* and *IFN-γ* together with the production of macrophage-derived effector molecules like nitric oxide played a vital role in opposing acute babesiosis. *Serpin family B member 6* (*SERPINB6*), is one of 4 *SERPIN* genes more highly expression in healthy bovine follicle than atretic follicles (Hayashi *et al*., 2011).

An animal's ability to acclimatize to a changing environment is an important factor in deferring heat stress that ultimately results in impaired liver function and reproductive performance while causing oxidative stress and jeopardizes the immune response (Bernabucci *et al*., 2010). Wang *et al*. (2015) proposes the possible variation in gene copy number of heat shock protein and transcription factor genes to possibly play a role the variation in the climatic adaptability of different cattle breeds. Recent studies in Sahiwal cattle demonstrate a SNP at the *heat shock protein family B* (*small) member 8* (*HSPB8*) locus to play a role in their ability to tolerate heat (Verma *et al*., 2016). *HSPB8* presented as a deletion in Nguni cattle (Additional file 5.4). (Kijas

et al. (2011), Bickhart *et al.* (2012) and Cicconardi *et al*. (2013) report CNVs covering heat shock transcription factor (*heat shock transcription factor 1* (*HSF1*) and *heat shock transcription factor 4* (*HSF4*)) and heat shock protein genes (*heat shock protein 1* (*HSP1*), *heat shock protein family A member 6* (*HSPA6*), *heat shock protein family A* (*Hsp70) member 12B* (*HSPA12B*), *alpha haemoglobin stabilizing protein* (*AHSP*) and *heat shock protein family B* (*small) member 1* (*HSPB1*)). A number of genes involved in various metabolic functions were identified within or lying close to CNVRs identified (Additional file 5.3). Long term heat acclimatization and hence thermal tolerance is distinguished by an enhanced efficiency in metabolic processes and signalling pathways which may primarily be mediated by heat shock proteins and altered gene expression (Bernabucci *et al.,* 2010; Horowitz, 2002). Fibroblast growth factor genes, solute carrier protein genes, interleukin and tick resistant genes have also demonstrated importance in thermal stress (Collier *et al.*, 2008; Collier *et al.,* 2006). Fibroblast growth factor genes were identified within close proximity of CNVRs reported here (Additional file 5.4). These were *protein sprouty homolog 1* (*SPRY1*), *thrombospondin-1* (*THBS1*) and *hedgehog interacting protein* (*HHIP*) protein involved in the regulation of fibroblast growth factor receptor signalling pathway and *40S ribosomal protein S19* (*RPS19*) involved in fibroblast growth factor binding. *THBS1* also plays a role in the activation of MAPK activity. The mitogen activated protein kinases comprise protein kinases involved in directing cellular responses to a diverse array of stimuli, including osmotic stress and heat shock. Other CNVR genes (*TAO kinase 3* (*TAOK3*), *adrenoceptor alpha 1B* (*ADRA1B*), *insulin like growth factor binding protein 3* (*IFGBP3*), *purinergic receptor P2X 7* (*P2RX7*) and *dual specificity phosphatase 18* (*DUSP18*)) involved in various aspects of MAPK activity were also identified (Additional file 5.3).

5.4.3.3 Correlation With Bovine 50K Beadchip Data

Chapter 3 reported CNVs identified in Nguni cattle by means of the Bovine 50K Beadchip. Findings demonstrate a similar pattern with only 231 of the 492 animals presenting CNVs within their genome. No exact CNV was, however detected by both this NGS study and that of the prior studies using the 50K beadchip. This reflects prior findings by authors who propose the respective biases of the two methodologies make for complementary results that generate a more conclusive picture. Although array comparative hybridisation and low-resolution NGS technologies have demonstrated consistent results with CNV identification, beadchip and next generation data are yet to exhibit the same level of consistency for CNV detection (Hayes *et al*., 2013). Although popular for their lower cost, beadchip array approaches are evidenced to have shortcomings that include limited genome coverage and hybridisation noise (Carter 2007). Array based analyses have been suggested to have difficulty identifying novel or rare variants, chapter 3 however reports a high number of variants that have not yet been characterized within cattle breeds (Additional file 3.4). CNVs identified using the Bovine 50K Beadchip were considerably larger than those identified by means of NGS technologies. Next generation sequencing acts as a suitable complementary methodology, with greater accuracy in identifying breakpoints while also allowing for a more detailed characterization of CNVs (Mills *et al*., 2011). For the most part, different CNVRs were identified by the two methodologies. CNVRs identified by sequencing technologies comprised of smaller regions, that were either

covered by no 50K beadchip SNPs, few 50K beadchip SNPs or were covered by 50K SNPs that were filtered out of the SNP analyses during quality control (Additional file 5.2). Both SNP and NGS analyses demonstrate a greater propensity for identifying deletions (Table 3.4 and Table 5.1). Detecting duplications is more challenging for both methodologies (Teo *et al*., 2012). For paired end mapping methods, duplications/insertions are detected when the mapped reads are placed at a distance shorter than the fragment length (Hormozdiari *et al.,* 2009). Insertions larger than the insert size of the reference library are therefore undetectable (Hormozdiari *et al.,* 2009). One region on chromosome 4 and another on chromosome 17 demonstrated CNVR overlap between the Nguni NGS and Nguni SNP studies. These specific CNVR were identified in single animals for the SNP data as a deletion (chr4) and duplication (chr17) and as a duplication (chr 4) and deletion (chr 17) for the NGS studies.

Table 5.5 depicts the number of genes that overlap between chapters in this study. Chapter 3 exhibits the greatest degree of overlap with this chapter. This is not surprising considering both chapters were in Nguni cattle only. A single gene, the *smoothelin* (*SMTN*) gene is reported in all chapters. *SMTN* gene is a protein coding gene alternately expressed from distinct promoters to produce two separate structural proteins found predominantly in visceral (SMTN-A) and vascular (SMTN-B) smooth muscle (Rensen *et al.,* 2002). Although little is currently understood about the SMTN protein family, knock-out mouse models demonstrate these proteins to provide critical contributions to normal muscle function (Niessen *et al.,* 2005; Rensen *et al.,* 2002; Wooldridge *et al.,* 2008). Blood flow in resistance vascular and systemic blood pressure are regulated by vascular smooth muscle contraction and myogenic responses (Turner and Macdonald 2014). The SMTN family proteins are along a number of regulatory smooth muscle proteins involved in fine tuning the myogenic response and facilitating the adaptations of vascular physiologies (Turner and Macdonald 2014). A deletion of the SMTN-B protein is characterized by an altered vascular phenotype (Bär *et al.,* 2002), while a knock-out mouse model of both the *SMTN* proteins is associated with a lethal gastrointestinal phenotype (Niessen *et al.,* 2005). In addition to *SMTN*, chapter 3 and chapter 5 also shared *protocadherin 10* (*PCDH10*), *ankyrin repeat domain 50* (*ANKRD50*), *adenylate cyclase 1* (*ADCY1*) and *williams-beuren syndrome chromosome region 17* (*WBSCR17*) genes. All five genes shared between NGS and SNP data of Nguni are also reported by Choi *et al.* (2013) to be copy number variable in cattle. The *PCDH10* gene forms part of the cadherin gene family with *PCDH10* specifically representing one of the non-clustered *PCDH*s of the *PCDH delta 2 group* (Kim *et al.,* 2011). Involved in the cadherin and Wnt signalling pathways and *PCDH10* also plays a role in calcium ion binding, cell communication, cell-cell adhesion, ectoderm development and nervous system development. The *PCDH10* protein is one of the most extensively studied protocadhedrins and is expressed in specific local circuits of functional systems like the visual and olfactory systems (Hirano *et al.,* 1999). *PCDH10* is also a tumor suppressor gene that plays a role in inhibiting cancer cell motility and cell migration (Yagi, 2008; Yu *et al.,* 2009). Cadherin genes tend to have a highly repetitive structure with cadherin, laminin A and G, EGF and mucin repeats that may cause genomic instability and a subsequent accumulation of CNVs (Seroussi *et al*., 2010). Protocadherins demonstrate considerable variations in copy number in human studies, with the *PCDH* cluster on the human chromosome 5 being

particularly prone to frequent copy number mutation and gene conversion events throughout mammal and vertebrate species, with variations in copy number and sequence content reflecting adaptive difference in protocadherin function (Cooper *et al.,* 2007; Noonan *et al.,* 2004). *ADCY1* gene is involved in adenylate and guanylate cyclase activity, playing a role in the gonadotropin releasing hormone receptor pathway, GABA-B_receptor_II_signaling, endothelin signaling pathway and the heterotrimeric G-protein signaling pathway, Gi alpha and Gs alpha mediated pathway and in cyclic nucleotide metabolic processes (Additional file 5.3). Adenylyl cyclase comprises a key cellular enzyme involved in catalysing the conversion of ATP to cAMP and pyrophosphate (Tang and Gilman 1992). The cAMP generated by adenylate cyclase enzymes is then utilized as a regulatory signal via specific cAMP binding proteins namely ion transporters, transcription factors or enzyme (Dwivedi and Pandey 2008). One such enzyme activated by cAMP is the phosphorylation enzyme protein kinase A which, once activated, phosphorylates various intracellular proteins modifying hormonal and neurotransmitter response, including the down regulation or desensitization of receptors, modification of the release of neurotransmitters and the activation or repression of gene expressions (Dwivedi and Pandey 2008; Nestler and Greengard 1994). The adenylyl cyclase 1 gene encodes one of the ten adenylyl cyclase isoforms recognized in mammals (Hanoune and Defer 2001). *ADCY1*, primarily expressed inn the brain and adrenal gland is involved in a number of immune system and DAG and IP3 signalling pathways and may play a role in regulating processes in the central nervous system, in memory and learning and in regulating circadian rhythms (Felder 1995; De Faria Poloni *et al.,* 2011; Hanoune and Defer 2001; Tang and Gilman 1992). Porto-Neto *et al.* (2014) report the *ADCY1* to be one of the closest candidate gene to the SNP that explained the greatest amount of genetic variations for body condition score in tropical composite cattle breeds. Body condition score is one of the measurements capturing the overall response of an animal to environmental conditions and is hence important in tropical adaptation (Porto-Neto *et al.,* 2014). *ANKRD50* comprises a protein coding gene that has recently been recognized as playing a role in the endosome-to-plasma membrane sorting and recycling of SNX27-retromer-dependent cargo proteins, such as glucose transporter GLUT1 and the Menkes disease copper transporter ATP7A (McGough *et al.,* 2014). The correct management of protein sorting for degradation or recycling is essential for cellular homeostasis (Gallon and Cullen 2015). Shortcomings in retromer function are increasingly being associated with human diseases like Alzheimer's disease (Gallon and Cullen 2015). Conte *et al.* (2016) recently reported a deletion covering *ANKRD50* gene in humans. *SNX27* knockout studies demonstrate *SNX27* to be important for cell motility and proliferation evident in wound healing (Li *et al.,* 2015). *WBSCR17* gene, currently uncharacterized in bovine, encodes N-acetylgalactosaminyltransferasen that may play a role in membrane trafficking in humans (Nakayama *et al.,* 2012). The implications of variable copy numbers over these genes or within close proximity to these genes should be further investigated. *RTDR1* was shared between chapters 4 and 5.

Table 5.5 The number (GEN) and name (GEN NM) of CNVR genes shared by chapters 3, 4 and 5 (CHPT).

CHPT*	GEN	GEN NM
DIV SNP NGS	1	*SMTN*
SNP DIV	39	*LOC527441 DDT GGT1 LYAR SPECC1L RNF185 SUSD2 FUBP3 WDR1 UPB1 SMARCB1 MSX1 ASS1 VPREB3 PATZ1 INPP5J ZNF70 ZNF280B CHCHD10 IGLL1 TMEM128 SLC5A1 SNRPD3 SELM MIF GSTT4 DRG1 ZBTB49 C17H22orf13 ZNF280A GSTT3 GSTT1 GGT5 LIMK2 NUP210 OTOP1 SLC2A11 PIK3IP1 DERL3*
SNP NGS	4	*ANKRD50 PCDH10 WBSCR17 ADCY1*
NGS DIV	1	*RTDR1*

*chapter 3 - SNP, chapter 4 -DIV and chapter 5 - NGS.

Comparisons of *PANTHER* pathways demonstrating representation by genes reveal that despite the different technologies failing to identify exact CNVRs, the same pattern of pathways are being represented by genes identified (Figure 5.8, 5.9, 5.10 and 5.11). Nearly all biological processes, molecular functions, cellular components and proteins represented by CNVR genes identified in this study, were already reported in chapter 3 from the SNP data. Despite only 5 genes being shared between the SNP and NGS CNVRs, the same processes, functions and components are for the most being captured across studies. Although the relative weights of the different components, processes, functions and proteins, does however differ to some degree CNVR appear to be acting on different genes involved in the same specific functions. The proposition that CNVRs occur as a result of possible genomic instability caused by external forces that exerts pressure on specific processes, functions, components and proteins may well stand true. This result corresponds with the presence of correlated CNVRs overrepresenting pathways reported in chapter 4. Both results indicate CNVR to be acting in specific processes, functions, components and proteins in cattle as a whole and in the Nguni specifically. Gaining insight into the implications of CNVRs at gene sites involved in biological processes reported in Figure 5.8 may hold important clues in the driving forces of CNVR formation.

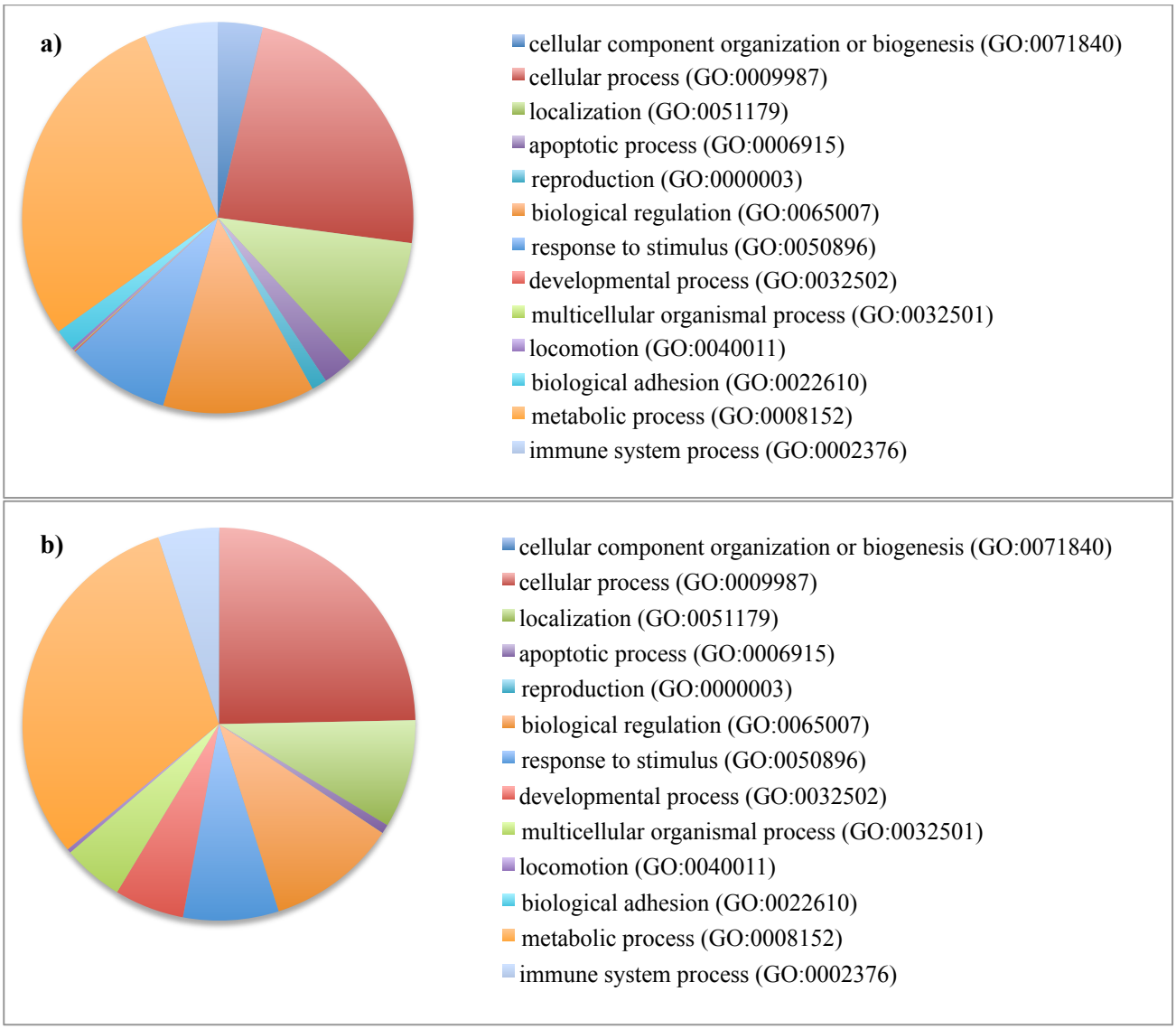

Figure 5.8 Percentage distribution of CNVR genes across biological processes for **a)** Nguni SNP and **b)** Nguni NGS CNV data.

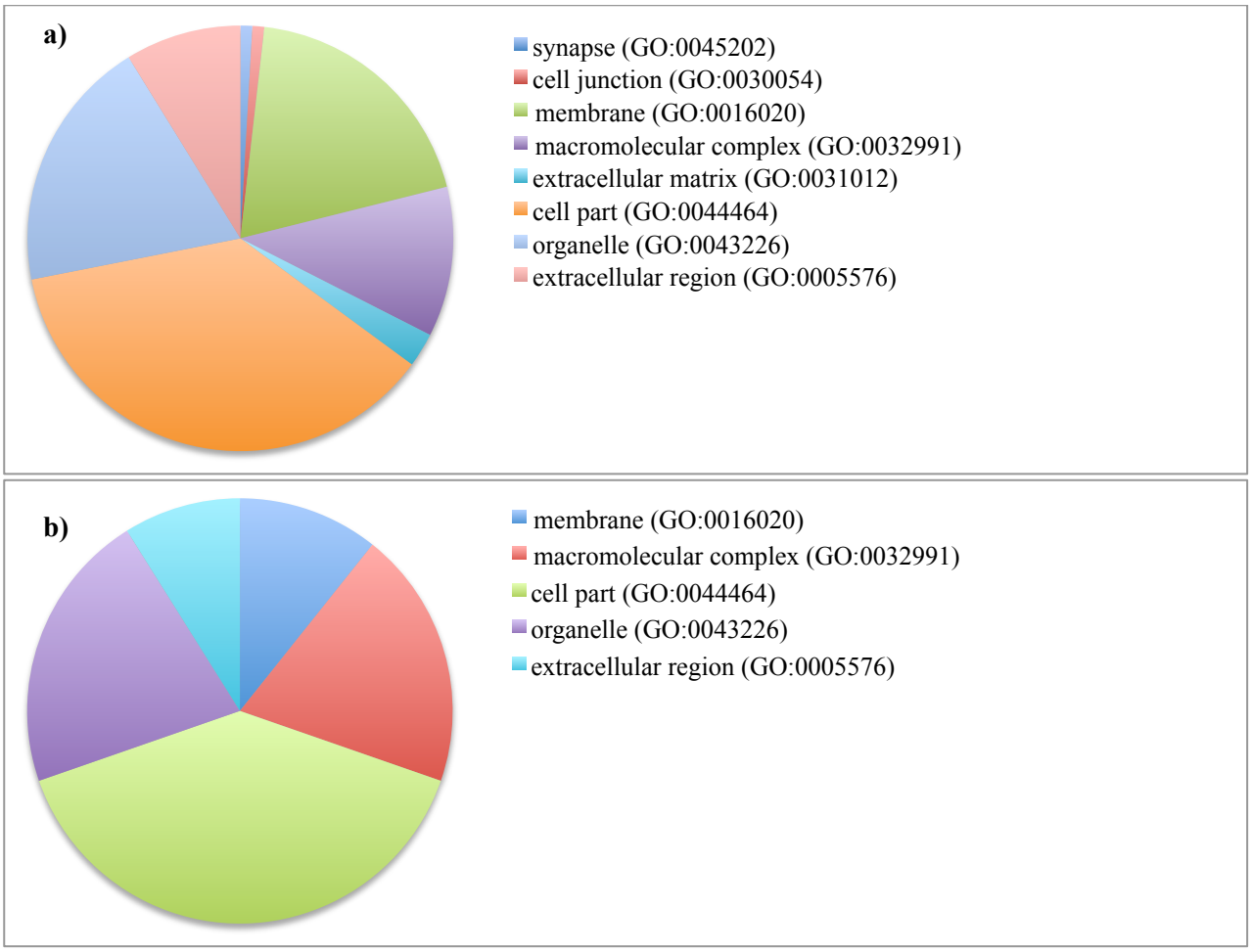

Figure 5.9 Percentage distribution of CNVR genes across cellular components for **a)** SNP and **b)** NGS CNV data.

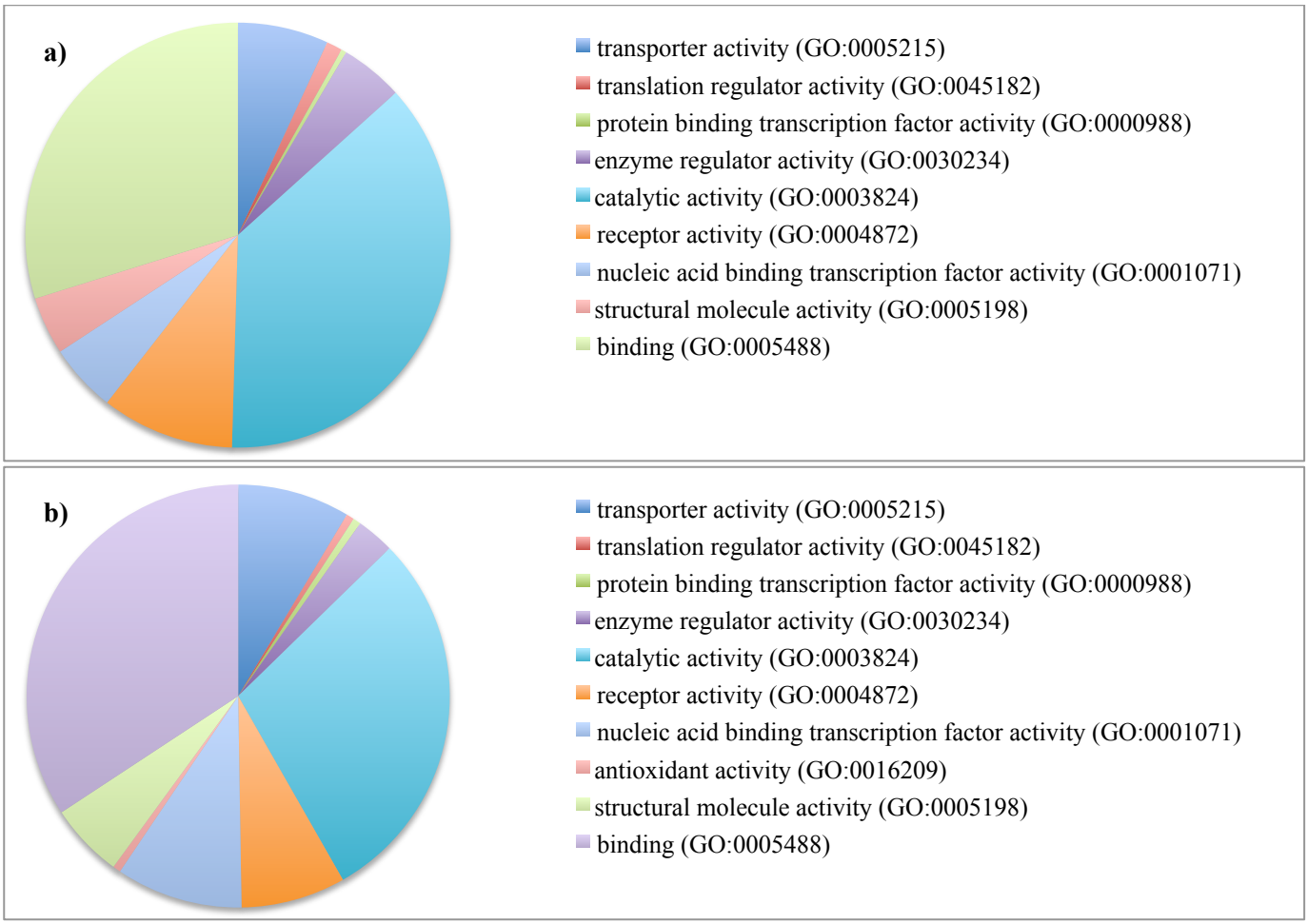

Figure 5.10 Percentage distribution of CNVR genes across molecular function for **a)** Nguni SNP and **b)** Nguni NGS CNV data.

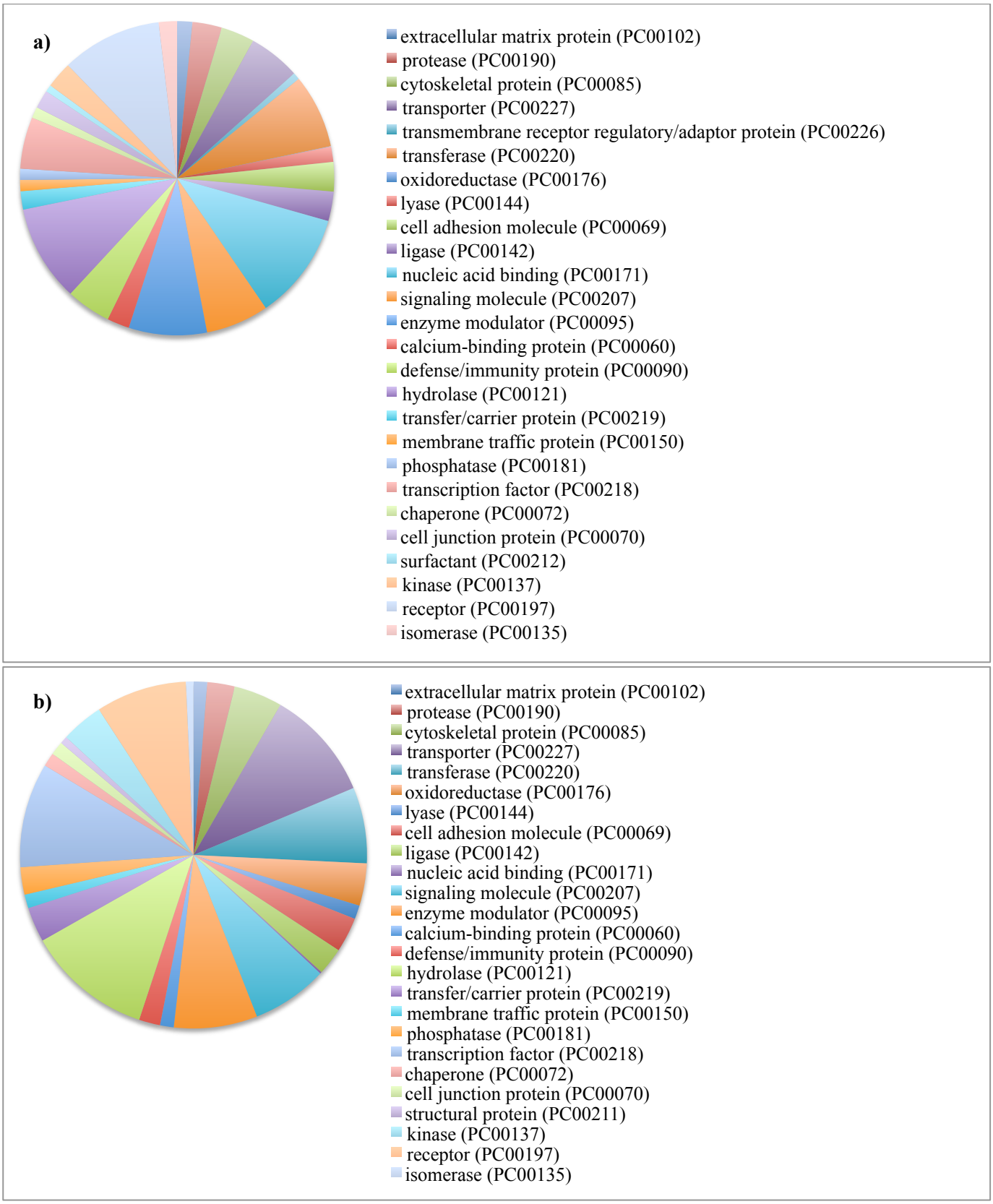

Figure 5.11 Percentage distribution of CNVR gene representation of proteins for **a)** Nguni SNP and **b)** Nguni NGS CNV data.

5.5 Conclusions

This study represents the first analyses of CNVs in South African Nguni cattle using whole genome NGS data. Twenty four Nguni cattle were sequenced at low to medium coverages and assessed for CNVs using the hybrid split read and paired end read mapping method, *RAPTR-SV* of Bickhart *et al.* (2015). Three hundred and twenty seven CNVRs were identified on the 29 autosomes of thirteen of the twenty three animals sequenced at the lowest stringency of F10. Chromosome 17 demonstrated a considerable amount for variations in copy number. Overall more deletion events were detected in alignment with the potential biases reported to be distinct to both NGS and SNP methodologies. Fourteen CNVRs overlapped or lay within close proximity of CNVRs reported by Choi *et al.* (2013) in Hanwoo cattle. Two hundred and fifty genes were covered or lay within close proximity of CNVRs reported at the lowest stringency. No specific biological pathways, molecular functions or cellular components demonstrated a statistically significant overrepresentation by CNVR genes, however genes involved in a number of interesting processes, functions or components were presented. The implications of CNVRs at these locations on the adaptive ability exhibited by Nguni cattle needs to be further explored. Relative to CNVRs reported in chapter 3 using the Bovine 50K Beadchip, CNVRs reported in this study were smaller, capturing different regions of the genome. Methods for the most part may be considered complementary to one another, identifying different CNVRs across the genome. Despite there being few CNVR genes that overlap across SNP and NGS results, the same biological processes, molecular functions, cellular components and proteins are represented by CNVR genes identified by both methods. This indicates specific processes, functions and components to be more prone to variable events, corresponding to findings in chapter 4 where correlated CNVRs were overrepresented by specific ontologies. Five genes, namely *SMTN, PCDH10, ANKRD50, ADCY1* and *WBSCR17* involved in a number were shared covered or lay within close proximity of CNVRs identified by both SNP and NGS methodologies. Genes shared were involved in an number of biological processes, molecular functions and cellular components including cellular response to stimulus, cellular metabolic process, regulation of cellular and biological process, calcium- and calmodium-responsive adenylate cyclase activity, ATP binding and protein binding.

5.6 References

Abyzov, A., A.E. Urban, M. Snyder, and M. Gerstein. 2011. “CNVnator: An Approach to Discover, Genotype, and Characterize Typical and Atypical CNVs from Family and Population Genome Sequencing.” *Genome Research* 21 (6): 974–84. doi:10.1101/gr.114876.110.

Aguilar-Delfin, I., P.J. Wettstein, and D.H. Persing. 2003. “Resistance to Acute Babesiosis Is Associated with Interleukin-12- and Gamma Interferon-Mediated Responses and Requires Macrophages and Natural Killer Cells.” *Infection and Immunity* 71 (4): 2002–8. doi:10.1128/IAI.71.4.2002-2008.2003.

Alkan, C., B.P. Coe, and E.E. Eichler. 2011. “Genome Structural Variation Discovery and Genotyping.” *Nature Reviews. Genetics* 12: 363–76. doi:10.1038/nrg2958.

Alkan, C., J.M. Kidd, T. Marques-bonet, G. Aksay, F. Hormozdiari, J.O. Kitzman, C. Baker, et al. 2009. “Personalized Copy-Number and Segmental Duplication Maps Using Next-Generation Sequencing.”

Nature Genetics 41 (10): 1061–67. doi:10.1038/ng.437.

Andrews, S. 2010. FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Araujo, R.N., T. Padilha, D. Zarlenga, T. Sonstegard, E.E. Connor, C. Van Tassel, W.S. Lima, Evaldo Nascimento, and Louis C. Gasbarre. 2009. "Use of a Candidate Gene Array to Delineate Gene Expression Patterns in Cattle Selected for Resistance or Susceptibility to Intestinal Nematodes." *Veterinary Parasitology* 162 (1–2): 106–15. doi:10.1016/j.vetpar.2008.12.017.

Aspatwar, A., M.E. Tolvanen, and S. Parkkila. 2010. "Phylogeny and Expression of Carbonic Anhydrase-Related Proteins." *BMC Molecular Biology* 11: 25. doi:10.1186/1471-2199-11-25.

Bär, H., P. Wende, L. Watson, S. Denger, G. Van Eys, J. Kreuzer, and L. Jahn. 2002. "Smoothelin Is an Indicator of Reversible Phenotype Modulation of Smooth Muscle Cells in Balloon-Injured Rat Carotid Arteries." *Basic Research in Cardiology* 97 (1): 9–16. doi:10.1007/s395-002-8382-z.

Bernabucci, U., N. Lacetera, L.H. Baumgard, R.P. Rhoads, B. Ronchi, and A. Nardone. 2010. "Metabolic and Hormonal Acclimation to Heat Stress in Domesticated Ruminants." *Animal* 4 (7): 1167–83. doi:10.1017/S175173111000090X.

Bester, J., I.E. Matjuda, J.M. Rust, H.J. Fourie. 2001. The Nguni: case study. In: *FAO Community-based management of animal genetic resources.* Rome: UNDP, GTZ, CTA, FAO. pp. 45-68.

Bickhart, D.M., Y. Hou, and S.G. Schroeder. 2012. "Copy Number Variation of Individual Cattle Genomes Using next-Generation Sequencing." *Genome*. doi:10.1101/gr.133967.111.

Bickhart, D.M., J.L. Hutchison, L. Xu, R.D. Schnabel, J.F. Taylor, J.M. Reecy, S. Schroeder, C.P. Van Tassell, T.S. Sonstegard, and G.E. Liu. 2015. "RAPTR-SV : A Hybrid Method for the Detection of Structural Variants." *Bioinformatics*, 1–7.

Biscarini, F., H. Bovenhuis, J.A.M. Van Arendonk, H.K. Parmentier, A.P. Jungerius, and J.J. Van Der Poel. 2010. "Across-Line SNP Association Study of Innate and Adaptive Immune Response in Laying Hens." *Animal Genetics* 41 (1): 26–38. doi:10.1111/j.1365-2052.2009.01960.x.

Bolger, A.M., M. Lohse, and B. Usadel. 2014. "Trimmomatic: A Flexible Trimmer for Illumina Sequence Data." *Bioinformatics* 30 (15): 2114–20. doi:10.1093/bioinformatics/btu170.

Cantsilieris, S., P.N. Baird, and S.J. White. 2013. "Molecular Methods for Genotyping Complex Copy Number Polymorphisms." *Genomics* 101 (2). Elsevier Inc.: 86–93. doi:10.1016/j.ygeno.2012.10.004.

Carter, N.P. 2007. "Methods and Strategies for Analyzing Copy Number Variation Using DNA Microarrays." *Nature Genetics* 39 (JULY): 1–11. doi:10.1038/ng2028.Methods.

Choi, J.W., K.T. Lee, X. Liao, P. Stothard, H.S. An, S. Ahn, S. Lee, S.Y. Lee, S.S. Moore, and T.H. Kim. 2013. "Genome-Wide Copy Number Variation in Hanwoo, Black Angus, and Holstein Cattle." *Mammalian Genome* 24 (3–4): 151–63. doi:10.1007/s00335-013-9449-z.

Cicconardi, F., G. Chillemi, A. Tramontano, C. Marchitelli, A. Valentini, P. Ajmone-Marsan, and A. Nardone. 2013. "Massive Screening of Copy Number Population-Scale Variation in Bos Taurus Genome." *BMC Genomics* 14 (1): 124. doi:10.1186/1471-2164-14-124.

Collier, R.J., J.L. Collier, R.P. Rhoads, and L.H. Baumgard. 2008. "Invited Review: Genes Involved in the

Bovine Heat Stress Response." *Journal of Dairy Science* 91 (2). Elsevier: 445–54. doi:10.3168/jds.2007-0540.

Collier, R.J., G.E. Dahl, and MJ. VanBaale. 2006. "Major Advances Associated with Environmental Effects on Dairy Cattle." *Journal of Dairy Science* 89 (4). Elsevier: 1244–53. doi:http://dx.doi.org/10.3168/jds.S0022-0302(06)72193-2.

Conte, F., M. Oti, J. Dixon, C.E.L. Carels, M. Rubini, and H. Zhou. 2016. "Systematic Analysis of Copy Number Variants of a Large Cohort of Orofacial Cleft Patients Identifies Candidate Genes for Orofacial Clefts." *Human Genetics* 135 (1). Springer Berlin Heidelberg: 41–59. doi:10.1007/s00439-015-1606-x.

Cooper, G.M., D.A. Nickerson, and E.E. Eichler. 2007. "Mutational and Selective Effects on Copy-Number Variants in the Human Genome." *Nature Genetics* 39 (7 Suppl): S22-9. doi:10.1038/ng2054.

De Faria Poloni, J., B.C. Feltes, and D. Bonatto. 2011. "Melatonin as a Central Molecule Connecting Neural Development and Calcium Signaling." *Functional and Integrative Genomics* 11 (3): 383–88. doi:10.1007/s10142-011-0221-8.

Dwivedi, Y., and G.N. Pandey. 2008. "Adenylyl Cyclase-cyclicAMP Signaling in Mood Disorders: Role of the Crucial Phosphorylating Enzyme Protein Kinase A." *Neuropsychiatric Disease and Treatment* 4 (1 A): 161–76. doi:10.2147/NDT.S2380.

Fadista, J., B. Thomsen, L. Holm, and C. Bendixen. 2010. "Copy Number Variation in the Bovine Genome." *BMC Genomics* 11: 284. doi:10.1186/1471-2164-11-284.

Felder, C C. 1995. "Muscarinic Acetylcholine Receptors: Signal Transduction through Multiple Effectors." *The FASEB Journal : Official Publication of the Federation of American Societies for Experimental Biology* 9 (8): 619–25. doi:10.1016/j.autneu.2006.10.006.

Ferreri, M., J. Gao, Z. Wang, L. Chen, J. Su, and B Han. 2011. "Chinese Holstein Cattle Shows a Genetic Contribution from Native Asian Cattle Breeds: A Study of Shared Haplotypes and Demographic History." *Asian-Australasian Journal of Animal Sciences* 24 (8): 1048–52. doi:10.5713/ajas.2011.10461.

Gallon, M., and P.J. Cullen. 2015. "Retromer and Sorting Nexins in Endosomal Sorting." *Biochemical Society Transactions* 43 (1): 33–47. doi:10.1042/BST20140290.

Gao, X., M.Y. Shi, X.R. Xu, J.Y. Li, H.Y. Ren, and S.Z. Xu. 2009. "Sequence Variations in the Bovine IGF-I and IGFBP3 Genes and Their Association with Growth and Development Traits in Chinese Beef Cattle." *Agricultural Sciences in China* 8 (6). Chinese Academy of Agricultural Sciences: 717–22. doi:10.1016/S1671-2927(08)60270-9.

Gautier, M., L. Flori, A. Riebler, F. Jaffrézic, D. Laloé, I. Gut, K. Moazami-Goudarzi, and J. Foulley. 2009. "A Whole Genome Bayesian Scan for Adaptive Genetic Divergence in West African Cattle." *BMC Genomics* 10: 550. doi:10.1186/1471-2164-10-550.

Gifford-Gonzalez, D., and O. Hanotte. 2011. "Domesticating Animals in Africa: Implications of Genetic and Archaeological Findings." *Journal of World Prehistory* 24: 1–23. doi:10.1007/s10963-010-9042-2.

Goff, W.L., W.C. Johnson, S.M. Parish, G.M. Barrington, and W. Tuo. 2001. "The Age-Related Immunity in Cattle to Babesia Bovis Infection Involves the Rapid Induction of Interleukin-12 , Interferon- G and

Inducible Nitric Oxide Synthase mRNA Expression in the Spleen." *Parasite Immunology*, no. 23: 463–71.

Hach, F., F. Hormozdiari, C. Alkan, F. Hormozdiari, I. Birol, E.E. Eichler, and S.C. Sahinalp. 2010. "mrsFast: A Cache-Oblivious Algorithm for Short-Read Mapping." *Nature Methods* 7 (8): 576–77. doi:10.1038/nmeth0810-576.mrsFast.

Hanoune, J., and N. Defer. 2001. "Regulation and Role of Adenylyl." *Annual Review of Pharmacology and Toxicology* 41: 145–74.

Hayashi, K., K. Ushizawa, M. Hosoe, and T. Takahashi. 2011. "Differential Gene Expression of Serine Protease Inhibitors in Bovine Ovarian Follicle: Possible Involvement in Follicular Growth and Atresia." *Reproductive Biology and Endocrinology : RB&E* 9 (1): 72. doi:10.1186/1477-7827-9-72.

Hayes, J. L., A. Tzika, H. Thygesen, S. Berri, H. M. Wood, S. Hewitt, M. Pendlebury, et al. 2013. "Diagnosis of Copy Number Variation by Illumina next Generation Sequencing Is Comparable in Performance to Oligonucleotide Array Comparative Genomic Hybridisation." *Genomics* 102 (3). Elsevier B.V.: 174–81. doi:10.1016/j.ygeno.2013.04.006.

Heo, J., R. Eki, and T. Abbas. 2016. "Deregulation of F-Box Proteins and Its Consequence on Cancer Development, Progression and Metastasis." *Seminars in Cancer Biology* 36: 33–51. doi:10.1016/j.semcancer.2015.09.015.

Hirano, S, Q Yan, and S T Suzuki. 1999. "Expression of a Novel Protocadherin, OL-Protocadherin, in a Subset of Functional Systems of the Developing Mouse Brain." *Journal of Neuroscience* 19 (3): 995–1005.

Hormozdiari, F., C. Alkan, E.E. Eichler, and S.C. Sahinalp. 2009. "Combinatorial Algorithms for Structural Variation Detection in High-Throughput Sequenced Genomes." *Genome Research* 19 (7): 1270–78. doi:10.1101/gr.088633.108.

Hormozdiari, Fereydoun, Iman Hajirasouliha, Phuong Dao, Faraz Hach, Deniz Yorukoglu, Can Alkan, Evan E. Eichler, and S. Cenk Sahinalp. 2010. "Next-Generation VariationHunter: Combinatorial Algorithms for Transposon Insertion Discovery." *Bioinformatics* 26 (12): 350–57. doi:10.1093/bioinformatics/btq216.

Horsburgh, K.A., S. Prost, A. Gosling, J. Stanton, C. Rand, and E.A. Matisoo-Smith. 2013. "The Genetic Diversity of the Nguni Breed of African Cattle (Bos Spp.): Complete Mitochondrial Genomes of Haplogroup T1." *PloS One* 8 (8): e71956. doi:10.1371/journal.pone.0071956.

Hou, Y., D.M. Bickhart, H. Chung, J.L. Hutchison, H.D Norman, E.E. Connor, G. Liu, and E. 2012. "Analysis of Copy Number Variations in Holstein Cows Identify Potential Mechanisms Contributing to Differences in Residual Feed Intake." *Functional & Integrative Genomics*, 717–23. doi:10.1007/s10142-012-0295-y.

Hou, Y., D.M. Bickhart, M.L. Hvinden, C. Li, J. Song, D.A. Boichard, S. Fritz, et al. 2012. "Fine Mapping of Copy Number Variations on Two Cattle Genome Assemblies Using High Density SNP Array." *BMC Genomics* 13: 376. doi:10.1186/1471-2164-13-376.

Hou, Y., G.E. Liu, D.M. Bickhart, L.K. Matukumalli, C. Li, J. Song, L.C. Gasbarre, C.P. Van Tassell, and

T.S. Sonstegard. 2012. “Genomic Regions Showing Copy Number Variations Associate with Resistance or Susceptibility to Gastrointestinal Nematodes in Angus Cattle.” *Functional and Integrative Genomics* 12: 81–92. doi:10.1007/s10142-011-0252-1.

Jansen, S., B. Aigner, H. Pausch, M. Wysocki, S. Eck, A. Benet-Pagès, E. Graf, et al. 2013. “Assessment of the Genomic Variation in a Cattle Population by Re-Sequencing of Key Animals at Low to Medium Coverage.” *BMC Genomics* 14: 446. doi:10.1186/1471-2164-14-446.

Karakoc, E., C. Alkan, B.J. O’Roak, M.Y. Dennis, L. Vives, K. Mark, M.J. Rieder, D.A. Nickerson, and E.E. Eichler. 2011. “Detection of Structural Variants and Indels within Exome Data.” *Nature Methods* 9 (2): 176–78. doi:10.1038/nmeth.1810.

Kijas, J.W., W. Barendse, W. Barris, B. Harrison, R. McCulloch, S. McWilliam, and V. Whan. 2011. “Analysis of Copy Number Variants in the Cattle Genome.” *Gene* 482. Elsevier B.V.: 73–77. doi:10.1016/j.gene.2011.04.011.

Kim, S.Y., S. Yasuda, H. Tanaka, K. Yamagata, and H. Kim. 2011. “Non-Clustered Protocadherin.” *Cell Adhesion and Migration* 5 (2): 97–105. doi:10.4161/cam.5.2.14374.

Koboldt, D.C., L. Ding, E.R. Mardis, and R.K. Wilson. 2010. “Challenges of Sequencing Human Genomes.” *Briefings in Bioinformatics*. doi:10.1093/bib/bbq016.

Korbel, J.O., A.E. Urban, J.P. Affourtit, B. Godwin, F. Grubert, J.F. Simons, P.M. Kim, et al. 2009. “Paired-End Mapping Reveals Extensive Structural Variation in the Human Genome.” *October* 318 (5849): 420–26. doi:10.1126/science.1149504.Paired-End.

Li, H., and R. Durbin. 2010. “Fast and Accurate Long-Read Alignment with Burrows-Wheeler Transform.” *Bioinformatics* 26 (5): 589–95. doi:10.1093/bioinformatics/btp698.

Li, K., S. Wu, M. Playford, and J. Sun. 2015. “Deletion of Sorting Nexin 27 Reverses Epithelial - Mesenchymal - Transition in Highly Aggressive Breast Cancer Cells” 29 (1).

Liu, G.E., and D.M. Bickhart. 2012. “Copy Number Variation in the Cattle Genome.” *Functional and Integrative Genomics* 12: 609–24. doi:10.1007/s10142-012-0289-9.

Liu, G.E., Y. Hou, B. Zhu, M.F. Cardone, L. Jiang, A. Cellamare, A. Mitra, et al. 2010. “Analysis of Copy Number Variations among Diverse Cattle Breeds.” *Genome Research* 20: 693–703. doi:10.1101/gr.105403.110.

Liu, G.E., C.P. Van Tassel, T.S. Sonstegard, R.W. Li, L.J. Alexander, J.W. Keele, L.K. Matukumalli, T.P. Smith, and L.C. Gasbarre. 2008. “Detection of Germline and Somatic Copy Number Variations in Cattle.” *Developments in Biologicals* 132: 231–37.

Liu, G.E., M. Ventura, A. Cellamare, L. Chen, Z. Cheng, B. Zhu, C. Li, J. Song, and E.E. Eichler. 2009. “Analysis of Recent Segmental Duplications in the Bovine Genome.” *BMC Genomics* 10: 571. doi:10.1186/1471-2164-10-571.

Makina, S.O., F.C. Muchadeyi, E. van Marle-KÃster, M.D. MacNeil, and A. Maiwashe. 2014. “Genetic Diversity and Population Structure among Six Cattle Breeds in South Africa Using a Whole Genome SNP Panel.” *Frontiers in Genetics* 5: 1–7. doi:10.3389/fgene.2014.00333.

Mapiye, C., M. Chimonyo, K. Dzama, J. G. Raats, and M. Mapekula. 2009. “Opportunities for Improving

Nguni Cattle Production in the Smallholder Farming Systems of South Africa." *Livestock Science* 124 (1–3). Elsevier B.V.: 196–204. doi:10.1016/j.livsci.2009.01.013.

Marufu, M.C., L. Qokweni, M. Chimonyo, and K. Dzama. 2011. "Relationships between Tick Counts and Coat Characteristics in Nguni and Bonsmara Cattle Reared on Semiarid Rangelands in South Africa." *Ticks and Tick-Borne Diseases* 2 (3). Elsevier GmbH.: 172–77. doi:10.1016/j.ttbdis.2011.07.001.

McGough, I.J., F. Steinberg, M. Gallon, A. Yatsu, N. Ohbayashi, K.J. Heesom, M. Fukuda, and P.J. Cullen. 2014. "Identification of Molecular Heterogeneity in SNX27-Retromer-Mediated Endosome-to-Plasma-Membrane Recycling." *Journal of Cell Science* 127 (22): 4940–53. doi:10.1242/jcs.156299.

Medvedev, P., M. Stanciu, and M. Brudno. 2009. "Computational Methods for Discovering Structural Variation with next-Generation Sequencing." *Nature Methods* 6 (11): S13–20. doi:10.1038/nmeth.1374.

Mills, R.E., K. Walter, C. Stewart, R.E. Handsaker, K. Chen, C. Alkan, A. Abyzov, et al. 2011. "Mapping Copy Number Variation by Population-Scale Genome Sequencing." *Nature* 470 (July 2010): 59–65. doi:10.1038/nature09708.Mapping.

Nakayama, Y., N. Nakamura, S. Oki, M. Wakabayashi, Y. Ishihama, A. Miyake, N. Itoh, and A. Kurosaka. 2012. "A Putative Polypeptide N-acetylgalactosaminyltransferase/Williams-Beuren Syndrome Chromosome Region 17 (WBSCR17) Regulates Lamellipodium Formation and Macropinocytosis." *Journal of Biological Chemistry* 287 (38): 32222–35. doi:10.1074/jbc.M112.370932.

Niessen, P., S. Rensen, J. Van Deursen, J. De Man, A. De Laet, J.M. Vanderwinden, T. Wedel, et al. 2005. "Smoothelin-A Is Essential for Functional Intestinal Smooth Muscle Contractility in Mice." *Gastroenterology* 129 (5): 1592–1601. doi:10.1053/j.gastro.2005.08.018.

Noonan, J.P., J. Grimwood, J. Schmutz, M. Dickson, and R.M. Myers. 2004. "Gene Conversion and the Evolution of Protocadherin Gene Cluster Diversity." *Genome Research* 14 (3): 354–66. doi:10.1101/gr.2133704.

Norris, B.J., and V.A. Whan. 2008. "A Gene Duplication Affecting Expression of the Ovine ASIP Gene Is Responsible for White and Black Sheep." *Genome Research* 18: 1282–93. doi:10.1101/gr.072090.107.

Porto-Neto, L.R., A. Reverter, K.C. Prayaga, E.K.F. Chan, D.J. Johnston, Rachel J. Hawken, Geoffry Fordyce, et al. 2014. "The Genetic Architecture of Climatic Adaptation of Tropical Cattle." *PLoS ONE* 9 (11): e113284. doi:10.1371/journal.pone.0113284.

Rensen, S.S.M., V.L.J.L. Thijssen, C.J. De Vries, P.A. Doevendans, S.D. Detera-Wadleigh, and G.J.J.M. Van Eys. 2002. "Expression of the Smoothelin Gene Is Mediated by Alternative Promoters." *Cardiovascular Research* 55 (4): 850–63. doi:10.1016/S0008-6363(02)00491-1.

Seroussi, E., G. Glick, A. Shirak, E. Yakobson, J.I. Weller, E. Ezra, and Y. Zeron. 2010. "Analysis of Copy Loss and Gain Variations in Holstein Cattle Autosomes Using BeadChip SNPs." *BMC Genomics* 11: 673. doi:10.1186/1471-2164-11-673.

Stothard, P., J. Choi, U. Basu, J.M. Sumner-Thomson, Y. Meng, X. Liao, and Stephen S Moore. 2011. "Whole Genome Resequencing of Black Angus and Holstein Cattle for SNP and CNV Discovery." *BMC Genomics* 12: 559. doi:10.1186/1471-2164-12-559.

Tada, O., V. Muchenje, and K. Dzama. 2013. "Preferential Traits for Breeding Nguni Cattle in Low-Input in-Situ Conservation Production Systems." *SpringerPlus* 2 (1): 195. doi:10.1186/2193-1801-2-195.

Tang, W.J., and A.G. Gilman. 1992. "Adenylyl Cyclases." *Cell* 70: 869–72. doi:10.3998/panc.2013.8.

Teo, S.M., Y. Pawitan, C.S. Ku, K.S. Chia, and A. Salim. 2012. "Sequence Analysis Statistical Challenges Associated with Detecting Copy Number Variations with next-Generation Sequencing" 28 (21): 2711–18. doi:10.1093/bioinformatics/bts535.

The Bovine Genome sequencing and analysis consortium, C.G. Elsik, R.L. Tellam, and K.C. Worley. 2009. "The Genome Sequence of Taurine Cattle: A Window to Ruminant Biology and Evolution." *Science* 324: 522–28. doi:10.1126/science.1169588.

Trinchieri, G., and F. Gerosa. 1996. "Immunoregulation by Interleukin-12." *Journal of Leukocyte Biology* 59 (4): 505–11.

Turner, S.R., and J.A. Macdonald. 2014. "Novel Contributions of the Smoothelin-like 1 Protein in Vascular Smooth Muscle Contraction and Its Potential Involvement in Myogenic Tone." *Microcirculation* 21 (3): 249–58. doi:10.1111/micc.12108.

Tuzun, E., A.J. Sharp, J.A. Bailey, R. Kaul, V.A. Morrison, L.M. Pertz, E. Haugen, et al. 2005. "Fine-Scale Structural Variation of the Human Genome." *Nature Genetics* 37 (7): 727–32. doi:10.1038/ng1562.

Uddin, Shahab, Ajaz A. Bhat, Roopesh Krishnankutty, Fayaz Mir, Michal Kulinski, and Ramzi M. Mohammad. 2016. "Involvement of F-BOX Proteins in Progression and Development of Human Malignancies." *Seminars in Cancer Biology* 36. Elsevier Ltd: 18–32. doi:10.1016/j.semcancer.2015.09.008.

Verma, N., I.D. Gupta, A. Verma, R. Kumar, R. Das, and M.R. Vineeth. 2016. "Novel SNPs in HSPB8 Gene and Their Association with Heat Tolerance Traits in Sahiwal Indigenous Cattle." *Tropical Animal Health and Production* 48 (1): 175–80. doi:10.1007/s11250-015-0938-9.

Wang, M. D., K. Dzama, C. A. Hefer, and F. C. Muchadeyi. 2015. "Genomic Population Structure and Prevalence of Copy Number Variations in South African Nguni Cattle." *BMC Genomics*. BMC Genomics, 1–16. doi:10.1186/s12864-015-2122-z.

Wooldridge, A.A., C.N. Fortner, B. Lontay, T. Akimoto, R.L. Neppl, C. Facemire, M.B. Datto, et al. 2008. "Deletion of the Protein Kinase A/protein Kinase G Target SMTNL1 Promotes an Exercise-Adapted Phenotype in Vascular Smooth Muscle." *Journal of Biological Chemistry* 283 (17): 11850–59. doi:10.1074/jbc.M708628200.

Yagi, T. 2008. "Clustered Protocadherin Family." *Development Growth and Differentiation* 50 (SUPPL. 1). doi:10.1111/j.1440-169X.2008.00991.x.

Ye, K., M.H. Schulz, Q. Long, R. Apweiler, and Z. Ning. 2009. "Pindel : A Pattern Growth Approach to Detect Break Points of Large Deletions and Medium Sized Insertions from Paired-End Short Reads" 25 (21): 2865–71. doi:10.1093/bioinformatics/btp394.

Yoon, S., Z. Xuan, V. Makarov, K. Ye, and J. Sebat. 2009. "Sensitive and Accurate Detection of Copy Number Variants Using Read Depth of Coverage." *Genome Research* 19 (9): 1586–92. doi:10.1101/gr.092981.109.

Yu, C., C. Chang, and J. Zhang. 2013. "Immunologic and Genetic Considerations of Cutaneous Lupus Erythematosus: A Comprehensive Review." *Journal of Autoimmunity* 41. Elsevier Ltd: 34–45. doi:10.1016/j.jaut.2013.01.007.

Yu, J., Y.Y. Cheng, Q. Tao, K.F. Cheung, C.N.Y. Lam, H. Geng, L.W. Tian, et al. 2009. "Methylation of Protocadherin 10, a Novel Tumor Suppressor, Is Associated With Poor Prognosis in Patients With Gastric Cancer." *Gastroenterology* 136 (2). AGA Institute American Gastroenterological Association: 640–651.e1. doi:10.1053/j.gastro.2008.10.050.

Zhao, M., Q.Q. Wang, P. Jia, and Z. Zhao. 2013. "Computational Tools for Copy Number Variation (CNV) Detection Using next-Generation Sequencing Data: Features and Perspectives - Springer." *BMC Bioinformatics* 14 (Suppl 11). BioMed Central Ltd: S1. doi:10.1186/1471-2105-14-S11-S1.

Zhou, X., R.M. Baron, M. Hardin, M.H. Cho, J. Zielinski, I. Hawrylkiewicz, P. Sliwinski, et al. 2012. "Identification of a Chronic Obstructive Pulmonary Disease Genetic Determinant That Regulates HHIP." *Human Molecular Genetics* 21 (6): 1325–35. doi:10.1093/hmg/ddr569.

Zimin, Aleksey V, Arthur L Delcher, Liliana Florea, David R Kelley, Michael C Schatz, Daniela Puiu, Finnian Hanrahan, et al. 2009. "A Whole-Genome Assembly of the Domestic Cow, Bos Taurus." *Genome Biology* 10 (4): R42. doi:10.1186/gb-2009-10-4-r42.

Chapter 6: General Discussion And Conclusion

6.1 Summary Of Findings

Comprising of deletions, duplications and insertions, CNVs are increasingly being shown to play a pivotal role in genetic diversity and subsequent phenotypic variation. A number of studies have demonstrated cattle to contain breed specific CNVs that may explain phenotypic discrepancies evident among breeds. Tropically adapted cattle comprise of a great variety of breeds with huge potential for studies relating to genetic diversity, prevalence and role in adaptation and disease resistance. The South African Nguni cattle is one such breed proven to withstand a variety of disease agents and harsh climatic conditions while having undergone little synthetic breeding (Bester *et al.,* 2001; Marufu *et al.,* 2011; Makina *et al.,* 2014). Gaining insight into the presence and prevalence of CNVs within the genome of South African Nguni cattle may give more insight on genetic forces behind adaptation and diversity.

This study set out to ascertain the prevalence and genetic diversity of CNVs in South African Nguni cattle. The possibility that CNVs may play a role in adaptation was also assessed. Three separate analyses were performed assessing CNVs within the genome of South African Nguni cattle. The Bovine 50K Beadchip was first utilized to screen for CNVRs in 492 South African Nguni cattle. CNVRs were then investigated as a measure of genetic diversity using an additional 6 South African breeds so as to determine between and within breed CNVR diversity.

Three hundred and thirty four CNVRs of between 30kb and 1Mb in length were identified using the Bovine 50K Beadchip and *PLINK* and *PennCNV* in 492 Nguni animals. Multiple stringency models were implemented so as to determine the most suitable model for CNV identification. Studies utilizing the high density bovine SNP chip to identify CNVs have been performed (Salomon-torres *et al.,* 2016; Sasaki *et al.,* 2016; Xu *et al.,* 2013; Zhang *et al.,* 2014). Despite the higher density of the beadchip, the number of CNVRs reported do not differ greatly to those reported using the 50K beadchip. Sasaki *et al*. (2016) report 861 CNVRs representing 1.74% of the bovine autosomes in 1 481 Japanese Black cattle using the high density beadchip. Salomon-torres *et al.* (2016) on the other hand report 56 CNVRs representing 0.33% of the genome in 12 Holstein cows, also using the high density beadchip. Zhang *et al.* (2015) report 0.4% of the genome of Qinchuan cattle to comprise CNV events in a study utilizing the high density beadchip in 6 cattle. Using the 50K beadchip Jiang *et al.* (2012) report 99 CNVRs in 2 047 Holstein cattle, while Hou *et al.* (2011) report 682 CNVRs in 539 cattle from different breeds. Discrepancies in the number of CNVs present among individuals and breeds have hence been reported in both high density and 50K beadchip studies. The higher density beadchip does cost more per sample and the benefits of having more probes were thus weighed against the benefits of a greater number of samples. The higher density beadchip was developed with the aim to enhance the accuracy of genomic predictions by decreasing the physical distance between SNP markers and QTLs (Harris and Creagh 2011). The utilization of the medium density SNP arrays do however capture CNVs efficiently and a number of recent studies reporting CNVs in cattle breeds using the

50K beadchip have been published (Hou *et al.,* 2012; Jiang *et al.,* 2012; Seroussi *et al.,* 2010). At the time of the studies design, few Nguni specific genomic analyses had been performed. With the discrepancies in CNV incidence within and across breeds, the use of more animals sampled from across the country analyzed with the 50K beadchip was deemed the most appropriate model for screening Nguni cattle for CNVs. A comparison of this chapter with the results in chapter 5 acquired from NGS analyses demonstrated more CNVs events that were on average considerably smaller than those identified from the 50K SNP array. Using the high density beadchip may add additional information for future studies if budgets allow for greater numbers to be sampled. The majority of the CNVRs captured by the next generation dataset in chapter 5 represented regions of the genome not captured on the Bovine 50K Beadchip. CNVRs are however known to be associated with segmental duplications (Hou *et al.,* 2011). The bias against segmental duplications of array based studies may however be more responsible for this discrepancy (Xu *et al.,* 2013). The utilization of the high density beadchip may therefore not be as beneficial as supplementing the 50K beadchip with NGS data which is more suited to capturing complementary regions of the genome that may not be well represented by the array dataset.

No prior information regarding population structure or Nguni ecotypes was available. *ADMIXTURE* software was thus utilized to perform population structure analyses using SNPs. This presented 5 Nguni sub populations within the dataset in which a degree of CNVR segregation was evident. CNVRs covered or lay within 10Mb of 289 genes of which 149, 28, 44, 2 and 14 genes were exclusive to the 5 sub-populations identified. The segregation of CNVRs within Nguni subpopulations indicates a possible role of CNVRs in breed history, diversity and formation. Population structure analyses also presented possible subpopulations within the 5 populations identified. An investigation into CNVRs within Nguni cattle from a larger dataset sampled from across ecotypes using prior information will contribute to a gaining insight into the sub structure present within Nguni cattle. For the present study, the current dataset was sufficient in demonstrating the prevalence of CNVRs within Nguni cattle and segregation of CNVRs across the different subpopulations present within the dataset. Haplotype blocks were also assessed and 541 HPBs were identified across the 492 animals. HPBs were between 84 and 199, 730 bp. Thirty four HPBs lay either within, across or adjacent to CNVRs identified within the Nguni cattle population. Of the CNVRs that were present near, across or within HPBs, half of them occurred in multiple individuals. HPBs may play a role in CNV formation that may relate to CNVs coinciding with breed formation and genetic diversity. It was hence determined to study CNVRs as a measure of genetic diversity. In order to further determine the presence of Nguni specific CNVRs and the factors influencing CNVR prevalence, multiple breeds from South Africa, representing the different breed groups were investigated for CNVRs and CNVR genetic diversity.

Chapter 4 reported CNVs in 7 different cattle breeds of South Africa. The pipeline developed for CNVR identification and characterization in chapter 3 was thus utilized to identify CNVR in two Taurine cattle breeds (Angus and Holstein), two composite cattle breeds (Bonsmara and Drakensberger), two Sanga cattle breeds (Nguni and Afrikaner and one crossbreed (Nguni Angus cross). 356 unique CNVRs were reported in

287 animals from 7 different South African cattle breeds representing Taurine, Sanga, composite and cross bred breed groups using the Bovine 50K Beadchip. Twenty-two of the 163 CNVR loci present in more than 1 animal constituted 74 significant correlations in all 7 breeds. Within the two exotic Taurine breeds, 906 significant CNVR correlations were determined, while only 6 significant CNVR correlations were identified in the indigenous Sanga and composite breeds. Most of the associations were between CNVR loci of the same type. PhiPT within breed group values were 2.510, 6.115 and 4.233 for the Sanga, Taurine and composite breeds respectively. The pure breeds at 0.085 (Sanga) and 0.113 (Taurine) demonstrated the lowest among breed group genetic variation with the composite breeds at a higher value of 3.897. The greater among breed genetic diversity of CNVs in these composite breeds should be further investigated. Unfortunately for this study only 10 suitable cross breed animals were available for inclusion in the study. Despite this, the Nguni Angus cross demonstrated considerably more CNVs per animal. Despite being a 50 50 cross between Nguni and Angus, these animals shared 30 CNVs with pure Angus animals and only 1 with pure Nguni animals, despite the pure Nguni and Angus animals sharing 11 CNVs. CNVR population structure displayed the segregation of breed type by CNVRs with Nguni Angus cross animals separating at K=3 and the Afrikaner, Drakensberger and Nguni breeds ghettoizing at K=7. The evolution of the CNV population structure with increasing K values depicts breed history patterns with CNVs segregating breeds groups. The Drakensberger is considered to be one of the earliest composite breeds developed. Its segregation with the Sanga type breeds is hence not surprising considering the possible role of adaptation on CNV prevalence. Although it was developed with a Taurine component, CNV evolution may reflect the selection pressures of adaptation that is evident in the Sanga breeds. The greater number of CNVs present in the Taurine breeds may suggest CNVs representing a response of the genome to selection pressures imposed by adverse climatic conditions on animals that have been bred for production and not necessarily for their innate ability to survive harsh conditions. Composite breeds were developed from multiple breeds with the aim to combine the adaptive ability of the local breeds with the productive capabilities of the exotic breeds (Bonsma 1980). The inclusion of the composite breeds as well as the Taurine Sanga crossbreed in this study provided insight into the age and evolution of CNVs and the translation of CNVs when breed groups are amalgamated in a composite breed and cross breed. The study of CNVs in crossbred and composite breeds may hold clues in gaining greater insight into CNV formation and the possible role of CNVs in factors like hybrid vigor.

Chapter 5 denotes the first whole genome NGS analyses of CNVs in South African Nguni cattle. Twenty three Nguni cattle, sequenced at average of 7.08x coverage were mapped to the UMD3.1 reference genome and subsequently studied for CNVs using the hybrid split read and paired end read mapping method of Bickhart *et al*. (2015). Bickhart *et al.* (2015) recommend a coverage of 10x to be suitable for CNV detection using their tool *RAPTR-SV*. Multiple stringencies were utilized to determine the most appropriate filtering for CNV identification in South African Nguni cattle. Three hundred and twenty seven CNVRs were identified on the 29 autosomes of thirteen of the twenty four animals sequenced at the lowest filtering of F10. The use of multiple stringencies highlighted CNVs with greater confidence. The comparison of CNV

events with those previously identified showed the greatest correspondence with CNVs identified by Choi *et al.* (2013) in Hanwoo cattle. Two hundred and fifty genes were covered or lay within close proximity of CNVRs reported at the lowest stringency. No specific biological pathways, molecular functions or cellular components demonstrated a statistically significant overrepresentation by CNVR genes, however genes involved in a number of interesting processes, functions or components were presented.

SNPs captured by array data tend to be sparse in regions that are segmentally duplicated or that contain complex CNVs (Carter 2007). Here, CNVs identified by NGS were on average smaller than those reported in chapters 3 and 4 from array data and tended to comprise regions not captured in the Bovine 50K Beadchip. Few CNVs were shared between NGS and SNP dataset. Capturing different regions NGS and SNP datasets may be suitable complementary methods for presenting a comprehensive whole genome CNV map. Despite chapter 5 only sharing 6 CNVR genes with chapters 3 and 4 the same biological processes, molecular functions, cellular components and proteins represented by CNVR genes. This implies CNVs to possibly be acting on regions of the genome involved in specific ontologies. This however needs to be further investigated using additional epigenetic and transcrptomic tools.

6.3 General Discussion

CNVRs comprise a prominent feature in the genome of South African Nguni cattle. Comprising of a number of subpopulations, CNVR distribution among the 5 subpopulations presented in chapter 3 demonstrated CNVRs as exhibiting population distinctions together with individual discrepancies. The presence of CNVRs at haplotype blocks in Nguni cattle (Section 3.4.4.5) together with the population and breed type distinction of CNVR discussed in sections 3.4.4.4 and 4.4.3.1 demonstrate CNVRs to comprise an important component of genetic diversity. The prevalence of CNVRs within regions of the genome involved in specific molecular functions, cellular components, biological processes and proteins also indicate CNVRs to have a specific driving force. The incidence of significantly associated CNVR loci involved in the same molecular functions, cellular components and biological processes indicate selection pressures being exerted on different genomic regions involved in specific processes. The simultaneous occurrence of such associations at different frequencies in Sanga, Taurine and composite breed groups suggested CNVs to possibly be a means by which the genomes respond to selection pressures and subsequently adapts. With Sanga and Taurine breeds having undergone different selection pressures, the variation in CNV incidence between these groups combined with the CNV correlations designate CNVRs to be genomic features prevalent in selection and adaptation. Choi *et al.* (2013) propose recent intensive artificial selection that has played a role in the improved productivity of economically important cattle breeds to influence CNVs. The two exotic breeds of South Africa (Holstein and Angus), which have both been developed as commercial breeds, demonstrated considerably more CNVs than the indigenous South African breeds in this study. This study found Nguni cattle to demonstrate a greater CNVR preponderance on chromosome 17, while the 7 South African breeds as a whole exhibited the most CNVRs on chromosomes 4 and 6. Commercially relevant cattle breeds like the Holstein and Black Angus cattle are reported to have significantly higher CNVR gains on chromosomes 14

and 6 (Choi *et al*., 2013; Jiang *et al*., 2012). These chromosomes have been extensively probed for quantitative trait loci concerning economically relevant dairy and carcass traits (Choi *et al*., 2013). Breeds also demonstrated breed specific CNV regions within a larger region reported across all breeds.

Despite discrepancies in CNVRs detected by different methodologies the same molecular functions, cellular components, biological processes and proteins are primarily represented by genes covered or lying within 10Mb of CNVRs identified across all three studies. These include biological regulation, reproduction, response to stimuli, immune system process, receptor activity, catalytic activity and defense/immunity protein. *SMTN* is reported in all three chapters, while chapters 3 and 4 shared an additional 39 genes, chapter 4 and 5 shared an additional gene and chapters 3 and 5 shared 5 additional genes. Shared genes were involved in a number of processes of interest. The occurrence of multiple CNVRs covering genes involved in the same pathways, functions and components, raises questions regarding the driving force of CNV formation within the genomes, and the possible role of selection pressures on CNV formation. This is further augmented by the notable discrepancy in CNV presence across breeds and even Nguni subpopulations presented in Chapters 3 and 4. The distinction CNVRs for Sanga and Taurine type breeds that have been exposed to similar types of selections pressures exemplifies the possible relationship between CNVRs and selection pressures. The presence of CNVs overlapping HPBs in chapter 3 also raises questions about the formation of CNVRs and their pattern of inheritance. Nguni cattle are well adapted to their environment having undergone years of natural selection. CNVs are prevalent within their genome. Multiple regions were present in several animals and represent possible breed, population and bovine specific CNVRs that should be further assessed. Chromosome 17 presented prevalent CNVRs across chapters. Only 14 of the 29 bovine autosomes presented CNVRs in chapter 5. In chapters 3 and 4, CNVRs were detected on all autosomes, although noticeable disparities in CNVR count across chromosomes is evident. In depth analyses into specific CNVRs identified should be performed. This may require the use of multiomics approaches which incorporate the transcriptome, methalome and proteome spheres in order to ascertain the effect of CNVRs present as well as to decipher the driving forces for CNV formation.

The addition of sequence data to array data provided a more comprehensive picture of CNV prevalence within the genome. It has been hypothesized that the two technologies may rather be used complementary to one another, demonstrating different strengths and weaknesses (Zhang *et al.,* 2011). NGS data identified smaller CNVRs than those identified using the Bovine 50K Beadchip. The greater number of smaller CNVRs identified in sequenced animals may be indicative of sequence data's greater ability to break up large CNVRs that comprise of multiple complex CNVRs lying within close proximity of each other and hence for identifying more complex CNVs. SNP data may report multiple CNVRs as a single region. Polymorphism ascertainment biases derived from detection from SNPs of a minimum minor allele frequency that segregate in multiple breeds are prevalent in array based technologies (Zhang *et al.,* 2011). Failure to distinguish the disequilibrium between genotyped SNPs and causal mutations complicates the discovery of rare causal mutations (Zhang *et al.,* 2011). Discrepancies in CNVRs identified by the two methodologies is

in keeping with prior findings (Zhang *et al.*, 2011). It has been proposed that NGS tools for CNV identification demonstrate greater accuracy in ascertaining CNV breakpoints while the SNP genotyping pipeline correspond to lower cost, denser coverage, and higher throughput (Hou *et al*., 2012). It has thus been hypothesized that the two technologies may rather be used in conjunction to complement each other by identifying different CNVRs. NGS technologies are able to identify CNVs that are too small for detection by array based methodologies (Zhang *et al*., 2009). Despite the discrepancies in the two tools identifying the exact same CNVRs, CNVR genes represented the same molecular functions, cellular components, biological processes and proteins

6.3 Conclusions

CNVRs present a prevalent measure of genetic diversity within South African Nguni cattle. Present in specific regions of the genome involved in molecular functions, cellular components, biological processes and proteins important to adaptation indicate CNVRs to form part of genomic adaptation to environmental and intensive selection pressures. Discrepancies in CNVR presence and CNVR correlations in the indigenous Sanga and composite breeds and exotic Taurine breeds demonstrate CNVRs to correspond to breed formation patterns, playing a fundamental role in genetic diversity and adaptation. The use of both array and sequencing methodologies for the detection of CNVs and CNVRs generates a more comprehensive picture of CNV distribution within the genome, although CNVR genes are involved in the same fundamental processes, functions and components across the genome.

6.4 Summary Of Contributions

6.4.1 Publications

6.4.1.1 Peer Reviewed Journal Articles

i. Wang, M.D., Dzama, K., Rees, J. and Muchadeyi, F.C., 2015. Genomic population structure and prevalence of copy number variations in South African Nguni cattle. *BMC Genomics* 16:894.

ii. Wang, M.D., Dzama, K., Rees, J. and Muchadeyi, F.C., 2016. Tropically adapted cattle of Africa: Perspectives on potential role of copy number variations. *Animal Genetics* 47(2): 154-164.

6.4.1.2 Submitted For Publication In Peer Reviewed Journal

i. Wang, M.D., Dzama, K. and Muchadeyi, F.C., 2016. Genetic diversity of South African cattle inferred using Copy number variations. Target journal: *Animal Genetics*.

6.4.1.3 Under Review For Submission

i. Wang, M.D., Dzama, K., Rees, G.J.R. and Muchadeyi, F.C., 2016. Whole genome sequencing of 24 South African Nguni cattle: Copy number variation prevalence and genetic diversity.

6.4.1.4 Conference Proceedings

i. Wang, M.D., Dzama, K., Rees, G.J.R. and Muchadeyi, F.C., 2015. South African cattle: Copy number variations and genetic diversity. *SASAS congress*.

ii. Wang, M.D., Dzama, K., Rees, G.J.R. and Muchadeyi, F.C., 2014. Preliminary identification and characterization of copy number variations in the genome of South African Nguni cattle. *Proceedings, 10th World Congress of Genetics Applied to Livestock Production*.

6.4.1.5 Popular Articles

i. Wang, M.D., Rees, R. and Muchadeyi, F.C., 2015. Genomics in light of wildlife. *WS2 Wildlife Breeders Journal*.19-21.

ii. Wang, M.D., Rees, R. and Muchadeyi, F.C., 2014. Global genome project. *Red Meat Rooivleis*. 5, 62-65.

6.4.2 Conference Participation

6.4.2.1 Oral Presentations

i. South African cattle: Copy number variations and genetic diversity. South African Society for Animal Science conference. Empangeni (2015).

ii. Preliminary identification and characterization of copy number variations in the Genome of South African Nguni cattle. World Congress on Genetics Applied to Livestock. Vancouver, Canada (2014).

iii. Preliminary identification and characterization of CNVs in the Genome of South African Nguni cattle. Agricultural Research Council Personal Development Program presentation day (2014).

6.4.2.2 Poster Presentations

i. Screening and characterization of copy number variation in South African Nguni cattle using next generation sequencing data. International Symposium of Animal Genetics. Salt Lake City, United States of America (2016).

ii. Copy number variations in the genome of South African cattle: correlations, pathways and genetic diversity. International Symposium on Animal Functional Genomics. Piacenza, Italy (2015).

iii. Population Structure and Prevalence of Copy Number Variations within the Genome of South African Nguni Cattle. Plant and Animal Genomics conference. San Diego, United States of America (2015).

iv. Preliminary identification and characterization of CNVs in the Genome of South African Nguni cattle. South African Society for Animal Science conference. Pretoria (2014).

v. Preliminary identification and characterization of copy number variations in the Genome of South African Nguni cattle. Joint South African Society of Bioinformatics and South African Genomics Society Congress. Pretoria (2013).

6.5 Future Research

Multi-omics approaches are gaining increasing attention as a powerful and accurate means to draw a complete and accurate picture of the dynamics of biological, cellular and molecular systems (Bersanelli *et al.,* 2016; Suravajhala *et al.,* 2016). Such an approach enables an integrative analyses of variations implicated in complex traits and have recently been successfully implemented in human studies (Miao *et al.,* 2014; Thingholm *et al.,* 2016). Multi-omics approaches should be utilized to explore the relationship of CNVRs and adaptation. This should include a genome wide association of CNVs with specific adaptive traits such as tick burden in South African Nguni cattle as well as an investigation into the prevalence of epigenetics in CNVRs formation/presence and role in adaptation (Skinner *et al.,* 2015).

6.6. References

Bersanelli, M., E. Mosca, D. Remondini, E. Giampieri, C. Sala, G. Castellani, and L. Milanesi. 2016. "Methods for the Integration of Multi-Omics Data: Mathematical Aspects." *BMC Bioinformatics* 17 (S2): 15. doi:10.1186/s12859-015-0857-9.

Bester, J., I.E. Matjuda, J.M. Rust, H.J. Fourie. 2001. The Nguni: case study. In: *FAO Community-based management of animal genetic resources.* Rome: UNDP, GTZ, CTA, FAO. pp. 45-68.

Bickhart, D.M., J.L. Hutchison, L. Xu, R.D. Schnabel, J.F. Taylor, J.M. Reecy, S. Schroeder, C.P. Van Tassell, T.S. Sonstegard, and G.E. Liu. 2015. "RAPTR-SV : A Hybrid Method for the Detection of Structural Variants." *Bioinformatics*, 1–7.

Carter, N.P. 2007. "Methods and Strategies for Analyzing Copy Number Variation Using DNA Microarrays." *Nature Genetics* 39 (JULY): 1–11. doi:10.1038/ng2028.Methods.

Choi, J.W., K.T. Lee, X. Liao, P. Stothard, H.S. An, S. Ahn, S. Lee, S.Y. Lee, S.S. Moore, and T.H. Kim. 2013. "Genome-Wide Copy Number Variation in Hanwoo, Black Angus, and Holstein Cattle." *Mammalian Genome* 24 (3–4): 151–63. doi:10.1007/s00335-013-9449-z.

Harris, BL, and FE Creagh. 2011. "Experiences with the Illumina High Density Bovine Beadchip." *Interbull Bulletin*, no. 44: 3–7.

Hou, Y., D.M. Bickhart, M.L. Hvinden, C. Li, J. Song, D.A. Boichard, S. Fritz, et al. 2012. "Fine Mapping of Copy Number Variations on Two Cattle Genome Assemblies Using High Density SNP Array." *BMC Genomics* 13: 376. doi:10.1186/1471-2164-13-376.

Hou, Y., G.E. Liu, D.M. Bickhart, M.F. Cardone, K. Wang, E. Kim, L.K. Matukumalli, et al. 2011. "Genomic Characteristics of Cattle Copy Number Variations." *BMC Genomics* 12: 127. doi:10.1186/1471-2164-12-127.

Jiang, L., J. Jiang, J. Wang, X. Ding, J. Liu, and Q. Zhang. 2012. "Genome-Wide Identification of Copy Number Variations in Chinese Holstein." *PLoS ONE* 7 (11). doi:10.1371/journal.pone.0048732.

Makina, S.O., F.C. Muchadeyi, E. van Marle-KÃster, M.D. MacNeil, and A. Maiwashe. 2014. "Genetic Diversity and Population Structure among Six Cattle Breeds in South Africa Using a Whole Genome SNP Panel." *Frontiers in Genetics* 5: 1–7. doi:10.3389/fgene.2014.00333.

Miao, R., H. Luo, H. Zhou, G. Li, D. Bu, X. Yang, X. Zhao, et al. 2014. "Identification of Prognostic

Biomarkers in Hepatitis B Virus-Related Hepatocellular Carcinoma and Stratification by Integrative Multi-Omics Analysis." *Journal of Hepatology* 61 (4): 840–49. doi:10.1016/j.jhep.2014.05.025.

Salomon-torres, R., C. Villa-angulo, and R. Villa-angulo. 2016. "Genome-Wide Identification of Copy Number Variations in Holstein Cattle from Baja California, Mexico, Using High-Density SNP Genotyping Arrays." *Genetics and Molecular Research* 14 (4): 11848–59. doi:10.4238/2015.October.2.18.

Sasaki, S., T. Watanabe, S. Nishimura, and Y. Sugimoto. 2016. "Genome-Wide Identification of Copy Number Variation Using High-Density Single-Nucleotide Polymorphism Array in Japanese Black Cattle." *BMC Genetics* 17 (1). BMC Genetics: 26. doi:10.1186/s12863-016-0335-z.

Seroussi, E., G. Glick, A. Shirak, E. Yakobson, J.I. Weller, E. Ezra, and Y. Zeron. 2010. "Analysis of Copy Loss and Gain Variations in Holstein Cattle Autosomes Using BeadChip SNPs." *BMC Genomics* 11: 673. doi:10.1186/1471-2164-11-673.

Skinner, M.K., C. Guerrero-Bosagna, and M.M. Haque. 2015. "Environmentally Induced Epigenetic Transgenerational Inheritance of Sperm Epimutations Promote Genetic Mutations." *Epigenetics* 10 (8): 762–71. doi:10.1080/15592294.2015.1062207.

Suravajhala, P., L.J.A. Kogelman, and H.N. Kadarmideen. 2016. "Multi-Omic Data Integration and Analysis Using Systems Genomics Approaches: Methods and Applications in Animal Production, Health and Welfare." *Genetics, Selection, Evolution : GSE* 48 (1). BioMed Central: 38. doi:10.1186/s12711-016-0217-x.

Thingholm, L.B., L. Andersen, E. Makalic, M.C. Southey, Mads Thomassen, and Lise Lotte Hansen. 2016. "Strategies for Integrated Analysis of Genetic, Epigenetic, and Gene Expression Variation in Cancer: Addressing the Challenges." *Frontiers in Genetics* 7 (February): 2. doi:10.3389/fgene.2016.00002.

Xu, Lingyang, Yali Hou, Derek Bickhart, Jiuzhou Song, and George Liu. 2013. "Comparative Analysis of CNV Calling Algorithms: Literature Survey and a Case Study Using Bovine High-Density SNP Data." *Microarrays* 2: 171–85. doi:10.3390/microarrays2030171.

Zhang, D., Y. Qian, N. Akula, N. Alliey-Rodriguez, J. Tang, E.S. Gershon, and C. Liu. 2011. "Accuracy of CNV Detection from GWAS Data." *PLoS ONE* 6 (1): 1–9. doi:10.1371/journal.pone.0014511.

Zhang, F., W. Gu, M.E. Hurles, and J.R. Lupski. 2009. "Copy Number Variation in Human Health, Disease, and Evolution." *Annual Review of Genomics and Human Genetics* 10: 451–81. doi:10.1146/annurev.genom.9.081307.164217.

Zhang, L., S. Jia, M. Plath, Y. Huang, C. Li, C. Lei, X. Zhao, and H. Chen. 2015. "Impact of Parental Bos Taurus and Bos Indicus Origins on Copy Number Variation in Traditional Chinese Cattle Breeds." *Genome Biology and Evolution* 7 (8): 2352–61. doi:10.1093/gbe/evv151.

Zhang, Q., Y. Ma, X. Wang, Y. Zhang, and X. Zhao. 2014. "Identification of Copy Number Variations in Qinchuan Cattle Using BovineHD Genotyping Beadchip Array." *Molecular Genetics and Genomics* 290 (1): 319–27. doi:10.1007/s00438-014-0923-4.

Addendum A

Additional file 2.1 Gene ontology categories that may play a role in adaptation that are reported (REF) to be overrepresented by genes covered by CNVs identified in cattle (BRD) from around the glove (OR).

REF	BRD	OR*	CNV	CNV adaptation gene ontologies	Adaptation relevance
Bickhart *et al.,* 2012	ANG HER HOL	EUR	1030	Signal transduction, Phenylethylamine degradation, Regulation of biological process, Viral coat protein, Viral protein, Antibacterial response protein, Antigen binding, B cell mediated immunity, Cellular defense response, Cytokine receptor, Defense response to bacterium, Defense/immunity protein, Immune response, Immune system process, Immunoglobulin complex, Immunoglobulin receptor family member, Interferon superfamily, Macrophage activation, Major histocompatibility complex antigen, MHC protein complex, Natural killer cell activation, T cell activation, Apolipoprotein, Enzyme inhibitor activity, Cadherin signaling pathway, G-protein coupled receptor, G-protein coupled receptor activity, Heterotrimeric G-protein, Heterotrimeric G-protein complex, Response to stimulus, Response to stress, Response to toxin, Sensory perception, Regulation of vasoconstriction, Immunoglobulin, Lipid transporter activity, Cytokine, G-protein modulator, Heterotrimeric G-protein signaling pathway-Gi alpha and Gs alpha mediated pathway, Heterotrimeric G-protein signaling pathway-Gq alpha and Go alpha mediated pathway, Visual perception	Cellular metabolism, Energy production, General regulation and control, Immune response, Metabolism - temperature regulation, Neurogenesis, Smooth muscle contraction/relaxation, heart rate, taste, vision, neuronal activity, Survival, Temperature regulation,
Bickhart *et al.,* 2012	NEL	IND	813	Signal transduction, Phenylethylamine degradation, Regulation of biological process, Viral coat protein, Viral protein, Antibacterial response protein, B cell mediated immunity, Cellular defense response, Cytokine, Cytokine receptor, Defense response to bacterium, Defense/immunity protein, Immune response, Immune system process, Immunoglobulin receptor family member, Interferon superfamily, Macrophage activation, Major histocompatibility complex antigen, MHC protein complex, Natural killer cell activation, Blood circulation, Cellular amino acid and derivative metabolic process, Enzyme inhibitor activity, Cadherin signaling pathway, G-protein coupled receptor, G-protein coupled receptor activity, Heterotrimeric G-protein signaling pathway-Gi alpha and Gs alpha mediated pathway, Heterotrimeric G-protein signaling pathway-Gq alpha and Go alpha mediated pathway, Response to stimulus, Response to stress, Sensory perception, Visual perception, Regulation of vasoconstriction	Cellular metabolism, Energy production, General regulation and control, Immune response, Metabolism - temperature regulation, Neurogenesis, Signal transduction - cellular metabolism, Smooth muscle contraction/relaxation, heart rate, taste, vision, neuronal activity, Survival, Temperature regulation
Hou *et al.,* 2012b	27 Breeds	GLB	674	Signal transduction, Antibacterial response protein, Antigen binding, B cell mediated immunity, Cellular defense response, Cytokine receptor, Cytokinesis, Defense/immunity protein, Immune response, Immune system process, Immunoglobulin complex, Immunoglobulin receptor family	Cellular metabolism, Immune response, Metabolism - temperature

REF	BRD	OR*	CNV	CNV adaptation gene ontologies	Adaptation relevance
				member, Interferon superfamily, Macrophage activation, Major histocompatibility complex antigen, MHC protein complex, Natural killer cell activation, T cell activation, Apolipoprotein, Atpase activity, coupled to transmembrane movement of substances, Blood circulation, Generation of precursor metabolites and energy, Metabolic process, Cadherin signaling pathway, Gut mesoderm development, Mesoderm development, G-protein coupled receptor, G-protein coupled receptor activity, Heterotrimeric G-protein, Heterotrimeric G-protein complex, Response to stimulus, Response to toxin, Sensory perception, Visual perception, Immunoglobulin, Lipid transporter activity	regulation, Neurogenesis, Nutrition, Smooth muscle contraction/relaxation, heart rate, taste, vision, neuronal activity, Survival,
Hou *et al.,* 2012c	ANG	USA	811	Defense/immunity protein, Immunity and defence, Immunoglobulin receptor family member, Protein metabolism and modification, T-cell mediated immunity	Immune response, metabolism
Seroussi *et al.,* 2010	HOL	ISR	418	Signal transduction, Cadherin signalling pathway1, Cadherin1, G-protein coupled receptor, G-protein mediated signaling, Olfaction, Sensory perception	Cellular metabolism, Neurogenesis, Smooth muscle contraction/relaxation, heart rate, taste, vision, neuronal activity, Survival

*OR – Origin, EUR – Europe, IND – India, GLB – Global, USA – United States of America and ISR - Israel

Additional file 2.2 CNV classification, formation and properties

Copy number variations comprise of quantitative variations in the genome, which together with orientational inversions and positional translocations fall under the umbrella term of structural variations (Scherer et al. 2007; Bae et al. 2010; Liu et al. 2010). CNVs however specifically refer to structural changes that cause deletions, duplications or insertions that ultimately change the genomic copy number (McCarroll 2008; Alkan *et al.,* 2011). Two major classes of polymorphic and de novo, pathogenic CNVs that demonstrate distinct discrepancies in both their structure and cellular origins have been uncovered (Arlt et al. 2013). Those CNVs primarily occurring within close proximity to segmental duplications have been given the broad term of recurrent CNVs (R-CNVs) (Arlt et al. 2013; Verdin et al. 2013). The rest, that occur independent of these segmentally duplicated regions of the genome are subsequently termed non-recurrent CNVs (NR-CNVs) (Arlt et al. 2013; Verdin et al. 2013). The mechanisms underlying the formation of novel or deleterious CNV mutations are vital in fully understanding and defining the environmental and genetic factors thereof (Arlt *et al*., 2012). Recurrent CNVs in all likelihood arise via non-allelic homologous recombination (NAHR) that is preempted during meiosis and genomic innovations by the genomic instability caused by misalignment of large contiguous SDs or LCRs (Alkan *et al*., 2011; Bickhart *et al.,* 2012; Arlt *et al.*, 2012). Non-recurrent CNVs (NR-CNVs), on the other hand comprise of those CNVs demonstrating inimitable breakpoints that are independent of segmental duplications and are observed to be subsequent to a diverse array of mechanism that can be collectively termed non-replicative (NAHR, NHEJ and MMEJ) or replicative-based repair mechanisms (FoSTeS, SRS, BISRS and MMBIR) (Verdin *et al*. 2013). Durkin *et al*. (2012) report the first CNV generating translocation mechanism incorporating circular intermediates that underly the colour sidedness in Belgium Blue cattle. Verdin *et al.* (2013) propose the non-recurrent deletions at the FOXL2 locus to be caused by microhomology mediated mechanisms such as microhomology mediated end-joining (MMEJ), fork stalling and template switching (FoSTeS), microhomology-mediated break-induced replication (MMBIR), serial replication slippage (SRS) or break-induced SRS (BISRS). The genomic architecture of specific regions may increase the susceptibility for DNA breakage or augment DNA replication fork stalling, thereby driving the formation of rare, locus-specific CNVs (Verdin *et al*., 2013).

The initial notion that CNVs are primarily rare events only associated with genomic diseases changed, in 2004 when two separate research groups reported the first genome wide CNV maps of apparently healthy individuals (Iafrate *et al*. 2004; Sebat *et al*. 2004). Tuzun *et al*. (2005) subsequently compared the genomes of two seemingly healthy individuals using in silico methodologies and identified 241 CNVs that were primarily between 8 and 40 kb in size. These findings have since been verified on multiple occasions with CNVs demonstrating particular enrichment in protein secretory, immunity and olfactory gene sites (McCarroll *et al*., 2008; Freeman *et al*., 2006; Wong *et al*., 2007). The phenotypic impact of CNVs is, however related to a large extent to the locations of the variants in relation to the genes with CNVs demonstrating effect via dosage of a single gene (Yang *et al*., 2007), a contiguous set of genes (Henrichsen *et al.,* 2009; Chaignat *et al*., 2011) or allele combinations (Buchanan and Scherer 2008). CNVs may alter

gene structure or dosage, disrupt coding sequences or long range regulation or potentially expose recessive alleles and thereby modify gene functioning (Zhang *et al*., 2009; Liu and Bickhart 2012).

Additional file 2.2 References

Alkan, C., B.P. Coe, and E.E. Eichler. 2011. "Genome Structural Variation Discovery and Genotyping." *Nature Reviews. Genetics* 12: 363–76. doi:10.1038/nrg2958.

Arlt, M.F., T.E. Wilson, and T.W. Glover. 2012. "Replication Stress and Mechanisms of CNV Formation." *Current Opinion in Genetic Developments* 22 (3): 204–10. doi:10.1016/j.gde.2012.01.009.

Bae, J.S., H.S. Cheong, L.H. Kim, S. NamGung, T.J. Park, J. Chun, J.Y. Kim, C.F.A. Pasaje, J.S. Lee, and H.D. Shin. 2010. "Identification of Copy Number Variations and Common Deletion Polymorphisms in Cattle." *BMC Genomics* 11: 232. doi:10.1186/1471-2164-11-232.

Bickhart, D.M., Y. Hou, and S.G. Schroeder. 2012. "Copy Number Variation of Individual Cattle Genomes Using next-Generation Sequencing." *Genome*. doi:10.1101/gr.133967.111.

Buchanan, J.A., and S.W. Scherer. 2008. "Contemplating Effects of Genomic Structural Variation." *Genetics in Medicine : Official Journal of the American College of Medical Genetics* 10 (9): 639–47. doi:10.1097/GIM.0b013e318183f848.

Chaignat, Evelyne, Emilie Aït Yahya-Graison, Charlotte N. Henrichsen, Jacqueline Chrast, Frédéric Schütz, Sylvain Pradervand, and Alexandre Reymond. 2011. "Copy Number Variation Modifies Expression Time Courses." *Genome Research* 21 (1): 106–13. doi:10.1101/gr.112748.110.

Durkin, K., W. Coppieters, C. Drögemüller, N. Ahariz, N. Cambisano, T. Druet, C. Fasquelle, et al. 2012. "Serial Translocation by Means of Circular Intermediates Underlies Colour Sidedness in Cattle." *Nature* 482 (7383): 81–84. doi:10.1038/nature10757.

Freeman, Jennifer L., George H. Perry, Lars Feuk, Richard Redon, Steven A. McCarroll, David M. Altshuler, Hiroyuki Aburatani, et al. 2006. "Copy Number Variation: New Insights in Genome Diversity." *Genome Research* 16 (8): 949–61. doi:10.1101/gr.3677206.

Henrichsen, C.N., E. Chaignat, and A. Reymond. 2009. "Copy Number Variants, Diseases and Gene Expression." *Human Molecular Genetics* 18 (1): 1–8. doi:10.1093/hmg/ddp011.

Hou, Y., G.E. Liu, D.M. Bickhart, M.F. Cardone, K. Wang, E. Kim, L.K. Matukumalli, et al. 2011. "Genomic Characteristics of Cattle Copy Number Variations." *BMC Genomics* 12: 127. doi:10.1186/1471-2164-12-127.

Iafrate, A.J., L. Feuk, M.N. Rivera, M.L. Listewnik, P.K. Donahoe, Y. Qi, S.W. Scherer, and C. Lee. 2004. "Detection of Large-Scale Variation in the Human Genome." *Nature Genetics* 36 (9): 949–51. doi:10.1038/ng1416.

Liu, G.E., and D.M. Bickhart. 2012. "Copy Number Variation in the Cattle Genome." *Functional and Integrative Genomics* 12: 609–24. doi:10.1007/s10142-012-0289-9.

Liu, G.E., Y. Hou, B. Zhu, M.F. Cardone, L. Jiang, A. Cellamare, A. Mitra, et al. 2010. "Analysis of Copy Number Variations among Diverse Cattle Breeds." *Genome Research* 20: 693–703. doi:10.1101/gr.105403.110.

McCarroll, Steven A. 2008. “Extending Genome-Wide Association Studies to Copy-Number Variation.” *Human Molecular Genetics* 17 (R2): 135–42. doi:10.1093/hmg/ddn282.

McCarroll, Steven a, Finny G Kuruvilla, Joshua M Korn, Simon Cawley, James Nemesh, Alec Wysoker, Michael H Shapero, et al. 2008. “Integrated Detection and Population-Genetic Analysis of SNPs and Copy Number Variation.” *Nature Genetics* 40 (10): 1166–74. doi:10.1038/ng.238.

Scherer, Stephen W, Charles Lee, Ewan Birney, David M Altshuler, Evan E Eichler, Nigel P Carter, Matthew E Hurles, and Lars Feuk. 2007. “Challenges and Standards in Integrating Surveys of Structural Variation.” *Nature Genetics* 39 (7 Suppl): S7–15. doi:10.1038/ng2093.

Sebat, J., B. Lakshmi, J. Troge, J. Alexander, J. Young, P. Lundin, S. Maner, et al. 2004. “Large-Scale Copy Number Polymorphism in the Human Genome.” *Science* 305: 525–28. doi:10.1126/science.1098918.

Tuzun, E., A.J. Sharp, J.A. Bailey, R. Kaul, V.A. Morrison, L.M. Pertz, E. Haugen, et al. 2005. “Fine-Scale Structural Variation of the Human Genome.” *Nature Genetics* 37 (7): 727–32. doi:10.1038/ng1562.

Verdin, Hannah, Barbara D’haene, Diane Beysen, Yana Novikova, Björn Menten, Tom Sante, Pablo Lapunzina, et al. 2013. “Microhomology-Mediated Mechanisms Underlie Non-Recurrent Disease-Causing Microdeletions of the FOXL2 Gene or Its Regulatory Domain.” *PLoS Genetics* 9 (3). doi:10.1371/journal.pgen.1003358.

Wong, KK, RJ deLeeuw, NS Dosanjh, LR Kimm, Z Cheng, DE Horsman, C MacAulay, et al. 2007. “A Comprehensive Analysis of Common Copy-Number Variations in the Human Genome.” *Am J Hum Genet* 80 (1): 91–104. doi:10.1086/510560.

Yang, Y., E.K. Chung, Y.L. Wu, S.L. Savelli, H.N. Nagaraja, B. Zhou, M. Hebert, et al. 2007. “Gene Copy-Number Variation and Associated Polymorphisms of Complement Component C4 in Human Systemic Lupus Erythematosus (SLE): Low Copy Number Is a Risk Factor for and High Copy Number Is a Protective Factor against SLE Susceptibility in European America.” *American Journal of Human Genetics* 80 (6): 1037–54. doi:10.1086/518257.

Zhang, F., W. Gu, M.E. Hurles, and J.R. Lupski. 2009. “Copy Number Variation in Human Health, Disease, and Evolution.” *Annual Review of Genomics and Human Genetics* 10: 451–81. doi:10.1146/annurev.genom.9.081307.164217.

Addendum B

Additional file 3.1 The python script developed to merge adjacent and overlapping CNVs into CNVRs.

```
import sys,os,csv, getopt, collections
import numpy as np
import subprocessdef rename_cnv(ll_cnv):
    ld_cnv_name = {}
    li_counter = 1
    for ld_cnv in ll_cnv:
        if not ld_cnv_name.has_key("%s:%s"%(ld_cnv["Start Position"],ld_cnv["End Position"])):
            ld_cnv_name["%s:%s"%(ld_cnv["Start Position"],ld_cnv["End Position"])] = li_counter
            li_counter += 1
    ll_named_cnv = []
    for ld_cnv in ll_cnv:
        ld_cnv["CNV ID"] = ld_cnv_name["%s:%s"%(ld_cnv["Start Position"],ld_cnv["End Position"])]
        ll_named_cnv.append(ld_cnv)
    return ll_named_cnvdef merge_intervals(intervals):
    intervals = iter(sorted(intervals))     current_lo, current_hi = next(intervals)
    for lo, hi in intervals:
        if lo <= current_hi:
            if hi > current_hi:
                current_hi = hi
        else:
            yield [current_lo, current_hi]
            current_lo = lo
            current_hi = hi
    yield [current_lo, current_hi]     def create_csv_file(ll_fieldnames,outputfile,ls_delimiter,ll_data):
          print "Writing new csv file of %s cnv" % len(ll_data)
        test_file = open(outputfile,'wb')
        csvwriter = csv.DictWriter(test_file, delimiter=ls_delimiter, fieldnames=ll_fieldnames)
        csvwriter.writerow(dict((fn,fn) for fn in ll_fieldnames))
        for row in ll_data:
            csvwriter.writerow(row)
        test_file.close()def cluster_cnv(inputfile,outputfile,ls_delimiter):
    ld_chr_clusters = {}
    lo_cnv = csv.DictReader(open(inputfile,'rU'),delimiter=ls_delimiter)
    results_list = split(inputfile,ls_delimiter)
    ld_chr_range = create_ranges(results_list)
    for key in ld_chr_range:
        ll_intervals = list(merge_intervals(ld_chr_range[key]))
        #print key,ll_intervals
        ld_chr_clusters[key] = ll_intervals
    ll_all_cnv = []
    for ll_cnv in results_list:
        for ld_cnv in ll_cnv:
            for ll_range in ld_chr_clusters[ld_cnv["Chr"]]:
                if (  int(ld_cnv['Start Position']) >= int(ll_range[0]) ) and (  int( ld_cnv['End Position'] ) <= int(ll_range[1]) ) :
                    ld_cnv["CNV Region"] = "Chr:%s:%s-%s" % (ld_cnv["Chr"],ll_range[0],ll_range[1])
                    ll_all_cnv.append(ld_cnv)
                #else:
                    #if ld_cnv["Chr"] == "4":
                        #print ld_cnv["Chr"],ld_cnv['Start Position'],ld_cnv['End Position']
    print "%s where clustered" % (len(ll_all_cnv))
    ll_named_cnv = rename_cnv(ll_all_cnv)
    ll_fieldnames = lo_cnv.fieldnames
    ll_fieldnames.append("CNV Region")
    ll_fieldnames.append("CNV ID")
    create_csv_file(ll_fieldnames,outputfile,ls_delimiter,ll_named_cnv)
    venn_scripts(outputfile,ls_delimiter)def create_ranges(results_list):
    ld_chr_range = {}
```

```
for ll_chr in results_list:
    ls_chr = "%s"%ll_chr[0]["Chr"]
    ld_chr_range[ls_chr] = []
    for ld_cnv in ll_chr:
        if [ int(ld_cnv['Start Position']) , int(ld_cnv['End Position']) ] not in ld_chr_range[ls_chr]:
            ld_chr_range[ls_chr].append( [ int(ld_cnv['Start Position']) ,int(ld_cnv['End Position']) ] )
        return ld_chr_range
            def split(inputfile,ls_delimiter):
lo_cnv = csv.DictReader(open(inputfile,'rU'),delimiter=ls_delimiter)
ll_cnv = list(lo_cnv)
print "Total of %s CNV found" % len(ll_cnv)
result = collections.defaultdict(list)
for d in ll_cnv:
    result[d['Chr']].append(d)    result_list = result.values()
return result_listdef venn_scripts(outputfile,ls_delimiter):
ld_venn_data = {}
lo_cnv = csv.DictReader(open(outputfile,'rU'),delimiter=ls_delimiter)
for ld_row in lo_cnv:
    if ld_venn_data.has_key(ld_row["Analyses"]):
        ld_venn_data[ld_row["Analyses"]].append(int(ld_row["CNV ID"]))
    else:
        ld_venn_data[ld_row["Analyses"]] = [ int(ld_row["CNV ID"])]
ll_data = []
for key in sorted(ld_venn_data.keys()):
    ll_data.append(ld_venn_data[key])
ll_rows = zip(*ll_data)
lo_file = open("ven.csv",'wb')
csvwriter = csv.writer(lo_file, delimiter=";")
csvwriter.writerow(tuple(sorted(ld_venn_data.keys())))
for lt_row in ll_rows:
    csvwriter.writerow(lt_row)
lo_file.close()
ls_sets = ""
li_count = 1
ll_sets = []
ls_col_names = "colnames(counts) <- c("
ls_universe = "universe <- sort(unique(c("
ls_counts = ""
for key in sorted(ld_venn_data.keys()):
    ls_sets += "set%s <- data$%s\n" % (li_count,key)
    ls_col_names += "'%s',"% key
    ll_sets.append("set%s"%li_count)
    li_count += 1
ls_col_names = ls_col_names[:-1]
ls_col_names += ")"
for ls_set in ll_sets:
    ls_universe += "%s,"% ls_set
    ls_counts += "counts[i,%s] <- universe[i] %%in%% %s\n" % (ls_set[-1],ls_set)
ls_universe = ls_universe[:-1]
ls_universe += ")))"
ls_counts += "}"
ls_pdf_file_name =  "Venn.pdf"
ls_rscript = '''
library(limma)
data <- read.csv("ven.csv",header=TRUE,sep=";")
%s
%s
counts <- matrix(0, nrow=length(universe), ncol=%s)
for (i in 1:length(universe)) {
%s
%s
cols<-c(rainbow(%s))
pdf("%s")
```

```
    par( las=2,  cex.axis=0.5, cex.lab=1, cex.main=2, cex.sub=1)
    vennDiagram(vennCounts(counts), circle.col=cols)
    dev.off()
    ''' % (ls_sets, ls_universe, len(ll_sets), ls_counts, ls_col_names,len(ll_sets),ls_pdf_file_name)
    ls_rscript_name =  "./RScript.R"
    f_ofile = open( ls_rscript_name, 'wb' )
    f_ofile.writelines(ls_rscript)
    f_ofile.close()
    cmd = r"Rscript --vanilla --verbose ./RScript.R"
    print cmd
    subprocess.call( cmd, stdout=subprocess.PIPE, stderr=subprocess.PIPE, shell=True )
def main(argv):
    opts, args = getopt.getopt(argv,"i:o:d:",["ifile=","ofile=","delimiter="])
    for opt, arg in opts:
        if opt in ("-i", "--ifile"):
            inputfile = arg
        elif opt in ("-o", "--ofile"):
            outputfile = arg
        elif opt in ("-d", "--delimiter"):
            ls_delimiter = arg
            cluster_cnv(inputfile,outputfile,ls_delimiter)    if __name__ == "__main__":
   main(sys.argv[1:])
```

Additional file 3.2 HPBs identified in 492 Nguni cattle that covered or lay within close proximity of genes (HPB), the number of genes (Num GEN) covered and the gene names (GEN).

HPB	Num GEN	GEN
chr28:44261945-44261945	1	*MARCH8*
chr1:84679123-84850448	1	*ABCC5*
chr23:27305227-27383176	3	*ABCF1, GNL1, PRR3*
chr19:15980813-15997977	1	*ACCN1*
chr16:44076390-44223723	3	*ACOT7, HES2, TNFRSF25*
chr3:21371931-21549827	7	*ADAMTSL4, ECM1, ENSA, GOLPH3L, MCL1, RPRD2, TARS2*
chr4:24436431-24445075	1	*AGMO*
chr1:147812230-147978297	3	*AGPAT3, CSTB, RRP1*
chr2:20011011-20023792	1	*AGPS*
chr2:131196339-131305612	2	*AHDC1, WASF2*
chr16:30756856-30935096	1	*AKT3*
chr1:50359829-50465233	1	*ALCAM*
chr2:4587203-4680618	2	*AMMECR1L, POLR2D*
chr21:25274580-25404906	2	*ANKRD34C, LOC539132*
chr3:31729269-31843839	3	*AP4B1, BCL2L15, PTPN22*
chr21:27887072-28055227	1	*APBA2*
chr11:68662418-68759678	3	*APLF, FBXO48, PLEK*
chr3:15404930-15525599	5	*APOA1BP, BCAN, GPATCH4, HAPLN2, IQGAP3*
chr19:28204745-28366979	7	*ARHGEF15, NDEL1, ODF4, PFAS, RANGRF, RPL26, SLC25A35*
chr9:43595237-43710426	1	*ARMC2*
chr15:38078775-38078775	1	*ARNTL*
chr6:94141326-94158559	1	*ART3*
chr9:34405240-34488147	1	*ASF1A*
chr2:22372855-22451660	1	*ATF2*
chr29:44740917-44756502	1	*ATG2A*
chr22:56526462-56526462	1	*ATG7*
chr24:632760-706868	1	*ATP9B*
chr13:60002265-60198826	5	*AURKA, CSTF1, FAM209B, FAM210B, RTFDC1*
chr2:27505377-27658055	1	*BBS5*
chr9:20347622-20356212	1	*BCKDHB*
chr5:75627333-75757551	1	*BTBD11*
chr26:22526369-22628269	1	*BTRC*
chr16:48694547-48776445	3	*C16H1orf170, KLHL17, NOC2L*
chr17:64164801-64308786	4	*C17H12orf52, CCDC42B, IQCD, SLC24A6*
chr1:147484581-147504216	2	*C1H21orf2, PFKL*
chr5:30639518-30659497	1	*C5H12orf44*
chr18:38645105-38756763	1	*CALB2*
chr3:57811390-58003570	1	*CCBL2*
chr3:58030374-58040470	1	*CCBL2*
chr4:45283780-45473258	2	*CCDC146, FGL2*
chr1:52409229-52537365	1	*CCDC54*
chr10:51307984-51502722	2	*CCNB2, RNF111*
chr13:75558137-75567844	1	*CD40*
chr8:87402415-87598677	2	*CDC14B, HABP4*
chr2:136417593-136506232	1	*CDC42*
chr20:51480878-51569900	1	*CDH10*
chr20:56867611-56869138	1	*CDH18*
chr24:20163874-20359135	1	*CELF4*
chr18:23426214-23436682	1	*CES1*
chr5:76286670-76385743	3	*CHRNA3, SYN3, TIMP3*

HPB	Num GEN	GEN
chr21:49290972-49362705	1	*CLEC14A*
chr23:19749211-19767455	1	*CLIC5*
chr21:33948119-34121218	4	*CLK3, CSK, CYP1A2, EDC3*
chr1:130517998-130684886	1	*CLSTN2*
chr8:28795833-28799249	1	*CNTLN*
chr25:3676450-3801478	1	*CREBBP*
chr10:45687660-45834171	1	*CSNK1G1*
chr7:69588814-69776569	2	*CYFIP2, ITK*
chr21:65198296-65198296	1	*DEGS2*
chr11:97936469-98127670	1	*DENND1A*
chr16:37748202-37936588	2	*DHRS3, VPS13D*
chr8:46144093-46154811	1	*DOCK8*
chr8:77028125-77204243	1	*EBF2*
chr16:63248646-63424839	2	*EDEM3, FAM129A*
chr20:38059359-38252896	2	*EGFLAM, LIFR*
chr24:21571435-21743914	3	*ELP2, FHOD3, MOCOS*
chr3:69752915-69870365	1	*ELTD1*
chr5:114355659-114473220	3	*ERC1, FBXL14, WNT5B*
chr19:18351965-18467274	3	*EVI2A, EVI2B, OMG*
chr11:12201969-12363892	1	*EXOC6B*
chr14:41741387-41897082	3	*FABP4, FABP9, PMP2*
chr12:80629629-80629629	1	*FAM155A*
chr12:80892109-81084869	1	*FAM155A*
chr14:10171919-10174410	1	*FAM49B*
chr7:24571599-24655689	1	*FBN2*
chr7:46537357-46553716	1	*FBXL21*
chr7:17913294-18022614	3	*FEM1A, MIR7, TICAM1*
chr13:48305310-48318066	1	*FERMT1*
chr18:33749406-33856062	4	*FHOD1, KCTD19, LRRC36, SLC9A5*
chr6:71421017-71552977	2	*FIP1L1, SCFD2*
chr12:17890808-18076789	1	*FNDC3A*
chr4:55566787-55687238	1	*FOXP2*
chr18:21237887-21250173	1	*FTO*
chr18:21675881-21869056	1	*FTO*
chr19:49173286-49329111	5	*FTSJ3, GH1, PSMC5, SMARCD2, TCAM1*
chr6:66991502-67070334	1	*GABRG1*
chr8:4304423-4372100	1	*GALNTL6*
chr8:5036995-5176744	1	*GALNTL6*
chr24:1854858-1854953	1	*GALR1*
chr25:1955733-2060211	8	*GFER, NDUFB10, RNF151, RPL3L, RPS2, SEPX1, SYNGR3, TBL3*
chr17:43956001-44065527	1	*GLRB*
chr23:16951579-17045647	2	*GLTSCR1L, RPL7L1*
chr8:55961041-56151592	1	*GNAQ*
chr7:52224595-52419683	5	*GNPDA1, KIAA0141, PCDH1, PCDH12, RNF14*
chr12:64631815-64712358	1	*GPC5*
chr12:64745304-64761912	1	*GPC5*
chr27:7962333-8077538	1	*GPM6A*
chr13:25606469-25631340	1	*GPR158*
chr9:50922485-51096382	1	*GRIK2*
chr4:94153292-94226966	1	*GRM8*
chr22:57410486-57430055	1	*H1FOO*
chr4:70643370-70839756	2	*HIBADH, TAX1BP1*
chr23:30244691-30369967	2	*HIST1H2BN, POM121L2*

HPB	Num GEN	GEN
chr26:19825454-20012464	3	*HPS1, MIR1287, PYROXD2*
chr12:60832218-61026341	1	*HTATSF1*
chr1:83030921-83132079	2	*IGF2BP2, SENP2*
chr7:71614369-71808779	1	*IL12B*
chr4:59291348-59428540	1	*IMMP2L*
chr3:89320961-89327246	1	*INADL*
chr16:69323428-69342974	1	*INTS7*
chr22:11035354-11063911	1	*ITGA9*
chr21:58060052-58072849	1	*ITPK1*
chr7:54632239-54735242	1	*KCTD16*
chr13:34909187-35029615	1	*KIAA1462*
chr12:48780336-48794617	1	*KLF12*
chr19:42613950-42752348	6	*KRT31, KRT32, KRT34, KRT35, KRT36, LOC618455*
chr9:43945908-44075848	1	*LACE1*
chr5:76501658-76691828	4	*LARGE, LOC511240, MGC137014, MGC137211*
chr10:13239142-13317919	1	*LCTL*
chr10:59812472-59948769	2	*LEO1, TMOD3*
chr4:85328669-85458623	1	*LOC613630*
chr7:59686629-59700374	1	*LOC777593*
chr10:55510249-55611885	2	*LOC788201, MIR628*
chr13:70523797-70656675	1	*LPIN3*
chr1:61450475-61540639	1	*LSAMP*
chr28:7001292-7013666	1	*LYST*
chr19:47010170-47200732	1	*MAPT*
chr1:81540249-81623098	2	*MASP1, RTP1*
chr2:65044427-65069112	1	*MGAT5*
chr15:47429234-47605562	2	*MGC137098, UBQLN3*
chr25:30927675-31107742	2	*MIR2386, WBSCR17*
chr19:9352943-9354310	2	*MIR454, SKA2*
chr14:62676800-62794502	2	*MIR599, MIR875*
chr25:14257075-14261980	1	*MKL2*
chr19:44799390-44808197	1	*MPP2*
chr27:22604390-22757505	1	*MSR1*
chr24:38307677-38449638	3	*MYL12A, MYL12B, MYOM1*
chr29:18419356-18592194	1	*NARS2*
chr3:89718673-89738009	1	*NFIA*
chr28:7048297-7188752	1	*NID1*
chr9:27946591-27965979	1	*NKAIN2*
chr9:27991255-28070840	1	*NKAIN2*
chr26:23167656-23180667	1	*NOLC1*
chr6:108934953-109022523	2	*NSG1, STX18*
chr5:74348477-74518588	2	*NUAK1, TCP11L2*
chr10:45351906-45488421	2	*OAZ2, ZNF609*
chr10:70871943-71022679	1	*OTX2*
chr20:34728244-34743583	1	*OXCT1*
chr17:56512519-56514551	1	*P2RX4*
chr13:2585410-2711744	1	*PAK7*
chr11:45130713-45267174	1	*PAPOLG*
chr25:14676885-14686647	1	*PARN*
chr2:14713525-14717851	1	*PDE1A*
chr3:104930456-105043063	3	*PDZK1IP1, STIL, TAL1*
chr1:89625133-89784824	2	*PIK3CA, ZMAT3*
chr16:65695370-65704897	1	*PLA2G4A*

HPB	Num GEN	GEN
chr5:41476949-41601867	2	*PPHLN1, ZCRB1*
chr2:15001586-15021561	1	*PPP1R1C*
chr10:45013558-45207315	1	*PTGDR*
chr4:44952454-44967890	1	*PTPN12*
chr6:105390830-105588440	2	*PTPN13, UBE2I*
chr10:81549606-81648739	3	*RAD51B, RDH12, ZFYVE26*
chr25:7381787-7393865	1	*RBFOX1*
chr16:54846459-54853614	1	*RFWD2*
chr6:94262697-94439594	1	*SCARB2*
chr14:24482969-24643266	1	*SDCBP*
chr6:101399909-101414694	1	*SEC31A*
chr21:59173696-59248804	1	*SERPINA10*
chr13:53618441-53618942	1	*SIRPA*
chr26:37919585-38060272	2	*SLC18A2, VAX1*
chr28:26667266-26786093	2	*SLC29A3, UNC5B*
chr5:44059505-44187145	1	*SLC2A13*
chr1:127429635-127591192	1	*SLC9A9*
chr16:66959283-67067114	1	*SMYD2*
chr18:35294024-35306248	1	*SNTB2*
chr3:11674665-11770282	1	*SPTA1*
chr11:15008020-15173418	1	*SRD5A2*
chr11:47283116-47300300	1	*ST6GAL2*
chr4:76286055-76482235	2	*STEAP1, STEAP2*
chr15:42949577-42960757	1	*STK33*
chr1:66519111-66668755	1	*STXBP5L*
chr5:9263262-9437669	1	*SYT1*
chr12:50320265-50324576	1	*TBC1D4*
chr10:53289822-53484074	1	*TCF12*
chr7:42948148-43132401	2	*TCF3, UQCR11*
chr23:23675301-23692539	1	*TFAP2B*
chr19:21716537-21733030	1	*TIMM22*
chr8:59778455-59954021	1	*TLE1*
chr17:1787723-1802505	1	*TLL1*
chr17:42567839-42584965	1	*TMEM144*
chr2:65004490-65017334	1	*TMEM163*
chr8:49981054-50074108	1	*TMEM2*
chr16:37309624-37391539	1	*TNFSF18*
chr17:4700529-4868261	1	*TRIM2*
chr6:23562312-23565493	1	*UBE2D3*
chr16:40282077-40396751	1	*UBE4B*
chr10:37971937-38097257	1	*UBR1*
chr27:32800374-32811897	1	*UNC5D*
chr18:39521360-39624096	1	*URI1*
chr16:42686848-42843345	1	*VAMP3*
chr27:39440886-39457953	1	*VDAC3*
chr18:4232138-4397257	1	*WWOX*
chr1:120479376-120496926	1	*WWTR1*
chr16:33504380-33521338	1	*XCL1*
chr8:63815716-63956656	1	*ZCCHC7*

Addendum B

Additional file 3.1 The python script developed to merge adjacent and overlapping CNVs into CNVRs.

```
import sys,os,csv, getopt, collections
import numpy as np
import subprocessdef rename_cnv(ll_cnv):
    ld_cnv_name = {}
    li_counter = 1
    for ld_cnv in ll_cnv:
        if not ld_cnv_name.has_key("%s:%s"%(ld_cnv["Start Position"],ld_cnv["End Position"])):
            ld_cnv_name["%s:%s"%(ld_cnv["Start Position"],ld_cnv["End Position"])] = li_counter
            li_counter += 1
    ll_named_cnv = []
    for ld_cnv in ll_cnv:
        ld_cnv["CNV ID"] = ld_cnv_name["%s:%s"%(ld_cnv["Start Position"],ld_cnv["End Position"])]
        ll_named_cnv.append(ld_cnv)
    return ll_named_cnvdef merge_intervals(intervals):
    intervals = iter(sorted(intervals))     current_lo, current_hi = next(intervals)
    for lo, hi in intervals:
        if lo <= current_hi:
            if hi > current_hi:
                current_hi = hi
        else:
            yield [current_lo, current_hi]
            current_lo = lo
            current_hi = hi
    yield [current_lo, current_hi]     def create_csv_file(ll_fieldnames,outputfile,ls_delimiter,ll_data):
          print "Writing new csv file of %s cnv" % len(ll_data)
        test_file = open(outputfile,'wb')
        csvwriter = csv.DictWriter(test_file, delimiter=ls_delimiter, fieldnames=ll_fieldnames)
        csvwriter.writerow(dict((fn,fn) for fn in ll_fieldnames))
        for row in ll_data:
            csvwriter.writerow(row)
        test_file.close()def cluster_cnv(inputfile,outputfile,ls_delimiter):
    ld_chr_clusters = {}
    lo_cnv = csv.DictReader(open(inputfile,'rU'),delimiter=ls_delimiter)
    results_list = split(inputfile,ls_delimiter)
    ld_chr_range = create_ranges(results_list)
    for key in ld_chr_range:
        ll_intervals = list(merge_intervals(ld_chr_range[key]))
        #print key,ll_intervals
        ld_chr_clusters[key] = ll_intervals
    ll_all_cnv = []
    for ll_cnv in results_list:
        for ld_cnv in ll_cnv:
            for ll_range in ld_chr_clusters[ld_cnv["Chr"]]:
                if (  int(ld_cnv['Start Position']) >= int(ll_range[0]) ) and (  int( ld_cnv['End Position'] ) <= int(ll_range[1]) ) :
                    ld_cnv["CNV Region"] = "Chr:%s:%s-%s" % (ld_cnv["Chr"],ll_range[0],ll_range[1])
                    ll_all_cnv.append(ld_cnv)
                #else:
                    #if ld_cnv["Chr"] == "4":
                        #print ld_cnv["Chr"],ld_cnv['Start Position'],ld_cnv['End Position']
    print "%s where clustered" % (len(ll_all_cnv))
    ll_named_cnv = rename_cnv(ll_all_cnv)
    ll_fieldnames = lo_cnv.fieldnames
    ll_fieldnames.append("CNV Region")
    ll_fieldnames.append("CNV ID")
    create_csv_file(ll_fieldnames,outputfile,ls_delimiter,ll_named_cnv)
    venn_scripts(outputfile,ls_delimiter)def create_ranges(results_list):
    ld_chr_range = {}
```

```
for ll_chr in results_list:
    ls_chr = "%s"%ll_chr[0]["Chr"]
    ld_chr_range[ls_chr] = []
    for ld_cnv in ll_chr:
        if [ int(ld_cnv['Start Position']) , int(ld_cnv['End Position']) ] not in ld_chr_range[ls_chr]:
            ld_chr_range[ls_chr].append( [ int(ld_cnv['Start Position']) ,int(ld_cnv['End Position']) ] )
        return ld_chr_range
            def split(inputfile,ls_delimiter):
lo_cnv = csv.DictReader(open(inputfile,'rU'),delimiter=ls_delimiter)
ll_cnv = list(lo_cnv)
print "Total of %s CNV found" % len(ll_cnv)
result = collections.defaultdict(list)
for d in ll_cnv:
    result[d['Chr']].append(d)    result_list = result.values()
return result_listdef venn_scripts(outputfile,ls_delimiter):
ld_venn_data = {}
lo_cnv = csv.DictReader(open(outputfile,'rU'),delimiter=ls_delimiter)
for ld_row in lo_cnv:
    if ld_venn_data.has_key(ld_row["Analyses"]):
        ld_venn_data[ld_row["Analyses"]].append(int(ld_row["CNV ID"]))
    else:
        ld_venn_data[ld_row["Analyses"]] = [ int(ld_row["CNV ID"])]
ll_data = []
for key in sorted(ld_venn_data.keys()):
    ll_data.append(ld_venn_data[key])
ll_rows = zip(*ll_data)
lo_file = open("ven.csv",'wb')
csvwriter = csv.writer(lo_file, delimiter=";")
csvwriter.writerow(tuple(sorted(ld_venn_data.keys())))
for lt_row in ll_rows:
    csvwriter.writerow(lt_row)
lo_file.close()
ls_sets = ""
li_count = 1
ll_sets = []
ls_col_names = "colnames(counts) <- c("
ls_universe = "universe <- sort(unique(c("
ls_counts = ""
for key in sorted(ld_venn_data.keys()):
    ls_sets += "set%s <- data$%s\n" % (li_count,key)
    ls_col_names += "'%s',"% key
    ll_sets.append("set%s"%li_count)
    li_count += 1
ls_col_names = ls_col_names[:-1]
ls_col_names += ")"
for ls_set in ll_sets:
    ls_universe += "%s,"% ls_set
    ls_counts += "counts[i,%s] <- universe[i] %%in%% %s\n" % (ls_set[-1],ls_set)
ls_universe = ls_universe[:-1]
ls_universe += ")))"
ls_counts += "}"
ls_pdf_file_name =  "Venn.pdf"
ls_rscript = '''
library(limma)
data <- read.csv("ven.csv",header=TRUE,sep=";")
%s
%s
counts <- matrix(0, nrow=length(universe), ncol=%s)
for (i in 1:length(universe)) {
%s
%s
cols<-c(rainbow(%s))
pdf("%s")
```

```
    par( las=2,  cex.axis=0.5, cex.lab=1, cex.main=2, cex.sub=1)
    vennDiagram(vennCounts(counts), circle.col=cols)
    dev.off()
    ''' % (ls_sets, ls_universe, len(ll_sets), ls_counts, ls_col_names,len(ll_sets),ls_pdf_file_name)
    ls_rscript_name =  "./RScript.R"
    f_ofile = open( ls_rscript_name, 'wb' )
    f_ofile.writelines(ls_rscript)
    f_ofile.close()
    cmd = r"Rscript --vanilla --verbose ./RScript.R"
    print cmd
    subprocess.call( cmd, stdout=subprocess.PIPE, stderr=subprocess.PIPE, shell=True )
def main(argv):
    opts, args = getopt.getopt(argv,"i:o:d:",["ifile=","ofile=","delimiter="])
    for opt, arg in opts:
        if opt in ("-i", "--ifile"):
            inputfile = arg
        elif opt in ("-o", "--ofile"):
            outputfile = arg
        elif opt in ("-d", "--delimiter"):
            ls_delimiter = arg
            cluster_cnv(inputfile,outputfile,ls_delimiter)    if __name__ == "__main__":
  main(sys.argv[1:])
```

Additional file 3.2 HPBs identified in 492 Nguni cattle that covered or lay within close proximity of genes (HPB), the number of genes (Num GEN) covered and the gene names (GEN).

HPB	Num GEN	GEN
chr28:44261945-44261945	1	*MARCH8*
chr1:84679123-84850448	1	*ABCC5*
chr23:27305227-27383176	3	*ABCF1, GNL1, PRR3*
chr19:15980813-15997977	1	*ACCN1*
chr16:44076390-44223723	3	*ACOT7, HES2, TNFRSF25*
chr3:21371931-21549827	7	*ADAMTSL4, ECM1, ENSA, GOLPH3L, MCL1, RPRD2, TARS2*
chr4:24436431-24445075	1	*AGMO*
chr1:147812230-147978297	3	*AGPAT3, CSTB, RRP1*
chr2:20011011-20023792	1	*AGPS*
chr2:131196339-131305612	2	*AHDC1, WASF2*
chr16:30756856-30935096	1	*AKT3*
chr1:50359829-50465233	1	*ALCAM*
chr2:4587203-4680618	2	*AMMECR1L, POLR2D*
chr21:25274580-25404906	2	*ANKRD34C, LOC539132*
chr3:31729269-31843839	3	*AP4B1, BCL2L15, PTPN22*
chr21:27887072-28055227	1	*APBA2*
chr11:68662418-68759678	3	*APLF, FBXO48, PLEK*
chr3:15404930-15525599	5	*APOA1BP, BCAN, GPATCH4, HAPLN2, IQGAP3*
chr19:28204745-28366979	7	*ARHGEF15, NDEL1, ODF4, PFAS, RANGRF, RPL26, SLC25A35*
chr9:43595237-43710426	1	*ARMC2*
chr15:38078775-38078775	1	*ARNTL*
chr6:94141326-94158559	1	*ART3*
chr9:34405240-34488147	1	*ASF1A*
chr2:22372855-22451660	1	*ATF2*
chr29:44740917-44756502	1	*ATG2A*
chr22:56526462-56526462	1	*ATG7*
chr24:632760-706868	1	*ATP9B*
chr13:60002265-60198826	5	*AURKA, CSTF1, FAM209B, FAM210B, RTFDC1*
chr2:27505377-27658055	1	*BBS5*
chr9:20347622-20356212	1	*BCKDHB*
chr5:75627333-75757551	1	*BTBD11*
chr26:22526369-22628269	1	*BTRC*
chr16:48694547-48776445	3	*C16H1orf170, KLHL17, NOC2L*
chr17:64164801-64308786	4	*C17H12orf52, CCDC42B, IQCD, SLC24A6*
chr1:147484581-147504216	2	*C1H21orf2, PFKL*
chr5:30639518-30659497	1	*C5H12orf44*
chr18:38645105-38756763	1	*CALB2*
chr3:57811390-58003570	1	*CCBL2*
chr3:58030374-58040470	1	*CCBL2*
chr4:45283780-45473258	2	*CCDC146, FGL2*
chr1:52409229-52537365	1	*CCDC54*
chr10:51307984-51502722	2	*CCNB2, RNF111*
chr13:75558137-75567844	1	*CD40*
chr8:87402415-87598677	2	*CDC14B, HABP4*
chr2:136417593-136506232	1	*CDC42*
chr20:51480878-51569900	1	*CDH10*
chr20:56867611-56869138	1	*CDH18*
chr24:20163874-20359135	1	*CELF4*
chr18:23426214-23436682	1	*CES1*
chr5:76286670-76385743	3	*CHRNA3, SYN3, TIMP3*

HPB	Num GEN	GEN
chr21:49290972-49362705	1	*CLEC14A*
chr23:19749211-19767455	1	*CLIC5*
chr21:33948119-34121218	4	*CLK3, CSK, CYP1A2, EDC3*
chr1:130517998-130684886	1	*CLSTN2*
chr8:28795833-28799249	1	*CNTLN*
chr25:3676450-3801478	1	*CREBBP*
chr10:45687660-45834171	1	*CSNK1G1*
chr7:69588814-69776569	2	*CYFIP2, ITK*
chr21:65198296-65198296	1	*DEGS2*
chr11:97936469-98127670	1	*DENND1A*
chr16:37748202-37936588	2	*DHRS3, VPS13D*
chr8:46144093-46154811	1	*DOCK8*
chr8:77028125-77204243	1	*EBF2*
chr16:63248646-63424839	2	*EDEM3, FAM129A*
chr20:38059359-38252896	2	*EGFLAM, LIFR*
chr24:21571435-21743914	3	*ELP2, FHOD3, MOCOS*
chr3:69752915-69870365	1	*ELTD1*
chr5:114355659-114473220	3	*ERC1, FBXL14, WNT5B*
chr19:18351965-18467274	3	*EVI2A, EVI2B, OMG*
chr11:12201969-12363892	1	*EXOC6B*
chr14:41741387-41897082	3	*FABP4, FABP9, PMP2*
chr12:80629629-80629629	1	*FAM155A*
chr12:80892109-81084869	1	*FAM155A*
chr14:10171919-10174410	1	*FAM49B*
chr7:24571599-24655689	1	*FBN2*
chr7:46537357-46553716	1	*FBXL21*
chr7:17913294-18022614	3	*FEM1A, MIR7, TICAM1*
chr13:48305310-48318066	1	*FERMT1*
chr18:33749406-33856062	4	*FHOD1, KCTD19, LRRC36, SLC9A5*
chr6:71421017-71552977	2	*FIP1L1, SCFD2*
chr12:17890808-18076789	1	*FNDC3A*
chr4:55566787-55687238	1	*FOXP2*
chr18:21237887-21250173	1	*FTO*
chr18:21675881-21869056	1	*FTO*
chr19:49173286-49329111	5	*FTSJ3, GH1, PSMC5, SMARCD2, TCAM1*
chr6:66991502-67070334	1	*GABRG1*
chr8:4304423-4372100	1	*GALNTL6*
chr8:5036995-5176744	1	*GALNTL6*
chr24:1854858-1854953	1	*GALR1*
chr25:1955733-2060211	8	*GFER, NDUFB10, RNF151, RPL3L, RPS2, SEPX1, SYNGR3, TBL3*
chr17:43956001-44065527	1	*GLRB*
chr23:16951579-17045647	2	*GLTSCR1L, RPL7L1*
chr8:55961041-56151592	1	*GNAQ*
chr7:52224595-52419683	5	*GNPDA1, KIAA0141, PCDH1, PCDH12, RNF14*
chr12:64631815-64712358	1	*GPC5*
chr12:64745304-64761912	1	*GPC5*
chr27:7962333-8077538	1	*GPM6A*
chr13:25606469-25631340	1	*GPR158*
chr9:50922485-51096382	1	*GRIK2*
chr4:94153292-94226966	1	*GRM8*
chr22:57410486-57430055	1	*H1FOO*
chr4:70643370-70839756	2	*HIBADH, TAX1BP1*
chr23:30244691-30369967	2	*HIST1H2BN, POM121L2*

HPB	Num GEN	GEN
chr26:19825454-20012464	3	*HPS1, MIR1287, PYROXD2*
chr12:60832218-61026341	1	*HTATSF1*
chr1:83030921-83132079	2	*IGF2BP2, SENP2*
chr7:71614369-71808779	1	*IL12B*
chr4:59291348-59428540	1	*IMMP2L*
chr3:89320961-89327246	1	*INADL*
chr16:69323428-69342974	1	*INTS7*
chr22:11035354-11063911	1	*ITGA9*
chr21:58060052-58072849	1	*ITPK1*
chr7:54632239-54735242	1	*KCTD16*
chr13:34909187-35029615	1	*KIAA1462*
chr12:48780336-48794617	1	*KLF12*
chr19:42613950-42752348	6	*KRT31, KRT32, KRT34, KRT35, KRT36, LOC618455*
chr9:43945908-44075848	1	*LACE1*
chr5:76501658-76691828	4	*LARGE, LOC511240, MGC137014, MGC137211*
chr10:13239142-13317919	1	*LCTL*
chr10:59812472-59948769	2	*LEO1, TMOD3*
chr4:85328669-85458623	1	*LOC613630*
chr7:59686629-59700374	1	*LOC777593*
chr10:55510249-55611885	2	*LOC788201, MIR628*
chr13:70523797-70656675	1	*LPIN3*
chr1:61450475-61540639	1	*LSAMP*
chr28:7001292-7013666	1	*LYST*
chr19:47010170-47200732	1	*MAPT*
chr1:81540249-81623098	2	*MASP1, RTP1*
chr2:65044427-65069112	1	*MGAT5*
chr15:47429234-47605562	2	*MGC137098, UBQLN3*
chr25:30927675-31107742	2	*MIR2386, WBSCR17*
chr19:9352943-9354310	2	*MIR454, SKA2*
chr14:62676800-62794502	2	*MIR599, MIR875*
chr25:14257075-14261980	1	*MKL2*
chr19:44799390-44808197	1	*MPP2*
chr27:22604390-22757505	1	*MSR1*
chr24:38307677-38449638	3	*MYL12A, MYL12B, MYOM1*
chr29:18419356-18592194	1	*NARS2*
chr3:89718673-89738009	1	*NFIA*
chr28:7048297-7188752	1	*NID1*
chr9:27946591-27965979	1	*NKAIN2*
chr9:27991255-28070840	1	*NKAIN2*
chr26:23167656-23180667	1	*NOLC1*
chr6:108934953-109022523	2	*NSG1, STX18*
chr5:74348477-74518588	2	*NUAK1, TCP11L2*
chr10:45351906-45488421	2	*OAZ2, ZNF609*
chr10:70871943-71022679	1	*OTX2*
chr20:34728244-34743583	1	*OXCT1*
chr17:56512519-56514551	1	*P2RX4*
chr13:2585410-2711744	1	*PAK7*
chr11:45130713-45267174	1	*PAPOLG*
chr25:14676885-14686647	1	*PARN*
chr2:14713525-14717851	1	*PDE1A*
chr3:104930456-105043063	3	*PDZK1IP1, STIL, TAL1*
chr1:89625133-89784824	2	*PIK3CA, ZMAT3*
chr16:65695370-65704897	1	*PLA2G4A*

HPB	Num GEN	GEN
chr5:41476949-41601867	2	*PPHLN1, ZCRB1*
chr2:15001586-15021561	1	*PPP1R1C*
chr10:45013558-45207315	1	*PTGDR*
chr4:44952454-44967890	1	*PTPN12*
chr6:105390830-105588440	2	*PTPN13, UBE2I*
chr10:81549606-81648739	3	*RAD51B, RDH12, ZFYVE26*
chr25:7381787-7393865	1	*RBFOX1*
chr16:54846459-54853614	1	*RFWD2*
chr6:94262697-94439594	1	*SCARB2*
chr14:24482969-24643266	1	*SDCBP*
chr6:101399909-101414694	1	*SEC31A*
chr21:59173696-59248804	1	*SERPINA10*
chr13:53618441-53618942	1	*SIRPA*
chr26:37919585-38060272	2	*SLC18A2, VAX1*
chr28:26667266-26786093	2	*SLC29A3, UNC5B*
chr5:44059505-44187145	1	*SLC2A13*
chr1:127429635-127591192	1	*SLC9A9*
chr16:66959283-67067114	1	*SMYD2*
chr18:35294024-35306248	1	*SNTB2*
chr3:11674665-11770282	1	*SPTA1*
chr11:15008020-15173418	1	*SRD5A2*
chr11:47283116-47300300	1	*ST6GAL2*
chr4:76286055-76482235	2	*STEAP1, STEAP2*
chr15:42949577-42960757	1	*STK33*
chr1:66519111-66668755	1	*STXBP5L*
chr5:9263262-9437669	1	*SYT1*
chr12:50320265-50324576	1	*TBC1D4*
chr10:53289822-53484074	1	*TCF12*
chr7:42948148-43132401	2	*TCF3, UQCR11*
chr23:23675301-23692539	1	*TFAP2B*
chr19:21716537-21733030	1	*TIMM22*
chr8:59778455-59954021	1	*TLE1*
chr17:1787723-1802505	1	*TLL1*
chr17:42567839-42584965	1	*TMEM144*
chr2:65004490-65017334	1	*TMEM163*
chr8:49981054-50074108	1	*TMEM2*
chr16:37309624-37391539	1	*TNFSF18*
chr17:4700529-4868261	1	*TRIM2*
chr6:23562312-23565493	1	*UBE2D3*
chr16:40282077-40396751	1	*UBE4B*
chr10:37971937-38097257	1	*UBR1*
chr27:32800374-32811897	1	*UNC5D*
chr18:39521360-39624096	1	*URI1*
chr16:42686848-42843345	1	*VAMP3*
chr27:39440886-39457953	1	*VDAC3*
chr18:4232138-4397257	1	*WWOX*
chr1:120479376-120496926	1	*WWTR1*
chr16:33504380-33521338	1	*XCL1*
chr8:63815716-63956656	1	*ZCCHC7*

Additional file 3.3 Biological process (BP), molecular function (MF) and cellular component (CC) representation of those genes covered by HPBs identified in Nguni cattle.

BP	BosT (19799)	NG	NG EXP	NG REP	NG ENR	Pval
Detection of stimulus involved in sensory perception	998	1	13.56	-	< 0.2	8.70E-02
Detection of chemical stimulus	977	0	13.27	-	< 0.2	7.68E-03
Sensory perception	1 270	4	17.25	-	.23	6.70E-01
Sensory perception of chemical stimulus	1 031	1	14.01	-	< 0.2	5.60E-02
Cellular process	11 535	188	156.72	+	1.20	3.40E-01
Detection of stimulus	1 070	1	14.54	-	< 0.2	3.32E-02
Detection of chemical stimulus involved in sensory perception of smell	934	0	12.69	-	< 0.2	1.42E-02
Sensory perception of smell	954	0	12.96	-	< 0.2	1.07E-02
Detection of chemical stimulus involved in sensory perception	957	0	13.00	-	< 0.2	1.02E-02
CC	**BosT (19799)**	**NG**	**NG EXP**	**NG REP**	**NG ENR**	**Pval**
Organelle	9 823	169	133.46	+	1.27	8.62E-03
Synaptic vesicle membrane	30	4	0.41	+	> 5	8.01E-01
Intracellular organelle	8 784	146	119.34	+	1.22	6.78E-01
Vesicle	2 742	57	37.25	+	1.53	6.39E-01
Nuclear lumen	1 436	35	19.51	+	1.79	5.97E-01
Synapse part	244	11	3.32	+	3.32	5.86E-01
Membrane-bounded vesicle	2 656	56	36.09	+	1.55	5.11E-01
Cytoplasmic part	4 982	97	67.69	+	1.43	4.65E-02
Organelle part	4 778	94	64.92	+	1.45	4.38E-02
Protein complex	3 314	67	45.03	+	1.49	4.11E-01
Intracellular organelle part	4 610	88	62.63	+	1.40	2.64E-01
Cell	12 532	209	170.27	+	1.23	2.62E-04
Cell part	12 532	209	170.27	+	1.23	2.62E-04
Membrane-bounded organelle	8 923	155	121.23	+	1.28	2.35E-02
Intracellular part	10 117	179	137.46	+	1.30	2.09E-04
Synapse	371	17	5.04	+	3.37	1.66E-02
Cell junction	763	24	10.37	+	2.32	1.39E-01
Cellular_component	15 002	238	203.83	+	1.17	1.18E-04
Cytoplasm	7 380	139	100.27	+	1.39	1.06E-03
Intracellular	10 464	181	142.17	+	1.27	1.05E-03

MF	BosT (19799)	NG	NG EXP	NG REP	NG ENR	Pval
Protein binding	4 643	92	63.08	+	1.46	8.49E-02
Binding	10 052	170	136.57	+	1.24	5.29E-02
Olfactory receptor activity	933	0	12.68	-	< 0.2	4.54E-03
Bhlh transcription factor binding	21	4	0.29	+	> 5	4.27E-01
G-protein coupled receptor activity	1 352	5	18.37	-	0.27	3.41E-01
Molecular_function	14 111	218	191.72	+	1.14	3.08E-01

**Bos taurus* genes – BosT, Nguni genes – NG, Expected Nguni genes – NG (EXP), representation (over (+) /under (-)) – NG REP, Nguni fold enrichment – NG ENR, significance - Pval

Additional file 3.4 CNVRs distribution across individuals (IND) and genes covered (GEN) for CNVRs identified in more than 1 animal.

CNVR	IND	GEN
chr17:73713062-74998349	21	*CHCHD10, IGLL1, LOC527441, SLC5A1, VPREB3, ZNF280A, ZNF280B, ZNF70, DERL3, GSTT1, GSTT3, GSTT4, MIF, SLC2A11, SMARCB1, DDT, GGT1, GGT5, SUSD2, C17H22orf13, LOC531152, MIR2323, RTDR1, SNRPD3, SPECC1L, UPB1*
chr1:104798012-105264358	17	*SI*
chr24:28154039-28497174	16	*CDH2*
chr7:75305297-75370366	14	*GABRG2*
chr5:3260057-3434356	13	*ATXN7L3B*
chr6:43037439-43089739	13	*GBA3*
chr6:108998175-109951981	12	*LYAR, NSG1, OTOP1, STX18, TMEM128, WDR1, ZBTB49*
chr9:3651455-4439872	12	
chr19:49657396-49784054	12	*PECAM1, POLG2*
chr1:32509969-32781614	11	
chr6:71910076-72118486	11	*CHIC2*
chr28:21101833-21762976	10	*CTNNA3*
chr6:53514737-53692295	9	
chr22:59487979-60960603	9	*ACAD9, C22H3orf37, CNBP, COPG1, EFCC1, GATA2, ISY1, MIR2374, RAB7A, RPN1, EFCC1, IQSEC1, ISY1, CHCHD4, HDAC11, NUP210, TMEM43, WNT7A, XPC*
chr14:54875898-55141942	8	*ANGPT1*
chr25:41191025-42687812	8	*BRAT1, CARD11, GNA12, GRIFIN, LFNG, MIR2390, MIR2890*
chr26:837967-1012643	7	*CISD1, IPMK*
chr28:25060861-25352987	7	*C28H10orf35*
chr28:25060861-25352987	7	*C28H10orf35, COL13A1*
chr28:25060861-25352987	7	*COL13A1*
chr3:92144760-92229630	6	*FGGY*
chr3:120501439-121275236	6	*ARL4C*
chr3:120501439-121275236	6	*SH3BP4*
chr6:50887205-51102728	6	*PCDH7*
chr13:12587622-12808180	6	*CELF2*
chr21:51733686-51857151	6	*LRFN5*
chr22:24007619-24219999	6	*TRNT1*
chr27:15141168-15291371	6	*DCTD*
chr6:14419369-14633122	5	*C6H4orf32*
chr7:37070486-37289420	5	*ARL10, CLTB, COMMD10, HIGD2A, NOP16*
chr7:37070486-37289420	5	*CLTB, FAF2, HIGD2A, NOP16, RNF44*
chr17:6380171-6612637	5	*PET112*
chr17:23285669-23431642	5	*MDK*
chr21:61310103-61370773	5	*ATG2B, BDKRB1, BDKRB2*
chr26:2001199-2519854	5	*ZWINT*
chr29:49979913-50586068	5	*DHCR7, NADSYN1*
chr29:49979913-50586068	5	*NADSYN1*
chr29:49979913-50586068	5	*CARS, CDKN1C, KCNQ1, NAP1L4, PHLDA2, SLC22A18*
chr2:56456865-56694744	4	
chr3:73895121-74636373	4	*CRYZ, TNNI3K, TYW3*
chr4:38548833-38908495	4	*CACNA2D1*
chr11:6701679-6777332	4	*IL1R2*

CNVR	IND	GEN
chr11:103615735-104117370	4	*C11H9orf78, FNBP1, GPR107, PTGES, TOR1A, USP20*
chr11:103615735-104117370	4	*GPR107*
chr12:45002070-45196682	4	*KLHL1*
chr14:74991009-75145124	4	*WWP1*
chr26:17163979-17307507	4	*PDLIM1*
chr27:9096031-9512004	4	*AGA, NEIL3*
chr27:9096031-9512004	4	*AGA*
chr1:31923335-32340451	3	
chr1:42142033-42265516	3	*ARL6*
chr1:77647196-77714385	3	*OSTN*
chr1:104257251-104344681	3	*SI*
chr2:39976359-40155558	3	*ACVR1C*
chr2:39976359-40155558	3	*ACVR1C, CYTIP*
chr2:39976359-40155558	3	*CYTIP*
chr3:33598353-33864274	3	*KCND3*
chr3:66296159-66381935	3	*LPHN2*
chr4:21118823-21516258	3	*ARL4A, SCIN*
chr4:21118823-21516258	3	*ARL4A*
chr4:55566787-55656122	3	*FOXP2*
chr4:89568067-89648770	3	*CADPS2*
chr4:91080419-91227550	3	*SPAM1*
chr5:119820680-120329153	3	*TCF20*
chr5:119820680-120329153	3	*RRP7A*
chr5:119820680-120329153	3	*ARFGAP3, CYB5R3, PACSIN2, POLDIP3, RRP7A*
chr6:51769371-51831698	3	*PCDH7*
chr6:115749351-115813300	3	*PROM1*
chr7:33572841-33819177	3	*TNFAIP8*
chr9:31644651-31933710	3	*GJA1*
chr9:61719292-61773207	3	
chr11:66974370-67059602	3	*MEIS1*
chr12:31368562-31639399	3	*FLT1, MIR2300A, MIR2300B*
chr12:43551814-43799317	3	*KLHL1*
chr12:61573515-61686477	3	*HTATSF1*
chr14:1616618-2468020	3	*LY6H*
chr14:1616618-2468020	3	*PTK2*
chr14:1616618-2468020	3	*CHRAC1, EIF2C2*
chr17:25745000-25895194	3	*PCDH10*
chr20:51449833-51685293	3	*CDH10*
chr20:53097198-53171907	3	*CDH12*
chr20:53674655-54052768	3	*CDH12*
chr21:20216308-20356312	3	*ACAN, HAPLN3*
chr21:20216308-20356312	3	*ACAN, HAPLN3, MFGE8*
chr24:24302542-24499452	3	*NOL4*
chr26:25880226-25982293	3	*SORCS3*
chr26:42959100-43150604	3	*HTRA1*
chr26:42959100-43150604	3	*SPADH2*
chr29:51396010-51502868	3	*CTSD*
chr1:5351369-5541297	2	*GRIK1*
chr1:13979316-14102864	2	
chr1:16715975-17160556	2	
chr1:17296638-17504974	2	*TMPRSS15*
chr1:20165566-20213558	2	*MIR99A*

CNVR	IND	GEN
chr1:39083240-39339779	2	*STX19*
chr2:78433731-78556325	2	
chr3:1843353-1937626	2	*POU2F1*
chr3:14811080-14919029	2	*ETV3*
chr3:64428546-64599121	2	*TTLL7*
chr3:86467404-86551027	2	*UBE2U*
chr4:190619-565142	2	*VSTM2A*
chr4:29528497-29677476	2	*TMEM196*
chr4:49717334-49857470	2	*PRKAR2B*
chr4:52624584-52849283	2	*CFTR*
chr4:52624584-52849283	2	*ASZ1, WNT2*
chr4:78440044-78523846	2	*ADCY1*
chr4:84872989-84963191	2	*AMPH*
chr4:108834886-108974924	2	*LOC780933*
chr5:15149224-15224660	2	
chr5:17995212-18075032	2	*MGAT4C*
chr5:26621180-26809399	2	*PLXNC1*
chr5:72044945-72142911	2	*STAB2*
chr5:75722589-75794378	2	*BTBD11, PWP1*
chr6:4156416-4217935	2	*QRFPR*
chr6:10068059-10199636	2	
chr6:10716501-10838635	2	*NDST4*
chr6:35147153-35360248	2	*CCSER1*
chr6:89962889-90075383	2	*GC*
chr7:34559205-34678536	2	*DTWD2*
chr7:87556048-87862183	2	*COX7C*
chr8:35434141-35671830	2	
chr9:5079903-5148301	2	
chr9:5901981-5949799	2	
chr9:54190285-54317953	2	*POU3F2*
chr10:41672195-41849457	2	*RPS29*
chr10:59812472-59948769	2	*LEO1, TMOD3*
chr11:10364456-10437469	2	*DOK1*
chr11:59622185-59724651	2	
chr11:104293559-104493462	2	*ASS1, FUBP3*
chr11:104293559-104493462	2	*FUBP3*
chr11:104737799-104897001	2	*AIF1L, NUP214*
chr11:105699664-106108993	2	*MED27, NTNG2, TTF1*
chr11:105699664-106108993	2	*TTF1*
chr12:21279986-21352699	2	*MRPS31*
chr12:26434681-26558724	2	*RFC3*
chr12:39801280-39925909	2	
chr12:57661486-57803744	2	*MIR1256*
chr12:58335403-58461348	2	*MIR1256*
chr13:44750541-44876436	2	*KLF6*
chr14:15879588-15969787	2	*WDYHV1*
chr14:43242051-43430880	2	*EXT1, SAMD12*
chr14:43242051-43430880	2	*EXT1*
chr15:5400560-5791533	2	*BIRC3*
chr15:5400560-5791533	2	*C15H11orf70*
chr15:11439502-11710409	2	*PPP1R14C*
chr15:66629434-66681363	2	*PRR5L*
chr16:10308240-10540543	2	*CDC73*

CNVR	IND	GEN
chr16:50670749-50862929	2	*PDPN*
chr17:39963957-40071626	2	
chr17:73118011-73257794	2	*INPP5J, RNF185, SELM, SMTN*
chr20:10233876-10486993	2	*MAP1B*
chr20:46121445-46179978	2	
chr21:26620013-26662447	2	*MESDC2*
chr24:11079679-11311755	2	*CDH7*
chr25:279528-472458	2	*POLR3K*
chr26:2687667-2849216	2	*ZWINT*
chr26:3884506-4160853	2	
chr27:17035351-17086309	2	*SORBS2*
chr28:22226165-22355840	2	*CTNNA3*
chr28:22941657-23075883	2	*CTNNA3*
chr29:27880841-28248785	2	*LOC504623*
chr29:27880841-28248785	2	*TMEM225*

Additional file 3.5 Comparisons of genes within CNVRs (genes – GEN and number of genes – GEN Num) identified within this study with those identified within other breeds as reported by Bae *et al.,* 2010; Bickhart *et al.,* 2012 and Hou *et al.,* 2011 respectively (REF) revealed 402 genes that were unique to the Nguni.

REF	GEN Num	GEN
Bae Bickhart Hou Wang	1	*IGLL1*
Bickhart Hou Wang	1	*MGC157405*
Bae Hou Wang	10	*RTDR1 RFWD2 KLF6 R3HDM2 CTNNA3 LFNG CDH9 PWWP2B ARL6 BIRC3*
Bae Bickhart Hou	3	*CFH ANKRD26 ALDH1L1*
Hou Wang	13	*FAM5C LRFN5 CHRAC1 MIR99A SLC9A9 CECR5 OSTN CTSD ZWINT ATP8A1 EIF2C2 MIR1256 GULP1*
Bickhart Wang	2	*ADCY1 GABRA5*
Bae Wang	29	*CNBP GPC5 CARD11 SMARCB1 STX19 ARHGAP15 RAB21 ARL10 RPN1 ATG7 NOP16 RPS29 CDH6 TMEM225 POU3F2 AGA MYOM2 MDK HDAC11 CHCHD10 CADM1 SMTN MIF PPP1R14C COX7C HIGD2A ZADH2 RAB7A DERL3*
Bickhart Hou	55	*ACAD8 CYP2D6 DEFB1 LOC539042 CEACAM8 GML MGC157082 GIMAP1 SCP2 RAET1G ULBP3 RHOBTB2 MGC139169 LOC404103 COG2 HLA-A GBP4 LOC510904 DEFB7 MGC127055 SYT1 LOC790886 BoLA GAT LOC100126815 SOX5 GIMAP7 LOC785621 GIMAP4 LOC537366 PAG20 MICB CD163L1 CGN1 GBP6 GLYAT CYP11B1 ALPI PTI ECHDC2 GIMAP5 LOC512150 PPIE DEFB5 IFN1AT MGC154956 PAG16 LOC100125266 CA1 FABP2 PAG11 LOC513767 LAP MGC138914 ART5*
Bae Hou	117	*ARHGEF10L ATP9B ARVCF TCF25 ATP5D CYLC2 PCOLCE SEPT5 MACROD1 PLCG1 HAGHL GALNT13 SLC25A1 TNNT3 PI4KA YME1L1 ALCAM SBNO2 UFD1L METRN MGC127919 ANTXR2 SLC6A18 MASTL TUBA3E TOM1L2 EGFL7 ARHGEF16 STUB1 EXOC1 EFNA2 DEF8 TXNRD2 WDR24 CARD9 CYHR1 SMARCA5 ERCC1 PLA2G4A DGAT1 RANBP1 H19 EDF1 GPX4 MC1R NARFL KLC3 RBL2 CNTN5 PHPT1 TUBB3 GRAMD4 RNPC3 GNB2 CCDC37 SERPIND1 CDC45L LOC515651 KCNH1 SNAP29 CPNE8 DGCR8 CNN2 C25H16orf14 PPP1R13L LOC100125578 PPARA ACTL6B LSP1 MOSPD3 CLDN5 PTGDS C17H22orf25 CPSF1 CLIC3 FAM128B C11H9ORF142 AGPAT2 THAP7 PSMG1 CLPTM1L VPS28 TPK1 RHOT2 TTC29 TERT GPR172B HSF1 DGCR14 ZEB2 MED15 MINA KIAA1984 FBXL16 RAB40C MAMDC4 KLF15 CIRBP FRZB WDR18 TMEM141 PARF MRPL23 SREBF1 OLFM3 POLR2E DGCR2 DPP10 LRPAP1 ADCK5 C25H16ORF13 LOC788610 COMT BRWD1 ERCC2 CD3EAP NFKBIL2*
Bae Bickhart	11	*TUBA3C SLC3A1 TUBA1B PRSS7 STAT5B GSTM3 IMMP2L TPST1 CIST1 B3GNT2 MGC139164*

REF	GEN Num	GEN
Wang	402	*PTGER3 ABRA LOC527441 GRAP2 RTKN2 HSPB1 DDT LCORL MNAT1 C15H11orf70 MESDC2 PACSIN2 CACNA2D1 C8A GPR107 SH3BP4 HLTF POLG2 GDA MMS22L NUP214 SPECC1L LYAR CLRN1 RNF185 C6H4orf32 LRRN1 RRP7A USP20 SUSD2 GJA1 POLR3K ATG2B FBXO8 RFC3 GK2 ARL4A CLN8 ACYP2 SLC22A18 GRM8 FUBP3 SAMD12 FGGY DNAJC15 TMEM150B GSTA2 TM4SF4 C23H6orf141 WDR1 SRSF6 ATXN7L3B SYT5 CNKSR3 SEMA3A PTK2 MIR454 NEIL3 BCHE SCIN PCDH7 CDH10 COPG1 BDKRB1 CCSER1 PPP6R1 BDKRB2 RNF180 ASS1 NDST4 SEC23IP CAMK2D NTNG2 FAM204A GRIK2 LTBP1 MIR2323 CDH2 QKI MED27 PTGES POLDIP3 ARL4C SPAM1 C28H10orf35 INPP5J ZNF70 XPC TCF4 PLXNC1 CARS TRIM37 NADSYN1 LOC780933 LACTB2 RSPO2 SORBS2 ASPH HCK LTBP2 FAM181B C15H11orf96 GC FAF1 P2RY1 CHL1 MIR2300A TTLL7 CADPS2 MEIS1 TMEM14A PDGFD TRAM2 KCND3 01-Mar PDGFB ZMYM5 DCT HPS3 RHAG TM9SF4 XPO4 TNK2 PTGS2 C1QTNF7 USH2A REXO1 NAP1L4 OXR1 ACVR1C HNF4G TAS2R1 ASZ1 ADAMTS20 WBSCR17 IPMK LUZP2 C15H11orf58 PRMT6 HS3ST5 SMAD4 DHCR7 PROS1 GSTT4 ADCY8 AIF1L ACAN TFRC NSUN3 CRISP2 DTWD2 LONP2 C17H22orf13 LOC514194 SPADH2 FARS2 PET112 FAF2 CRISP3 ZRANB2 WWP1 HMGXB4 COL13A1 EPHB1 KIF3B ANGPT1 DPP6 CHMP4C KCNQ1 TMPRSS15 TOR1A NUP210 TNRC6B ISY1 GRIN3A UCHL3 TMEM86B RGS2 NR3C2 PTPN23 BOD1L PDPN MIR2890 DPH5 PRKAA2 GGT1 AMPH PRKAR2B GRB10 SPAG17 GRAMD1B FOXP2 FLT1 CNNM1 EXT1 PECAM1 TYW3 NOL4 ZC3H7B LEO1 MIR30B NGEF HTR1F IGSF10 ACAD9 KLHL1 MIR2313 ADCYAP1 MFGE8 ENTHD1 MIR2390 ETV3 PDLIM1 GRIK1 HMGB4 DDX31 GLYATL3 OTX2 PLXDC2 DDI1 T AK8 PIK3CG SPTY2D1 GJA3 BRAT1 GBE1 PPAPDC1A LOC504623 DOK1 IL1R2 CHD2 MBNL1 CYB5R1 WNT2 FGF9 ATP5G3 TMEM196 IZUMO3 UPB1 STX18 SORCS3 SELPLG WNT7A MIR30D OPCML TMEM43 MEOX2 AMMECR1L MGAT4C ATF2 ENDOD1 CNTNAP3 COMMD10 HSPBP1 NRG3 TNFAIP8 GNPDA2 VPREB3 ACSL1 GTF3C4 TRNT1 IQSEC1 TTF1 CHIC2 PCDH10 PRR5L TNNI3K BTBD11 MRPS31 ZNF423 DIO2 CYTIP CRYZ KCNJ3 PPAP2B KCMF1 QTRTD1 LY6H HTRA1 IFT88 P2RX2 PROM1 C22H3orf75 LOC615200 CDC73 CFTR ZNF280B PHLDA2 PWP1 DHX29 HTATSF1 CELF2 CYB5R3 VSTM2A LRP12 ARFGAP3 INPP5D CDKN1C POLE LRRC1 SEC62 TCF3 TACR3 LMO3 VCAN MAP1B MED12L TMEM128 TM4SF1 CDH7 RSAD2 SRGAP3 METTL4 ZMYM2 HERC4 EFCC1 AMY2B C11H9orf78 KIRREL3 GABRG2 QRFPR ACSL6 MBP CPS1 LOC527409 LYRM4 GATA2 ZNF462 HAPLN3 GBA3 CISD1 SLC5A1 POFUT1 TCF20 SNRPD3 SELM SERINC1 UBE2U AREL1 CDH12 NSG1 KCND2 ATP8B3 ERICH1 MIR2300B MIR2374 PSPC1 DCTD ZNF703 TMEM119 SLIT2 WDYHV1 CLTB LIN7C POU2F1 PMP22 GNA12 N6AMT2 ANKRD55 CAPN7 ALKBH3 NAA11 ETV1 STAB2 SERPINI2 CTXN3 MIR186 REG3A C21H14orf2 NPL LPHN2 ZBTB49 SCAP GRIFIN ZNF280A CHCHD4 DTNA CELF4 TMOD3 LOC531152 GSTT3 MIR708 GSTT1 GGT5 SI SKA2 RNF44 SEMA3C SCN9A NXPH2 BET1 FNBP1 OTOP1 CRISP1 RANGAP1 HAPLN1 PUM1 SLC2A11 DRD1 C22H3orf37 XRCC2*

REF	GEN Num	GEN
Hou	291	*PNMA1 CCDC116 ECHS1 PMPCA C17H5orf52 AMZ2 KIAA1279 RTN1 CENPQ SLC33A1 ADCK1 TMX4 C10H14orf38 B3GAT1 MNS1 C10H14ORF53 SLC25A26 ABL2 PSMA3 MIR218-1 SQSTM1 MIR2389 CCR9 NFATC1 TXNDC8 anemia TMEM45A MPP7 KIAA1715 MGAT4B AMN1 ZDHHC6 MGC151949 LPGAT1 RBPJ ALDH18A1 RNF4 C23H6ORF142 EIF1 CCAR1 TRAPPC9 MXD4 GLTSCR2 YWHAQ MGC152007 MTHFSD MAEL PRKG2 LMF1 MIR196B ABO ZNF331 LETM1 CCHCR1 MIRLET7C MRGPRF RNF145 THAP1 EEFSEC FGFR3 AGT LRRTM3 MIR125B-2 LZTFL1 C23H6orf64 CHMP7 TBC1D9B ZW10 MIR301B DUSP18 GUCY1B3 SLBP CDSN EIF4E2 CRABP1 RDX TMEM170A CDH18 VTI1A LOC407171 MS4A13 SLA HCRTR1 AKR7A2 SCRN1 INVS HOXA10 MEMO1 KIT PRKAR1A HSPA2 YDJC DPYSL4 PQLC2 IFNT OOEP GLI1 IL34 MGC133692 LOC510651 WIPI2 IFNAR1 MKLN1 EGFLAM EHD2 TCF19 SULT2A1 MCHR1 LRCH4 DEPDC6 TCTEX1D2 STRN3 C6H4orf22 PEF1 RNF168 KCNAB1 CALY SFN YPEL1 MTP18 ACSL5 TRIM9 GAL3ST1 MIR220D C25H7ORF50 TMEM129 SLC6A20 MIR551A PARVB MRPL17 MAN2C1 GAST COX19 UBE2L3 MCTP2 IL10RB GTSE1 SOAT1 GAL COCH FAM92A1 C7H5orf45 WDFY1 CHD1L GPM6A WHSC2 SNX29 EIF2AK3 TLR3 SNW1 IGJ STK39 MRPL44 ANGPTL1 ACER3 EFEMP1 MIR335 LOC512887 IGHMBP2 VAC14 TMEM45B MS4A1 COMMD4 SCFD2 TFAM CRX TBCC TG POMP MMD2 SERPINE2 FRMD8 PLCB1 PRPH2 MAPK1 LOXL2 ATPIF1 JKAMP TACC3 SCG2 C19H17orf64 SORBS1 TRIM49 CHST9 TUBGCP2 B, MIRLET7B SLC25A17 FADD SYK DDIT3 SLC37A2 WIPI1 MUT BRP44L membrane CLSTN3 SPDEF COPG2 COG5 PPIL2 CALCRL RNASEH2B MOCOS CHRND ARSG SLC25A33 CHEK2 TBX19 ADAP1 DDX21 HOXA9 SPON2 TMEM201 SEC14L2 TMX1 BCNT2 TMEM218 PRMT3 MRPS2 PDE4DIP MIR124A-2 HCCA2 GRM7 NUP54 PEX5 MRPL21 LOC616908 PSORS1C2 LOC615685 MARK4 CTU1 ACCN1 MEST ZNF32 SAPS3 MMD NIT2 PTBP2 C6orf106 RADIL CSRP2 SLC46A3 SEC14L4 PTGR2 C11H9ORF116 MGC139000 POU5F1 ALG10 MARS TECRL ELP2 ACOT4 AGXT LOC524749 HEPACAM2 TDO2 TXN SDF2L1 LOC751809 DNAL1 SFT2D1 NINJ1 COL11A1 C9H12orf49 TBCD IFNAR2 LOC783012 GPC6 transporter), MACROD2 ZBTB25 TCN2 CHRNG SESN2 MUM1 HBG MAB21L1 TTC38 SDCCAG3 LRRC3B MIR202 CAPZB SCN5A PPP1R7 MAEA TRMU MIR33B S1PR3 SPOCK1 SLC39A6*

REF	GEN Num	GEN
Bickhart	315	*MMS19 DBNL KRTCAP2 VPS33A CATHL4 DECR1 SLC7A6 UBE2G1 LILRA4 LOC617875 FBXO16 SERPINB5 PRP5 IMP4 PROP1 TP-1 ITLN1 TPD52L2 LOC514330 BNBD-4 PTCD3 PAG4 PLA2G2D3 FCGR2B PAPSS2 IFN-a SLC23A2 CNN1 IQCF2 LOC100124497 SERPINA3-7 SLC16A7 ZDHHC16 PCDHA13 SERPINB9 LOC509513 HIST1H1E TMEM66 LMBRD1 CYP21A2 BTRC YEATS2 RPA2 UGT2B10 TUBB TOR3A C10H14ORF1 PRP3 LOC507082 BCAM IMMP1L DHDDS GPI RNASE2 ZNF547 LOC100125946 DEFB8 MGC152344 MAPK10 LOC618633 MGC137405 PLD1 SULT1C2 BOLA-DQB TMEM11 PLA2R1 LOC100125916 IMMT IFT80 IPO13 MYH1 KRTAP10-2 BOLA-DQA5 SLC12A2 LOC510320 LITAF PRG3 JSP.1 PLA2G2D5 CYP4A22 TUBA1 IL8RA PAG12 AOC3 PCDHGB4 AP2M1 C13H20ORF12 LOC618367 LOC781494 ZYX WDR51A AKR1C4 COQ5 CATHL5 H2B RRAGA TKDP1 MGC134066 CD7 RGS7 CREB3 HRASLS3 LOC515336 MGC139448 RPS26 SFTPD DNAJA1 SECTM1 B2M CYB561D2 H2AFY2 FSHR CD97 ZCCHC10 ECSIT LOC780781 CAMP MGC134040 MED6 BXDC1 BOLA-NC1 THEM4 UGT1A6 PAG7 OR10H1 RPL6 CES KRT33B DHRS7B CYP4A11 KRT34 ATP6V1E1 RPS4Y1 KRT31 NUMBL LOC751563 LOC617104 OR12D2 ISCA1 PI3 RCC1 AUH KRT6A PRP9 PAG21 ATAD3A KRT81 MOGAT2 MMP1 RAN PAG19 EEF1A1 TRIM6 ZNF215 CYP21 BSP30A PAG1B GNLY GRAMD1C SLC35A4 LOC781146 GSR PAG5 GFM2 PCDHGA2 GLO1 P4HA2 TUBB2A SAO SERPINA3-5 CCR1 DNAJA2 CL-46 ACADSB GSTM1 LOC100124518 APBB3 PSMB7 MESDC1 LOC100125302 PRKG1 CALCB LOC510193 TUSC4 CD5L MGC152321 PACSIN1 RRP1 SAA3 SERPINB4 LEPROTL1 RPS3A MGC134093 LOC511106 SERPINB6 PMP22CD LOC786254 SERPINA3-8 DYNLL1 PDE5A MGC157408 PLA2G2D4 IFNW1 GZMB ABCF3 SP2 CCBL1 SLC3A2 MTHFD1L PCDHGA8 PAG2 ADH4 SUHW2 ALS2CR8 TMEM22 MC5R PAG15 BOLA PRP-VII RTP4 SF4 ALG8 PSKH1 KRTAP9-2 LOC517016 PRP6 ARL3 NOL5A KCTD19 LOC615103 PTPLB SYNJ2BP LOC527068 LOC512741 TMEM163 ATF6 PRP1 LOC508153 RNASE1 STK19 MGC152278 IFN-tau-c1 ILVBL CATHL1 DEFB CATHL6 TMEM115 MX1 LOC100125415 IFNB3 PAG18 TAP BSP30B BNBD10 LYZ1 LY6E LOC515697 FLOT1 KRTAP4-7 PDZK1 IGBP1 SLC6A9 ADAT1 PCTP KRT33A S100A7 IL8RB ADH6 RNF8 LCAT DEFB103B AKR1C3 HTR2A CTNNA2 LOC529196 IFNB1 PGAM2 CBLC PLA2G2D1 C21H15ORF26 KPNA6 SERPINA3 TSHR MGC148762 CA6 PSMB10 HAVCR2 ZNF133 TFF3 IQCC APOL3 FCRL1 AOX1 ST8SIA6 CCDC115 CES1 MGC157043 PITPNB PAG6 LOC751562 CYP4B1 VISA MGC126945 ISG12(A) MGC152202 GARS BCKDHB LOC780876 CCDC9 SAFB ACTR2 EXOC6 WDR63 CTSC MPST*
Bae	533	*GPAA1 TBC1D19 ERGIC1 VAMP2 PKNOX2 ACTR3 CXCR4 C15H11orf74 TOMM20 PRKRA RAP1A ROGDI C25H16orf71 FAM151B HMG20B MGC139698 FGR TOP3A STK16 BAK1 GABARAP FSCN2 NR4A2 RHOF PCYT2 SERPINF1 MED29 ZNF500 TMEM180 MGC137018 FARSB HNMT LOC506277 TMEM60 BMPR1A C26H10ORF6 YBX2 RTEL Magmas MXRA8 ATOH1 COL1A1 SLC8A1 TEKT1 PDCD4 HMOX2 POU1F1 CPLX3 DYSF PPP3R2 GLE1 SSBP4 PAFAH1B1 OSTF1 MYL12A LOC100137087 GYPC MYO1C YWHAE ATP1B1 MOSPD2 NPB EPB42 SLC6A15 LOC786620 MRPL12 ACBD5 CSF1 CCDC137 LLGL2 PREP C8G LIM2 CKB NFKB2 TMEM120B MGC157263 ABHD2 TMCO5A SLC6A6 LOC768237 ERLIN2 RAB3A SLCO3A1 MBTPS1 COX8A SFRS16 CDH15 SUPT5H FTMT GUCY2C MRPL28 F7 TMEM92 STMN3 SNX6 ZNHIT2 LSM4 LOC617776 CAPN1 PPAPDC3 DOHH LOC100139208 FERMT3 FAM82C SPTAN1 OPRK1 LRRC48 DBH DNAJA3 NDUFC1 CCRN4L SRD5A2L2 C4H7orf23 DULLARD DIDO1 OR4X1 TMEM8 SMTNL2 CLDN7 LOC516156 DUSP26 IL27RA LRRC25 PPP1R12A CC2D1A CHMP2B RNF20 ATP1A1 LLGL1 IP6K3 ACVR2A F11 PRPF39 SV2B PITPNA TIPIN MGC140224 C5H12orf44 RNF182 RPE65 MGC152531 HIST3H2A ARFRP1 CUEDC2 SLC17A9 TNFRSF4 CUGBP1 TTC25 RPUSD1 FAU CSK JUND ZFP36 COL6A2 C6orf173 CCDC77 CCDC32 LOC100126230 MPV17L2 NECAB2 SEPT12 JAKMIP1 GATAD1 HPS1 C22H3ORF37 RNF112 MGC128405 DES CSMD3 XRCC4*

REF	GEN Num	GEN
		C14H8ORF55 XRCC3 AURKAIP1 ACAP3 CYC1 TYRP1 FHIT HMP19 TTC5 ROPN1 NKIRAS1 PDK2 ACADVL FAM82A C3H1orf183 LOC777786 ASPSCR1 LRRC50 THOC7 DUSP1 PLRG1 SPG7 CHST14 TMEM69 STK3 CHN2 RSAD1 GAS8 NINJ2 ZNF644 ACO1 PYCR1 MAPK7 NUDT22 NEUROD1 TFF2 GABRA2 RHOV LOC524650 TNS1 MAT2B LENG8 DSC1 TTC15 C26H10ORF2 XYLT2 PDZD7 PPP1R14D C9ORF59 GCHFR SMCR7 ATG9A LOC509263 NFIC LHPP DECR2 GRB14 SOCS3 GGT6 SERPINF2 GABRA6 TRAPPC6A NLRP9 EPN1 UBAP1 MRPS30 HSDL1 CHODL LOC512391 GDPD1 TSPAN14 PSMC1 SHMT1 SFXN3 PDE6G ZC3H14 CHAD ASGR1 CORO6 NME4 LPHN3 DLL4 GHDC CHAC1 BANF1 ARHGEF5 ISYNA1 NMRAL1 ECT2 LHX1 EEF1A2 MRPL43 ITGA4 SPEG CD34 INPP5K WDR81 KIAA1737 RXRG LOC617922 PSMD9 GLUD1 TUBA4A THAP4 HPD SLC39A4 CITED2 MC4R ODF2 MFAP4 SIRT7 PDIA2 CLTLB PET112L ANAPC11 TMEM65 MAF1 MYL12B FAM19A5 KRT80 FBXL15 MGC137027 CYP1A2 B3GALNT1 IFI6 IRX4 EPN3 RNF126 XPNPEP1 GGNBP1 LOC513822 NFIB SDF4 LOC786832 OSGIN1 NDST3 CXCL2 THOC4 BOK ZC3H8 THBS4 MRPL46 LHX3 LSM5 TFAP4 SAMD14 SGCA MGC142811 NR4A1 DNPEP CREB1 FANCM PLCD1 PTGR1 SLC2A6 C19H17ORF48 CNTN1 PRPF8 PPP2R1A ITFG3 FKBP3 RPA1 IFI30 C28H10ORF116 DBNDD1 ST8SIA3 FEZ1 KLF5 ZNF784 EIF3J VPS18 GADD45GIP1 PPIF TEKT4 GEMIN7 C2H2orf24 LRRC68 WNT11 PIK3R2 SPINT1 PLBD1 ADK MGC127138 STIP1 USP33 ZNF483 ARF5 SLC5A10 MRPL49 SLC43A2 TNIP2 LOC617991 PMPCB STEAP2 SNCG NUDT16L1 LSM8 SLC2A4 C13H20orf149 LRRC59 SPATA2L NOS3 SEMA4G CHST12 KBTBD4 COPG CHMP1A KCNIP4 FGFR1OP ZFP2 PTPRN RAB5C CYLD SHOC2 MRPL27 VASN ACSF3 UNC50 MGAT4A CCNH BRE PRPH KRT7 RPL13 DNAJB2 DPEP1 CBARA1 NAT11 THUMPD1 TMUB1 LYNX1 PCCA SBDS IMPG1 TTYH1 KANK2 C1H3orf26 MAFG OTUB1 PRPSAP2 KLC1 EXOSC4 MSLN FIZ1 BAG5 EME1 C1H21ORF91 COBL MAP3K7 RAD51 EID2 GTF2I CDK5 FBXL3 TMEM30A LIPC C13H20orf195 RASA1 ARHGDIA LOC511424 GSTO1 CORO7 STRA13 DNAJC17 DAP SHARPIN FSTL3 PCP4 PIK3R1 PTPMT1 RIMS2 TAF1C MAF ABCB6 PRKACA GRAP SOCS5 TIMM50 ARHGDIG CNP HRNBP3 LMCD1 CD164L2 MAP2K1 FASTK TNFRSF6B CHRM3 USP50 SYVN1 P4HB C13H20ORF11 POMC DIS3L CERCAM RAB28 KLF13 DERL1 JMJD7-PLA2G4B ZZZ3 C16orf5 GLTPD1 SLC25A31 EPS15L1 PYCR2 MGC148329 CASP4 BRD3 PDIA4 NDUFS3 ZFYVE19 IGFBPL1 HEATR7A RESP18 MTCH2 MGRN1 FBXL4 PGS1 NXPH1 ACTG1 MTMR3 ZFAND2B UGT8 ZGPAT HINT1 TMEM168 ITPRIP RPS16 TMEM30C NCAPG LHX9 ATPAF2 HGS POLA2 ABTB1 DDX20 ABCB8 PRMT7 HSPB9 HGF FKBP2 TRPC3 PGPEP1 NKIRAS2 ARFGAP1 MLYCD ANXA11 VGLL4 SPC24 WASF2 VEGFB CDK10 MGC165793 DRG2 UBE2J2 PHF23 TMEM63A C14orf153 ANKZF1 ZNF622 ITPR3 LZTS2 FZR1

Additional file 3.6 Tables demonstrating the over-representation of biological processes (BP) and cellular components (CC) by those genes within and or 10Mb downstream of CNVRs identified in Nguni cattle.

CC	BosT (19 799)	NG	NG EXP	NG REP	NG ENR	Pval
Dendrite	166	13	3.38	+	3.85	4.68E-02
Neuron part	536	25	10.91	+	2.29	1.33E-01
Membrane region	772	32	15.71	+	2.04	1.40E-01
Cell	12 228	284	248.90	+	1.14	1.52E-01
Cell part	12 228	284	248.90	+	1.14	1.52E-01
Plasma membrane	3 153	91	64.18	+	1.42	2.94E-01
Cell periphery	3 239	93	65.93	+	1.41	2.96E-01
Synapse part	230	14	4.68	+	2.99	3.26E-01
Somatodendritic compartment	261	15	5.31	+	2.82	3.67E-01
Neuron projection	414	20	8.43	+	2.37	3.97E-01
Cellular_component	14 894	331	303.16	+	1.09	5.48E-01
Membrane-bounded organelle	8 750	211	178.10	+	1.18	5.86E-01
Cell projection	973	35	19.80	+	1.77	9.10E-01
BP	**BosT (19 799)**	**NG**	**NG EXP**	**NG REP**	**NG ENR**	**Pval**
Nervous system development	1 335	54	27.17	+	1.99	8.65E-03
Generation of neurons	901	36	18.34	+	1.96	7.07E-01
Regulation of localization	1 447	51	29.45	+	1.73	6.46E-01
Anatomical structure morphogenesis	1 637	56	33.32	+	1.68	6.30E-01
Somite development	65	8	1.32	+	> 5	4.28E-01
Neuron development	508	25	10.34	+	2.42	3.72E-01
Segmentation	81	9	1.65	+	> 5	3.32E-01
Neurogenesis	970	39	19.74	+	1.98	3.20E-01
Single-organism behavior	279	21	5.68	+	3.70	2.89E-03
Multicellular organismal development	3 085	94	62.79	+	1.50	1.96E-01
Neuron differentiation	641	30	13.05	+	2.30	1.74E-01
Somitogenesis	57	8	1.16	+	> 5	1.72E-01
Developmental process	3 612	107	73.52	+	1.46	1.51E-01
Behavior	390	22	7.94	+	2.77	1.45E-01
Single-organism developmental process	3 596	107	73.20	+	1.46	1.25E-01

Additional file 3.7 HPB CNVR overlap regions, the number of animals presenting the CNVR (NUM) and the genes covered (GEN) in Nguni cattle.

CNVR	CNVR-HPB No.	NUM	GEN
chr1:31923335-32340451	CNVR-HPB 1	3	
chr1:59409838-59463782	CNVR-HPB 2	1	*QTRTD1*
chr1:102538612-103219606	CNVR-HPB 3	1	*BCHE*
chr1:104798012-105264358	CNVR-HPB 4	17	*SI*
chr1:120572284-120752725	CNVR-HPB 5	1	*TM4SF1,TM4SF4*
chr2:4521411-4565625	CNVR-HPB 6	1	*AMMECR1L*
chr2:22283989-22451660	CNVR-HPB 7	1	*ATF2,ATP5G3*
chr2:56456865-56694744	CNVR-HPB 8	4	
chr2:78433731-78556325	CNVR-HPB 9	2	
chr3:120501439-121275236	CNVR-HPB 10	6	*SH3BP4, ARL4C*
chr4:55566787-55656122	CNVR-HPB 11	3	*FOXP2*
chr4:108834886-108974924	CNVR-HPB 12	2	*LOC780933*
chr5:75722589-75794378	CNVR-HPB 13	2	*BTBD11,PWP1*
chr6:38845992-38939012	CNVR-HPB 14	1	*LCORL*
chr6:52628477-52725432	CNVR-HPB 15	1	*PCDH7*
chr6:108998175-109951981	CNVR-HPB 16	12	*LYAR,NSG1,OTOP1,STX18,TMEM128,WDR1,ZBTB49*
chr7:43097791-43311132	CNVR-HPB 17	1	*TCF3*
chr7:43567130-43808593	CNVR-HPB 18	1	*ATP8B3,REXO1*
chr9:3651455-4439872	CNVR-HPB 19	12	
chr9:91439245-91469581	CNVR-HPB 20	1	*SPAG17*
chr10:59812472-59948769	CNVR-HPB 21	2	*LEO1,TMOD3*
chr10:71022679-71082204	CNVR-HPB 22	1	*OTX2*
chr12:39801280-39925909	CNVR-HPB 23	2	
chr12:45358430-45409287	CNVR-HPB 24	1	*KLHL1*
chr12:49733358-49781954	CNVR-HPB 25	1	*UCHL3*
chr12:57661486-57803744	CNVR-HPB 26	2	*MIR1256*
chr12:61573515-61686477	CNVR-HPB 27	3	*HTATSF1*
chr16:54846459-54886491	CNVR-HPB 28	1	*RFWD2*
chr19:9172300-9354310	CNVR-HPB 29	1	*MIR454,SKA2,TRIM37*
chr20:51449833-51685293	CNVR-HPB 30	3	*CDH10*
chr21:3061993-3089789	CNVR-HPB 31	1	*GABRA5*
chr22:56526462-56603472	CNVR-HPB 32	1	*ATG7*
chr27:9096031-9512004	CNVR-HPB 33	4	*AGA,NEIL3*
chr27:15141168-15291371	CNVR-HPB 34	6	*DCTD*

Addendum C

Additional file 4.1 Eigen values (EIV) of first 15 principle components (PC) generated from a genetic distance matrix of 197 animals (AN) from 7 South African cattle breeds (BRD).

		PC	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
AN	**BRD**	**EIV**	**356.82**	**20.785**	**7.890**	**5.744**	**5.224**	**4.549**	**4.168**	**2.690**	**2.501**	**2.351**	**2.134**	**2.026**	**1.790**	**1.640**	**1.587**
197	AFR		0.96	-0.05	0.01	0.03	-0.01	-0.02	-0.06	0.04	0.02	-0.06	-0.14	0.13	-0.10	0.10	0.04
196	AFR		0.83	-0.07	-0.21	0.05	-0.07	0.20	-0.04	0.09	0.34	-0.25	-0.19	0.15	-0.26	0.24	0.06
195	AFR		1.03	0.02	0.01	0.00	0.00	0.01	0.00	0.01	0.00	0.02	0.01	0.00	0.00	-0.01	0.01
194	AFR		1.02	-0.02	-0.02	0.03	-0.03	0.00	-0.02	-0.03	0.01	-0.05	-0.17	0.09	-0.09	0.10	0.06
193	AFR		0.90	-0.06	-0.13	0.01	-0.01	-0.31	0.10	-0.08	0.07	-0.01	0.02	-0.25	-0.10	0.09	0.02
192	AFR		0.85	-0.12	-0.35	0.12	-0.05	-0.19	0.04	0.05	0.19	-0.15	0.12	-0.10	-0.14	0.07	0.06
191	AFR		0.78	-0.27	-0.21	0.07	0.04	0.09	0.16	0.06	0.04	0.21	-0.02	0.07	-0.17	0.04	0.23
190	AFR		0.82	-0.20	-0.03	0.07	0.05	0.16	0.07	-0.09	0.11	0.46	0.19	-0.08	-0.09	-0.18	0.03
189	AFR		0.96	-0.08	0.00	-0.01	-0.01	-0.03	0.00	0.00	0.12	0.02	0.20	0.03	0.11	-0.09	-0.03
188	AFR		0.91	-0.10	-0.05	-0.15	-0.02	0.20	0.03	-0.15	-0.01	0.11	0.10	-0.18	-0.06	-0.11	0.07
187	AFR		0.77	-0.35	-0.02	0.02	0.04	0.00	0.07	-0.13	0.10	-0.11	0.15	-0.19	-0.15	0.13	-0.12
186	AFR		1.00	-0.01	0.01	0.01	-0.01	0.00	-0.05	-0.01	0.10	-0.13	0.06	0.07	-0.06	0.06	-0.06
185	AFR		0.95	-0.13	0.00	-0.01	0.00	-0.01	0.07	-0.06	0.00	0.10	0.09	-0.07	-0.02	-0.03	0.01
184	AFR		0.83	-0.32	-0.03	0.03	0.00	-0.02	0.08	-0.16	0.12	0.07	-0.03	0.08	-0.25	0.14	0.06
183	AFR		1.01	-0.04	-0.01	0.02	-0.01	-0.04	-0.02	-0.06	0.09	-0.01	0.10	-0.07	-0.07	0.03	0.14
182	AFR		1.05	0.05	0.01	0.00	-0.03	-0.01	-0.06	0.06	-0.07	-0.03	0.07	0.08	-0.03	-0.06	0.06
157	AFR		1.10	0.23	-0.03	0.01	-0.07	0.07	-0.22	0.03	0.04	0.05	-0.05	-0.01	-0.09	-0.17	-0.14
156	AFR		1.05	0.05	0.00	0.01	-0.01	0.01	-0.04	-0.01	0.03	-0.04	-0.03	-0.03	-0.08	0.04	0.01
155	AFR		1.02	0.10	-0.20	0.07	-0.02	0.18	0.02	0.09	0.08	0.01	0.00	-0.03	-0.06	0.00	0.04
154	AFR		0.70	-0.36	-0.25	0.31	-0.01	0.04	0.16	-0.03	-0.13	-0.05	-0.12	-0.17	0.11	-0.11	0.18
153	AFR		1.01	-0.03	0.00	0.01	-0.02	0.01	0.01	0.02	-0.02	0.02	-0.14	0.13	-0.01	0.05	0.05
152	AFR		0.97	-0.08	0.02	0.00	0.00	-0.01	0.05	0.02	-0.04	0.00	-0.08	0.11	0.02	0.04	0.05
151	AFR		1.03	0.03	0.02	0.00	0.00	0.01	0.00	0.02	-0.01	0.00	0.00	0.01	-0.01	0.00	0.01
150	AFR		0.96	-0.09	0.00	-0.02	-0.01	0.00	0.03	-0.07	0.15	0.11	0.09	-0.09	0.01	-0.02	0.02
149	AFR		0.99	0.02	0.06	-0.03	0.03	0.02	0.10	-0.01	0.05	0.03	-0.03	-0.01	0.04	0.04	-0.07
148	AFR		1.05	0.05	0.01	0.00	-0.02	0.02	-0.04	0.01	0.02	0.01	-0.01	-0.01	-0.01	0.00	-0.01
147	AFR		1.03	0.02	0.02	0.00	0.00	0.01	0.00	0.02	-0.01	0.00	0.00	0.01	0.00	0.00	0.00

		PC	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
AN	**BRD**	**EIV**	**356.82**	**20.785**	**7.890**	**5.744**	**5.224**	**4.549**	**4.168**	**2.690**	**2.501**	**2.351**	**2.134**	**2.026**	**1.790**	**1.640**	**1.587**
146	AFR		1.04	0.05	0.01	0.00	-0.01	0.01	-0.03	0.01	0.00	0.01	-0.01	0.00	0.00	-0.01	-0.01
145	AFR		0.95	-0.14	0.00	0.01	0.00	-0.04	0.06	-0.03	0.05	0.03	-0.03	0.06	0.01	0.02	0.12
144	AFR		1.00	-0.01	0.00	0.03	-0.01	0.00	-0.06	-0.04	0.12	-0.17	0.04	0.04	-0.15	0.11	-0.05
143	AFR		1.07	0.13	0.03	-0.01	0.00	-0.01	0.04	-0.01	0.03	0.01	-0.02	-0.01	0.01	-0.01	-0.05
140	ANG		0.95	-0.13	-0.01	0.01	0.01	-0.06	0.05	-0.03	-0.01	-0.03	0.08	-0.09	0.03	-0.03	0.05
139	ANG		0.90	-0.22	0.00	0.02	0.02	-0.06	0.09	-0.01	-0.03	-0.03	-0.03	0.03	0.04	0.02	0.09
131	ANG		1.03	0.02	0.01	0.00	-0.01	0.01	0.00	0.02	-0.01	0.01	0.00	0.00	0.01	-0.01	0.01
103	ANG		1.07	0.08	0.15	0.10	0.00	0.06	-0.19	-0.13	0.02	-0.07	-0.12	-0.11	-0.13	-0.03	-0.01
101	ANG		1.04	0.05	0.01	0.00	-0.01	0.01	-0.03	0.01	0.00	0.01	-0.01	0.00	0.00	-0.01	-0.01
100	ANG		-4.78	-0.45	0.45	-1.42	-0.40	-0.94	-0.22	0.19	-0.04	-0.06	-0.11	-0.09	-0.07	0.12	-0.06
98	ANG		-3.20	-1.01	0.32	0.09	-0.01	0.55	0.43	0.99	0.29	0.06	0.07	-0.18	-0.09	0.16	0.02
97	ANG		-5.32	-0.10	-0.40	-0.98	0.17	1.05	0.17	-0.39	-0.09	-0.18	0.13	0.16	-0.02	-0.03	0.13
96	ANG		-6.49	0.49	-1.24	0.35	1.24	-0.39	-0.27	0.14	0.09	0.09	-0.16	-0.05	0.11	-0.06	0.02
94	ANG		0.95	-0.05	0.13	-0.07	-0.18	0.12	0.01	0.07	-0.16	-0.09	0.02	-0.08	-0.09	0.04	-0.02
93	ANG		-7.28	0.85	-0.47	0.87	-1.24	-0.07	0.23	-0.16	-0.03	0.10	0.08	0.05	-0.06	0.00	-0.08
92	ANG		1.02	0.10	0.02	0.00	-0.01	0.05	-0.14	0.05	0.01	0.05	-0.03	-0.03	-0.05	0.00	-0.13
90	ANG		1.07	0.11	-0.01	0.01	-0.03	0.03	-0.10	0.00	0.02	0.02	-0.04	-0.02	-0.01	-0.04	-0.05
89	ANG		1.11	0.33	0.04	0.01	-0.02	0.02	-0.03	0.04	0.06	-0.02	-0.02	0.02	-0.14	-0.05	0.01
88	ANG		1.03	0.02	0.02	0.00	0.00	0.01	0.00	0.02	-0.01	0.00	0.00	0.01	0.00	0.00	0.01
86	ANG		1.03	0.06	0.04	-0.01	0.03	-0.02	0.08	0.00	0.00	-0.01	0.00	0.00	0.03	0.01	0.00
85	ANG		1.03	0.02	0.02	0.00	0.00	0.01	0.00	0.02	-0.01	0.00	0.00	0.01	0.00	0.00	0.01
84	ANG		1.04	0.05	0.01	0.00	-0.02	0.01	-0.03	0.01	0.00	0.01	-0.01	0.00	0.00	-0.01	-0.01
63	ANG		1.06	0.09	0.00	0.01	-0.03	0.03	-0.09	0.02	-0.03	0.04	0.01	-0.02	0.01	0.04	-0.01
61	ANG		1.05	0.09	0.04	-0.01	0.02	-0.01	0.06	0.00	0.01	0.00	-0.01	-0.01	0.03	0.00	-0.02
59	ANG		1.03	0.06	0.04	-0.01	0.03	-0.02	0.08	0.00	0.00	-0.01	0.00	0.00	0.03	0.01	0.00
56	ANG		0.91	-0.19	-0.01	0.00	0.01	-0.02	0.06	-0.09	0.08	-0.01	0.17	-0.03	-0.05	0.02	-0.03
54	ANG		1.03	0.02	0.02	0.00	0.00	0.01	0.00	0.02	-0.01	0.00	0.00	0.01	0.00	0.00	0.01
53	ANG		1.03	0.02	0.02	0.00	0.00	0.01	0.00	0.02	-0.01	0.00	0.00	0.01	0.00	0.00	0.01
49	ANG		1.00	-0.02	0.16	-0.05	0.05	0.25	0.00	0.04	0.01	-0.05	0.04	0.00	-0.01	-0.02	0.06
142	BON		1.05	0.09	0.04	-0.01	0.02	-0.01	0.06	0.00	0.01	0.00	-0.01	-0.01	0.03	0.00	-0.02
141	BON		0.19	-0.81	-0.09	0.13	0.03	0.07	0.16	-0.18	0.02	-0.37	-0.03	-0.01	0.19	0.02	-0.01
138	BON		1.08	0.19	0.06	0.00	0.02	-0.01	0.08	0.00	0.02	-0.01	-0.04	0.00	-0.02	0.00	-0.01
136	BON		1.02	0.15	-0.15	-0.14	-0.04	-0.07	0.04	-0.11	0.00	-0.01	-0.04	-0.01	0.06	0.00	-0.09
135	BON		1.09	0.16	-0.01	0.02	-0.06	0.05	-0.18	0.00	0.02	0.05	0.02	-0.03	0.04	0.07	0.03

		PC	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
AN	**BRD**	**EIV**	**356.82**	**20.785**	**7.890**	**5.744**	**5.224**	**4.549**	**4.168**	**2.690**	**2.501**	**2.351**	**2.134**	**2.026**	**1.790**	**1.640**	**1.587**
134	BON		1.11	0.33	0.05	0.01	-0.02	0.02	-0.04	0.02	0.05	0.03	-0.04	-0.03	-0.08	-0.01	-0.07
133	BON		1.03	0.06	0.04	-0.01	0.03	-0.02	0.08	0.00	0.00	-0.01	0.00	0.00	0.03	0.01	0.00
130	BON		1.05	0.09	0.04	0.00	0.02	0.00	0.05	0.00	0.00	-0.01	-0.01	0.00	0.01	0.00	0.00
129	BON		1.03	0.02	0.02	0.00	0.00	0.01	0.00	0.02	-0.01	0.00	0.00	0.01	0.00	0.00	0.01
128	BON		1.03	0.02	0.02	0.00	0.00	0.01	0.00	0.02	-0.01	0.00	0.00	0.01	0.00	0.00	0.01
127	BON		1.11	0.27	-0.03	0.02	-0.09	0.08	-0.29	-0.01	0.07	0.10	-0.05	-0.08	0.00	0.05	-0.18
126	BON		1.05	0.09	0.04	-0.01	0.02	-0.01	0.06	0.00	0.01	0.00	-0.01	-0.01	0.03	0.00	-0.02
125	BON		0.99	0.03	0.09	0.06	-0.14	0.17	-0.01	-0.20	-0.07	-0.03	0.06	0.06	-0.01	-0.07	0.02
124	BON		1.03	0.06	0.04	-0.01	0.03	-0.02	0.08	0.00	0.00	-0.01	0.00	0.00	0.03	0.01	0.00
123	BON		1.03	0.02	0.02	0.00	0.00	0.01	0.00	0.02	-0.01	0.00	0.00	0.01	0.00	0.00	0.01
122	BON		1.03	0.02	0.02	0.00	0.00	0.01	0.00	0.02	-0.01	0.00	0.00	0.01	0.00	0.00	0.01
82	BON		1.06	0.12	0.03	-0.01	0.01	-0.01	0.04	-0.01	0.02	0.00	-0.02	-0.02	0.03	-0.01	-0.04
81	BON		0.82	2.63	0.84	-0.06	0.48	-0.32	1.15	-0.06	0.00	-0.12	-0.02	-0.01	0.04	0.06	0.08
80	BON		1.04	0.05	0.01	0.00	-0.01	0.01	-0.03	0.01	0.00	0.01	-0.01	0.00	0.00	-0.01	-0.01
79	BON		1.04	0.05	0.01	0.00	-0.02	0.01	-0.03	0.01	0.00	0.01	-0.01	0.00	0.00	-0.01	-0.01
78	BON		1.05	0.09	0.04	-0.01	0.02	-0.01	0.06	0.00	0.01	0.00	-0.01	-0.01	0.03	0.00	-0.02
77	BON		1.03	0.02	0.02	0.00	0.00	0.01	0.00	0.02	-0.01	0.00	0.00	0.01	0.00	0.00	0.01
76	BON		1.05	0.05	0.01	0.01	-0.02	-0.01	-0.04	0.03	0.00	-0.01	0.01	0.03	0.02	-0.03	0.00
75	BON		1.03	0.02	0.02	0.00	0.00	0.01	0.00	0.02	-0.01	0.00	0.00	0.01	0.00	0.00	0.01
74	BON		1.04	0.05	0.01	0.00	-0.01	0.01	-0.03	0.01	0.00	0.01	-0.01	0.00	0.00	-0.01	-0.01
73	BON		1.09	0.13	-0.02	0.01	-0.05	0.01	-0.12	0.01	0.03	-0.03	-0.01	0.01	0.07	-0.09	-0.09
72	BON		0.99	-0.04	0.00	0.01	0.02	0.00	0.01	0.05	-0.13	-0.06	0.02	-0.04	-0.01	0.02	0.00
71	BON		1.04	0.05	0.01	0.00	-0.02	0.01	-0.03	0.01	0.00	0.01	-0.01	0.00	0.00	-0.01	-0.01
70	BON		0.91	-0.18	-0.01	-0.04	0.01	-0.03	0.08	-0.06	0.02	0.19	0.17	-0.04	0.00	-0.07	-0.01
69	BON		1.05	0.05	0.00	0.00	-0.03	-0.01	-0.06	0.06	-0.07	-0.03	0.07	0.08	-0.01	-0.06	0.07
68	BON		1.03	0.18	-0.16	-0.14	-0.06	-0.07	0.02	-0.12	0.01	0.00	-0.03	-0.03	0.06	-0.01	-0.13
67	BON		1.06	0.12	0.03	-0.01	0.01	-0.01	0.04	-0.01	0.02	0.00	-0.02	-0.02	0.03	-0.01	-0.04
66	BON		1.04	0.05	0.01	0.00	-0.02	0.01	-0.04	0.02	0.00	0.01	-0.01	0.00	-0.01	-0.01	-0.01
65	BON		1.03	0.02	0.01	0.00	0.00	0.01	0.00	0.01	0.00	0.02	0.01	0.00	0.00	-0.01	0.01
64	BON		1.03	0.02	0.02	0.00	0.00	0.01	0.00	0.02	-0.01	0.00	0.00	0.01	0.00	0.00	0.01
181	DRK		0.91	-0.14	0.03	-0.05	0.05	-0.06	0.18	-0.07	0.04	0.17	0.15	-0.04	0.04	-0.06	-0.03
180	DRK		1.03	0.02	0.01	0.00	0.00	0.01	0.00	0.01	0.00	0.02	0.01	0.00	0.00	-0.01	0.01
179	DRK		1.03	0.02	0.02	0.00	0.00	0.01	0.00	0.02	-0.01	0.00	0.00	0.01	0.00	0.00	0.01
178	DRK		1.05	0.09	0.04	-0.01	0.02	-0.01	0.06	0.00	0.01	0.00	-0.01	-0.01	0.03	0.00	-0.02

		PC	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
AN	**BRD**	**EIV**	**356.82**	**20.785**	**7.890**	**5.744**	**5.224**	**4.549**	**4.168**	**2.690**	**2.501**	**2.351**	**2.134**	**2.026**	**1.790**	**1.640**	**1.587**
177	DRK		1.03	0.02	0.01	0.00	0.00	0.00	0.00	0.01	-0.01	0.00	0.01	0.00	0.00	0.00	0.00
176	DRK		0.95	-0.14	0.00	0.01	0.00	-0.04	0.06	-0.03	0.05	0.03	-0.03	0.06	0.01	0.02	0.12
175	DRK		0.90	-0.08	0.06	-0.04	0.03	0.01	0.11	0.05	0.05	0.09	-0.14	0.25	0.10	0.03	-0.07
174	DRK		0.82	-0.08	-0.22	0.27	0.20	0.07	-0.22	-0.18	-0.09	-0.49	-0.13	-0.23	-0.25	0.07	-0.02
173	DRK		1.00	-0.03	0.00	0.02	0.00	-0.01	-0.01	0.02	-0.09	-0.05	-0.03	-0.03	0.02	0.00	-0.05
172	DRK		0.99	-0.06	0.01	0.00	0.00	-0.03	0.03	-0.03	0.07	0.03	0.08	-0.04	0.01	-0.03	0.08
171	DRK		0.95	-0.10	0.02	-0.01	-0.01	0.04	0.04	0.01	0.01	0.04	-0.15	0.11	0.02	0.09	-0.01
170	DRK		0.99	-0.02	0.03	-0.02	0.00	0.05	0.01	0.01	0.04	0.04	-0.02	0.00	0.02	0.04	-0.06
169	DRK		0.83	-0.28	-0.01	0.00	-0.01	-0.03	0.23	-0.04	-0.11	-0.01	-0.09	-0.09	0.11	-0.05	0.18
168	DRK		0.97	0.07	-0.01	-0.07	-0.14	0.12	-0.15	0.11	0.19	-0.05	-0.20	-0.13	0.19	-0.18	0.04
167	DRK		0.99	-0.06	0.00	0.00	0.00	-0.04	0.03	-0.03	0.08	0.05	0.09	-0.05	0.00	-0.04	0.08
166	DRK		0.93	-0.02	-0.19	-0.14	-0.14	-0.05	-0.06	-0.11	0.13	0.00	0.04	-0.15	0.24	-0.06	0.18
165	DRK		1.04	0.05	0.01	0.00	-0.01	0.01	-0.03	0.01	0.00	0.01	-0.01	0.00	0.00	-0.01	-0.01
164	DRK		0.98	-0.10	-0.03	0.00	-0.02	-0.01	0.04	-0.10	0.01	0.15	0.12	-0.13	-0.02	-0.06	-0.01
163	DRK		0.95	-0.07	0.03	0.00	0.01	-0.04	-0.03	0.01	0.10	0.03	0.08	0.00	-0.02	-0.02	0.06
162	DRK		1.04	0.05	0.01	0.00	-0.01	0.01	-0.03	0.01	0.00	0.01	-0.01	0.00	0.00	-0.01	-0.01
161	DRK		1.07	0.11	-0.01	0.01	-0.04	0.03	-0.09	0.01	0.02	0.02	-0.03	-0.01	-0.01	-0.03	-0.04
160	DRK		0.71	-0.42	0.13	0.03	0.00	0.10	0.05	-0.20	0.10	0.22	-0.25	-0.04	0.03	-0.11	0.15
159	DRK		1.06	0.08	0.00	0.01	-0.02	0.05	-0.06	-0.01	-0.01	0.01	-0.07	-0.04	-0.02	-0.04	0.00
158	DRK		0.95	-0.12	0.00	0.01	0.01	-0.04	0.01	-0.06	0.16	-0.09	0.15	0.01	-0.04	0.03	0.04
47	HOL		1.02	-0.01	-0.01	0.02	-0.03	0.01	-0.02	0.01	-0.03	0.02	-0.19	0.16	0.00	0.06	0.02
46	HOL		1.03	0.05	-0.01	0.03	-0.03	0.02	-0.12	-0.03	0.13	-0.14	0.03	0.08	-0.09	0.09	-0.15
45	HOL		0.91	-0.19	-0.02	0.03	0.03	-0.05	0.02	-0.05	0.07	-0.19	0.15	-0.07	-0.14	0.09	0.06
44	HOL		0.95	-0.04	-0.04	0.29	0.00	-0.09	-0.15	-0.25	0.25	-0.13	0.06	-0.12	-0.18	0.00	0.03
43	HOL		1.03	0.06	0.02	0.02	-0.01	-0.01	0.05	0.00	-0.03	0.00	-0.23	0.18	0.00	0.08	0.03
42	HOL		0.97	0.03	0.05	-0.22	-0.06	-0.16	0.02	-0.03	0.15	0.01	0.03	-0.20	0.03	0.00	-0.03
41	HOL		0.93	-0.18	-0.04	0.04	-0.02	0.00	0.07	-0.03	-0.24	0.03	-0.18	0.08	0.02	0.06	-0.09
40	HOL		0.93	-0.19	-0.04	0.05	-0.01	-0.08	0.07	-0.05	-0.02	-0.03	-0.06	0.00	0.03	0.02	0.11
39	HOL		0.48	-0.69	0.14	-0.09	0.04	-0.19	0.12	-0.10	-0.15	0.29	-0.06	0.08	-0.07	-0.05	0.14
38	HOL		1.06	0.08	0.00	0.00	-0.03	0.02	-0.06	0.01	0.01	0.02	-0.02	-0.01	-0.01	-0.02	-0.02
36	HOL		0.96	-0.04	0.02	0.00	-0.01	0.04	-0.03	0.07	-0.03	0.03	-0.13	0.12	-0.01	0.08	-0.03
34	HOL		0.81	-0.27	-0.03	0.02	0.02	-0.05	0.02	0.00	-0.02	-0.26	0.18	-0.19	-0.06	0.05	0.00
33	HOL		1.05	0.05	0.01	0.00	-0.02	0.01	-0.03	0.01	0.00	0.01	-0.01	0.00	0.01	-0.01	-0.01
31	HOL		0.96	0.08	-0.27	-0.32	-0.33	0.01	0.02	-0.29	-0.13	0.11	0.02	0.02	-0.05	0.02	-0.24

		PC	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
AN	**BRD**	**EIV**	**356.82**	**20.785**	**7.890**	**5.744**	**5.224**	**4.549**	**4.168**	**2.690**	**2.501**	**2.351**	**2.134**	**2.026**	**1.790**	**1.640**	**1.587**
30	HOL		1.02	0.09	0.04	-0.02	0.01	0.00	0.03	0.05	0.07	-0.05	0.05	0.10	0.17	-0.03	-0.18
29	HOL		1.05	0.05	0.01	0.00	-0.02	0.01	-0.03	0.02	0.00	0.01	-0.01	0.00	0.00	-0.01	-0.01
28	HOL		1.02	0.07	0.00	-0.01	-0.04	0.04	-0.10	0.07	0.07	-0.04	0.05	0.13	0.15	-0.05	-0.19
27	HOL		1.01	0.02	0.08	-0.15	0.20	0.06	-0.14	0.11	0.03	0.00	-0.02	-0.02	-0.03	0.03	-0.04
26	HOL		1.06	0.12	0.03	-0.01	0.01	-0.01	0.04	-0.01	0.02	0.00	-0.02	-0.02	0.03	-0.01	-0.04
25	HOL		1.04	0.05	0.01	0.00	-0.02	0.01	-0.03	0.02	0.00	0.01	-0.01	0.00	0.00	-0.01	-0.01
24	HOL		1.01	-0.03	0.00	0.01	-0.02	0.01	0.00	0.05	-0.07	0.01	-0.11	0.14	-0.04	-0.01	0.04
23	HOL		0.98	-0.04	0.09	-0.14	0.21	0.04	-0.07	0.11	0.01	-0.01	0.00	-0.01	-0.02	0.04	-0.01
22	HOL		1.03	0.02	0.02	0.00	-0.01	0.01	0.00	0.02	-0.01	0.00	0.00	0.01	0.01	0.00	0.01
21	HOL		1.05	0.09	0.04	-0.01	0.02	-0.01	0.06	0.00	0.01	0.00	-0.01	-0.01	0.03	0.00	-0.02
20	HOL		1.05	0.05	0.00	0.00	-0.02	0.02	-0.04	0.04	-0.03	0.01	0.01	0.02	-0.04	-0.06	-0.01
19	HOL		1.03	0.02	0.02	0.00	-0.01	0.01	0.00	0.02	-0.01	0.00	0.00	0.01	0.01	0.00	0.01
18	HOL		1.03	0.02	0.02	0.00	0.00	0.01	0.00	0.02	-0.01	0.00	0.00	0.01	0.00	0.00	0.01
17	HOL		1.00	0.01	0.01	-0.01	-0.02	0.03	-0.04	0.07	0.05	-0.04	0.06	0.11	0.13	-0.03	-0.13
137	NGU		1.00	-0.01	0.08	-0.15	0.21	0.05	-0.10	0.11	0.02	0.00	-0.01	-0.02	-0.02	0.04	-0.02
132	NGU		0.96	0.02	-0.13	-0.02	-0.06	-0.26	-0.07	0.08	0.04	0.16	0.02	0.03	-0.02	-0.05	-0.05
121	NGU		0.83	-0.19	-0.20	-0.11	-0.06	-0.03	0.07	-0.12	-0.23	0.17	-0.25	0.07	0.00	0.04	-0.11
120	NGU		0.49	-0.50	0.07	0.05	-0.19	0.36	0.13	0.03	-0.34	0.08	-0.50	-0.14	-0.01	-0.06	0.09
119	NGU		1.01	0.11	-0.18	-0.13	-0.08	-0.04	-0.06	-0.09	0.00	0.03	-0.02	-0.01	0.03	-0.02	-0.08
118	NGU		1.02	0.00	0.17	0.07	0.02	0.05	-0.08	-0.07	-0.04	-0.02	-0.06	-0.04	0.00	-0.06	0.02
117	NGU		0.87	-0.10	0.06	-0.17	0.14	0.14	-0.15	0.23	-0.04	-0.18	-0.14	-0.24	0.13	-0.07	0.04
116	NGU		1.01	0.01	0.02	0.01	0.02	0.06	-0.03	-0.02	0.05	0.10	-0.04	0.05	-0.01	0.07	-0.15
115	NGU		1.04	0.05	0.01	0.00	-0.01	0.01	-0.03	0.01	0.00	0.01	-0.01	0.00	0.00	-0.01	-0.01
114	NGU		0.87	-0.18	-0.01	0.00	-0.03	0.00	0.00	0.02	0.10	-0.10	0.02	-0.24	0.15	-0.13	0.11
113	NGU		0.92	-0.12	0.00	0.01	0.00	-0.02	-0.04	-0.10	0.28	0.05	0.11	0.00	-0.17	0.02	0.07
112	NGU		1.00	0.04	0.03	-0.01	0.01	0.06	-0.07	0.05	0.02	0.00	0.00	-0.04	-0.02	0.02	-0.06
111	NGU		1.02	-0.01	-0.01	0.02	-0.03	0.01	-0.02	0.01	-0.03	0.02	-0.17	0.15	0.01	0.05	0.04
110	NGU		1.06	0.08	0.00	0.00	-0.03	0.02	-0.06	0.01	-0.01	0.05	0.02	-0.02	-0.02	-0.04	-0.02
109	NGU		1.03	0.02	-0.02	0.03	-0.04	0.02	-0.04	0.01	-0.03	0.02	-0.23	0.18	0.00	0.07	0.01
108	NGU		0.97	-0.08	0.02	-0.01	0.01	-0.04	0.05	-0.02	0.05	0.02	0.08	-0.03	0.01	-0.01	0.08
107	NGU		0.88	-0.09	-0.18	0.09	0.03	0.19	0.07	0.17	-0.15	-0.05	-0.16	0.05	-0.06	0.13	0.03
106	NGU		0.91	-0.15	0.08	-0.16	0.21	0.04	-0.05	0.04	0.17	0.07	0.05	-0.12	0.01	0.02	0.02
105	NGU		0.96	-0.09	0.01	0.00	0.00	-0.05	-0.01	-0.04	0.13	0.08	0.07	-0.01	0.00	-0.08	0.13
104	NGU		0.83	-0.19	-0.15	0.15	-0.01	-0.03	0.10	0.24	-0.07	0.12	0.15	0.06	0.18	-0.08	-0.28

		PC	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
AN	**BRD**	**EIV**	**356.82**	**20.785**	**7.890**	**5.744**	**5.224**	**4.549**	**4.168**	**2.690**	**2.501**	**2.351**	**2.134**	**2.026**	**1.790**	**1.640**	**1.587**
102	NGU		0.76	-0.34	-0.04	-0.03	0.02	-0.04	0.10	-0.01	-0.18	0.04	0.22	-0.25	-0.01	-0.07	-0.06
99	NGU		1.03	0.01	0.03	0.02	0.03	-0.01	0.00	0.01	-0.02	0.04	0.04	0.05	-0.01	0.03	-0.04
95	NGU		-0.32	-1.10	0.35	0.40	0.70	-0.23	0.05	-0.10	-0.07	0.29	0.26	0.32	-0.08	0.15	-0.31
91	NGU		-5.06	-0.58	1.87	0.53	0.24	0.19	-0.50	-0.29	-0.08	-0.05	-0.05	-0.05	0.04	-0.14	-0.02
87	NGU		0.83	-0.25	-0.03	0.26	0.02	0.02	0.06	0.10	-0.13	-0.05	-0.07	0.05	-0.06	0.11	-0.13
83	NGU		0.81	-0.30	-0.02	0.02	0.04	-0.06	0.07	0.00	-0.20	-0.03	0.15	-0.19	-0.02	-0.03	-0.01
58	NGU		1.06	0.60	-0.15	-0.12	-0.13	0.05	-0.21	0.43	-0.09	-0.05	0.14	0.18	-0.49	-0.56	-0.05
55	NGU		1.06	0.12	0.03	0.00	0.00	0.00	0.02	0.03	0.00	0.00	-0.01	0.01	-0.04	-0.05	-0.01
52	NGU		0.95	0.06	-0.11	-0.03	-0.08	-0.21	-0.10	0.11	0.01	0.04	-0.07	-0.08	-0.05	0.00	-0.01
50	NGU		0.24	-0.73	-0.27	-0.14	-0.12	-0.07	0.23	-0.14	0.02	-0.25	0.08	-0.14	0.32	-0.01	-0.26
48	NGU		1.03	0.02	0.01	-0.01	-0.01	0.00	0.00	0.00	-0.01	-0.02	0.01	0.00	0.04	-0.01	-0.02
16	NGU		0.62	-0.54	-0.04	-0.01	0.00	-0.04	0.16	0.06	-0.15	0.00	0.03	-0.02	0.21	-0.10	-0.02
15	NGU		0.55	-0.53	-0.24	0.21	-0.19	0.14	0.38	0.21	-0.10	0.01	0.01	-0.11	0.27	0.02	-0.07
14	NGU		0.87	-0.14	0.14	0.17	-0.17	0.10	-0.02	0.13	0.22	-0.20	0.11	0.10	0.19	-0.03	-0.20
13	NGU		0.98	0.00	-0.07	-0.22	0.41	0.14	-0.13	0.03	0.02	-0.03	-0.02	0.04	0.03	-0.02	0.03
12	NGU		0.60	-0.58	-0.05	-0.23	0.03	-0.03	0.18	-0.14	-0.10	-0.05	0.12	0.03	-0.16	0.18	-0.05
11	NGU		0.54	-0.62	-0.19	-0.02	0.23	-0.03	0.15	-0.13	-0.06	-0.11	0.16	0.24	0.06	-0.03	0.02
10	NGU		0.59	-0.38	0.14	0.43	-0.24	-0.45	-0.18	-0.04	0.23	-0.19	-0.22	0.01	0.14	-0.12	0.16
9	NGU		0.80	-0.22	0.01	-0.06	-0.04	0.09	0.01	0.07	-0.07	0.09	-0.12	-0.15	0.07	-0.12	-0.03
8	NGU		1.01	0.20	-0.20	-0.13	0.16	0.13	-0.28	-0.26	0.00	-0.06	-0.09	0.03	0.10	-0.19	-0.07
7	NGU		1.10	0.45	-0.15	-0.10	0.24	0.15	-0.14	-0.11	0.02	-0.01	-0.01	0.08	0.15	-0.12	-0.04
6	NGU		0.32	-0.68	-0.01	0.05	-0.09	-0.60	-0.07	0.16	0.00	-0.34	0.17	0.38	0.16	-0.21	0.29
5	NGU		1.00	0.12	-0.04	0.03	-0.04	0.04	-0.07	0.10	-0.60	0.05	0.28	-0.04	-0.12	-0.09	0.01
4	NGU		1.11	0.66	0.03	0.04	-0.12	0.06	-0.47	0.15	-0.32	0.03	0.43	0.23	0.18	0.18	0.54
3	NGU		0.32	-0.82	-0.07	0.15	0.09	0.06	0.27	-0.14	0.02	0.12	0.00	-0.03	-0.04	-0.02	0.15
2	NGU		0.77	-0.35	-0.08	0.09	-0.01	-0.04	0.04	0.08	-0.42	-0.33	0.13	0.32	0.04	0.09	-0.09
1	NGU		1.11	0.62	-0.01	0.04	-0.15	0.16	-0.62	0.01	-0.05	0.34	0.17	-0.18	0.35	0.66	0.07
62	NxA		0.42	-0.59	-0.10	-0.05	0.12	-0.43	-0.01	0.04	-0.19	0.36	-0.18	-0.11	-0.14	-0.04	0.06
60	NxA		0.93	-0.11	0.01	-0.01	-0.03	0.01	0.02	0.07	-0.03	-0.12	-0.07	-0.07	0.15	-0.08	0.05
57	NxA		0.58	-0.49	-0.05	0.23	0.05	-0.10	0.04	-0.13	-0.14	-0.08	0.03	-0.07	-0.13	0.05	-0.01
51	NxA		0.18	-0.70	-0.05	-0.26	-0.28	0.05	0.17	-0.15	0.54	0.23	-0.03	0.42	0.09	-0.06	0.08
37	NxA		1.04	0.05	0.01	0.00	-0.02	0.01	-0.03	0.01	0.00	0.01	-0.01	0.00	0.00	-0.01	-0.01
35	NxA		0.80	-0.34	-0.19	0.07	-0.04	-0.31	0.11	-0.05	-0.21	-0.03	-0.07	-0.13	-0.12	0.10	0.06
32	NxA		0.67	-0.52	-0.03	0.03	0.04	-0.04	0.17	-0.08	-0.07	-0.09	0.04	0.01	-0.08	0.13	0.00

Additional file 4.2 Significant pairwise association Chi^2 and P-values of deletion and duplication (CN_A and CN_B) CNVR events (CNVR_LocA and CNVR_LocB) identified in all 7 South Africancattle breeds.

CNVR_LocB	CN_B	CNVR_LocA	CN_A	Chi2	Df	P-Value
chr25:40940951-42768470	DEL	chr17:73118011-74998349	DEL	48.3483	10	0
chr11:102861577-107043330	DEL	chr17:73118011-74998349	DEL	39.423	8	0
chr25:40940951-42768470	DEL	chr22:58873440-61283415	DEL	38.3375	8	0
chr22:58873440-61283415	DEL	chr17:73118011-74998349	DEL	37.5537	8	0
chr29:48948337-51502868	DEL	chr17:73118011-74998349	DEL	40.485	10	0
chr11:102861577-107043330	DUP	chr17:73118011-74998349	DUP	43.3063	10	0
chr18:62375495-63727709	DEL	chr17:73118011-74998349	DEL	28.9423	6	0.0001
chr25:40940951-42768470	DEL	chr11:102861577-107043330	DEL	32.561	8	0.0001
chr3:120122176-121403393	DEL	chr14:1514056-2553525	DEL	32.4661	8	0.0001
chr26:50817833-51680135	DEL	chr14:1514056-2553525	DEL	32.4238	8	0.0001
chr18:62375495-63727709	DEL	chr11:102861577-107043330	DEL	28.2878	6	0.0001
chr14:1514056-2553525	DEL	chr6:107678393-109951981	DEL	35.9174	10	0.0001
chr3:120122176-121403393	DEL	chr26:50817833-51680135	DEL	31.9913	8	0.0001
chr18:62375495-63727709	DEL	chr25:40940951-42768470	DEL	27.7483	6	0.0001
chr14:1514056-2553525	DEL	chr25:40940951-42768470	DEL	35.1045	10	0.0001
chr25:40940951-42768470	DUP	chr11:102861577-107043330	DUP	32.186	8	0.0001
chr14:1514056-2553525	DEL	chr17:73118011-74998349	DEL	33.9269	10	0.0002
chr29:48948337-51502868	DEL	chr25:40940951-42768470	DEL	33.6586	10	0.0002
chr11:102861577-107043330	DUP	chr6:107678393-109951981	DUP	34.1401	10	0.0002
chr18:62375495-63727709	DUP	chr17:73118011-74998349	DUP	30.3805	8	0.0002
chr3:120122176-121403393	DEL	chr6:107678393-109951981	DEL	29.446	8	0.0003
chr26:50817833-51680135	DEL	chr6:107678393-109951981	DEL	29.1246	8	0.0003
chr25:40940951-42768470	DEL	chr6:107678393-109951981	DEL	32.5181	10	0.0003
chr26:50817833-51680135	DEL	chr25:40940951-42768470	DEL	28.4356	8	0.0004
chr3:120122176-121403393	DEL	chr25:40940951-42768470	DEL	28.4082	8	0.0004
chr22:58873440-61283415	DEL	chr11:102861577-107043330	DEL	28.3064	8	0.0004
chr29:48948337-51502868	DEL	chr11:102861577-107043330	DEL	27.6979	8	0.0005
chr18:62375495-63727709	DEL	chr22:58873440-61283415	DEL	23.6167	6	0.0006
chr17:73118011-74998349	DEL	chr6:107678393-109951981	DEL	30.8271	10	0.0006
chr14:1514056-2553525	DEL	chr11:102861577-107043330	DEL	27.2527	8	0.0006
chr25:40940951-42768470	DUP	chr17:73118011-74998349	DUP	27.2361	8	0.0006
chr3:120122176-121403393	DEL	chr11:102861577-107043330	DEL	27.21	8	0.0007
chr26:50817833-51680135	DEL	chr17:73118011-74998349	DEL	27.1066	8	0.0007
chr3:120122176-121403393	DEL	chr17:73118011-74998349	DEL	27.0481	8	0.0007
chr26:50817833-51680135	DEL	chr11:102861577-107043330	DEL	26.9878	8	0.0007
chr3:120122176-121403393	DEL	chr18:62375495-63727709	DEL	23.123	6	0.0008
chr26:50817833-51680135	DEL	chr18:62375495-63727709	DEL	23.0606	6	0.0008
chr18:62375495-63727709	DEL	chr14:1514056-2553525	DEL	22.9879	6	0.0008
chr14:1514056-2553525	DEL	chr29:48948337-51502868	DEL	29.5806	10	0.001
chr22:58873440-61283415	DUP	chr6:107678393-109951981	DUP	29.5212	10	0.001
chr22:58873440-61283415	DUP	chr17:73118011-74998349	DUP	28.559	10	0.0015
chr14:1514056-2553525	DEL	chr22:58873440-61283415	DEL	24.9631	8	0.0016
chr3:120122176-121403393	DEL	chr22:58873440-61283415	DEL	24.8789	8	0.0016
chr26:50817833-51680135	DEL	chr22:58873440-61283415	DEL	24.877	8	0.0016
chr25:40940951-42768470	DUP	chr6:107678393-109951981	DUP	24.9115	8	0.0016
chr29:48948337-51502868	DEL	chr6:107678393-109951981	DEL	28.1857	10	0.0017
chr22:58873440-61283415	DUP	chr11:102861577-107043330	DUP	28.1635	10	0.0017
chr26:50817833-51680135	DEL	chr21:70089833-71136925	DEL	24.3825	8	0.002
chr3:120122176-121403393	DEL	chr21:70089833-71136925	DEL	24.3746	8	0.002
chr14:1514056-2553525	DEL	chr21:70089833-71136925	DEL	24.268	8	0.0021

CNVR_LocB	CN_B	CNVR_LocA	CN_A	Chi2	Df	P-Value
chr11:102861577-107043330	DEL	chr6:107678393-109951981	DEL	24.1678	8	0.0021
chr26:25880226-25982293	DEL	chr25:40940951-42768470	DEL	20.2225	6	0.0025
chr18:62375495-63727709	DEL	chr6:107678393-109951981	DEL	20.1898	6	0.0026
chr26:25880226-25982293	DEL	chr17:73118011-74998349	DEL	20.184	6	0.0026
chr29:48948337-51502868	DEL	chr22:58873440-61283415	DEL	23.3444	8	0.0029
chr22:58873440-61283415	DUP	chr25:40940951-42768470	DUP	23.1954	8	0.0031
chr29:48948337-51502868	DEL	chr21:70089833-71136925	DEL	23.0615	8	0.0033
chr21:70089833-71136925	DEL	chr6:107678393-109951981	DEL	22.9391	8	0.0034
chr3:120122176-121403393	DEL	chr29:48948337-51502868	DEL	22.8419	8	0.0036
chr26:50817833-51680135	DEL	chr29:48948337-51502868	DEL	22.7164	8	0.0037
chr6:53514737-53719693	DEL	chr29:48948337-51502868	DUP	19.2848	6	0.0037
chr21:70089833-71136925	DEL	chr25:40940951-42768470	DEL	22.1765	8	0.0046
chr18:62375495-63727709	DUP	chr11:102861577-107043330	DUP	22.1631	8	0.0046
chr22:58873440-61283415	DEL	chr6:107678393-109951981	DEL	22.0859	8	0.0048
chr21:70089833-71136925	DEL	chr17:73118011-74998349	DEL	20.6759	8	0.0081
chr21:70089833-71136925	DEL	chr11:102861577-107043330	DEL	20.5967	8	0.0083
chr26:25880226-25982293	DEL	chr22:58873440-61283415	DEL	16.7661	6	0.0102
chr20:70669729-71652724	DUP	chr18:62375495-63727709	DUP	16.4884	6	0.0114
chr26:25880226-25982293	DEL	chr29:48948337-51502868	DEL	16.0625	6	0.0134
chr21:70089833-71136925	DEL	chr22:58873440-61283415	DEL	18.5698	8	0.0173
chr18:62375495-63727709	DEL	chr29:48948337-51502868	DEL	15.3472	6	0.0177
chr18:62375495-63727709	DEL	chr21:70089833-71136925	DEL	15.2903	6	0.0181
chr20:70669729-71652724	DUP	chr11:102861577-107043330	DUP	18.377	8	0.0186
chr29:48948337-51502868	DUP	chr22:58873440-61283415	DUP	17.7409	8	0.0233
chr3:120122176-121403393	DUP	chr6:107678393-109951981	DUP	17.36	8	0.0266
chr3:120122176-121403393	DUP	chr17:73118011-74998349	DUP	17.293	8	0.0272
chr26:25880226-25982293	DEL	chr11:102861577-107043330	DEL	13.9901	6	0.0297
chr26:25880226-25982293	DEL	chr18:62375495-63727709	DEL	13.829	6	0.0316
chr29:48948337-51502868	DUP	chr17:73118011-74998349	DUP	16.7919	8	0.0324
chr18:62375495-63727709	DUP	chr6:107678393-109951981	DUP	15.7646	8	0.0459

*Deletion- DEL and Duplication - DUP

Additional file 4.3 Associated CNVRs and number of animals (IND) in which they were identified across 7 South African cattle breeds.

Correlation CNVR	IND	Correlation CNVR	IND
chr1:104798012-105264358_DEL	22	chr25:38171850-38377594_DEL	4
chr1:48701612-48866528_DEL	2	chr25:39286957-40282215_DEL	5
chr10:102887596-103470001_DEL	7	chr25:39286957-40282215_DUP	2
chr10:14129663-14288475_DUP	2	chr25:40940951-42768470_DEL	17
chr11:102861577-107043330_DEL	24	chr25:40940951-42768470_DUP	36
chr11:102861577-107043330_DUP	52	chr26:23167656-23414945_DEL	4
chr12:34505806-34662470_DEL	3	chr26:25880226-25982293_DEL	8
chr12:89006700-91017026_DEL	12	chr26:42933219-43087057_DUP	3
chr12:89006700-91017026_DUP	2	chr26:49532894-49762633_DEL	5
chr13:16901756-17025364_DEL	3	chr26:50817833-51680135_DEL	9
chr13:54496419-54829615_DEL	4	chr26:50817833-51680135_DUP	7
chr14:1514056-2553525_DEL	14	chr27:36664-405328_DEL	6
chr14:1514056-2553525_DUP	7	chr28:25060861-25352987_DEL	3
chr14:2803998-3342470_DEL	7	chr28:41674187-41737604_DEL	2
chr14:3885798-4672500_DUP	2	chr29:2324336-2396643_DUP	4
chr15:70597937-70921852_DEL	3	chr29:42897144-43269744_DEL	3
chr15:84865591-85112130_DUP	3	chr29:44372611-44416282_DEL	2
chr16:47654206-47780813_DEL	2	chr29:44969518-45023665_DEL	2
chr16:49386191-49568812_DEL	4	chr29:48012818-48355723_DEL	6
chr16:70816380-71125864_DEL	4	chr29:48948337-51502868_DEL	18
chr17:73118011-74998349_DEL	18	chr29:48948337-51502868_DUP	23
chr17:73118011-74998349_DUP	60	chr3:117575562-118346051_DEL	5
chr18:11121144-11813752_DEL	5	chr3:118813014-119077206_DEL	4
chr18:11121144-11813752_DUP	2	chr3:120122176-121403393_DEL	10
chr18:62375495-63727709_DEL	11	chr3:120122176-121403393_DUP	16
chr18:62375495-63727709_DUP	15	chr3:85764516-85944553_DEL	2
chr18:64188147-64382705_DEL	5	chr4:113079474-113532717_DEL	4
chr18:65819321-65978584_DEL	3	chr4:114326665-114640077_DEL	7
chr19:50336021-50447799_DEL	3	chr4:114326665-114640077_DUP	2
chr19:51395684-52234974_DEL	7	chr4:117831202-120555019_DEL	10
chr19:56607168-57213764_DEL	8	chr4:117831202-120555019_DUP	5
chr19:63424825-63734072_DEL	4	chr5:114543256-114764090_DEL	2
chr2:133816808-134465054_DEL	8	chr5:116915398-118353758_DEL	14
chr2:136386853-136531159_DEL	4	chr5:119221776-120378417_DEL	8
chr20:49090857-49631380_DEL	2	chr5:119221776-120378417_DUP	18
chr20:54180787-54785723_DEL	3	chr5:120553341-121175859_DEL	7
chr20:60639566-60783431_DEL	1	chr5:14770370-14953446_DEL	4
chr20:70669729-71652724_DEL	9	chr6:106495683-107186270_DEL	6
chr20:70669729-71652724_DUP	12	chr6:107678393-109951981_DEL	22
chr21:70089833-71136925_DEL	12	chr6:107678393-109951981_DUP	46
chr21:70089833-71136925_DUP	12	chr6:11671868-11746541_DEL	2
chr22:58873440-61283415_DEL	17	chr6:52628477-52828988_DEL	3
chr22:58873440-61283415_DUP	31	chr6:53514737-53719693_DEL	14
chr23:49094579-52091670_DEL	16	chr7:21462645-21677064_DEL	4
chr23:49094579-52091670_DUP	6	chr7:33722644-33868759_DEL	2
chr24:1282069-1582182_DEL	3	chr7:4226753-4655753_DEL	4
chr24:61455723-62320145_DEL	6	chr7:44658442-46025089_DEL	8
chr25:104438-1365841_DEL	12	chr9:102574022-103180423_DUP	2
chr25:104438-1365841_DUP	4	chr9:103383683-105462864_DEL	11
chr25:1665327-1808056_DEL	5	chr9:56003900-56370974_DEL	2

Correlation CNVR	IND
chr25:1955733-2606575_DEL	7
chr25:36448529-36514994_DEL	2

Correlation CNVR	IND
chr9:5901981-5981648_DEL	5

Additional file 4.4 Significant pairwise association Chi^2 and P-values of deletion and duplication (CN_A and CN_B) CNVR events (CNVR_LocA and CNVR_LocB) identified in 2 South African taurine cattle breeds.

CNVR_LocA	CN_A	CNVR_LocB	CN_B	Chi2	Df	P-Value
chr11:102861577-107043330	DUP	chr6:107678393-109951981	DUP	28.3813	4	0
chr25:40940951-42768470	DUP	chr11:102861577-107043330	DUP	28.5774	4	0
chr22:58873440-61283415	DUP	chr11:102861577-107043330	DUP	24.1208	4	0.0001
chr25:40940951-42768470	DEL	chr17:73118011-74998349	DEL	22.2278	4	0.0002
chr22:58873440-61283415	DUP	chr17:73118011-74998349	DUP	21.6396	4	0.0002
chr18:62375495-63727709	DEL	chr11:102861577-107043330	DEL	16.1579	2	0.0003
chr25:40940951-42768470	DUP	chr6:107678393-109951981	DUP	21.4098	4	0.0003
chr25:40940951-42768470	DEL	chr22:58873440-61283415	DEL	15.5035	2	0.0004
chr22:58873440-61283415	DEL	chr17:73118011-74998349	DEL	15.7507	2	0.0004
chr22:58873440-61283415	DEL	chr11:102861577-107043330	DEL	15.1234	2	0.0005
chr18:62375495-63727709	DEL	chr17:73118011-74998349	DEL	15.1234	2	0.0005
chr25:40940951-42768470	DEL	chr11:102861577-107043330	DEL	15.2422	2	0.0005
chr18:62375495-63727709	DEL	chr22:58873440-61283415	DEL	15.3686	2	0.0005
chr22:58873440-61283415	DUP	chr25:40940951-42768470	DUP	20.0827	4	0.0005
chr11:102861577-107043330	DEL	chr17:73118011-74998349	DEL	14.8708	2	0.0006
chr18:62375495-63727709	DEL	chr25:40940951-42768470	DEL	14.5005	2	0.0007
chr26:50817833-51680135	DUP	chr29:48948337-51502868	DUP	14.3644	2	0.0008
chr25:1665327-1808056	DEL	chr14:1514056-2553525	DEL	18.5261	4	0.001
chr25:1665327-1808056	DEL	chr25:104438-1365841	DEL	18.2473	4	0.0011
chr25:1665327-1808056	DEL	chr29:48948337-51502868	DEL	18.3284	4	0.0011
chr25:104438-1365841	DEL	chr29:48948337-51502868	DEL	18.0158	4	0.0012
chr29:48948337-51502868	DEL	chr6:107678393-109951981	DEL	18.0746	4	0.0012
chr25:104438-1365841	DEL	chr14:1514056-2553525	DEL	18.1388	4	0.0012
chr25:104438-1365841	DEL	chr6:107678393-109951981	DEL	17.8423	4	0.0013
chr25:1665327-1808056	DEL	chr6:107678393-109951981	DEL	17.8939	4	0.0013
chr14:1514056-2553525	DEL	chr6:107678393-109951981	DEL	17.9446	4	0.0013
chr14:1514056-2553525	DEL	chr29:48948337-51502868	DEL	17.7196	4	0.0014
chr18:62375495-63727709	DUP	chr17:73118011-74998349	DUP	17.6068	4	0.0015
chr11:102861577-107043330	DUP	chr17:73118011-74998349	DUP	17.0344	4	0.0019
chr18:62375495-63727709	DUP	chr6:107678393-109951981	DUP	16.7631	4	0.0021
chr25:40940951-42768470	DUP	chr17:73118011-74998349	DUP	16.6543	4	0.0023
chr12:89006700-91017026	DEL	chr6:107678393-109951981	DEL	11.9749	2	0.0025
chr25:104438-1365841	DEL	chr3:120122176-121403393	DEL	11.999	2	0.0025
chr25:1665327-1808056	DEL	chr5:116915398-118353758	DEL	11.8664	2	0.0027
chr18:64188147-64382705	DEL	chr27:36664-405328	DEL	11.6587	2	0.0029
chr16:49386191-49568812	DEL	chr5:116915398-118353758	DEL	11.6587	2	0.0029
chr18:64188147-64382705	DEL	chr14:1514056-2553525	DEL	11.6655	2	0.0029
chr12:89006700-91017026	DEL	chr14:1514056-2553525	DEL	11.6723	2	0.0029
chr14:2803998-3342470	DEL	chr21:70089833-71136925	DEL	11.6723	2	0.0029
chr25:1665327-1808056	DEL	chr23:49094579-52091670	DEL	11.6723	2	0.0029
chr16:70816380-71125864	DEL	chr27:36664-405328	DEL	11.6723	2	0.0029
chr20:70669729-71652724	DEL	chr14:1514056-2553525	DEL	11.6861	2	0.0029
chr29:44372611-44416282	DEL	chr16:49386191-49568812	DEL	11.6999	2	0.0029
chr5:116915398-118353758	DEL	chr14:1514056-2553525	DEL	11.7069	2	0.0029
chr14:2803998-3342470	DEL	chr26:50817833-51680135	DEL	11.7069	2	0.0029
chr14:2803998-3342470	DEL	chr5:116915398-118353758	DEL	11.7069	2	0.0029
chr16:70816380-71125864	DEL	chr14:1514056-2553525	DEL	11.7069	2	0.0029
chr26:50817833-51680135	DEL	chr14:1514056-2553525	DEL	11.5918	2	0.003
chr6:11671868-11746541	DEL	chr28:25060861-25352987	DEL	11.5984	2	0.003

CNVR_LocA	CN_A	CNVR_LocB	CN_B	Chi2	Df	P-Value
chr27:36664-405328	DEL	chr12:89006700-91017026	DEL	11.6116	2	0.003
chr16:49386191-49568812	DEL	chr25:1665327-1808056	DEL	11.6183	2	0.003
chr16:70816380-71125864	DEL	chr21:70089833-71136925	DEL	11.6183	2	0.003
chr16:70816380-71125864	DEL	chr25:1665327-1808056	DEL	11.6183	2	0.003
chr29:44372611-44416282	DEL	chr12:89006700-91017026	DEL	11.6183	2	0.003
chr20:70669729-71652724	DEL	chr26:50817833-51680135	DEL	11.6317	2	0.003
chr26:50817833-51680135	DEL	chr21:70089833-71136925	DEL	11.6384	2	0.003
chr29:44372611-44416282	DEL	chr25:104438-1365841	DEL	11.6384	2	0.003
chr5:116915398-118353758	DEL	chr26:50817833-51680135	DEL	11.6451	2	0.003
chr12:89006700-91017026	DEL	chr23:49094579-52091670	DEL	11.6519	2	0.003
chr29:44372611-44416282	DEL	chr5:116915398-118353758	DEL	11.6519	2	0.003
chr25:104438-1365841	DEL	chr17:73118011-74998349	DEL	16.0258	4	0.003
chr23:49094579-52091670	DEL	chr14:1514056-2553525	DEL	11.5271	2	0.0031
chr23:49094579-52091670	DEL	chr5:116915398-118353758	DEL	11.5334	2	0.0031
chr14:2803998-3342470	DEL	chr12:89006700-91017026	DEL	11.5463	2	0.0031
chr16:49386191-49568812	DEL	chr25:104438-1365841	DEL	11.5463	2	0.0031
chr27:36664-405328	DEL	chr20:70669729-71652724	DEL	11.5527	2	0.0031
chr25:1665327-1808056	DEL	chr14:2803998-3342470	DEL	11.5527	2	0.0031
chr29:44372611-44416282	DEL	chr29:48948337-51502868	DEL	11.5527	2	0.0031
chr16:49386191-49568812	DEL	chr29:48948337-51502868	DEL	11.5592	2	0.0031
chr16:70816380-71125864	DEL	chr6:107678393-109951981	DEL	11.5592	2	0.0031
chr23:49094579-52091670	DEL	chr3:120122176-121403393	DEL	11.5657	2	0.0031
chr27:36664-405328	DEL	chr23:49094579-52091670	DEL	11.5657	2	0.0031
chr29:44372611-44416282	DEL	chr20:70669729-71652724	DEL	11.5657	2	0.0031
chr29:44372611-44416282	DEL	chr18:64188147-64382705	DEL	11.5657	2	0.0031
chr25:104438-1365841	DEL	chr5:116915398-118353758	DEL	11.5787	2	0.0031
chr25:40940951-42768470	DEL	chr6:107678393-109951981	DEL	15.9482	4	0.0031
chr25:1665327-1808056	DEL	chr25:40940951-42768470	DEL	15.9494	4	0.0031
chr29:48948337-51502868	DEL	chr25:40940951-42768470	DEL	15.953	4	0.0031
chr26:50817833-51680135	DUP	chr17:73118011-74998349	DUP	11.5463	2	0.0031
chr27:36664-405328	DEL	chr14:2803998-3342470	DEL	11.4644	2	0.0032
chr16:49386191-49568812	DEL	chr14:1514056-2553525	DEL	11.4705	2	0.0032
chr20:70669729-71652724	DEL	chr3:120122176-121403393	DEL	11.4767	2	0.0032
chr16:70816380-71125864	DEL	chr23:49094579-52091670	DEL	11.4767	2	0.0032
chr29:44372611-44416282	DEL	chr16:70816380-71125864	DEL	11.4767	2	0.0032
chr3:120122176-121403393	DEL	chr26:50817833-51680135	DEL	11.4892	2	0.0032
chr14:2803998-3342470	DEL	chr3:120122176-121403393	DEL	11.4955	2	0.0032
chr18:64188147-64382705	DEL	chr23:49094579-52091670	DEL	11.4955	2	0.0032
chr25:1665327-1808056	DEL	chr18:64188147-64382705	DEL	11.4955	2	0.0032
chr3:120122176-121403393	DEL	chr21:70089833-71136925	DEL	11.5017	2	0.0032
chr16:70816380-71125864	DEL	chr3:120122176-121403393	DEL	11.5017	2	0.0032
chr21:70089833-71136925	DEL	chr6:107678393-109951981	DEL	11.508	2	0.0032
chr12:89006700-91017026	DEL	chr29:48948337-51502868	DEL	11.5144	2	0.0032
chr16:70816380-71125864	DEL	chr25:104438-1365841	DEL	11.5207	2	0.0032
chr14:1514056-2553525	DEL	chr17:73118011-74998349	DEL	15.8418	4	0.0032
chr25:1665327-1808056	DEL	chr17:73118011-74998349	DEL	15.8437	4	0.0032
chr23:49094579-52091670	DEL	chr29:48948337-51502868	DEL	11.4036	2	0.0033
chr14:2803998-3342470	DEL	chr6:107678393-109951981	DEL	11.4036	2	0.0033
chr16:49386191-49568812	DEL	chr6:107678393-109951981	DEL	11.4036	2	0.0033
chr16:49386191-49568812	DEL	chr27:36664-405328	DEL	11.4036	2	0.0033
chr5:116915398-118353758	DEL	chr6:107678393-109951981	DEL	11.4096	2	0.0033
chr25:1665327-1808056	DEL	chr21:70089833-71136925	DEL	11.4096	2	0.0033
chr27:36664-405328	DEL	chr21:70089833-71136925	DEL	11.4156	2	0.0033

CNVR_LocA	CN_A	CNVR_LocB	CN_B	Chi2	Df	P-Value
chr27:36664-405328	DEL	chr3:120122176-121403393	DEL	11.4156	2	0.0033
chr16:70816380-71125864	DEL	chr16:49386191-49568812	DEL	11.4156	2	0.0033
chr29:44372611-44416282	DEL	chr27:36664-405328	DEL	11.4156	2	0.0033
chr16:49386191-49568812	DEL	chr3:120122176-121403393	DEL	11.4216	2	0.0033
chr29:48948337-51502868	DEL	chr21:70089833-71136925	DEL	11.4277	2	0.0033
chr29:44372611-44416282	DEL	chr23:49094579-52091670	DEL	11.4277	2	0.0033
chr18:64188147-64382705	DEL	chr6:107678393-109951981	DEL	11.4398	2	0.0033
chr16:70816380-71125864	DEL	chr12:89006700-91017026	DEL	11.4398	2	0.0033
chr12:89006700-91017026	DEL	chr21:70089833-71136925	DEL	11.4521	2	0.0033
chr12:89006700-91017026	DEL	chr3:120122176-121403393	DEL	11.4521	2	0.0033
chr25:1665327-1808056	DEL	chr20:70669729-71652724	DEL	11.4521	2	0.0033
chr16:49386191-49568812	DEL	chr21:70089833-71136925	DEL	11.4521	2	0.0033
chr29:44372611-44416282	DEL	chr14:1514056-2553525	DEL	11.4521	2	0.0033
chr14:2803998-3342470	DEL	chr29:48948337-51502868	DEL	11.4582	2	0.0033
chr18:64188147-64382705	DEL	chr29:48948337-51502868	DEL	11.4582	2	0.0033
chr29:48948337-51502868	DEL	chr17:73118011-74998349	DEL	15.7806	4	0.0033
chr14:1514056-2553525	DEL	chr25:40940951-42768470	DEL	15.7812	4	0.0033
chr3:120122176-121403393	DEL	chr6:107678393-109951981	DEL	11.3504	2	0.0034
chr14:2803998-3342470	DEL	chr23:49094579-52091670	DEL	11.3504	2	0.0034
chr27:36664-405328	DEL	chr25:104438-1365841	DEL	11.3504	2	0.0034
chr16:70816380-71125864	DEL	chr14:2803998-3342470	DEL	11.3504	2	0.0034
chr20:70669729-71652724	DEL	chr5:116915398-118353758	DEL	11.3562	2	0.0034
chr29:44372611-44416282	DEL	chr14:2803998-3342470	DEL	11.3562	2	0.0034
chr3:120122176-121403393	DEL	chr14:1514056-2553525	DEL	11.3621	2	0.0034
chr5:116915398-118353758	DEL	chr21:70089833-71136925	DEL	11.3739	2	0.0034
chr12:89006700-91017026	DEL	chr26:50817833-51680135	DEL	11.3739	2	0.0034
chr12:89006700-91017026	DEL	chr25:104438-1365841	DEL	11.3739	2	0.0034
chr18:64188147-64382705	DEL	chr20:70669729-71652724	DEL	11.3739	2	0.0034
chr25:1665327-1808056	DEL	chr27:36664-405328	DEL	11.3739	2	0.0034
chr14:2803998-3342470	DEL	chr25:104438-1365841	DEL	11.3798	2	0.0034
chr27:36664-405328	DEL	chr29:48948337-51502868	DEL	11.3798	2	0.0034
chr25:104438-1365841	DEL	chr23:49094579-52091670	DEL	11.3857	2	0.0034
chr20:70669729-71652724	DEL	chr29:48948337-51502868	DEL	11.3916	2	0.0034
chr27:36664-405328	DEL	chr26:50817833-51680135	DEL	11.3976	2	0.0034
chr16:49386191-49568812	DEL	chr14:2803998-3342470	DEL	11.3976	2	0.0034
chr17:73118011-74998349	DEL	chr6:107678393-109951981	DEL	15.7197	4	0.0034
chr16:49386191-49568812	DEL	chr18:64188147-64382705	DEL	11.2929	2	0.0035
chr29:44372611-44416282	DEL	chr26:50817833-51680135	DEL	11.31	2	0.0035
chr29:44372611-44416282	DEL	chr25:1665327-1808056	DEL	11.31	2	0.0035
chr27:36664-405328	DEL	chr14:1514056-2553525	DEL	11.3214	2	0.0035
chr18:64188147-64382705	DEL	chr25:104438-1365841	DEL	11.3214	2	0.0035
chr14:2803998-3342470	DEL	chr14:1514056-2553525	DEL	11.3272	2	0.0035
chr25:1665327-1808056	DEL	chr26:50817833-51680135	DEL	11.3272	2	0.0035
chr5:116915398-118353758	DEL	chr3:120122176-121403393	DEL	11.333	2	0.0035
chr16:70816380-71125864	DEL	chr5:116915398-118353758	DEL	11.333	2	0.0035
chr16:49386191-49568812	DEL	chr26:50817833-51680135	DEL	11.3388	2	0.0035
chr27:36664-405328	DEL	chr6:107678393-109951981	DEL	11.2481	2	0.0036
chr26:50817833-51680135	DEL	chr6:107678393-109951981	DEL	11.2536	2	0.0036
chr29:44372611-44416282	DEL	chr21:70089833-71136925	DEL	11.2536	2	0.0036
chr18:64188147-64382705	DEL	chr26:50817833-51680135	DEL	11.2592	2	0.0036
chr18:64188147-64382705	DEL	chr3:120122176-121403393	DEL	11.2592	2	0.0036
chr26:50817833-51680135	DEL	chr29:48948337-51502868	DEL	11.2648	2	0.0036
chr25:104438-1365841	DEL	chr26:50817833-51680135	DEL	11.2704	2	0.0036

CNVR_LocA	CN_A	CNVR_LocB	CN_B	Chi2	Df	P-Value
chr29:44372611-44416282	DEL	chr3:120122176-121403393	DEL	11.2816	2	0.0036
chr25:104438-1365841	DEL	chr25:40940951-42768470	DEL	15.6047	4	0.0036
chr12:89006700-91017026	DEL	chr5:116915398-118353758	DEL	11.1827	2	0.0037
chr14:1514056-2553525	DEL	chr21:70089833-71136925	DEL	11.1934	2	0.0037
chr18:64188147-64382705	DEL	chr12:89006700-91017026	DEL	11.1934	2	0.0037
chr12:89006700-91017026	DEL	chr20:70669729-71652724	DEL	11.1988	2	0.0037
chr18:64188147-64382705	DEL	chr5:116915398-118353758	DEL	11.1988	2	0.0037
chr25:104438-1365841	DEL	chr20:70669729-71652724	DEL	11.2043	2	0.0037
chr25:1665327-1808056	DEL	chr12:89006700-91017026	DEL	11.2043	2	0.0037
chr23:49094579-52091670	DEL	chr26:50817833-51680135	DEL	11.2151	2	0.0037
chr25:104438-1365841	DEL	chr21:70089833-71136925	DEL	11.2206	2	0.0037
chr20:70669729-71652724	DEL	chr6:107678393-109951981	DEL	11.2261	2	0.0037
chr25:1665327-1808056	DEL	chr3:120122176-121403393	DEL	11.1298	2	0.0038
chr18:64188147-64382705	DEL	chr21:70089833-71136925	DEL	11.135	2	0.0038
chr23:49094579-52091670	DEL	chr21:70089833-71136925	DEL	11.1508	2	0.0038
chr3:120122176-121403393	DEL	chr29:48948337-51502868	DEL	11.1561	2	0.0038
chr29:44372611-44416282	DEL	chr6:107678393-109951981	DEL	11.1561	2	0.0038
chr16:49386191-49568812	DEL	chr23:49094579-52091670	DEL	11.1667	2	0.0038
chr16:70816380-71125864	DEL	chr29:48948337-51502868	DEL	11.1667	2	0.0038
chr16:70816380-71125864	DEL	chr18:64188147-64382705	DEL	11.1667	2	0.0038
chr26:50817833-51680135	DUP	chr14:1514056-2553525	DUP	11.1298	2	0.0038
chr14:1514056-2553525	DUP	chr18:62375495-63727709	DUP	15.4575	4	0.0038
chr27:36664-405328	DEL	chr5:116915398-118353758	DEL	11.0731	2	0.0039
chr16:70816380-71125864	DEL	chr26:50817833-51680135	DEL	11.0782	2	0.0039
chr14:2803998-3342470	DEL	chr20:70669729-71652724	DEL	11.0884	2	0.0039
chr18:64188147-64382705	DEL	chr14:2803998-3342470	DEL	11.0987	2	0.0039
chr16:70816380-71125864	DEL	chr20:70669729-71652724	DEL	11.0987	2	0.0039
chr23:49094579-52091670	DEL	chr6:107678393-109951981	DEL	11.063	2	0.004
chr23:49094579-52091670	DEL	chr20:70669729-71652724	DEL	11.0681	2	0.004
chr5:116915398-118353758	DEL	chr29:48948337-51502868	DEL	10.9887	2	0.0041
chr20:70669729-71652724	DEL	chr21:70089833-71136925	DEL	10.9984	2	0.0041
chr16:49386191-49568812	DEL	chr12:89006700-91017026	DEL	11.0131	2	0.0041
chr16:49386191-49568812	DEL	chr20:70669729-71652724	DEL	10.9693	2	0.0042
chr26:50817833-51680135	DUP	chr6:107678393-109951981	DUP	10.7418	2	0.0047
chr29:48948337-51502868	DUP	chr17:73118011-74998349	DUP	14.4669	4	0.0059
chr23:49094579-52091670	DEL	chr17:73118011-74998349	DEL	9.5166	2	0.0086
chr18:62375495-63727709	DEL	chr21:70089833-71136925	DEL	9.4981	2	0.0087
chr16:49386191-49568812	DEL	chr18:62375495-63727709	DEL	9.4434	2	0.0089
chr14:2803998-3342470	DEL	chr18:62375495-63727709	DEL	9.4502	2	0.0089
chr20:70669729-71652724	DEL	chr11:102861577-107043330	DEL	9.4166	2	0.009
chr18:62375495-63727709	DEL	chr29:48948337-51502868	DEL	9.4188	2	0.009
chr11:102861577-107043330	DEL	chr6:107678393-109951981	DEL	9.4277	2	0.009
chr22:58873440-61283415	DEL	chr6:107678393-109951981	DEL	9.3728	2	0.0092
chr26:25880226-25982293	DEL	chr25:40940951-42768470	DEL	9.349	2	0.0093
chr5:116915398-118353758	DEL	chr11:102861577-107043330	DEL	9.349	2	0.0093
chr14:1514056-2553525	DEL	chr22:58873440-61283415	DEL	9.3619	2	0.0093
chr29:44372611-44416282	DEL	chr18:62375495-63727709	DEL	9.3277	2	0.0094
chr14:2803998-3342470	DEL	chr25:40940951-42768470	DEL	9.332	2	0.0094
chr25:1665327-1808056	DEL	chr18:62375495-63727709	DEL	9.3341	2	0.0094
chr3:120122176-121403393	DEL	chr17:73118011-74998349	DEL	9.3362	2	0.0094
chr12:89006700-91017026	DEL	chr17:73118011-74998349	DEL	9.3045	2	0.0095
chr26:25880226-25982293	DEL	chr22:58873440-61283415	DEL	9.3193	2	0.0095
chr20:70669729-71652724	DEL	chr17:73118011-74998349	DEL	9.3193	2	0.0095

CNVR_LocA	CN_A	CNVR_LocB	CN_B	Chi2	Df	P-Value
chr14:2803998-3342470	DEL	chr17:73118011-74998349	DEL	9.2878	2	0.0096
chr20:70669729-71652724	DEL	chr22:58873440-61283415	DEL	9.2899	2	0.0096
chr26:25880226-25982293	DEL	chr18:62375495-63727709	DEL	9.3003	2	0.0096
chr18:64188147-64382705	DEL	chr11:102861577-107043330	DEL	9.263	2	0.0097
chr16:70816380-71125864	DEL	chr22:58873440-61283415	DEL	9.2651	2	0.0097
chr29:44372611-44416282	DEL	chr17:73118011-74998349	DEL	9.2651	2	0.0097
chr20:70669729-71652724	DEL	chr18:62375495-63727709	DEL	9.2671	2	0.0097
chr25:104438-1365841	DEL	chr11:102861577-107043330	DEL	9.2671	2	0.0097
chr18:64188147-64382705	DEL	chr25:40940951-42768470	DEL	9.2692	2	0.0097
chr29:44372611-44416282	DEL	chr11:102861577-107043330	DEL	9.2692	2	0.0097
chr25:1665327-1808056	DEL	chr11:102861577-107043330	DEL	9.2713	2	0.0097
chr27:36664-405328	DEL	chr25:40940951-42768470	DEL	9.2775	2	0.0097
chr26:50817833-51680135	DEL	chr18:62375495-63727709	DEL	9.2426	2	0.0098
chr5:116915398-118353758	DEL	chr18:62375495-63727709	DEL	9.2487	2	0.0098
chr25:104438-1365841	DEL	chr22:58873440-61283415	DEL	9.2528	2	0.0098
chr27:36664-405328	DEL	chr22:58873440-61283415	DEL	9.2548	2	0.0098
chr27:36664-405328	DEL	chr18:62375495-63727709	DEL	9.2548	2	0.0098
chr16:49386191-49568812	DEL	chr22:58873440-61283415	DEL	9.2224	2	0.0099
chr14:2803998-3342470	DEL	chr22:58873440-61283415	DEL	9.2244	2	0.0099
chr25:104438-1365841	DEL	chr18:62375495-63727709	DEL	9.2264	2	0.0099
chr29:48948337-51502868	DEL	chr22:58873440-61283415	DEL	9.2365	2	0.0099
chr5:116915398-118353758	DEL	chr17:73118011-74998349	DEL	9.2365	2	0.0099
chr23:49094579-52091670	DEL	chr22:58873440-61283415	DEL	9.2385	2	0.0099
chr16:49386191-49568812	DEL	chr11:102861577-107043330	DEL	9.2043	2	0.01
chr27:36664-405328	DEL	chr17:73118011-74998349	DEL	9.2063	2	0.01
chr3:120122176-121403393	DEL	chr11:102861577-107043330	DEL	9.2083	2	0.01
chr5:116915398-118353758	DEL	chr25:40940951-42768470	DEL	9.2163	2	0.01
chr18:62375495-63727709	DEL	chr14:1514056-2553525	DEL	9.1825	2	0.0101
chr18:64188147-64382705	DEL	chr17:73118011-74998349	DEL	9.1825	2	0.0101
chr21:70089833-71136925	DEL	chr22:58873440-61283415	DEL	9.1865	2	0.0101
chr29:44372611-44416282	DEL	chr22:58873440-61283415	DEL	9.1865	2	0.0101
chr27:36664-405328	DEL	chr11:102861577-107043330	DEL	9.1885	2	0.0101
chr18:62375495-63727709	DEL	chr6:107678393-109951981	DEL	9.1904	2	0.0101
chr26:25880226-25982293	DEL	chr17:73118011-74998349	DEL	9.1924	2	0.0101
chr12:89006700-91017026	DEL	chr11:102861577-107043330	DEL	9.1984	2	0.0101
chr21:70089833-71136925	DEL	chr11:102861577-107043330	DEL	9.2004	2	0.0101
chr3:120122176-121403393	DEL	chr18:62375495-63727709	DEL	9.2004	2	0.0101
chr23:49094579-52091670	DEL	chr25:40940951-42768470	DEL	9.1766	2	0.0102
chr16:49386191-49568812	DEL	chr25:40940951-42768470	DEL	9.1766	2	0.0102
chr21:70089833-71136925	DEL	chr25:40940951-42768470	DEL	9.1806	2	0.0102
chr29:48948337-51502868	DEL	chr11:102861577-107043330	DEL	9.1806	2	0.0102
chr16:70816380-71125864	DEL	chr17:73118011-74998349	DEL	9.1551	2	0.0103
chr18:64188147-64382705	DEL	chr18:62375495-63727709	DEL	9.161	2	0.0103
chr25:1665327-1808056	DEL	chr22:58873440-61283415	DEL	9.13	2	0.0104
chr3:120122176-121403393	DEL	chr22:58873440-61283415	DEL	9.1319	2	0.0104
chr3:118813014-119077206	DEL	chr5:119221776-120378417	DEL	13.189	4	0.0104
chr26:50817833-51680135	DEL	chr25:40940951-42768470	DEL	9.109	2	0.0105
chr14:2803998-3342470	DEL	chr11:102861577-107043330	DEL	9.1109	2	0.0105
chr16:70816380-71125864	DEL	chr11:102861577-107043330	DEL	9.1109	2	0.0105
chr12:89006700-91017026	DEL	chr25:40940951-42768470	DEL	9.1147	2	0.0105
chr26:50817833-51680135	DEL	chr22:58873440-61283415	DEL	9.1185	2	0.0105
chr26:25880226-25982293	DEL	chr11:102861577-107043330	DEL	9.1223	2	0.0105
chr16:49386191-49568812	DEL	chr17:73118011-74998349	DEL	9.0995	2	0.0106

CNVR_LocA	CN_A	CNVR_LocB	CN_B	Chi2	Df	P-Value
chr3:120122176-121403393	DEL	chr25:40940951-42768470	DEL	9.075	2	0.0107
chr23:49094579-52091670	DEL	chr18:62375495-63727709	DEL	9.0769	2	0.0107
chr20:70669729-71652724	DEL	chr25:40940951-42768470	DEL	9.0788	2	0.0107
chr18:64188147-64382705	DEL	chr22:58873440-61283415	DEL	9.0825	2	0.0107
chr29:44372611-44416282	DEL	chr25:40940951-42768470	DEL	9.0825	2	0.0107
chr5:116915398-118353758	DEL	chr22:58873440-61283415	DEL	9.0546	2	0.0108
chr26:50817833-51680135	DEL	chr17:73118011-74998349	DEL	9.0601	2	0.0108
chr16:70816380-71125864	DEL	chr18:62375495-63727709	DEL	9.0453	2	0.0109
chr3:118813014-119077206	DEL	chr25:38171850-38377594	DEL	13.0762	4	0.0109
chr26:50817833-51680135	DEL	chr11:102861577-107043330	DEL	9.0288	2	0.011
chr25:38171850-38377594	DEL	chr5:119221776-120378417	DEL	13.0646	4	0.011
chr14:1514056-2553525	DEL	chr11:102861577-107043330	DEL	8.9962	2	0.0111
chr12:89006700-91017026	DEL	chr22:58873440-61283415	DEL	9.0016	2	0.0111
chr12:89006700-91017026	DEL	chr18:62375495-63727709	DEL	8.989	2	0.0112
chr16:70816380-71125864	DEL	chr25:40940951-42768470	DEL	8.9677	2	0.0113
chr21:70089833-71136925	DEL	chr17:73118011-74998349	DEL	8.9395	2	0.0115
chr23:49094579-52091670	DEL	chr11:102861577-107043330	DEL	8.8692	2	0.0119
chr29:48948337-51502868	DUP	chr6:107678393-109951981	DUP	12.5424	4	0.0137
chr3:120122176-121403393	DUP	chr17:73118011-74998349	DUP	12.2939	4	0.0153
chr4:117831202-120555019	DUP	chr18:62375495-63727709	DUP	8.1739	2	0.0168
chr14:1514056-2553525	DUP	chr20:70669729-71652724	DUP	11.916	4	0.018
chr29:2324336-2396643	DUP	chr17:73118011-74998349	DUP	7.9638	2	0.0187
chr3:118813014-119077206	DEL	chr14:1514056-2553525	DEL	11.7631	4	0.0192
chr3:118813014-119077206	DEL	chr25:1665327-1808056	DEL	11.7501	4	0.0193
chr5:119221776-120378417	DEL	chr25:104438-1365841	DEL	11.7544	4	0.0193
chr25:38171850-38377594	DEL	chr25:1665327-1808056	DEL	11.741	4	0.0194
chr5:119221776-120378417	DEL	chr14:1514056-2553525	DEL	11.7323	4	0.0195
chr25:38171850-38377594	DEL	chr25:104438-1365841	DEL	11.7125	4	0.0196
chr25:38171850-38377594	DEL	chr14:1514056-2553525	DEL	11.7077	4	0.0197
chr3:118813014-119077206	DEL	chr29:48948337-51502868	DEL	11.6896	4	0.0198
chr5:119221776-120378417	DEL	chr6:107678393-109951981	DEL	11.6836	4	0.0199
chr3:118813014-119077206	DEL	chr25:104438-1365841	DEL	11.6716	4	0.02
chr5:119221776-120378417	DEL	chr29:48948337-51502868	DEL	11.6436	4	0.0202
chr25:38171850-38377594	DEL	chr6:107678393-109951981	DEL	11.6327	4	0.0203
chr3:118813014-119077206	DEL	chr6:107678393-109951981	DEL	11.635	4	0.0203
chr25:38171850-38377594	DEL	chr29:48948337-51502868	DEL	11.6199	4	0.0204
chr20:49090857-49631380	DEL	chr9:5901981-5981648	DEL	7.3194	2	0.0257
chr15:70597937-70921852	DEL	chr9:5901981-5981648	DEL	7.3256	2	0.0257
chr25:38171850-38377594	DEL	chr25:40940951-42768470	DEL	10.9655	4	0.027
chr25:38171850-38377594	DEL	chr17:73118011-74998349	DEL	10.9396	4	0.0273
chr5:119221776-120378417	DEL	chr17:73118011-74998349	DEL	10.9228	4	0.0274
chr3:118813014-119077206	DEL	chr25:40940951-42768470	DEL	10.9186	4	0.0275
chr3:118813014-119077206	DEL	chr17:73118011-74998349	DEL	10.8549	4	0.0282
chr5:119221776-120378417	DEL	chr25:40940951-42768470	DEL	10.8499	4	0.0283
chr4:114326665-114640077	DUP	chr14:1514056-2553525	DUP	7.1094	2	0.0286
chr3:120122176-121403393	DUP	chr6:107678393-109951981	DUP	10.7853	4	0.0291
chr18:62375495-63727709	DUP	chr29:48948337-51502868	DUP	10.7119	4	0.03
chr21:70089833-71136925	DUP	chr18:62375495-63727709	DUP	10.5532	4	0.0321
chr14:3885798-4672500	DUP	chr22:58873440-61283415	DUP	6.858	2	0.0324
chr5:114543256-114764090	DEL	chr29:48948337-51502868	DEL	6.7972	2	0.0334
chr5:114543256-114764090	DEL	chr5:119221776-120378417	DEL	6.7781	2	0.0337
chr21:70089833-71136925	DUP	chr29:48948337-51502868	DUP	10.4353	4	0.0337
chr21:70089833-71136925	DUP	chr5:119221776-120378417	DUP	10.417	4	0.034

CNVR_LocA	CN_A	CNVR_LocB	CN_B	Chi2	Df	P-Value
chr9:56003900-56370974	DEL	chr20:54180787-54785723	DEL	6.7342	2	0.0345
chr25:39286957-40282215	DUP	chr23:49094579-52091670	DUP	6.7353	2	0.0345
chr5:114543256-114764090	DEL	chr12:34505806-34662470	DEL	6.7301	2	0.0346
chr3:118813014-119077206	DEL	chr12:34505806-34662470	DEL	6.7215	2	0.0347
chr25:39286957-40282215	DUP	chr10:14129663-14288475	DUP	6.7203	2	0.0347
chr25:38171850-38377594	DEL	chr12:34505806-34662470	DEL	6.7134	2	0.0349
chr12:34505806-34662470	DEL	chr29:48948337-51502868	DEL	6.6985	2	0.0351
chr5:114543256-114764090	DEL	chr6:107678393-109951981	DEL	6.6985	2	0.0351
chr12:89006700-91017026	DUP	chr15:84865591-85112130	DUP	6.702	2	0.0351
chr12:34505806-34662470	DEL	chr25:40940951-42768470	DEL	6.6957	2	0.0352
chr12:34505806-34662470	DEL	chr25:104438-1365841	DEL	6.6855	2	0.0353
chr5:114543256-114764090	DEL	chr25:1665327-1808056	DEL	6.6832	2	0.0354
chr12:34505806-34662470	DEL	chr5:119221776-120378417	DEL	6.6776	2	0.0355
chr12:34505806-34662470	DEL	chr25:1665327-1808056	DEL	6.6691	2	0.0356
chr7:33722644-33868759	DEL	chr24:61455723-62320145	DEL	6.6613	2	0.0358
chr20:60639566-60783431	DEL	chr6:52628477-52828988	DEL	6.6485	2	0.036
chr5:114543256-114764090	DEL	chr14:1514056-2553525	DEL	6.649	2	0.036
chr12:34505806-34662470	DEL	chr14:1514056-2553525	DEL	6.6496	2	0.036
chr5:114543256-114764090	DEL	chr17:73118011-74998349	DEL	6.6457	2	0.0361
chr5:114543256-114764090	DEL	chr25:104438-1365841	DEL	6.6346	2	0.0363
chr1:48701612-48866528	DEL	chr1:104798012-105264358	DEL	6.6275	2	0.0364
chr5:114543256-114764090	DEL	chr3:118813014-119077206	DEL	6.628	2	0.0364
chr24:1282069-1582182	DEL	chr7:44658442-46025089	DEL	6.6143	2	0.0366
chr5:114543256-114764090	DEL	chr25:40940951-42768470	DEL	6.6154	2	0.0366
chr12:34505806-34662470	DEL	chr17:73118011-74998349	DEL	6.6023	2	0.0368
chr5:114543256-114764090	DEL	chr25:38171850-38377594	DEL	6.5942	2	0.037
chr29:48012818-48355723	DEL	chr19:56607168-57213764	DEL	6.585	2	0.0372
chr12:34505806-34662470	DEL	chr6:107678393-109951981	DEL	6.5743	2	0.0374
chr28:41674187-41737604	DEL	chr4:114326665-114640077	DEL	6.561	2	0.0376
chr3:117575562-118346051	DEL	chr25:39286957-40282215	DEL	6.5636	2	0.0376
chr26:23167656-23414945	DEL	chr5:119221776-120378417	DEL	6.5567	2	0.0377
chr3:85764516-85944553	DEL	chr5:14770370-14953446	DEL	6.5504	2	0.0378
chr25:36448529-36514994	DEL	chr6:106495683-107186270	DEL	6.553	2	0.0378
chr19:63424825-63734072	DEL	chr5:120553341-121175859	DEL	6.544	2	0.0379
chr24:1282069-1582182	DEL	chr4:117831202-120555019	DEL	6.5382	2	0.038
chr19:63424825-63734072	DEL	chr6:106495683-107186270	DEL	6.5419	2	0.038
chr7:4226753-4655753	DEL	chr18:11121144-11813752	DEL	6.534	2	0.0381
chr18:65819321-65978584	DEL	chr10:102887596-103470001	DEL	6.5356	2	0.0381
chr19:50336021-50447799	DEL	chr19:51395684-52234974	DEL	6.5356	2	0.0381
chr4:114326665-114640077	DEL	chr19:56607168-57213764	DEL	6.5367	2	0.0381
chr28:41674187-41737604	DEL	chr3:117575562-118346051	DEL	6.5377	2	0.0381
chr3:117575562-118346051	DEL	chr19:51395684-52234974	DEL	6.5277	2	0.0382
chr19:63424825-63734072	DEL	chr7:44658442-46025089	DEL	6.5293	2	0.0382
chr26:49532894-49762633	DEL	chr18:11121144-11813752	DEL	6.5298	2	0.0382
chr25:36448529-36514994	DEL	chr7:21462645-21677064	DEL	6.5298	2	0.0382
chr7:4226753-4655753	DEL	chr3:117575562-118346051	DEL	6.5309	2	0.0382
chr3:117575562-118346051	DEL	chr19:56607168-57213764	DEL	6.523	2	0.0383
chr19:50336021-50447799	DEL	chr24:61455723-62320145	DEL	6.5257	2	0.0383
chr25:1955733-2606575	DEL	chr19:51395684-52234974	DEL	6.5173	2	0.0384
chr25:38171850-38377594	DEL	chr26:49532894-49762633	DEL	6.5184	2	0.0384
chr7:44658442-46025089	DEL	chr19:56607168-57213764	DEL	6.5194	2	0.0384
chr18:65819321-65978584	DEL	chr18:11121144-11813752	DEL	6.5204	2	0.0384
chr19:50336021-50447799	DEL	chr5:119221776-120378417	DEL	6.5215	2	0.0384

CNVR_LocA	CN_A	CNVR_LocB	CN_B	Chi2	Df	P-Value
chr24:1282069-1582182	DEL	chr4:114326665-114640077	DEL	6.5215	2	0.0384
chr3:118813014-119077206	DEL	chr7:44658442-46025089	DEL	6.5126	2	0.0385
chr16:47654206-47780813	DEL	chr19:63424825-63734072	DEL	6.5132	2	0.0385
chr29:48012818-48355723	DEL	chr24:61455723-62320145	DEL	6.5137	2	0.0385
chr25:36448529-36514994	DEL	chr7:4226753-4655753	DEL	6.5142	2	0.0385
chr18:65819321-65978584	DEL	chr4:113079474-113532717	DEL	6.5147	2	0.0385
chr2:133816808-134465054	DEL	chr10:102887596-103470001	DEL	6.5075	2	0.0386
chr16:47654206-47780813	DEL	chr7:4226753-4655753	DEL	6.508	2	0.0386
chr28:41674187-41737604	DEL	chr29:48012818-48355723	DEL	6.509	2	0.0386
chr13:54496419-54829615	DEL	chr18:11121144-11813752	DEL	6.5095	2	0.0386
chr3:117575562-118346051	DEL	chr24:61455723-62320145	DEL	6.5106	2	0.0386
chr25:36448529-36514994	DEL	chr4:114326665-114640077	DEL	6.5106	2	0.0386
chr26:49532894-49762633	DEL	chr2:133816808-134465054	DEL	6.5111	2	0.0386
chr29:48012818-48355723	DEL	chr25:1955733-2606575	DEL	6.5116	2	0.0386
chr18:65819321-65978584	DEL	chr7:44658442-46025089	DEL	6.5116	2	0.0386
chr25:1955733-2606575	DEL	chr19:56607168-57213764	DEL	6.5018	2	0.0387
chr19:50336021-50447799	DEL	chr18:65819321-65978584	DEL	6.5018	2	0.0387
chr24:61455723-62320145	DEL	chr5:120553341-121175859	DEL	6.5028	2	0.0387
chr18:65819321-65978584	DEL	chr26:23167656-23414945	DEL	6.5038	2	0.0387
chr26:23167656-23414945	DEL	chr19:51395684-52234974	DEL	6.5054	2	0.0387
chr16:47654206-47780813	DEL	chr4:117831202-120555019	DEL	6.5059	2	0.0387
chr29:48012818-48355723	DEL	chr4:114326665-114640077	DEL	6.5064	2	0.0387
chr19:63424825-63734072	DEL	chr26:49532894-49762633	DEL	6.5064	2	0.0387
chr2:136386853-136531159	DEL	chr29:48012818-48355723	DEL	6.4966	2	0.0388
chr18:11121144-11813752	DEL	chr19:51395684-52234974	DEL	6.4971	2	0.0388
chr29:42897144-43269744	DEL	chr5:119221776-120378417	DEL	6.4997	2	0.0388
chr25:39286957-40282215	DEL	chr25:1955733-2606575	DEL	6.5002	2	0.0388
chr7:44658442-46025089	DEL	chr5:119221776-120378417	DEL	6.5007	2	0.0388
chr29:42897144-43269744	DEL	chr13:54496419-54829615	DEL	6.5007	2	0.0388
chr16:47654206-47780813	DEL	chr10:102887596-103470001	DEL	6.5007	2	0.0388
chr18:65819321-65978584	DEL	chr6:106495683-107186270	DEL	6.5012	2	0.0388
chr28:41674187-41737604	DEL	chr3:118813014-119077206	DEL	6.5012	2	0.0388
chr25:39286957-40282215	DEL	chr18:11121144-11813752	DEL	6.4915	2	0.0389
chr7:44658442-46025089	DEL	chr2:133816808-134465054	DEL	6.492	2	0.0389
chr7:44658442-46025089	DEL	chr24:61455723-62320145	DEL	6.4925	2	0.0389
chr13:54496419-54829615	DEL	chr9:103383683-105462864	DEL	6.4925	2	0.0389
chr25:36448529-36514994	DEL	chr25:1955733-2606575	DEL	6.493	2	0.0389
chr9:103383683-105462864	DEL	chr10:102887596-103470001	DEL	6.4935	2	0.0389
chr13:54496419-54829615	DEL	chr19:51395684-52234974	DEL	6.4935	2	0.0389
chr7:4226753-4655753	DEL	chr25:1955733-2606575	DEL	6.4935	2	0.0389
chr3:117575562-118346051	DEL	chr5:119221776-120378417	DEL	6.494	2	0.0389
chr19:63424825-63734072	DEL	chr2:133816808-134465054	DEL	6.494	2	0.0389
chr19:50336021-50447799	DEL	chr25:38171850-38377594	DEL	6.494	2	0.0389
chr2:133816808-134465054	DEL	chr4:114326665-114640077	DEL	6.4946	2	0.0389
chr26:49532894-49762633	DEL	chr5:120553341-121175859	DEL	6.4946	2	0.0389
chr26:23167656-23414945	DEL	chr13:54496419-54829615	DEL	6.4946	2	0.0389
chr28:41674187-41737604	DEL	chr26:49532894-49762633	DEL	6.4946	2	0.0389
chr26:49532894-49762633	DEL	chr6:106495683-107186270	DEL	6.4951	2	0.0389
chr3:117575562-118346051	DEL	chr29:48012818-48355723	DEL	6.4961	2	0.0389
chr25:36448529-36514994	DEL	chr3:118813014-119077206	DEL	6.4961	2	0.0389
chr3:117575562-118346051	DEL	chr26:49532894-49762633	DEL	6.4868	2	0.039
chr2:136386853-136531159	DEL	chr26:49532894-49762633	DEL	6.4868	2	0.039
chr19:50336021-50447799	DEL	chr26:49532894-49762633	DEL	6.4868	2	0.039

CNVR_LocA	CN_A	CNVR_LocB	CN_B	Chi2	Df	P-Value
chr25:39286957-40282215	DEL	chr9:103383683-105462864	DEL	6.4874	2	0.039
chr4:113079474-113532717	DEL	chr3:118813014-119077206	DEL	6.4874	2	0.039
chr24:1282069-1582182	DEL	chr26:49532894-49762633	DEL	6.4874	2	0.039
chr24:1282069-1582182	DEL	chr29:48012818-48355723	DEL	6.4879	2	0.039
chr2:136386853-136531159	DEL	chr9:103383683-105462864	DEL	6.4884	2	0.039
chr19:50336021-50447799	DEL	chr6:106495683-107186270	DEL	6.4884	2	0.039
chr25:38171850-38377594	DEL	chr9:103383683-105462864	DEL	6.4889	2	0.039
chr26:23167656-23414945	DEL	chr7:44658442-46025089	DEL	6.4899	2	0.039
chr19:50336021-50447799	DEL	chr7:4226753-4655753	DEL	6.4899	2	0.039
chr28:41674187-41737604	DEL	chr7:21462645-21677064	DEL	6.4904	2	0.039
chr3:117575562-118346051	DEL	chr6:106495683-107186270	DEL	6.4812	2	0.0391
chr26:23167656-23414945	DEL	chr4:114326665-114640077	DEL	6.4812	2	0.0391
chr29:42897144-43269744	DEL	chr2:136386853-136531159	DEL	6.4817	2	0.0391
chr7:44658442-46025089	DEL	chr6:106495683-107186270	DEL	6.4828	2	0.0391
chr7:21462645-21677064	DEL	chr5:119221776-120378417	DEL	6.4828	2	0.0391
chr4:117831202-120555019	DEL	chr2:133816808-134465054	DEL	6.4833	2	0.0391
chr25:39286957-40282215	DEL	chr5:119221776-120378417	DEL	6.4833	2	0.0391
chr13:54496419-54829615	DEL	chr5:119221776-120378417	DEL	6.4838	2	0.0391
chr3:118813014-119077206	DEL	chr6:106495683-107186270	DEL	6.4838	2	0.0391
chr10:102887596-103470001	DEL	chr19:56607168-57213764	DEL	6.4848	2	0.0391
chr29:42897144-43269744	DEL	chr25:39286957-40282215	DEL	6.4848	2	0.0391
chr7:4226753-4655753	DEL	chr25:38171850-38377594	DEL	6.4858	2	0.0391
chr29:42897144-43269744	DEL	chr25:38171850-38377594	DEL	6.4858	2	0.0391
chr3:118813014-119077206	DEL	chr5:120553341-121175859	DEL	6.4761	2	0.0392
chr2:133816808-134465054	DEL	chr5:119221776-120378417	DEL	6.4766	2	0.0392
chr25:36448529-36514994	DEL	chr5:119221776-120378417	DEL	6.4766	2	0.0392
chr28:41674187-41737604	DEL	chr16:47654206-47780813	DEL	6.4766	2	0.0392
chr25:38171850-38377594	DEL	chr4:114326665-114640077	DEL	6.4771	2	0.0392
chr3:117575562-118346051	DEL	chr5:120553341-121175859	DEL	6.4776	2	0.0392
chr13:54496419-54829615	DEL	chr29:48012818-48355723	DEL	6.4776	2	0.0392
chr4:113079474-113532717	DEL	chr9:103383683-105462864	DEL	6.4776	2	0.0392
chr24:1282069-1582182	DEL	chr18:11121144-11813752	DEL	6.4776	2	0.0392
chr7:4226753-4655753	DEL	chr5:120553341-121175859	DEL	6.4782	2	0.0392
chr4:114326665-114640077	DEL	chr10:102887596-103470001	DEL	6.4787	2	0.0392
chr9:103383683-105462864	DEL	chr25:1955733-2606575	DEL	6.4787	2	0.0392
chr18:11121144-11813752	DEL	chr9:103383683-105462864	DEL	6.4787	2	0.0392
chr19:63424825-63734072	DEL	chr4:114326665-114640077	DEL	6.4787	2	0.0392
chr28:41674187-41737604	DEL	chr19:56607168-57213764	DEL	6.4787	2	0.0392
chr13:54496419-54829615	DEL	chr10:102887596-103470001	DEL	6.4792	2	0.0392
chr4:113079474-113532717	DEL	chr4:114326665-114640077	DEL	6.4792	2	0.0392
chr4:113079474-113532717	DEL	chr2:133816808-134465054	DEL	6.4792	2	0.0392
chr3:118813014-119077206	DEL	chr9:103383683-105462864	DEL	6.4797	2	0.0392
chr19:50336021-50447799	DEL	chr3:118813014-119077206	DEL	6.4797	2	0.0392
chr24:61455723-62320145	DEL	chr2:133816808-134465054	DEL	6.4807	2	0.0392
chr7:44658442-46025089	DEL	chr5:120553341-121175859	DEL	6.4807	2	0.0392
chr3:118813014-119077206	DEL	chr25:1955733-2606575	DEL	6.471	2	0.0393
chr7:21462645-21677064	DEL	chr19:51395684-52234974	DEL	6.472	2	0.0393
chr4:113079474-113532717	DEL	chr10:102887596-103470001	DEL	6.472	2	0.0393
chr26:23167656-23414945	DEL	chr29:48012818-48355723	DEL	6.4731	2	0.0393
chr18:65819321-65978584	DEL	chr25:1955733-2606575	DEL	6.4731	2	0.0393
chr16:47654206-47780813	DEL	chr19:50336021-50447799	DEL	6.4731	2	0.0393
chr7:4226753-4655753	DEL	chr26:23167656-23414945	DEL	6.4736	2	0.0393
chr24:1282069-1582182	DEL	chr24:61455723-62320145	DEL	6.4736	2	0.0393

CNVR_LocA	CN_A	CNVR_LocB	CN_B	Chi2	Df	P-Value
chr24:1282069-1582182	DEL	chr25:38171850-38377594	DEL	6.4736	2	0.0393
chr16:47654206-47780813	DEL	chr5:119221776-120378417	DEL	6.4741	2	0.0393
chr28:41674187-41737604	DEL	chr25:1955733-2606575	DEL	6.4741	2	0.0393
chr24:1282069-1582182	DEL	chr6:106495683-107186270	DEL	6.4746	2	0.0393
chr26:23167656-23414945	DEL	chr6:106495683-107186270	DEL	6.4659	2	0.0394
chr25:36448529-36514994	DEL	chr3:117575562-118346051	DEL	6.467	2	0.0394
chr4:114326665-114640077	DEL	chr5:119221776-120378417	DEL	6.4675	2	0.0394
chr26:23167656-23414945	DEL	chr25:39286957-40282215	DEL	6.4685	2	0.0394
chr7:4226753-4655753	DEL	chr25:39286957-40282215	DEL	6.4685	2	0.0394
chr18:65819321-65978584	DEL	chr13:54496419-54829615	DEL	6.4685	2	0.0394
chr18:11121144-11813752	DEL	chr6:106495683-107186270	DEL	6.47	2	0.0394
chr26:49532894-49762633	DEL	chr5:119221776-120378417	DEL	6.47	2	0.0394
chr7:4226753-4655753	DEL	chr7:44658442-46025089	DEL	6.47	2	0.0394
chr26:49532894-49762633	DEL	chr19:51395684-52234974	DEL	6.4705	2	0.0394
chr29:44969518-45023665	DEL	chr10:14129663-14288475	DUP	6.4665	2	0.0394
chr5:120553341-121175859	DEL	chr5:119221776-120378417	DEL	6.4609	2	0.0395
chr26:49532894-49762633	DEL	chr25:39286957-40282215	DEL	6.4609	2	0.0395
chr26:23167656-23414945	DEL	chr9:103383683-105462864	DEL	6.4614	2	0.0395
chr18:65819321-65978584	DEL	chr26:49532894-49762633	DEL	6.4619	2	0.0395
chr19:50336021-50447799	DEL	chr5:120553341-121175859	DEL	6.4619	2	0.0395
chr18:65819321-65978584	DEL	chr9:103383683-105462864	DEL	6.4624	2	0.0395
chr29:42897144-43269744	DEL	chr26:49532894-49762633	DEL	6.4624	2	0.0395
chr29:42897144-43269744	DEL	chr19:50336021-50447799	DEL	6.4624	2	0.0395
chr25:36448529-36514994	DEL	chr2:136386853-136531159	DEL	6.4624	2	0.0395
chr29:42897144-43269744	DEL	chr24:61455723-62320145	DEL	6.4629	2	0.0395
chr6:106495683-107186270	DEL	chr10:102887596-103470001	DEL	6.4634	2	0.0395
chr7:4226753-4655753	DEL	chr19:56607168-57213764	DEL	6.4639	2	0.0395
chr29:48012818-48355723	DEL	chr2:133816808-134465054	DEL	6.4644	2	0.0395
chr7:21462645-21677064	DEL	chr2:133816808-134465054	DEL	6.4654	2	0.0395
chr26:23167656-23414945	DEL	chr18:11121144-11813752	DEL	6.4654	2	0.0395
chr4:114326665-114640077	DUP	chr26:50817833-51680135	DUP	6.4609	2	0.0395
chr18:11121144-11813752	DEL	chr7:44658442-46025089	DEL	6.4558	2	0.0396
chr2:136386853-136531159	DEL	chr4:117831202-120555019	DEL	6.4558	2	0.0396
chr25:38171850-38377594	DEL	chr2:136386853-136531159	DEL	6.4558	2	0.0396
chr16:47654206-47780813	DEL	chr4:114326665-114640077	DEL	6.4563	2	0.0396
chr7:4226753-4655753	DEL	chr9:103383683-105462864	DEL	6.4568	2	0.0396
chr4:113079474-113532717	DEL	chr18:11121144-11813752	DEL	6.4573	2	0.0396
chr9:103383683-105462864	DEL	chr2:133816808-134465054	DEL	6.4579	2	0.0396
chr3:118813014-119077206	DEL	chr19:56607168-57213764	DEL	6.4579	2	0.0396
chr7:4226753-4655753	DEL	chr4:114326665-114640077	DEL	6.4584	2	0.0396
chr18:65819321-65978584	DEL	chr7:4226753-4655753	DEL	6.4584	2	0.0396
chr13:54496419-54829615	DEL	chr25:39286957-40282215	DEL	6.4589	2	0.0396
chr25:38171850-38377594	DEL	chr6:106495683-107186270	DEL	6.4589	2	0.0396
chr24:1282069-1582182	DEL	chr2:133816808-134465054	DEL	6.4589	2	0.0396
chr18:11121144-11813752	DEL	chr24:61455723-62320145	DEL	6.4594	2	0.0396
chr25:38171850-38377594	DEL	chr5:120553341-121175859	DEL	6.4594	2	0.0396
chr4:113079474-113532717	DEL	chr24:61455723-62320145	DEL	6.4594	2	0.0396
chr25:36448529-36514994	DEL	chr29:42897144-43269744	DEL	6.4594	2	0.0396
chr28:41674187-41737604	DEL	chr24:61455723-62320145	DEL	6.4594	2	0.0396
chr26:49532894-49762633	DEL	chr4:117831202-120555019	DEL	6.4599	2	0.0396
chr4:113079474-113532717	DEL	chr5:120553341-121175859	DEL	6.4599	2	0.0396
chr28:41674187-41737604	DEL	chr25:36448529-36514994	DEL	6.4604	2	0.0396
chr26:50817833-51680135	DUP	chr18:62375495-63727709	DUP	6.4563	2	0.0396

CNVR_LocA	CN_A	CNVR_LocB	CN_B	Chi2	Df	P-Value
chr4:113079474-113532717	DEL	chr19:56607168-57213764	DEL	6.4508	2	0.0397
chr3:118813014-119077206	DEL	chr10:102887596-103470001	DEL	6.4513	2	0.0397
chr7:21462645-21677064	DEL	chr26:49532894-49762633	DEL	6.4518	2	0.0397
chr5:120553341-121175859	DEL	chr19:56607168-57213764	DEL	6.4523	2	0.0397
chr9:103383683-105462864	DEL	chr19:51395684-52234974	DEL	6.4523	2	0.0397
chr16:47654206-47780813	DEL	chr7:21462645-21677064	DEL	6.4523	2	0.0397
chr25:36448529-36514994	DEL	chr24:1282069-1582182	DEL	6.4523	2	0.0397
chr7:44658442-46025089	DEL	chr25:1955733-2606575	DEL	6.4528	2	0.0397
chr29:42897144-43269744	DEL	chr19:51395684-52234974	DEL	6.4538	2	0.0397
chr4:117831202-120555019	DEL	chr5:119221776-120378417	DEL	6.4548	2	0.0397
chr3:117575562-118346051	DEL	chr2:133816808-134465054	DEL	6.4548	2	0.0397
chr4:113079474-113532717	DEL	chr25:39286957-40282215	DEL	6.4548	2	0.0397
chr13:54496419-54829615	DEL	chr4:114326665-114640077	DEL	6.4458	2	0.0398
chr19:63424825-63734072	DEL	chr25:39286957-40282215	DEL	6.4458	2	0.0398
chr9:103383683-105462864	DEL	chr5:120553341-121175859	DEL	6.4463	2	0.0398
chr7:4226753-4655753	DEL	chr5:119221776-120378417	DEL	6.4463	2	0.0398
chr7:4226753-4655753	DEL	chr2:133816808-134465054	DEL	6.4463	2	0.0398
chr24:1282069-1582182	DEL	chr7:21462645-21677064	DEL	6.4463	2	0.0398
chr25:36448529-36514994	DEL	chr7:44658442-46025089	DEL	6.4463	2	0.0398
chr2:136386853-136531159	DEL	chr2:133816808-134465054	DEL	6.4468	2	0.0398
chr16:47654206-47780813	DEL	chr18:65819321-65978584	DEL	6.4468	2	0.0398
chr25:36448529-36514994	DEL	chr19:63424825-63734072	DEL	6.4468	2	0.0398
chr28:41674187-41737604	DEL	chr2:133816808-134465054	DEL	6.4468	2	0.0398
chr18:11121144-11813752	DEL	chr25:1955733-2606575	DEL	6.4473	2	0.0398
chr18:65819321-65978584	DEL	chr19:51395684-52234974	DEL	6.4473	2	0.0398
chr28:41674187-41737604	DEL	chr9:103383683-105462864	DEL	6.4473	2	0.0398
chr16:47654206-47780813	DEL	chr9:103383683-105462864	DEL	6.4478	2	0.0398
chr26:49532894-49762633	DEL	chr9:103383683-105462864	DEL	6.4483	2	0.0398
chr19:63424825-63734072	DEL	chr3:117575562-118346051	DEL	6.4483	2	0.0398
chr26:23167656-23414945	DEL	chr26:49532894-49762633	DEL	6.4483	2	0.0398
chr4:117831202-120555019	DEL	chr19:56607168-57213764	DEL	6.4488	2	0.0398
chr25:39286957-40282215	DEL	chr10:102887596-103470001	DEL	6.4488	2	0.0398
chr2:136386853-136531159	DEL	chr4:114326665-114640077	DEL	6.4488	2	0.0398
chr4:117831202-120555019	DEL	chr10:102887596-103470001	DEL	6.4493	2	0.0398
chr7:21462645-21677064	DEL	chr24:61455723-62320145	DEL	6.4493	2	0.0398
chr16:47654206-47780813	DEL	chr7:44658442-46025089	DEL	6.4493	2	0.0398
chr6:106495683-107186270	DEL	chr4:117831202-120555019	DEL	6.4498	2	0.0398
chr9:103383683-105462864	DEL	chr6:106495683-107186270	DEL	6.4498	2	0.0398
chr5:120553341-121175859	DEL	chr10:102887596-103470001	DEL	6.4503	2	0.0398
chr19:50336021-50447799	DEL	chr9:103383683-105462864	DEL	6.4408	2	0.0399
chr28:41674187-41737604	DEL	chr19:63424825-63734072	DEL	6.4408	2	0.0399
chr26:49532894-49762633	DEL	chr10:102887596-103470001	DEL	6.4413	2	0.0399
chr3:117575562-118346051	DEL	chr18:11121144-11813752	DEL	6.4413	2	0.0399
chr19:63424825-63734072	DEL	chr19:56607168-57213764	DEL	6.4413	2	0.0399
chr28:41674187-41737604	DEL	chr29:42897144-43269744	DEL	6.4413	2	0.0399
chr7:21462645-21677064	DEL	chr19:56607168-57213764	DEL	6.4418	2	0.0399
chr7:21462645-21677064	DEL	chr4:114326665-114640077	DEL	6.4418	2	0.0399
chr25:36448529-36514994	DEL	chr18:65819321-65978584	DEL	6.4418	2	0.0399
chr25:39286957-40282215	DEL	chr6:106495683-107186270	DEL	6.4423	2	0.0399
chr19:63424825-63734072	DEL	chr4:117831202-120555019	DEL	6.4423	2	0.0399
chr26:23167656-23414945	DEL	chr4:117831202-120555019	DEL	6.4423	2	0.0399
chr19:63424825-63734072	DEL	chr19:51395684-52234974	DEL	6.4428	2	0.0399
chr19:50336021-50447799	DEL	chr4:114326665-114640077	DEL	6.4428	2	0.0399

CNVR_LocA	CN_A	CNVR_LocB	CN_B	Chi2	Df	P-Value
chr24:1282069-1582182	DEL	chr13:54496419-54829615	DEL	6.4428	2	0.0399
chr9:103383683-105462864	DEL	chr4:114326665-114640077	DEL	6.4433	2	0.0399
chr19:63424825-63734072	DEL	chr18:11121144-11813752	DEL	6.4433	2	0.0399
chr24:61455723-62320145	DEL	chr10:102887596-103470001	DEL	6.4438	2	0.0399
chr26:49532894-49762633	DEL	chr24:61455723-62320145	DEL	6.4438	2	0.0399
chr25:39286957-40282215	DEL	chr29:48012818-48355723	DEL	6.4443	2	0.0399
chr26:49532894-49762633	DEL	chr7:44658442-46025089	DEL	6.4443	2	0.0399
chr3:118813014-119077206	DEL	chr13:54496419-54829615	DEL	6.4443	2	0.0399
chr25:36448529-36514994	DEL	chr25:38171850-38377594	DEL	6.4448	2	0.0399
chr4:117831202-120555019	DEL	chr24:61455723-62320145	DEL	6.4453	2	0.0399
chr6:106495683-107186270	DEL	chr19:51395684-52234974	DEL	6.4453	2	0.0399
chr6:106495683-107186270	DEL	chr2:133816808-134465054	DEL	6.4453	2	0.0399
chr25:39286957-40282215	DEL	chr19:51395684-52234974	DEL	6.4453	2	0.0399
chr9:102574022-103180423	DUP	chr3:120122176-121403393	DUP	6.4428	2	0.0399
chr18:11121144-11813752	DEL	chr19:56607168-57213764	DEL	6.4358	2	0.04
chr7:4226753-4655753	DEL	chr29:48012818-48355723	DEL	6.4358	2	0.04
chr7:4226753-4655753	DEL	chr19:63424825-63734072	DEL	6.4358	2	0.04
chr25:39286957-40282215	DEL	chr4:117831202-120555019	DEL	6.4363	2	0.04
chr3:118813014-119077206	DEL	chr3:117575562-118346051	DEL	6.4363	2	0.04
chr4:113079474-113532717	DEL	chr3:117575562-118346051	DEL	6.4363	2	0.04
chr29:42897144-43269744	DEL	chr4:117831202-120555019	DEL	6.4363	2	0.04
chr18:11121144-11813752	DEL	chr19:56607168-57213764	DEL	6.4358	2	0.04
chr7:4226753-4655753	DEL	chr29:48012818-48355723	DEL	6.4358	2	0.04
chr7:4226753-4655753	DEL	chr19:63424825-63734072	DEL	6.4358	2	0.04
chr25:39286957-40282215	DEL	chr4:117831202-120555019	DEL	6.4363	2	0.04
chr3:118813014-119077206	DEL	chr3:117575562-118346051	DEL	6.4363	2	0.04
chr4:113079474-113532717	DEL	chr3:117575562-118346051	DEL	6.4363	2	0.04
chr29:42897144-43269744	DEL	chr4:117831202-120555019	DEL	6.4363	2	0.04
chr19:51395684-52234974	DEL	chr5:119221776-120378417	DEL	6.4368	2	0.04
chr25:36448529-36514994	DEL	chr4:113079474-113532717	DEL	6.4368	2	0.04
chr19:51395684-52234974	DEL	chr4:114326665-114640077	DEL	6.4373	2	0.04
chr4:117831202-120555019	DEL	chr5:120553341-121175859	DEL	6.4373	2	0.04
chr3:118813014-119077206	DEL	chr4:114326665-114640077	DEL	6.4378	2	0.04
chr25:1955733-2606575	DEL	chr10:102887596-103470001	DEL	6.4383	2	0.04
chr19:50336021-50447799	DEL	chr25:1955733-2606575	DEL	6.4383	2	0.04
chr25:36448529-36514994	DEL	chr9:103383683-105462864	DEL	6.4383	2	0.04
chr6:106495683-107186270	DEL	chr19:56607168-57213764	DEL	6.4388	2	0.04
chr3:117575562-118346051	DEL	chr4:114326665-114640077	DEL	6.4388	2	0.04
chr7:4226753-4655753	DEL	chr7:21462645-21677064	DEL	6.4388	2	0.04
chr18:65819321-65978584	DEL	chr2:133816808-134465054	DEL	6.4388	2	0.04
chr29:42897144-43269744	DEL	chr6:106495683-107186270	DEL	6.4388	2	0.04
chr28:41674187-41737604	DEL	chr18:11121144-11813752	DEL	6.4388	2	0.04
chr13:54496419-54829615	DEL	chr5:120553341-121175859	DEL	6.4393	2	0.04
chr3:118813014-119077206	DEL	chr2:133816808-134465054	DEL	6.4398	2	0.04
chr9:103383683-105462864	DEL	chr7:44658442-46025089	DEL	6.4403	2	0.04
chr3:118813014-119077206	DEL	chr18:11121144-11813752	DEL	6.4403	2	0.04
chr16:47654206-47780813	DEL	chr2:136386853-136531159	DEL	6.4403	2	0.04
chr18:11121144-11813752	DUP	chr26:42933219-43087057	DUP	6.4368	2	0.04
chr16:47654206-47780813	DEL	chr24:1282069-1582182	DEL	6.4308	2	0.0401
chr16:47654206-47780813	DEL	chr19:51395684-52234974	DEL	6.4313	2	0.0401
chr18:65819321-65978584	DEL	chr2:136386853-136531159	DEL	6.4318	2	0.0401
chr13:54496419-54829615	DEL	chr3:117575562-118346051	DEL	6.4328	2	0.0401
chr7:4226753-4655753	DEL	chr10:102887596-103470001	DEL	6.4328	2	0.0401

CNVR_LocA	CN_A	CNVR_LocB	CN_B	Chi2	Df	P-Value
chr7:21462645-21677064	DEL	chr6:106495683-107186270	DEL	6.4333	2	0.0401
chr19:50336021-50447799	DEL	chr4:113079474-113532717	DEL	6.4333	2	0.0401
chr26:23167656-23414945	DEL	chr2:133816808-134465054	DEL	6.4338	2	0.0401
chr26:23167656-23414945	DEL	chr7:21462645-21677064	DEL	6.4338	2	0.0401
chr3:118813014-119077206	DEL	chr2:136386853-136531159	DEL	6.4338	2	0.0401
chr24:1282069-1582182	DEL	chr19:63424825-63734072	DEL	6.4338	2	0.0401
chr25:36448529-36514994	DEL	chr5:120553341-121175859	DEL	6.4338	2	0.0401
chr25:36448529-36514994	DEL	chr19:50336021-50447799	DEL	6.4343	2	0.0401
chr3:117575562-118346051	DEL	chr4:117831202-120555019	DEL	6.4348	2	0.0401
chr13:54496419-54829615	DEL	chr25:1955733-2606575	DEL	6.4348	2	0.0401
chr19:63424825-63734072	DEL	chr7:21462645-21677064	DEL	6.4348	2	0.0401
chr4:113079474-113532717	DEL	chr25:38171850-38377594	DEL	6.4348	2	0.0401
chr28:41674187-41737604	DEL	chr18:65819321-65978584	DEL	6.4348	2	0.0401
chr19:50336021-50447799	DEL	chr19:56607168-57213764	DEL	6.4353	2	0.0401
chr24:1282069-1582182	DEL	chr19:51395684-52234974	DEL	6.4353	2	0.0401
chr16:47654206-47780813	DEL	chr2:133816808-134465054	DEL	6.4353	2	0.0401
chr28:41674187-41737604	DEL	chr13:54496419-54829615	DEL	6.4353	2	0.0401
chr2:133816808-134465054	DEL	chr5:120553341-121175859	DEL	6.4258	2	0.0402
chr19:50336021-50447799	DEL	chr25:39286957-40282215	DEL	6.4258	2	0.0402
chr16:47654206-47780813	DEL	chr3:118813014-119077206	DEL	6.4258	2	0.0402
chr28:41674187-41737604	DEL	chr5:120553341-121175859	DEL	6.4258	2	0.0402
chr29:48012818-48355723	DEL	chr19:51395684-52234974	DEL	6.4263	2	0.0402
chr7:4226753-4655753	DEL	chr4:117831202-120555019	DEL	6.4263	2	0.0402
chr28:41674187-41737604	DEL	chr26:23167656-23414945	DEL	6.4263	2	0.0402
chr28:41674187-41737604	DEL	chr6:106495683-107186270	DEL	6.4268	2	0.0402
chr13:54496419-54829615	DEL	chr19:56607168-57213764	DEL	6.4273	2	0.0402
chr26:23167656-23414945	DEL	chr25:38171850-38377594	DEL	6.4273	2	0.0402
chr25:36448529-36514994	DEL	chr26:49532894-49762633	DEL	6.4273	2	0.0402
chr26:23167656-23414945	DEL	chr10:102887596-103470001	DEL	6.4278	2	0.0402
chr4:113079474-113532717	DEL	chr26:49532894-49762633	DEL	6.4278	2	0.0402
chr25:36448529-36514994	DEL	chr10:102887596-103470001	DEL	6.4278	2	0.0402
chr18:65819321-65978584	DEL	chr7:21462645-21677064	DEL	6.4283	2	0.0402
chr18:11121144-11813752	DEL	chr5:120553341-121175859	DEL	6.4288	2	0.0402
chr19:50336021-50447799	DEL	chr18:11121144-11813752	DEL	6.4288	2	0.0402
chr26:49532894-49762633	DEL	chr19:56607168-57213764	DEL	6.4293	2	0.0402
chr2:136386853-136531159	DEL	chr19:63424825-63734072	DEL	6.4293	2	0.0402
chr4:113079474-113532717	DEL	chr6:106495683-107186270	DEL	6.4293	2	0.0402
chr19:50336021-50447799	DEL	chr10:102887596-103470001	DEL	6.4293	2	0.0402
chr24:1282069-1582182	DEL	chr25:1955733-2606575	DEL	6.4293	2	0.0402
chr29:42897144-43269744	DEL	chr9:103383683-105462864	DEL	6.4293	2	0.0402
chr10:102887596-103470001	DEL	chr5:119221776-120378417	DEL	6.4298	2	0.0402
chr25:39286957-40282215	DEL	chr2:133816808-134465054	DEL	6.4298	2	0.0402
chr2:136386853-136531159	DEL	chr5:119221776-120378417	DEL	6.4298	2	0.0402
chr3:118813014-119077206	DEL	chr26:23167656-23414945	DEL	6.4303	2	0.0402
chr29:42897144-43269744	DEL	chr7:4226753-4655753	DEL	6.4218	2	0.0403
chr25:39286957-40282215	DEL	chr7:44658442-46025089	DEL	6.4223	2	0.0403
chr19:63424825-63734072	DEL	chr10:102887596-103470001	DEL	6.4228	2	0.0403
chr24:1282069-1582182	DEL	chr4:113079474-113532717	DEL	6.4228	2	0.0403
chr7:44658442-46025089	DEL	chr4:117831202-120555019	DEL	6.4238	2	0.0403
chr19:63424825-63734072	DEL	chr13:54496419-54829615	DEL	6.4238	2	0.0403
chr24:1282069-1582182	DEL	chr25:39286957-40282215	DEL	6.4238	2	0.0403
chr25:1955733-2606575	DEL	chr4:114326665-114640077	DEL	6.4243	2	0.0403
chr6:106495683-107186270	DEL	chr25:1955733-2606575	DEL	6.4243	2	0.0403

CNVR_LocA	CN_A	CNVR_LocB	CN_B	Chi2	Df	P-Value
chr18:11121144-11813752	DEL	chr4:114326665-114640077	DEL	6.4243	2	0.0403
chr2:136386853-136531159	DEL	chr24:61455723-62320145	DEL	6.4243	2	0.0403
chr24:1282069-1582182	DEL	chr5:120553341-121175859	DEL	6.4243	2	0.0403
chr29:42897144-43269744	DEL	chr4:114326665-114640077	DEL	6.4248	2	0.0403
chr29:42897144-43269744	DEL	chr19:63424825-63734072	DEL	6.4253	2	0.0403
chr29:42897144-43269744	DEL	chr3:118813014-119077206	DEL	6.4253	2	0.0403
chr4:117831202-120555019	DEL	chr19:51395684-52234974	DEL	6.4159	2	0.0404
chr25:36448529-36514994	DEL	chr2:133816808-134465054	DEL	6.4159	2	0.0404
chr9:103383683-105462864	DEL	chr5:119221776-120378417	DEL	6.4164	2	0.0404
chr19:63424825-63734072	DEL	chr24:61455723-62320145	DEL	6.4164	2	0.0404
chr25:38171850-38377594	DEL	chr19:63424825-63734072	DEL	6.4164	2	0.0404
chr19:50336021-50447799	DEL	chr19:63424825-63734072	DEL	6.4164	2	0.0404
chr2:136386853-136531159	DEL	chr25:39286957-40282215	DEL	6.4169	2	0.0404
chr4:113079474-113532717	DEL	chr19:63424825-63734072	DEL	6.4169	2	0.0404
chr4:117831202-120555019	DEL	chr29:48012818-48355723	DEL	6.4174	2	0.0404
chr7:4226753-4655753	DEL	chr3:118813014-119077206	DEL	6.4174	2	0.0404
chr9:103383683-105462864	DEL	chr29:48012818-48355723	DEL	6.4183	2	0.0404
chr7:21462645-21677064	DEL	chr9:103383683-105462864	DEL	6.4183	2	0.0404
chr25:38171850-38377594	DEL	chr10:102887596-103470001	DEL	6.4183	2	0.0404
chr26:49532894-49762633	DEL	chr4:114326665-114640077	DEL	6.4188	2	0.0404
chr26:49532894-49762633	DEL	chr25:1955733-2606575	DEL	6.4188	2	0.0404
chr26:23167656-23414945	DEL	chr24:61455723-62320145	DEL	6.4188	2	0.0404
chr29:42897144-43269744	DEL	chr4:113079474-113532717	DEL	6.4188	2	0.0404
chr7:21462645-21677064	DEL	chr10:102887596-103470001	DEL	6.4193	2	0.0404
chr24:1282069-1582182	DEL	chr19:56607168-57213764	DEL	6.4193	2	0.0404
chr29:42897144-43269744	DEL	chr7:44658442-46025089	DEL	6.4193	2	0.0404
chr28:41674187-41737604	DEL	chr24:1282069-1582182	DEL	6.4193	2	0.0404
chr5:119221776-120378417	DEL	chr19:56607168-57213764	DEL	6.4198	2	0.0404
chr4:113079474-113532717	DEL	chr2:136386853-136531159	DEL	6.4198	2	0.0404
chr19:50336021-50447799	DEL	chr4:117831202-120555019	DEL	6.4198	2	0.0404
chr25:1955733-2606575	DEL	chr5:120553341-121175859	DEL	6.4203	2	0.0404
chr7:21462645-21677064	DEL	chr4:117831202-120555019	DEL	6.4203	2	0.0404
chr19:51395684-52234974	DEL	chr19:56607168-57213764	DEL	6.4109	2	0.0405
chr6:106495683-107186270	DEL	chr29:48012818-48355723	DEL	6.4114	2	0.0405
chr25:39286957-40282215	DEL	chr5:120553341-121175859	DEL	6.4114	2	0.0405
chr29:42897144-43269744	DEL	chr10:102887596-103470001	DEL	6.4114	2	0.0405
chr18:11121144-11813752	DEL	chr5:119221776-120378417	DEL	6.4119	2	0.0405
chr13:54496419-54829615	DEL	chr2:133816808-134465054	DEL	6.4119	2	0.0405
chr18:65819321-65978584	DEL	chr3:117575562-118346051	DEL	6.4119	2	0.0405
chr26:23167656-23414945	DEL	chr2:136386853-136531159	DEL	6.4124	2	0.0405
chr4:113079474-113532717	DEL	chr7:44658442-46025089	DEL	6.4124	2	0.0405
chr29:42897144-43269744	DEL	chr26:23167656-23414945	DEL	6.4124	2	0.0405
chr25:38171850-38377594	DEL	chr7:21462645-21677064	DEL	6.4129	2	0.0405
chr25:36448529-36514994	DEL	chr29:48012818-48355723	DEL	6.4129	2	0.0405
chr9:103383683-105462864	DEL	chr4:117831202-120555019	DEL	6.4134	2	0.0405
chr26:23167656-23414945	DEL	chr5:120553341-121175859	DEL	6.4139	2	0.0405
chr7:4226753-4655753	DEL	chr19:51395684-52234974	DEL	6.4139	2	0.0405
chr18:65819321-65978584	DEL	chr4:117831202-120555019	DEL	6.4139	2	0.0405
chr7:21462645-21677064	DEL	chr29:48012818-48355723	DEL	6.4144	2	0.0405
chr2:136386853-136531159	DEL	chr19:51395684-52234974	DEL	6.4144	2	0.0405
chr28:41674187-41737604	DEL	chr25:38171850-38377594	DEL	6.4144	2	0.0405
chr25:36448529-36514994	DEL	chr19:51395684-52234974	DEL	6.4149	2	0.0405
chr25:38171850-38377594	DEL	chr29:48012818-48355723	DEL	6.4154	2	0.0405

CNVR_LocA	CN_A	CNVR_LocB	CN_B	Chi2	Df	P-Value
chr25:38171850-38377594	DEL	chr18:11121144-11813752	DEL	6.4154	2	0.0405
chr19:50336021-50447799	DEL	chr7:21462645-21677064	DEL	6.4154	2	0.0405
chr25:1955733-2606575	DEL	chr24:61455723-62320145	DEL	6.4065	2	0.0406
chr26:49532894-49762633	DEL	chr29:48012818-48355723	DEL	6.4065	2	0.0406
chr4:113079474-113532717	DEL	chr4:117831202-120555019	DEL	6.4065	2	0.0406
chr7:4226753-4655753	DEL	chr4:113079474-113532717	DEL	6.4065	2	0.0406
chr25:36448529-36514994	DEL	chr16:47654206-47780813	DEL	6.4065	2	0.0406
chr7:4226753-4655753	DEL	chr2:136386853-136531159	DEL	6.407	2	0.0406
chr18:65819321-65978584	DEL	chr4:114326665-114640077	DEL	6.4075	2	0.0406
chr25:36448529-36514994	DEL	chr13:54496419-54829615	DEL	6.4075	2	0.0406
chr3:117575562-118346051	DEL	chr7:44658442-46025089	DEL	6.409	2	0.0406
chr2:136386853-136531159	DEL	chr5:120553341-121175859	DEL	6.409	2	0.0406
chr4:117831202-120555019	DEL	chr4:114326665-114640077	DEL	6.4095	2	0.0406
chr7:44658442-46025089	DEL	chr4:114326665-114640077	DEL	6.4095	2	0.0406
chr29:42897144-43269744	DEL	chr5:120553341-121175859	DEL	6.4095	2	0.0406
chr13:54496419-54829615	DEL	chr26:49532894-49762633	DEL	6.4099	2	0.0406
chr3:118813014-119077206	DEL	chr4:117831202-120555019	DEL	6.4099	2	0.0406
chr13:54496419-54829615	DEL	chr7:21462645-21677064	DEL	6.4104	2	0.0406
chr25:38171850-38377594	DEL	chr19:51395684-52234974	DEL	6.4104	2	0.0406
chr13:16901756-17025364	DEL	chr25:104438-1365841	DUP	6.4099	2	0.0406
chr7:21462645-21677064	DEL	chr25:39286957-40282215	DEL	6.4011	2	0.0407
chr13:54496419-54829615	DEL	chr4:117831202-120555019	DEL	6.4011	2	0.0407
chr29:42897144-43269744	DEL	chr24:1282069-1582182	DEL	6.4011	2	0.0407
chr25:36448529-36514994	DEL	chr26:23167656-23414945	DEL	6.4011	2	0.0407
chr2:136386853-136531159	DEL	chr3:117575562-118346051	DEL	6.4021	2	0.0407
chr3:117575562-118346051	DEL	chr25:1955733-2606575	DEL	6.4031	2	0.0407
chr19:50336021-50447799	DEL	chr3:117575562-118346051	DEL	6.4031	2	0.0407
chr29:42897144-43269744	DEL	chr18:11121144-11813752	DEL	6.4031	2	0.0407
chr16:47654206-47780813	DEL	chr19:56607168-57213764	DEL	6.4031	2	0.0407
chr4:113079474-113532717	DEL	chr13:54496419-54829615	DEL	6.4035	2	0.0407
chr19:50336021-50447799	DEL	chr7:44658442-46025089	DEL	6.4035	2	0.0407
chr16:47654206-47780813	DEL	chr25:1955733-2606575	DEL	6.4035	2	0.0407
chr7:44658442-46025089	DEL	chr10:102887596-103470001	DEL	6.404	2	0.0407
chr9:103383683-105462864	DEL	chr19:56607168-57213764	DEL	6.404	2	0.0407
chr7:21462645-21677064	DEL	chr3:117575562-118346051	DEL	6.404	2	0.0407
chr18:65819321-65978584	DEL	chr5:120553341-121175859	DEL	6.404	2	0.0407
chr24:1282069-1582182	DEL	chr3:117575562-118346051	DEL	6.4045	2	0.0407
chr24:1282069-1582182	DEL	chr26:23167656-23414945	DEL	6.4045	2	0.0407
chr29:42897144-43269744	DEL	chr29:48012818-48355723	DEL	6.4045	2	0.0407
chr25:36448529-36514994	DEL	chr24:61455723-62320145	DEL	6.4045	2	0.0407
chr3:117575562-118346051	DEL	chr10:102887596-103470001	DEL	6.405	2	0.0407
chr25:38171850-38377594	DEL	chr19:56607168-57213764	DEL	6.405	2	0.0407
chr25:38171850-38377594	DEL	chr4:117831202-120555019	DEL	6.405	2	0.0407
chr26:23167656-23414945	DEL	chr19:56607168-57213764	DEL	6.405	2	0.0407
chr28:41674187-41737604	DEL	chr5:119221776-120378417	DEL	6.4055	2	0.0407
chr24:61455723-62320145	DEL	chr19:56607168-57213764	DEL	6.3967	2	0.0408
chr7:21462645-21677064	DEL	chr18:11121144-11813752	DEL	6.3967	2	0.0408
chr29:48012818-48355723	DEL	chr5:120553341-121175859	DEL	6.3972	2	0.0408
chr13:54496419-54829615	DEL	chr7:44658442-46025089	DEL	6.3972	2	0.0408
chr7:4226753-4655753	DEL	chr6:106495683-107186270	DEL	6.3972	2	0.0408
chr18:11121144-11813752	DEL	chr2:133816808-134465054	DEL	6.3977	2	0.0408
chr7:21462645-21677064	DEL	chr5:120553341-121175859	DEL	6.3977	2	0.0408
chr7:21462645-21677064	DEL	chr7:44658442-46025089	DEL	6.3977	2	0.0408

CNVR_LocA	CN_A	CNVR_LocB	CN_B	Chi2	Df	P-Value
chr3:118813014-119077206	DEL	chr7:21462645-21677064	DEL	6.3977	2	0.0408
chr18:65819321-65978584	DEL	chr25:39286957-40282215	DEL	6.3977	2	0.0408
chr29:42897144-43269744	DEL	chr19:56607168-57213764	DEL	6.3977	2	0.0408
chr29:42897144-43269744	DEL	chr7:21462645-21677064	DEL	6.3977	2	0.0408
chr6:106495683-107186270	DEL	chr4:114326665-114640077	DEL	6.3981	2	0.0408
chr7:4226753-4655753	DEL	chr26:49532894-49762633	DEL	6.3981	2	0.0408
chr3:118813014-119077206	DEL	chr19:51395684-52234974	DEL	6.3986	2	0.0408
chr24:1282069-1582182	DEL	chr5:119221776-120378417	DEL	6.3986	2	0.0408
chr16:47654206-47780813	DEL	chr26:49532894-49762633	DEL	6.3991	2	0.0408
chr25:1955733-2606575	DEL	chr2:133816808-134465054	DEL	6.3996	2	0.0408
chr25:38171850-38377594	DEL	chr25:1955733-2606575	DEL	6.3996	2	0.0408
chr3:118813014-119077206	DEL	chr19:63424825-63734072	DEL	6.3996	2	0.0408
chr4:113079474-113532717	DEL	chr26:23167656-23414945	DEL	6.3996	2	0.0408
chr18:65819321-65978584	DEL	chr25:38171850-38377594	DEL	6.4001	2	0.0408
chr18:65819321-65978584	DEL	chr19:63424825-63734072	DEL	6.4006	2	0.0408
chr29:42897144-43269744	DEL	chr2:133816808-134465054	DEL	6.4006	2	0.0408
chr16:47654206-47780813	DEL	chr5:120553341-121175859	DEL	6.4006	2	0.0408
chr2:136386853-136531159	DEL	chr25:1955733-2606575	DEL	6.3913	2	0.0409
chr4:113079474-113532717	DEL	chr7:21462645-21677064	DEL	6.3913	2	0.0409
chr25:38171850-38377594	DEL	chr3:117575562-118346051	DEL	6.3918	2	0.0409
chr19:51395684-52234974	DEL	chr5:120553341-121175859	DEL	6.3923	2	0.0409
chr25:36448529-36514994	DEL	chr4:117831202-120555019	DEL	6.3923	2	0.0409
chr29:48012818-48355723	DEL	chr10:102887596-103470001	DEL	6.3928	2	0.0409
chr19:63424825-63734072	DEL	chr25:1955733-2606575	DEL	6.3928	2	0.0409
chr28:41674187-41737604	DEL	chr4:113079474-113532717	DEL	6.3928	2	0.0409
chr16:47654206-47780813	DEL	chr24:61455723-62320145	DEL	6.3933	2	0.0409
chr5:120553341-121175859	DEL	chr4:114326665-114640077	DEL	6.3937	2	0.0409
chr18:65819321-65978584	DEL	chr24:61455723-62320145	DEL	6.3937	2	0.0409
chr25:38171850-38377594	DEL	chr7:44658442-46025089	DEL	6.3942	2	0.0409
chr26:23167656-23414945	DEL	chr25:1955733-2606575	DEL	6.3947	2	0.0409
chr25:36448529-36514994	DEL	chr19:56607168-57213764	DEL	6.3952	2	0.0409
chr7:44658442-46025089	DEL	chr29:48012818-48355723	DEL	6.3864	2	0.041
chr16:47654206-47780813	DEL	chr29:42897144-43269744	DEL	6.3864	2	0.041
chr13:54496419-54829615	DEL	chr24:61455723-62320145	DEL	6.3869	2	0.041
chr16:47654206-47780813	DEL	chr6:106495683-107186270	DEL	6.3869	2	0.041
chr19:63424825-63734072	DEL	chr5:119221776-120378417	DEL	6.3874	2	0.041
chr28:41674187-41737604	DEL	chr25:39286957-40282215	DEL	6.3874	2	0.041
chr3:118813014-119077206	DEL	chr29:48012818-48355723	DEL	6.3879	2	0.041
chr3:118813014-119077206	DEL	chr25:39286957-40282215	DEL	6.3879	2	0.041
chr24:1282069-1582182	DEL	chr18:65819321-65978584	DEL	6.3879	2	0.041
chr25:1955733-2606575	DEL	chr5:119221776-120378417	DEL	6.3884	2	0.041
chr2:136386853-136531159	DEL	chr18:11121144-11813752	DEL	6.3889	2	0.041
chr28:41674187-41737604	DEL	chr19:51395684-52234974	DEL	6.3893	2	0.041
chr16:47654206-47780813	DEL	chr29:48012818-48355723	DEL	6.3898	2	0.041
chr16:47654206-47780813	DEL	chr26:23167656-23414945	DEL	6.3898	2	0.041
chr19:50336021-50447799	DEL	chr26:23167656-23414945	DEL	6.3815	2	0.0411
chr9:103383683-105462864	DEL	chr24:61455723-62320145	DEL	6.382	2	0.0411
chr24:61455723-62320145	DEL	chr4:114326665-114640077	DEL	6.3825	2	0.0411
chr25:38171850-38377594	DEL	chr13:54496419-54829615	DEL	6.3825	2	0.0411
chr25:38171850-38377594	DEL	chr2:133816808-134465054	DEL	6.383	2	0.0411
chr28:41674187-41737604	DEL	chr10:102887596-103470001	DEL	6.384	2	0.0411
chr25:39286957-40282215	DEL	chr24:61455723-62320145	DEL	6.3854	2	0.0411
chr29:42897144-43269744	DEL	chr3:117575562-118346051	DEL	6.3859	2	0.0411

CNVR_LocA	CN_A	CNVR_LocB	CN_B	Chi2	Df	P-Value
chr19:50336021-50447799	DEL	chr2:133816808-134465054	DEL	6.3777	2	0.0412
chr28:41674187-41737604	DEL	chr7:4226753-4655753	DEL	6.3781	2	0.0412
chr4:113079474-113532717	DEL	chr29:48012818-48355723	DEL	6.3786	2	0.0412
chr16:47654206-47780813	DEL	chr25:39286957-40282215	DEL	6.3796	2	0.0412
chr25:36448529-36514994	DEL	chr18:11121144-11813752	DEL	6.3796	2	0.0412
chr2:133816808-134465054	DEL	chr19:56607168-57213764	DEL	6.3801	2	0.0412
chr24:61455723-62320145	DEL	chr19:51395684-52234974	DEL	6.3801	2	0.0412
chr18:11121144-11813752	DEL	chr4:117831202-120555019	DEL	6.3801	2	0.0412
chr24:1282069-1582182	DEL	chr10:102887596-103470001	DEL	6.3811	2	0.0412
chr18:65819321-65978584	DEL	chr29:48012818-48355723	DEL	6.3718	2	0.0413
chr2:136386853-136531159	DEL	chr7:44658442-46025089	DEL	6.3723	2	0.0413
chr4:113079474-113532717	DEL	chr5:119221776-120378417	DEL	6.3733	2	0.0413
chr16:47654206-47780813	DEL	chr25:38171850-38377594	DEL	6.3733	2	0.0413
chr6:106495683-107186270	DEL	chr5:119221776-120378417	DEL	6.3738	2	0.0413
chr6:106495683-107186270	DEL	chr24:61455723-62320145	DEL	6.3743	2	0.0413
chr19:63424825-63734072	DEL	chr9:103383683-105462864	DEL	6.3757	2	0.0413
chr16:47654206-47780813	DEL	chr3:117575562-118346051	DEL	6.368	2	0.0414
chr4:117831202-120555019	DEL	chr25:1955733-2606575	DEL	6.3714	2	0.0414
chr2:136386853-136531159	DEL	chr7:21462645-21677064	DEL	6.3714	2	0.0414
chr18:65819321-65978584	DEL	chr19:56607168-57213764	DEL	6.3714	2	0.0414
chr24:1282069-1582182	DEL	chr2:136386853-136531159	DEL	6.3622	2	0.0415
chr18:11121144-11813752	DEL	chr10:102887596-103470001	DEL	6.3627	2	0.0415
chr25:38171850-38377594	DEL	chr25:39286957-40282215	DEL	6.3636	2	0.0415
chr16:47654206-47780813	DEL	chr4:113079474-113532717	DEL	6.3636	2	0.0415
chr2:133816808-134465054	DEL	chr19:51395684-52234974	DEL	6.3661	2	0.0415
chr2:136386853-136531159	DEL	chr13:54496419-54829615	DEL	6.3661	2	0.0415
chr24:1282069-1582182	DEL	chr3:118813014-119077206	DEL	6.3661	2	0.0415
chr7:4226753-4655753	DEL	chr24:61455723-62320145	DEL	6.3574	2	0.0416
chr29:48012818-48355723	DEL	chr5:119221776-120378417	DEL	6.3593	2	0.0416
chr19:63424825-63734072	DEL	chr29:48012818-48355723	DEL	6.3612	2	0.0416
chr26:23167656-23414945	DEL	chr19:63424825-63734072	DEL	6.3612	2	0.0416
chr28:41674187-41737604	DEL	chr4:117831202-120555019	DEL	6.3612	2	0.0416
chr6:106495683-107186270	DEL	chr5:120553341-121175859	DEL	6.3531	2	0.0417
chr29:42897144-43269744	DEL	chr18:65819321-65978584	DEL	6.3531	2	0.0417
chr2:136386853-136531159	DEL	chr10:102887596-103470001	DEL	6.354	2	0.0417
chr4:113079474-113532717	DEL	chr19:51395684-52234974	DEL	6.354	2	0.0417
chr25:39286957-40282215	DEL	chr19:56607168-57213764	DEL	6.3545	2	0.0417
chr7:21462645-21677064	DEL	chr25:1955733-2606575	DEL	6.3545	2	0.0417
chr24:1282069-1582182	DEL	chr9:103383683-105462864	DEL	6.3545	2	0.0417
chr25:36448529-36514994	DEL	chr25:39286957-40282215	DEL	6.355	2	0.0417
chr3:117575562-118346051	DEL	chr9:103383683-105462864	DEL	6.3559	2	0.0417
chr3:118813014-119077206	DEL	chr26:49532894-49762633	DEL	6.3507	2	0.0418
chr19:50336021-50447799	DEL	chr2:136386853-136531159	DEL	6.3512	2	0.0418
chr7:44658442-46025089	DEL	chr19:51395684-52234974	DEL	6.3435	2	0.0419
chr29:42897144-43269744	DEL	chr25:1955733-2606575	DEL	6.3449	2	0.0419
chr18:65819321-65978584	DEL	chr5:119221776-120378417	DEL	6.3454	2	0.0419
chr2:136386853-136531159	DEL	chr6:106495683-107186270	DEL	6.3459	2	0.0419
chr25:39286957-40282215	DEL	chr4:114326665-114640077	DEL	6.3468	2	0.0419
chr19:50336021-50447799	DEL	chr29:48012818-48355723	DEL	6.3473	2	0.0419
chr2:136386853-136531159	DEL	chr19:56607168-57213764	DEL	6.3383	2	0.042
chr26:23167656-23414945	DEL	chr3:117575562-118346051	DEL	6.3383	2	0.042
chr18:65819321-65978584	DEL	chr3:118813014-119077206	DEL	6.3335	2	0.0421
chr28:41674187-41737604	DEL	chr7:44658442-46025089	DEL	6.3378	2	0.0421

CNVR_LocA	CN_A	CNVR_LocB	CN_B	Chi2	Df	P-Value
chr4:113079474-113532717	DEL	chr25:1955733-2606575	DEL	6.324	2	0.0423
chr7:4226753-4655753	DEL	chr13:54496419-54829615	DEL	6.325	2	0.0423
chr28:41674187-41737604	DEL	chr19:50336021-50447799	DEL	6.3259	2	0.0423
chr28:41674187-41737604	DEL	chr2:136386853-136531159	DEL	6.3236	2	0.0424
chr16:47654206-47780813	DEL	chr13:54496419-54829615	DEL	6.3156	2	0.0425
chr18:11121144-11813752	DEL	chr29:48012818-48355723	DEL	6.3174	2	0.0425
chr13:54496419-54829615	DEL	chr6:106495683-107186270	DEL	6.3095	2	0.0427
chr16:47654206-47780813	DEL	chr18:11121144-11813752	DEL	6.3006	2	0.0428
chr19:50336021-50447799	DEL	chr13:54496419-54829615	DEL	6.2926	2	0.043
chr24:1282069-1582182	DEL	chr19:50336021-50447799	DEL	6.2889	2	0.0431
chr24:1282069-1582182	DEL	chr7:4226753-4655753	DEL	6.2659	2	0.0436
chr4:117831202-120555019	DUP	chr20:70669729-71652724	DUP	6.2462	2	0.044
chr14:1514056-2553525	DUP	chr17:73118011-74998349	DUP	9.7357	4	0.0451
chr14:3885798-4672500	DUP	chr6:107678393-109951981	DUP	5.9999	2	0.0498

*Deletion- DEL and Duplication - DUP

Additional file 4.5 Significant pairwise association Chi^2 and P-values of deletion and duplication (CN_LocA and CN_LocB) CNVR events (CNVR_LocA and CNVR_LocB) identified in indigenous South Africansanga and composite cattle breeds.

CNVR_LocA	CN_LocA*	CNVR_LocB	CN_LocB	Chi2	Df	P-Value
chr11:102861577-107043330	DUP	chr17:73118011-74998349	DUP	27.0185	6	0.0001
chr6:53514737-53719693	DEL	chr29:48948337-51502868	DUP	19.2963	4	0.0007
chr18:62375495-63727709	DUP	chr11:102861577-107043330	DUP	13.4261	4	0.0094
chr18:62375495-63727709	DUP	chr17:73118011-74998349	DUP	12.6799	4	0.013
chr5:119221776-120378417	DUP	chr11:102861577-107043330	DUP	10.6075	4	0.0313
chr6:53514737-53719693	DEL	chr22:58873440-61283415	DUP	10.3585	4	0.0348

*Deletion- DEL and Duplication - DUP

Additional file 4.6 The number of CNVs (Num CNVs) and the CNVs identified that were present in more than 1 of the 7 South African cattle breeds (BRDs).

BRDs	Num CNVs	CNVs
AFR ANG HOL NGUXANG	1	chr6:108910274-109868839
AFR BON DRK NGU	1	chr1:105084197-105264358
ANG BON DRK HOL	1	chr18:63096692-63167945
AFR ANG HOL	3	chr17:74723634-74817054, chr17:74508803-74998349, chr29:51396010-51502868
AFR BON DRK	1	chr11:105778702-106019172
AFR DRK HOL	1	chr5:14770370-14894403
AFR DRK NGU	1	chr4:108168742-108313356
ANG BON NGU	5	chr2:135433480-135491609, chr20:71531915-71652724, chr6:10786656-10838635, chr6:109536093-109719477, chr11:104043185-104182498
ANG HOL NGU	5	chr25:41191025-41321020, chr22:60056909-60105535, chr26:51107888-51267717, chr25:42269092-42364359, chr11:106245832-106348964
ANG HOL NGUXANG	4	chr26:25880226-25982293, chr18:65819321-65978584, chr25:1665327-1808056, chr16:70816380-71125864
BON HOL NGU	1	chr6:53514737-53719693
AFR ANG	1	chr5:120718722-121175859
AFR BON	1	chr15:11439502-11604685
AFR DRK	2	chr6:440021-551383, chr1:105084197-105215796
AFR HOL	6	chr28:25175373-25352987, chr5:117738204-117823521, chr7:33722644-33868759, chr5:119567333-119853322, chr26:51107888-51181758, chr29:50126810-50240781
AFR NGU	1	chr12:80331512-80460375
AFR NGUXANG	1	chr14:1514056-2054457
ANG BON	5	chr29:2324336-2396643, chr6:10716501-10838635, chr21:70272221-70466564, chr22:60877108-61040701, chr17:74123863-74393620

BRDs	Num CNVs	CNVs
ANG DRK	3	chr13:12587622-12759014, chr10:88110028-88202054, chr27:9164768-9250049
ANG HOL	9	chr22:60056909-60130492, chr14:1616618-2054457, chr11:103742782-103856100, chr25:38283088-38357005, chr5:119795140-119880599, chr11:105677940-105797400, chr3:120325736-120573628, chr10:14129663-14200836, chr11:103742782-104182498
ANG NGU	11	chr11:104415459-104493462, chr18:62375495-62512168, chr6:109536093-109868839, chr29:50860475-50971886, chr3:121179950-121275236, chr25:1665327-1737669, chr4:44671762-44792807, chr20:71220728-71296022, chr13:16901756-16988665, chr16:50670749-50862929, chr6:109536093-109951981
ANG NGUXANG	30	chr19:50364787-50421201, chr17:73183467-74998349, chr19:51395684-51934105, chr29:44969518-45023665, chr6:106495683-107186270, chr3:120191150-121403393, chr26:50933887-51680135, chr28:41674187-41737604, chr29:50020743-51502868, chr5:117738204-118353758, chr16:49386191-49568812, chr18:63096692-63727709, chr19:56861094-57213764, chr18:64231273-64286141, chr18:62375495-62751093, chr12:89095085-89655152, chr4:114326665-114603252, chr15:84987785-85049720, chr21:70089833-71109676, chr16:47654206-47780813, chr29:48948337-49478288, chr5:116915398-117133270, chr14:2194228-2468020, chr25:1955733-2105645, chr14:2803998-3137184, chr25:36448529-36514994, chr25:279528-1365841, chr13:54496419-54829615, chr19:63424825-63734072, chr20:71220728-71531915
BON HOL	2	chr17:74292319-74344162, chr9:5901981-5949799
BON NGU	12	chr17:73804497-74393620, chr29:50020743-50126810, chr6:76407633-76754229, chr26:42959100-43087057, chr17:73559752-74393620, chr17:73678846-74393620, chr10:93759181-93823011, chr21:69279283-69395154, chr7:15869064-15921536, chr8:29379357-29464724, chr17:73915069-74393620, chr3:120547501-120622998
DRK HOL	4	chr20:55554731-55693187, chr9:5901981-5981648, chr27:9096031-9250049, chr17:73773784-73944911
DRK NGU	1	chr4:84872989-84963191
HOL NGU	2	chr21:70373409-70466564, chr6:109056666-109209793
HOL NGUXANG	1	chr12:34505806-34554241
NGU NGUXANG	1	chr9:92400217-92462210

Additional file 4.7 Table demonstrating the overlap (Num GEN) of CNVR genes (GEN) identified in 7 South African cattle breeds (BRD).

BRD	Num GEN	GEN
AFR ANG BON DRK HOL NG NGXANG	6	*LOC527441, WDR1, NTNG2, TMEM128, SLC5A1, OTOP1*
AFR ANG BON HOL NG NGXANG	17	*LYAR, ACAD9, MIR2390, CARD11, SMARCB1, TTF1, IGLL1, cnBP, EFCC1, MIF, GSTT4, ZBTB49, GSTT3, GSTT1, ISY1, SLC2A11, DERL3*
AFR ANG DRK HOL NG NGXANG	1	*MED27*
AFR ANG BON HOL NGXANG	7	*COPG1, EML5, RPN1, GATA2, MIR2374, RAB7A, C22H3orf37*
AFR ANG HOL NG NGXANG	18	*MGC127055, RTDR1, NDUFA6, SMDT1, SPECC1L, WBP2NL, UPB1, STX18, FAM109B, MIR2323, NAGA, LOC785804, TCF20, NSG1, C17H22orf13, MIR2442, LOC531152, CYP2D14*
ANG BON HOL NG NGXANG	10	*FUBP3, DDX31, AK8, ASS1, VPREB3, GTF3C4, ZNF70, ZNF280B, CHCHD10, ZNF280A*
AFR ANG HOL NGXANG	8	*GRAP2, ENTHD1, CHRAC1, CARS, EIF2C2, NAP1L4, GMDS, KCNQ1*
ANG BON HOL NGXANG	7	*RRP7A, NGEF, EXOSC2, ABL1, BARHL1, POLDIP3, PRDM12*
ANG BON NG NGXANG	2	*YWHAH, PISD*
ANG HOL NG NGXANG	35	*TMEM86B, DDT, GGT1, MIR33A, TMEM204, NUP214, RNF185, SUSD2, MIR2888-2, TMEM150B, PTK2, SORCS3, WNT7A, TMEM43, SLC6A6, PPP6R1, HSPBP1, INPP5J, LOC618516, CRAMP1L, XPC, HDAC11, SEPT3, SMTN, SNRPD3, SELM, SREBF2, AIF1L, CHCHD4, GGT5, CENPM, LIMK2, NUP210, PIK3IP1, XRCC2*
AFR ANG NGXANG	13	*PARVG, SLC22A18, BRAT1, PHLDA2, CDKN1C, ARFRP1, CDC42, PARVB, GNA12, ZGPAT, GRIFIN, ZBTB46, LFNG*
AFR DRK HOL	1	*CDH12*
AFR DRK NG	1	*CLEC5A*
ANG BON HOL	1	*GIGYF2*
ANG BON NG	2	*LUZP1, HTRA1*
ANG BON NGXANG	7	*SLC22A23, EEFSEC, ALKBH5, RUVBL1, QRFP, KBTBD12, SEC61A1*
ANG DRK HOL	1	*NEIL3*
ANG HOL NGXANG	41	*GPR107, USP20, FASN, HN1L, C25H16orf13, SERTAD4, KCNK9, BRI3, SYT5, TRAPPC9, MRPL28, CBX7, SLC16A3, ARL4C, RHOT2, GAL3ST4, APOBEC3B, TMEM8A, CBX6, GPC2, FAM195A, LSM3, LAMTOR4, STUB1, RAB40C, LOC781977, C11H9orf78, WFIKKN1, CD7, MAPK8IP3, APOBEC3A, PWWP2B, CSNK1D, DECR2, ZC3H14, NME4, GPR123, STAG3, TOR1A, FNBP1, LOC516108*
ANG NG NGXANG	7	*TSPAN32, EVC2, TEKT4, PATZ1, CCDC134, ERICH1, DRG1*
HOL NG NGXANG	2	*NADSYN1, DHCR7*
AFR BON	1	*MIR1256*
AFR DRK	1	*NPR3*
AFR HOL	3	*UFL1, FHL5, COL13A1*
AFR NG	2	*CLRN1, HNF4G*
ANG BON	1	*LLGL1*
ANG DRK	2	*FCF1, AREL1*

BRD	Num GEN	GEN
ANG HOL	1	*CTNNA2*
ANG NG	5	*TMEM60, YME1L1, ANKRD26, EPHB2, PHTF2*
ANG NGXANG	181	*LOC511094, HSPB1, MIR2890, NDUFB10, DMP1, TMC4, AXIN1, INS, PACSIN2, PDIA2, SH3BP4, RDH13, TMEM51, GALK1, KIAA0415, MRPL38, LSP1, TUBB2B, POLR3K, FOXC2, ZNF582, PXDC1, NTMT1, IGF2, TBL3, HPS6, SYNGR3, ZACN, T, HAGHL, ITGB4, CDKL3, SKP1, SERPINB9, ANXA8L1, TUBGCP2, SEPX1, TNNT3, FADD, CELA3B, PRPF31, SERPINB1, RPL3L, CHRNA4, SERPINB6, WIPI1, RGS11, ACOX1, MTHFSD, ITFG3, LMF1, STPG1, NLRP5, STMN3, ASB6, RIPK1, EXOC2, COMP, CAPN1, FAM135B, GBGT1, IQSEC1, RHBDF1, SNRNP25, C7H5orf15, H3F3B, MIR2896, METRN, PTGES, WBP2, NQO2, CEL, TSEN34, NIPAL3, GFI1B, CATSPER3, ADSL, PAEP, HBM, CYB5R3, PPP2CA, BPHL, ARFGAP3, ZNF583, SAP30BP, CCZ1, RSPH10B, GFER, RGS9, GALR2, ZNF667, SRP68, RPS2, RTEL1, EPS8L1, FBXL16, WDR24, RPUSD1, LENG1, NPRL3, PITX1, TFPT, TH, C11H9orf9, LUC7L, PEX10, ANO1, MSLN, SRMS, KIR2DS1, DCXR, JAKMIP1, GTF3C5, UNC13D, PPFIA1, RER1, C13H20orf195, FBF1, NDUFA3, UNK, ASCL2, RADIL, TNNI2, DPYSL4, GPS1, SECTM1, LOC515042, ARHGDIG, CDC42SE2, H19, TNFRSF6B, CAMLG, DUS1L, RNF151, MBOAT7, CHTF18, C7H5orf24, GNG13, TCF7, WRNIP1, MRPL23, EXOC7, CRTC1, NARFL, PPDPF, PTK6, HBA, VDAC1, CDKN2AIPNL, HBQ1, RALGDS, RCAN3, CTSD, SEC24A, ZNF787, SGSM3, UBE2B, ZNF444, SLC9A3R2, FAM173A, SLC22A10, MIR483, RPS9, SLC22A9, GLT6D1, MIR2345, EVPL, CTTN, FARS2, CDK3, MIR202, ARFGAP1, SAR1B, C9H6orf118, DPP6, GMEB2, RECQL5, TRIM47, TNRC6B, TNNT1, EEF1A2*
BON HOL	7	*TOP3A, FAM83G, SLC5A10, PRPSAP2, GRAP, SMCR7, SHMT1*
BON NGXANG	2	*USP42, DRG2*
DRK HOL	1	*TCF3*
DRK NG	1	*AMPH*
DRK NGXANG	1	*CCDC174*
HOL NG	4	*LEO1, CDH9, C28H10orf35, TMOD3*
HOL NGXANG	17	*VPS35, SHCBP1, SLC7A5, FXR2, EIF4A1, SHBG, TNFSF12, PPP2R5C, SAT2, CA5A, SENP3, MPDU1, DCP1B, SOX15, JPH3, TNFSF13, CD68*
NG NGXANG	3	*XRCC6, NHP2L1, OBSCN*
AFR	17	*FLT1, SDCCAG8, ARL4A, MGC134473, CEP170, UNC13B, CTBS, TBC1D9B, CD1E, MIR2300A, RNF130, MIR2300B, NRXN1, ETV1, C7H5orf45, WWP1, ADCY1*
ANG	57	*PTPN23, GAPVD1, MPZL2, FBXL12, AMICA1, ZC3H7B, L3MBTL2, MAP2K5, HMGB4, RTF1, BHMT, TSEN2, NECAP1, DCUN1D1, PIN1, HSPA5, UBL5, POMT1, PPAPDC3, MPZL3, PPARG, GLG1, C22H3orf75, KCNJ13, SLC12A5, RABEPK, IL10RA, PPP6R3, PPP6C, RAPGEF1, ZYX, CREBBP, C29H11orf84, ACSL6, SCN4B, TMPRSS4, STXBP5, PPARG-TSEN2, CLEC4A, FAM78A, SCAI, PLEKHA1, HRASLS5, NDUFAF1, FAM131B, SCN2B, UCK1, SCAP, CLEC6A, MIR2441, ZNF335, NUSAP1, EPHA1, RANGAP1, MMP9, PCIF1, H2AFY*
BON	26	*ANP32E, CA14, BIRC5, PLEKHO1, APH1A, LOC514490, ERLIN2, RASSF2, HNRNPAB, C3H1orf54, AFMID, MTR, ZNF804B, PRND, AGXT2L2, FNTA, MRPS21, QRFPR, POFUT1, C4H7orf62, COL11A1, ZNF703, SOCS3, PGS1, CTNNA3, KIF3B*
DRK	13	*IRAK1BP1, RBFOX3, CD36, MGAT4C, SLC25A21, UBXN7, GNAT3, SUB1, TTLL7, SEPHS1, PRKACB, PHYH, ZADH2*
HOL	19	*GGH, UQCR11, CITED2, FOXP2, MIR2306, GK2, FAM98A, GALNTL6, NDST4, KCNJ3, GATA5, RASGRP3, LOC524676, RPS21, CADPS2, DCT, TXLNB, TTPA, NAA11*
NG	26	*DPH5, PTPN18, GDA, GAT, DDI1, FGF9, MIR454, SLC9A9, NCDN, CHIC2, TCF12, GLYAT, MASTL, TRIM37, GLYATL2, FAF1, PDGFD, PSMB2, SLIT3, CISD1, IPMK, LOC518623, ANKRD55, STAB2, ADCYAP1R1, SKA2*

BRD	Num GEN	GEN
NGXANG	231	*TRIP6, HYAL3, ABCA3, TBC1D24, PDPN, FAM73B, PGP, RAC1, LILRA4, PACRG, SMCR7L, SIRT7, FAM195B, PPP2R4, RFC2, ANAPC11, CDH13, PCYT2, GNAI2, MMD2, DOLK, LOC786914, NUP188, SLC44A2, MIR2382, ABCG1, PSMB7, ZCWPW1, RPP40, GRID2IP, ELOVL3, COPE, TAP2, FOXJ1, CYGB, CDHR4, C25H7orf61, B4GALNT3, SYT3, EMC10, IP6K1, NPB, UBA7, NTHL1, MRPL12, TMPRSS3, RPL12, MIR1225, CCDC137, SLC38A3, GNAT1, ATP6V0C, MYADML2, QTRT1, MSX1, LOC618733, MIR2348, AGPAT3, UBA52, LRRC4B, C25H7orf26, MRPL34, LOC618591, RPL3, LRRC8A, DDX49, MIR940, KDM5A, ABHD8, DYNC1H1, ARSG, GTPBP3, ANKRD61, MLST8, IL22RA1, PILRA, LOC613393, RBM5, WDR20, LRRC48, C7H19orf60, DNM2, RAB26, CHST12, LRSAM1, EFNB3, TOM1L2, YIPF2, RPS19BP1, SSR1, TSC22D4, C7H19orf52, E4F1, ALYREF, RAC3, TAB1, NEK6, INPP5D, ABCA9, MAP2K6, TRAF7, PAXIP1, ZDHHC4, TTYH1, RNPS1, PRPF4B, EIF2AK1, DNASE1L2, SLC2A8, SNAI1, MIR2440, TP53, BEND6, MAFG, NPRL2, TMEM115, ST6GALNAC2, CCDC77, KLC1, LOC407171, SGSM1, ZFYVE21, RAD51, RSPH1, PDGFB, TMEM59L, SLC6A12, MEPCE, CAMKV, ARHGDIA, CERS1, XRCC3, TMEM189, NOLC1, DOLPP1, STRA13, LYRM4, ADIPOR2, STXBP1, TPCN2, C25H16orf59, GID4, RABGAP1L, CCNF, PPP1R35, MIR199A-2, KXD1, ASPSCR1, PITX3, DAGLB, FGF4, CYB561D2, RMDN3, ORC6, IER5L, LOC777692, MIR2347, SLC6A13, WRAP53, GCGR, MIR2346, NINJ2, CCND1, SH3GLB2, P4HB, UBE2V1, ATP1B2, PYCR1, KDELR2, TRAIP, RNF114, IL28RA, TFF2, SLC12A9, CLEC11A, LOC100196901, TNFRSF19, AMDHD2, DDA1, SREBF1, SRRT, RNF157, PLVAP, MIR2284K, CALCOCO2, LRRC45, LENG8, CD69, TMC6, MON1A, IFRD2, CRAT, SEMA3B, ELL, PRPSAP1, TMED1, SPATA2, LHX2, ECI1, ILF3, C21H14orf2, PHYHD1, MGC134105, TRAPPC10, AIMP2, TFF3, NTN3, ACHE, ATPAF2, ATF4, HGS, WHSC2, EPHB4, CDH20, RASSF1, SYNGR1, PAPOLB, HYAL2, ECI2, ZMYND10, HYAL1, PPP1R27, DNAH2, FAM212A, MIR33B, HSP90AA1, ADRBK2, UBASH3A, PMS2, FKBP8, CDYL*

Addendum D

Additional file 5.1 The number of animals exhibiting CNVRs of respective copy numbers (CN) identified at stringencies F10 (10), F45 (45) and F75 (75) in South African Nguni cattle.

CNVR	CN*	10	45	75
chr4:32303-41995	+	1		
chr4:34928-41983	-	1		
chr4:2441386-2466289	-	1		
chr4:12607955-12630750	-	1		
chr4:14860133-14881046	-	1		
chr4:23174389-23193715	-	1		
chr4:31134305-31263892	-	2		
chr4:32519439-33008799	-	1		
chr4:39226190-39265254	-	1		
chr4:57397783-57423041	+	2		
chr4:78073104-78496643	+	1		
chr4:95094842-95114551	+	2		
chr4:95095030-95114622	-	2	2	
chr4:106448509-106552670	-	2		
chr4:106453948-106551899	+	2		
chr4:106578490-106584134	-	1		
chr4:113759719-114140193	-	1		
chr4:119561206-119578416	-	1		
chr4:120788454-120794502	-	1		
chr5:56-4458	-	1		
chr5:267356-329198	-	1		
chr5:20166117-20181497	-	1		
chr5:34221627-34253800	-	1		
chr5:46549077-46574225	-	1		
chr5:48744516-48750151	+	1		
chr5:60011753-60118955	+	1		
chr5:60709806-60714502	-	1		
chr5:66573555-66580731	-	1		
chr5:102463144-102627616	-	2		
chr5:103280881-103399995	-	1		
chr6:5320385-6915590	-	1		
chr6:5320488-6570629	+	1		
chr6:5344519-6915218	-		1	
chr6:5350609-6915218	-			1
chr6:5356700-6570629	+		1	2
chr6:6575235-6915207	+	1	1	
chr6:7123271-7180476	-	1		
chr6:15612436-15619622	-	1		
chr6:77187566-77190382	-	1		
chr6:87455946-87458256	-	1		
chr6:89203532-89221220	-	1		
chr6:112186281-112188511	-	1		
chr6:117852853-117900534	-	1		
chr6:119226504-119234390	-	1		
chr7:4879830-4900814	-	1		
chr7:9879015-10071377	-	1		
chr7:10985788-11149404	+	1		

CNVR	CN*	10	45	75
chr7:18106496-18115538	-	1		
chr7:51681819-51684955	-	1		
chr7:56569111-56622106	-	1		
chr7:56571198-56622077	+	2		
chr7:56575684-56622106	-		1	
chr7:66692483-66696663	+	1	1	
chr7:66692498-66893174	-	1	1	
chr7:71568571-71952538	-	1		
chr7:71937757-71964157	+	1		
chr7:73810665-73832622	-	1		
chr7:84730601-84733863	-	1		
chr7:97267905-97281709	-	1		
chr7:103355416-103364632	-	1		
chr7:105826121-105889108	-	1		
chr7:105826613-105871051	+	1		
chr7:105894461-105925656	-	1		
chr7:105903679-105920464	+	1		
chr7:105921909-105923358	+	1		
chr10:1-9129	+	2		
chr10:101-9459	-	2		
chr10:1160-9129	+		1	
chr10:3588690-3639136	-	1		
chr10:4607232-4611131	-	1		
chr10:27726686-27737648	+	1		
chr10:34475549-34482054	-	1		
chr10:56753105-56784555	-	1		
chr10:58885992-59517111	-	1		
chr10:59480541-59498705	+	1		
chr10:80775099-80782762	-	1		
chr10:90625123-90626330	+	2		
chr10:91404851-91411384	-	1		
chr10:99258240-99270978	-	1		
chr11:58-69938	-		1	
chr11:58-89200	-	2		
chr11:51596-53837	+	1		
chr11:6522635-6608283	+	1		
chr11:63024570-63112241	-	1		
chr11:67539470-67543102	-	1		
chr11:70743818-70745609	-	1		
chr11:86451993-86456839	-	1		
chr11:89003371-89014916	-	1		
chr11:107284699-107310573	+	1		
chr11:107284820-107308548	-			1
chr11:107284820-107310284	-	1	1	
chr11:107286020-107310573	+		1	
chr12:70317527-70326517	-	1		
chr12:74559025-75480689	+	1		

CNVR	CN*	10	45	75
chr12:81159980-81171606	-	1		
chr17:43-294794	-	17		
chr17:115-270741	-		3	
chr17:244-294193	+	8		
chr17:269514-270605	-			1
chr17:2344774-2347403	-	2		
chr17:5578538-5579703	-	1		
chr17:5634899-5637236	-	1		
chr17:8964111-8991876	-	2		
chr17:13259846-13263263	-	1		
chr17:14199444-14400224	-	23		
chr17:14199697-14408231	+	18		
chr17:14199699-14408207	+		2	
chr17:14199713-14398261	-		7	
chr17:16932040-16939140	-	1		
chr17:18676790-18678062	+	4		
chr17:18758771-18762869	-	3		
chr17:23294853-23297089	-	1		
chr17:25036050-25049983	+		5	6
chr17:25036050-25053912	+	7		
chr17:25036050-25109391	-	6	5	6
chr17:26639936-26642136	-	1		
chr17:27153896-27155272	-	1		
chr17:28686550-28688790	-	1		
chr17:30659800-30663193	-	1		
chr17:32572974-32587529	-	1		
chr17:34092441-34216095	-	4		
chr17:35642369-35661942	-	6		
chr17:35642385-35656947	-			3
chr17:35642385-35661942	-		4	
chr17:35643527-35654427	+			1
chr17:35643527-35656500	+		2	
chr17:35643527-35660112	+	6		
chr17:37120308-37124192	-	2		
chr17:50578234-50585403	-	1		
chr17:50661428-51813221	-	7		
chr17:50666250-50835066	+	4		
chr17:50974351-51812865	+	2		
chr17:52637302-52639826	-	1		
chr17:52851602-52853703	-	1		
chr17:53567075-53568198	-	1		
chr17:54936312-54939450	-	1		
chr17:56539806-56544860	-	1		
chr17:57597023-57650232	+	1		
chr17:58587158-58591909	-	2		
chr17:59122579-59129557	-	2		
chr17:59635110-59726384	-	2		
chr17:61186147-61188455	-	1		
chr17:61926672-61932863	-	1		
chr17:63642816-63674501	-	5		
chr17:65682753-65685094	-	1		
chr17:66665571-66693826	-	1		
chr17:67134586-67135955	-	1		

CNVR	CN*	10	45	75
chr17:68726952-68730498	-	1		
chr17:70603798-70605114	-	1		
chr17:71002846-71004266	-	3		
chr17:71002936-71004119	-		1	1
chr17:72805796-72959600	-	8		
chr17:73139642-73158792	-	2		
chr17:74831218-74832755	-	1		
chr18:5-1552	+	1		
chr18:1304-19594	-	2		
chr18:4415-12334	+	1		
chr18:1918499-1920199	-	1		
chr18:9594472-9596563	-	1		
chr18:16667132-16673681	-	1		
chr18:43799171-43833210	-	2		
chr18:43802913-43832174	+	2		
chr18:43947911-43965808	-	1		
chr18:50941651-50945321	-	2		
chr18:50979723-50999074	-	1		
chr18:51002700-51114990	+	1		
chr18:51747266-51782049	+	1		
chr18:51749222-51779541	-	1		
chr18:61106911-61367707	-	1		
chr18:63752559-63927807	-	1		
chr19:5128-73174	-	2		
chr19:7174222-7185380	-	1		
chr19:8387443-8390682	-	1		
chr19:16942496-16944985	-	1		
chr19:19832938-19948198	-	2		
chr19:20331380-20335803	-	2		
chr19:27803406-27821442	-	1		
chr19:57721023-57730499	+	2		
chr19:57721034-57730499	+		1	
chr19:57721759-57842535	-	1		
chr22:135340-140629	+	1		
chr22:31068951-31077458	-	1		
chr22:42425957-42436768	-	1		
chr22:50601487-50609843	-	1		
chr22:54094522-54102915	-	1		
chr23:6842626-6846589	-	1		
chr23:23059995-23066223	-	1		
chr23:27082056-27173313	-	1		
chr23:27347991-27374716	-	1		
chr23:28455995-28510803	+	1		
chr23:28462489-28515089	-	1		
chr23:34676884-34735204	+	1		
chr23:39372200-39389412	-	1		
chr23:50571054-50626717	-	1		
chr24:7252288-7255684	-	1		
chr24:47067984-47071435	+	1		
chr24:61816078-62450202	+	2		
chr24:62364348-62450218	-	1		
chr25:9082923-9085304	+	1		

CNVR	CN*	10	45	75
chr25:9483302-9493089	-	1		
chr25:30396107-30416875	-	1		

CNVR	CN*	10	45	75
chr25:32363794-32382827	-	1	1	1

*CN: "-" represents a deletion and "+" represents a duplication

Additional file 5.2 The number of animals (ANMLs) presenting CNVR identified by NGS methodologies that cover SNPs captured on the Bovine 50K Beadchip (SNP Count).

CNVR	ANMLs	SNP Count
chr10:4607232-4611131	1	1
chr10:56753105-56784555	1	1
chr10:58885992-59517111	2	12
chr10:99258240-99270978	1	1
chr11:63024570-63112241	1	2
chr11:6522635-6608283	1	3
chr11:86451993-86456839	1	1
chr12:74559025-75480689	1	2
chr17:14199444-14400224	77	2
chr17:2344774-2347403	2	1
chr17:34092441-34216095	4	2
chr17:43-294794	31	4
chr17:50578234-51813221	18	5
chr17:52637302-52639826	1	1
chr17:57597023-57650232	1	2
chr17:59635110-59726384	2	2
chr17:66665571-66693826	1	1
chr18:16667132-16673681	1	1
chr18:61106911-61367707	2	5
chr18:63752559-63927807	4	2
chr19:19832938-19948198	3	2
chr19:57721759-57842535	3	3
chr22:50601487-50609843	1	1
chr23:27082056-27173313	1	1
chr23:28455995-28510803	2	1
chr23:39372200-39389412	1	1
chr23:50571054-50626717	1	1
chr24:61816078-62450202	2	14
chr24:62364348-62450218	1	1
chr25:30396107-30416875	1	1
chr4:113759719-114140193	1	2
chr4:119561206-119578416	1	1
chr4:31134305-31263892	1	3
chr4:32519439-33008799	1	6
chr4:39226190-39265254	1	1
chr4:78073104-78496643	1	10
chr5:20166117-20181497	1	1
chr5:60709806-60714502	1	1
chr5:66573555-66580731	1	1
chr6:5320385-6915590	48	2
chr6:7123271-7180476	2	1
chr7:66692483-66696663	2	6
chr7:71568571-71952538	3	8
chr7:73810665-73832622	1	1

Additional file 5.3 Molecular functions (MF), biological processes (BP) and cellular components (CC) of CNVR genes (GEN) identified in South African Nguni cattle.

GEN	MF	BP	CC
AACS	Lipid metabolic process, metabolic process	Acetoacetate-coa ligase activity, catalytic activity	
ADCK1	Protein phosphorylation, phosphorylation	Protein kinase activity, atp binding, kinase activity	Mitochondrion
ADCY1	Camp biosynthetic process, axonogenesis, long-term memory, intracellular signal transduction, regulation of circadian rhythm, cyclic nucleotide biosynthetic process, adenylate cyclase-activating G-protein coupled receptor signaling pathway, adenylate cyclase-inhibiting G-protein coupled receptor signaling pathway, camp-mediated signaling, rhythmic process, cellular response to calcium ion, cellular response to forskolin	Nucleotide binding, adenylate cyclase activity, lyase activity, phosphorus-oxygen lyase activity, calmodulin binding, ATP binding, calcium- and calmodulin-responsive adenylate cyclase activity, metal ion binding	Cytoplasm, plasma membrane, integral component of membrane, extracellular exosome, membrane, integral component of plasma membrane, membrane raft
ADRA1B	Norepinephrine-epinephrine vasoconstriction involved in regulation of systemic arterial blood pressure, phospholipase C-activating G-protein coupled receptor signaling pathway, positive regulation of cytosolic calcium ion concentration, cell-cell signaling, glucose homeostasis, positive regulation of MAPK cascade, positive regulation of vasoconstriction, positive regulation of smooth muscle contraction, regulation of cardiac muscle contraction, adenylate cyclase-activating adrenergic receptor signaling pathway, regulation of muscle contraction, signal transduction, G-protein coupled receptor signaling pathway, regulation of vasoconstriction	Alpha1-adrenergic receptor activity, protein heterodimerization activity, signal transducer activity, G-protein coupled receptor activity, adrenergic receptor activity	Integral component of plasma membrane, nuclear membrane, nucleus, plasma membrane, membrane, integral component of membrane
ADRBK2	Protein phosphorylation, signal transduction, receptor internalization, phosphorylation	G-protein coupled receptor kinase activity, atp binding, nucleotide binding, protein kinase activity, protein serine/threonine kinase activity, kinase activity, transferase activity, beta-adrenergic receptor kinase activity	
ANAPC10	Mitotic nuclear division, regulation of mitotic metaphase/anaphase transition, anaphase-promoting complex-dependent catabolic process, cell division, protein K11-linked ubiquitination, cell cycle, positive regulation of ubiquitin protein ligase activity, negative regulation of cyclin-dependent protein serine/threonine kinase by cyclin degradation, protein ubiquitination	Ubiquitin protein ligase activity	Anaphase-promoting complex, cytoplasm

GEN	MF	BP	CC
ANKRD 50	Retrograde transport, endosome to plasma membrane		Intracellular
ARHGE F25	Regulation of Rho protein signal transduction, positive regulation of gtpase activity	Rho guanyl-nucleotide exchange factor activity	
ASNA1	Transport, metabolic process, protein insertion into ER membrane	ATP binding, atpase activity, metal ion binding, nucleotide binding, hydrolase activity	Nucleolus, endoplasmic reticulum, extracellular exosome, nucleus, cytoplasm
ATP5A1	Negative regulation of endothelial cell proliferation, lipid metabolic process, ATP synthesis coupled proton transport, ATP hydrolysis coupled proton transport, ATP biosynthetic process, transport, ion transport, proton transport, ATP metabolic process	ATP binding, MHC class I protein binding, poly(A) RNA binding, proton-transporting ATP synthase activity, rotational mechanism, proton-transporting atpase activity, rotational mechanism, nucleotide binding, hydrolase activity, acting on acid anhydrides, catalyzing transmembrane movement of substances, atpase activity, protein binding	Mitochondrial proton-transporting ATP synthase complex, plasma membrane, COP9 signalosome, myelin sheath, proton-transporting ATP synthase complex, catalytic core F(1), extracellular exosome, mitochondrion, mitochondrial inner membrane, membrane, proton-transporting two-sector atpase complex, catalytic domain
ATP5I	ATP synthesis coupled proton transport, ATP biosynthetic process, transport, ion transport, proton transport	Hydrogen ion transmembrane transporter activity	Mitochondrial proton-transporting ATP synthase complex, coupling factor F(o), mitochondrial proton-transporting ATP synthase complex, mitochondrion, mitochondrial inner membrane, membrane, proton-transporting ATP synthase complex, coupling factor F(o)
BIVM	Nucleotide-excision repair, nucleic acid phosphodiester bond hydrolysis, DNA repair	Single-stranded DNA binding, endonuclease activity, DNA binding, catalytic activity, nuclease activity, hydrolase activity, acting on ester bonds	Nucleus
C22H3o rf18			Integral component of membrane, membrane
CACNA 2D1	Regulation of calcium ion transport, calcium ion transmembrane transport, calcium ion transport	Voltage-gated calcium channel activity	Sarcoplasmic reticulum, extracellular exosome, l-type voltage-gated calcium channel complex, voltage-gated calcium

GEN	MF	BP	CC
			channel complex, t-tubule, membrane, integral component of membrane
CALN1		Calcium ion binding, metal ion binding	Integral component of membrane, trans-Golgi network membrane, membrane
CAP2	Cell morphogenesis, cytoskeleton organization	Actin binding	Neuronal postsynaptic density
CARHSP1	Regulation of transcription, DNA-templated, regulation of mrna stability	DNA binding, mrna 3'-UTR binding, nucleic acid binding	Cytoplasmic exosome (rnase complex), cytosol, P granule, extracellular exosome
CC2D1A	Negative regulation of transcription from RNA polymerase II promoter, signal transduction, positive regulation of I-kappab kinase/NF-kappab signaling	RNA polymerase II core promoter proximal region sequence-specific DNA binding, RNA polymerase II transcription factor activity, sequence-specific DNA binding, transcriptional repressor activity, RNA polymerase II core promoter proximal region sequence-specific binding, signal transducer activity	Nucleus, membrane, extracellular exosome
CD79A	Adaptive immune response, B cell differentiation, B cell proliferation, B cell activation, B cell receptor signaling pathway, immune system process, cell surface receptor signaling pathway	Transmembrane signaling receptor activity	Cytoplasm, multivesicular body, integral component of plasma membrane, external side of plasma membrane, B cell receptor complex, membrane raft, plasma membrane, membrane, integral component of membrane
CDCA7L	Positive regulation of cell proliferation		Nucleolus, cytoplasm, nucleus
CDH20	Homophilic cell adhesion via plasma membrane adhesion molecules, cell adhesion	Calcium ion binding	Plasma membrane, integral component of membrane, membrane
CHEK2	G2/M transition of mitotic cell cycle, replicative cell aging, double-strand break repair, DNA damage induced protein phosphorylation, intrinsic apoptotic signaling pathway in response to DNA damage, response to gamma radiation, regulation of protein catabolic process, cellular protein catabolic process, positive regulation of transcription, DNA-templated, protein autophosphorylation, protein stabilization, signal	Protein serine/threonine kinase activity, ATP binding, protein kinase binding, ubiquitin protein ligase binding, protein homodimerization activity, nucleotide binding, protein kinase activity, kinase activity, transferase activity, identical protein binding	Chromosome, telomeric region, Golgi apparatus, PML body, nucleoplasm

GEN	MF	BP	CC
	transduction involved in intra-S DNA damage checkpoint, mitotic spindle assembly, regulation of transcription, DNA-templated, protein phosphorylation, apoptotic process, cellular response to DNA damage stimulus, phosphorylation, signal transduction in response to DNA damage		
CLDN10	Ion transport, transport	Structural molecule activity	Cytoplasm, plasma membrane, bicellular tight junction, integral component of membrane, membrane, cell junction
CNOT2	Negative regulation of transcription from RNA polymerase II promoter, nuclear-transcribed mrna catabolic process, deadenylation-dependent decay, trophectodermal cell differentiation, positive regulation of cytoplasmic mrna processing body assembly, negative regulation of translation, negative regulation of intracellular estrogen receptor signaling pathway, RNA phosphodiester bond hydrolysis, exonucleolytic, regulation of stem cell population maintenance, regulation of transcription, DNA-templated	RNA polymerase II transcription corepressor binding, poly(A)-specific ribonuclease activity	Cytoplasmic mrna processing body, nucleus, membrane, CCR4-NOT core complex, cytoplasm, CCR4-NOT complex
CNOT8	Nuclear-transcribed mrna catabolic process, deadenylation-dependent decay, transcription, DNA-templated, positive regulation of cell proliferation, exonucleolytic nuclear-transcribed mrna catabolic process involved in deadenylation-dependent decay, positive regulation of mrna catabolic process, RNA phosphodiester bond hydrolysis, exonucleolytic	Nucleic acid binding, poly(A)-specific ribonuclease activity, 3'-5'-exoribonuclease activity	Nucleus, CCR4-NOT complex, intracellular
COPS2	Cell proliferation, cullin deneddylation, skeletal muscle cell differentiation, negative regulation of transcription, DNA-templated	Transcription corepressor activity	Cytoplasm, COP9 signalosome, nucleus
CRISP1	Binding of sperm to zona pellucida, regulation of acrosome reaction	Calcium channel regulator activity	Extracellular region, nucleus
CRYBA1	Visual perception	Structural constituent of eye lens	
CTRB1	Proteolysis	Serine-type endopeptidase activity, peptidase activity, serine-type peptidase activity, hydrolase activity	
CUX2	Negative regulation of transcription from RNA polymerase II promoter, transcription, DNA-templated, short-term memory,	RNA polymerase II regulatory region sequence-specific DNA binding, RNA polymerase II core promoter proximal	Nucleus, extracellular exosome

GEN	MF	BP	CC
	positive regulation of gene expression, positive regulation of dendrite morphogenesis, positive regulation of synapse assembly, positive regulation of dendritic spine morphogenesis, cellular response to organic substance, positive regulation of excitatory postsynaptic potential, regulation of transcription, DNA-templated, negative regulation of transcription, DNA-templated, cognition	region sequence-specific DNA binding, transcriptional repressor activity, RNA polymerase II core promoter proximal region sequence-specific binding, DNA binding, sequence-specific DNA binding	
CYP19A1	Negative regulation of chronic inflammatory response, androgen metabolic process, negative regulation of macrophage chemotaxis, oxidation-reduction process, prostate gland growth	Iron ion binding, heme binding, aromatase activity, monooxygenase activity, oxidoreductase activity, oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, reduced flavin or flavoprotein as one donor, and incorporation of one atom of oxygen, metal ion binding	Endoplasmic reticulum, integral component of membrane, membrane
DCAF15	Protein ubiquitination		
DCK	Pyrimidine nucleotide metabolic process, deoxyribonucleoside monophosphate biosynthetic process, nucleotide biosynthetic process, phosphorylation, nucleobase-containing compound metabolic process, nucleosome assembly	Deoxycytidine kinase activity, atp binding, drug binding, protein homodimerization activity, nucleotide binding, kinase activity, transferase activity, nucleoside kinase activity, dna binding, protein heterodimerization activity	Nucleus, nucleosome, chromosome
DCTN2	Mitotic spindle organization, mitotic nuclear division, metabolic process, cell proliferation, melanosome transport, protein localization to centrosome, microtubule-based process	Motor activity, spectrin binding	Kinetochore, cytoplasm, centrosome, dynactin complex, microtubule, membrane, dynein complex, growth cone, vesicle, extracellular exosome, microtubule organizing center, cytoskeleton
DDR1	Regulation of cell growth, regulation of cell-matrix adhesion, embryo implantation, negative regulation of cell proliferation, regulation of extracellular matrix disassembly, smooth muscle cell migration, collagen-activated tyrosine kinase receptor signaling pathway, peptidyl-tyrosine autophosphorylation, ear development, wound healing, spreading of cells, branching involved in mammary gland duct morphogenesis, mammary gland alveolus development, smooth muscle cell-matrix adhesion,	Collagen binding, atp binding, protein tyrosine kinase collagen receptor activity, nucleotide binding, protein kinase activity, protein tyrosine kinase activity, transmembrane receptor protein tyrosine kinase activity, kinase activity, transferase activity	Extracellular space, integral component of plasma membrane, receptor complex, extracellular exosome, plasma membrane, membrane, integral component of membrane

GEN	MF	BP	CC
	protein phosphorylation, transmembrane receptor protein tyrosine kinase signaling pathway, phosphorylation, peptidyl-tyrosine phosphorylation, protein autophosphorylation		
DGKB	Phosphorylation, intracellular signal transduction	Diacylglycerol kinase activity, calcium ion binding, kinase activity	Intracellular, plasma membrane
DHPS	Peptidyl-lysine modification to peptidyl-hypusine, positive regulation of T cell proliferation, glucose homeostasis, deoxyhypusine biosynthetic process from spermidine	Deoxyhypusine synthase activity, transferase activity	
DMRTC2	Regulation of transcription, DNA-templated, male meiosis I, spermatid nucleus elongation, positive regulation of histone H3-K9 dimethylation, positive regulation of histone H3-K9 trimethylation, transcription, DNA-templated, sex differentiation, cell differentiation	Transcription factor activity, sequence-specific DNA binding, protein homodimerization activity, sequence-specific DNA binding, core promoter proximal region sequence-specific DNA binding, metal ion binding, DNA binding	XY body, nucleus
DNAH2	Microtubule-based movement, metabolic process	Microtubule motor activity, atp binding, atpase activity, nucleotide binding	Dynein complex
DPP6	Proteolysis, regulation of potassium ion transmembrane transport, protein localization to plasma membrane	Serine-type peptidase activity, potassium channel regulator activity	Voltage-gated potassium channel complex, extracellular exosome, integral component of membrane, plasma membrane, membrane
DTX3	Notch signaling pathway	Zinc ion binding, metal ion binding	
DUSP18	Inactivation of MAPK activity, peptidyl-tyrosine dephosphorylation, protein dephosphorylation, dephosphorylation	Protein tyrosine phosphatase activity, MAP kinase tyrosine/serine/threonine phosphatase activity, phosphoprotein phosphatase activity, protein tyrosine/serine/threonine phosphatase activity, hydrolase activity, phosphatase activity	Nucleoplasm, mitochondrial inner membrane, nucleus, cytoplasm, mitochondrion, membrane
E2F6	Regulation of transcription involved in G1/S transition of mitotic cell cycle, transcription, DNA-templated, regulation of transcription, DNA-templated, cell cycle	DNA binding, transcription factor activity, sequence-specific DNA binding	Transcription factor complex, MLL1 complex, nucleus
EIF4E	G1/S transition of mitotic cell cycle, behavioral fear response, translational initiation, regulation of translation, negative regulation of translation, stem cell population maintenance, negative regulation of neuron differentiation, positive regulation of mitotic cell cycle, translation	Translation initiation factor activity, enzyme binding, eukaryotic initiation factor 4G binding, poly(A) RNA binding, repressing transcription factor binding, RNA binding	Cytoplasmic mrna processing body, cytoplasm, mrna cap binding complex, cytoplasmic stress granule, eukaryotic translation initiation factor 4F complex, RISC complex, chromatoid body, perinuclear region

GEN	MF	BP	CC
			of cytoplasm, extracellular exosome
EMP2	Membrane raft assembly, T cell mediated cytotoxicity, positive regulation of cell-matrix adhesion, regulation of glomerular filtration, actin filament organization, cell-matrix adhesion, embryo implantation, cell death, positive regulation of cell proliferation, regulation of endothelial cell migration, bleb assembly, activation of protein kinase activity, protein localization to cell surface, blood vessel endothelial cell migration, early endosome to late endosome transport, regulation of angiogenesis, actin-mediated cell contraction, protein localization to plasma membrane, positive regulation of integrin-mediated signaling pathway, regulation of vasculogenesis, regulation of cell-matrix adhesion, cell adhesion, cell migration, regulation of kinase activity	Integrin binding, protein kinase binding, kinase binding	Golgi apparatus, cytosol, caveola, cell surface, integral component of membrane, apical plasma membrane, cytoplasmic vesicle, cytoplasm, plasma membrane, membrane, membrane raft, apical part of cell, Golgi membrane, nucleus
ETAA1			Cytoplasm
FABP2	Transport	Transporter activity, fatty acid binding, lipid binding	Intracellular, cytoplasm
FADS6	Fatty acid biosynthetic process, oxidation-reduction process, lipid metabolic process, fatty acid metabolic process	Oxidoreductase activity	Integral component of membrane, membrane
FAM155A			Integral component of membrane, membrane
FAM71D			Nucleoplasm, cytoplasm, nucleus
FBXW7	Cellular response to DNA damage stimulus, sister chromatid cohesion, protein ubiquitination, SCF-dependent proteasomal ubiquitin-dependent protein catabolic process, negative regulation of DNA endoreduplication, cellular response to UV, positive regulation of epidermal growth factor-activated receptor activity, protein stabilization, positive regulation of ubiquitin-protein transferase activity, positive regulation of ERK1 and ERK2 cascade, positive regulation of proteasomal protein catabolic process, regulation of mitophagy, positive regulation of oxidative stress-induced neuron intrinsic apoptotic signaling pathway, positive regulation of protein targeting to mitochondrion, positive regulation of protein ubiquitination involved in ubiquitin-	Cyclin binding, protein binding, bridging, ubiquitin protein ligase binding, identical protein binding, phosphothreonine binding, ubiquitin-protein transferase activator activity	Nucleoplasm, nucleolus, cytoplasm, SCF ubiquitin ligase complex, Parkin-FBXW7-Cul1 ubiquitin ligase complex, nucleus, protein complex

GEN	MF	BP	CC
	dependent protein catabolic process, positive regulation of protein ubiquitination		
FBXW9			
FDXR	Oxidation-reduction process, steroid biosynthetic process, cholesterol metabolic process, lipid metabolic process, steroid metabolic process	Oxidoreductase activity, ferredoxin-nadp+ reductase activity, protein binding, nadph-adrenodoxin reductase activity, flavin adenine dinucleotide binding, nadp binding	Mitochondrial matrix, mitochondrion, mitochondrial inner membrane
FHIT	Purine nucleotide metabolic process, transcription, DNA-templated, regulation of transcription, DNA-templated, negative regulation of proteasomal ubiquitin-dependent protein catabolic process, intrinsic apoptotic signaling pathway by p53 class mediator, apoptotic process, metabolic process	Nucleotide binding, ubiquitin protein ligase binding, identical protein binding, bis(5'-adenosyl)-triphosphatase activity, catalytic activity, hydrolase activity	Cytosol, extracellular exosome, cytoplasm
FOXP1	Negative regulation of transcription from RNA polymerase II promoter, in utero embryonic development, positive regulation of mesenchymal cell proliferation, pre-B cell differentiation, positive regulation of immunoglobulin production, transcription, DNA-templated, skeletal muscle tissue development, motor neuron axon guidance, ventral spinal cord development, immunoglobulin V(D)J recombination, sarcomere organization, negative regulation of transcription, DNA-templated, positive regulation of transcription from RNA polymerase II promoter, smooth muscle tissue development, positive regulation of epithelial cell proliferation, cardiac muscle cell differentiation, regulation of cardiac muscle cell proliferation, lung secretory cell differentiation, T follicular helper cell differentiation, interleukin-21 secretion, negative regulation of lung goblet cell differentiation, positive regulation of cardiac muscle cell differentiation, regulation of transcription, DNA-templated, heart development, lung development, positive regulation of transcription, DNA-templated, cardiovascular system development, regulation of lung goblet cell differentiation	RNA polymerase II core promoter proximal region sequence-specific DNA binding, RNA polymerase II transcription factor activity, sequence-specific DNA binding, chromatin binding, transcription factor activity, RNA polymerase II distal enhancer sequence-specific binding, protein homodimerization activity, protein self-association, metal ion binding, protein heterodimerization activity, DNA binding, transcription factor activity, sequence-specific DNA binding, sequence-specific DNA binding	Nucleus
GABRB2	Signal transduction, sensory perception of sound, ion transmembrane transport, negative regulation of neuron apoptotic process, inner ear receptor cell development, innervation, cochlea development, transport, ion transport, regulation of neuron	Gaba-a receptor activity, extracellular ligand-gated ion channel activity, ion channel activity, chloride channel activity	Plasma membrane, integral component of membrane, cell junction, synapse, extracellular exosome, membrane, integral

GEN	MF	BP	CC
	apoptotic process, neuron development, negative regulation of neuron death, cellular response to histamine, chloride transmembrane transport, chloride transport		component of plasma membrane, chloride channel complex, postsynaptic membrane, gaba-a receptor complex
GAL3ST1	Sphingolipid metabolic process, glycolipid biosynthetic process, lipid metabolic process	Galactosylceramide sulfotransferase activity, transferase activity	Golgi membrane, integral component of membrane, golgi apparatus, membrane
GALNT14	Protein glycosylation	Transferase activity, transferring glycosyl groups, carbohydrate binding, transferase activity	Golgi membrane, integral component of membrane, golgi apparatus, membrane
GEN	Bp	Mf	Extracellular exosome
GGACT	Cellular modified amino acid catabolic process	Gamma-glutamylcyclotransferase activity, transferase activity, transferase activity, transferring acyl groups	Cytoplasm
GLT1D1			Nucleus
GNL1	Cellular response to DNA damage stimulus, ribosome biogenesis	Gtpase activity, GTP binding	Integral component of plasma membrane, cytoplasmic, membrane-bounded vesicle, membrane, integral component of membrane
GPNMB	Cell adhesion, negative regulation of tumor necrosis factor production	Integrin binding, heparin binding	Intracellular, integral component of membrane, membrane
GPR182	G-protein coupled receptor signaling pathway, intracellular signal transduction, signal transduction	G-protein coupled receptor activity, signal transducer activity	NMDA selective glutamate receptor complex, cell junction, postsynaptic membrane, plasma membrane, membrane, integral component of membrane, synapse
GRIN2B	Ion transmembrane transport, ionotropic glutamate receptor signaling pathway, response to ethanol, transport, ion transport	Nmda glutamate receptor activity, extracellular-glutamate-gated ion channel activity, glycine binding, receptor activity, ionotropic glutamate receptor activity, ion channel activity	Postsynaptic density, nmda selective glutamate receptor complex, cell junction, postsynaptic membrane, plasma membrane, membrane, integral component of membrane, synapse
GRIN2C	Protein localization, response to wounding, directional locomotion, ionotropic glutamate receptor signaling pathway, negative regulation of protein catabolic process, neuromuscular	Nmda glutamate receptor activity, extracellular-glutamate-gated ion channel activity, cation channel activity, receptor activity, ionotropic glutamate receptor activity, ion channel	Microvillus

GEN	MF	BP	CC
	process controlling balance, excitatory postsynaptic potential, cation transmembrane transport, transport, ion transport, ion transmembrane transport, regulation of membrane potential	activity	
GRXCR2	Sensory perception of sound	Heat shock protein binding, unfolded protein binding	Nucleoplasm, mitochondrion, centrosome, haus complex, microtubule cytoskeleton
HAUS3	Spindle assembly, centrosome organization		Cytoplasm
HEBP1		Heme binding	Integral component of plasma membrane, cell surface
HHIP	Carbohydrate metabolic process, smoothened signaling pathway, neuroblast proliferation, dorsal/ventral pattern formation, regulation of fibroblast growth factor receptor signaling pathway, negative regulation of smoothened signaling pathway, skeletal system morphogenesis, oxidation-reduction process, epithelial tube branching involved in lung morphogenesis, signal transduction, organ morphogenesis, negative regulation of signal transduction	Zinc ion binding, oxidoreductase activity, acting on the ch-oh group of donors, quinone or similar compound as acceptor, quinone binding, hedgehog family protein binding, catalytic activity	Centrosome, cytosol, hops complex, fhf complex
HOOK2	Endosome organization, lysosome organization, endosome to lysosome transport, early endosome to late endosome transport	Identical protein binding	Nucleus, nucleoplasm, cytoplasm, golgi apparatus, intracellular
HSPB8		Identical protein binding	Cytoplasm
IGF2BP3		Nucleotide binding, mrna 3'-UTR binding, mrna 5'-UTR binding, nucleic acid binding, RNA binding, poly(A) RNA binding	Extracellular space, nucleus, insulin-like growth factor ternary complex, extracellular exosome, extracellular region
IGFBP3	Regulation of cell growth, osteoblast differentiation, negative regulation of protein phosphorylation, protein phosphorylation, regulation of glucose metabolic process, negative regulation of smooth muscle cell migration, response to insulin, positive regulation of apoptotic process, positive regulation of catalytic activity, positive regulation of MAPK cascade, regulation of insulin-like growth factor receptor signaling pathway, positive regulation of insulin-like growth factor receptor signaling pathway, type B pancreatic cell proliferation, positive regulation of myoblast differentiation, negative regulation of smooth muscle	Fibronectin binding, protein tyrosine phosphatase activator activity, insulin-like growth factor I binding, insulin-like growth factor II binding, insulin-like growth factor binding, growth factor binding	Extracellular space, cytoplasm, membrane, interleukin-12 complex, interleukin-23 complex, extracellular region

GEN	MF	BP	CC
	cell proliferation, negative regulation of cell proliferation, regulation of growth		
IL12B	Positive regulation of T cell mediated cytotoxicity, positive regulation of defense response to virus by host, positive regulation of T-helper 1 type immune response, positive regulation of natural killer cell mediated cytotoxicity directed against tumor cell target, negative regulation of inflammatory response to antigenic stimulus, immune response, cell cycle arrest, response to UV-B, positive regulation of activation of JAK2 kinase activity, cell migration, cytokine-mediated signaling pathway, natural killer cell activation, negative regulation of interleukin-10 production, negative regulation of interleukin-17 production, positive regulation of granulocyte macrophage colony-stimulating factor production, positive regulation of interleukin-10 production, positive regulation of interleukin-12 production, positive regulation of interleukin-17 production, positive regulation of tumor necrosis factor production, positive regulation of natural killer cell activation, positive regulation of natural killer cell proliferation, positive regulation of smooth muscle cell apoptotic process, T-helper cell differentiation, interferon-gamma biosynthetic process, positive regulation of activated T cell proliferation, regulation of tyrosine phosphorylation of Stat1 protein, positive regulation of tyrosine phosphorylation of Stat3 protein, positive regulation of tyrosine phosphorylation of Stat4 protein, positive regulation of tyrosine phosphorylation of Stat5 protein, defense response to protozoan, negative regulation of growth of symbiont in host, positive regulation of interferon-gamma biosynthetic process, positive regulation of osteoclast differentiation, negative regulation of smooth muscle cell proliferation, defense response to Gram-negative bacterium, positive regulation of NK T cell activation, positive regulation of NK T cell proliferation, defense response to virus, cellular response to lipopolysaccharide, cellular response to interferon-gamma, interferon-gamma secretion, cell surface receptor signaling pathway, response to organic substance,	Cytokine receptor activity, cytokine activity, interleukin-12 receptor binding, protein binding, growth factor activity, interleukin-12 alpha subunit binding, protein homodimerization activity, interleukin-23 receptor binding, protein heterodimerization activity, cytokine receptor binding, identical protein binding	Extracellular space, nucleoplasm, cytoplasm, extracellular region

GEN	MF	BP	CC
	positive regulation of interferon-gamma production, positive regulation of mononuclear cell proliferation, positive regulation of T cell proliferation, positive regulation of cell adhesion, positive regulation of lymphocyte proliferation		
IL15	Natural killer cell differentiation, NK T cell proliferation, immune response, signal transduction, positive regulation of cell proliferation, positive regulation of interleukin-17 production, positive regulation of natural killer cell proliferation, positive regulation of natural killer cell differentiation, positive regulation of T cell proliferation, tyrosine phosphorylation of Stat5 protein, extrathymic T cell selection, regulation of T cell differentiation, cell maturation, lymph node development, regulation of defense response to virus by host, positive regulation of immune response, positive regulation of protein O-linked glycosylation	Cytokine activity, cytokine receptor binding	
IL27RA	Positive regulation of T-helper 1 type immune response, negative regulation of type 2 immune response, positive regulation of interferon-gamma production, regulation of isotype switching to igg isotypes, defense response to Gram-positive bacterium, interleukin-27-mediated signaling pathway	Interleukin-27 receptor activity	SREBP-SCAP-Insig complex, endoplasmic reticulum, endoplasmic reticulum membrane, membrane, integral component of membrane
INSIG1	Triglyceride metabolic process, cholesterol biosynthetic process, negative regulation of steroid biosynthetic process, SREBP signaling pathway, inner ear morphogenesis, middle ear morphogenesis, negative regulation of fat cell differentiation, negative regulation of fatty acid biosynthetic process, palate development, cranial suture morphogenesis, negative regulation of cargo loading into COPII-coated vesicle, lipid metabolic process, response to sterol depletion, steroid metabolic process, cholesterol metabolic process, sterol biosynthetic process		Cytoplasm, cell surface
INTU	Nervous system development, regulation of smoothened signaling pathway, negative regulation of keratinocyte proliferation, spinal cord dorsal/ventral patterning, neural tube development, keratinocyte differentiation, regulation of ossification, hair follicle morphogenesis, cilium assembly, embryonic digit morphogenesis, positive regulation of		

GEN	MF	BP	CC
	smoothened signaling pathway, negative regulation of cell division, limb development, motile primary cilium assembly, multicellular organism development, cell projection organization, cilium morphogenesis		
ITGBL1			Nuclear chromatin, nucleoplasm, transcription factor complex, nucleus
JUNB	Negative regulation of transcription from RNA polymerase II promoter, vasculogenesis, osteoblast differentiation, trophectodermal cell differentiation, transcription from RNA polymerase II promoter, response to radiation, response to mechanical stimulus, regulation of cell death, osteoclast differentiation, response to lipopolysaccharide, cellular response to hormone stimulus, osteoblast proliferation, response to cytokine, regulation of cell proliferation, response to drug, positive regulation of cell differentiation, positive regulation of transcription from RNA polymerase II promoter, decidualization, response to camp, regulation of cell cycle, embryonic process involved in female pregnancy, labyrinthine layer blood vessel development, cellular response to calcium ion, in utero embryonic development, transcription, DNA-templated, regulation of transcription, DNA-templated, regulation of transcription from RNA polymerase II promoter, cellular process	RNA polymerase II core promoter proximal region sequence-specific DNA binding, RNA polymerase II transcription factor activity, sequence-specific DNA binding, transcriptional activator activity, RNA polymerase II core promoter proximal region sequence-specific binding, transcription coactivator activity, transcription factor binding, RNA polymerase II regulatory region sequence-specific DNA binding, DNA binding, transcription factor activity, sequence-specific DNA binding, sequence-specific DNA binding	Nuclear chromatin, transcription factor complex, nucleus
JUND	Negative regulation of transcription from RNA polymerase II promoter, transcription, DNA-templated, response to radiation, response to mechanical stimulus, regulation of cell death, response to lipopolysaccharide, cellular response to hormone stimulus, response to cytokine, regulation of cell proliferation, response to drug, positive regulation of cell differentiation, positive regulation of transcription from RNA polymerase II promoter, response to camp, regulation of cell cycle, regulation of transcription, DNA-templated, regulation of transcription from RNA polymerase II promoter	RNA polymerase II core promoter proximal region sequence-specific DNA binding, RNA polymerase II transcription factor activity, sequence-specific DNA binding, transcription coactivator activity, transcription factor binding, DNA binding, transcription factor activity, sequence-specific DNA binding, sequence-specific DNA binding	
KCTD15	Multicellular organism development, protein		Endoplasmic reticulum lumen

GEN	MF	BP	CC
	homooligomerization		
KDELC1	Glycolipid metabolic process	Glucosyltransferase activity	
KIAA0100			Cytoplasm, kinesin complex, microtubule, membrane, ciliary rootlet, neuron projection, neuronal cell body
KIF5A	Microtubule-based movement, axon guidance, protein localization, metabolic process, cytoskeleton-dependent intracellular transport	Atp binding, microtubule binding, atp-dependent microtubule motor activity, plus-end-directed, nucleotide binding, microtubule motor activity	Cytoplasm, actin cytoskeleton, integral component of membrane, cul3-ring ubiquitin ligase complex, membrane
KLHL2	Protein ubiquitination	Actin binding, ubiquitin-protein transferase activity	Nucleoplasm, nucleolus, cytoplasm, plasma membrane, cul3-ring ubiquitin ligase complex, nucleus
KLHL7	Protein ubiquitination involved in ubiquitin-dependent protein catabolic process, protein ubiquitination	Protein homodimerization activity, ubiquitin-protein transferase activity, identical protein binding	Extracellular space, extracellular region
LALBA	Lactose biosynthetic process, response to estradiol, response to progesterone, response to dehydroepiandrosterone, response to 11-deoxycorticosterone	Lactose synthase activity, calcium ion binding, identical protein binding, metal ion binding	Heterotrimeric g-protein complex, plasma membrane, membrane
LOC529425	G-protein coupled receptor signaling pathway, signal transduction	Signal transducer activity	
LRRC1			Cytoplasmic mrna processing body, cytoplasmic stress granule, intracellular membrane-bounded organelle, cytoplasm, intracellular ribonucleoprotein complex
LSM14A	Cytoplasmic mrna processing body assembly, RIG-I signaling pathway, positive regulation of type I interferon-mediated signaling pathway, defense response to virus, regulation of translation, multicellular organism development	Double-stranded DNA binding, double-stranded RNA binding, single-stranded RNA binding, poly(A) RNA binding	Cytoplasmic mrna processing body, spliceosomal complex, U6 snrnp, nucleolus, small nucleolar ribonucleoprotein complex, U4/U6 x U5 tri-snrnp complex, extracellular exosome, nucleus, intracellular ribonucleoprotein complex
LSM6	Mrna splicing, via spliceosome, maturation of SSU-rrna, mrna	RNA binding, poly(A) RNA binding	Plasma membrane, anchored

GEN	MF	BP	CC
	processing, RNA splicing		component of membrane, membrane
LYPD4			
MAD2L1	Mitotic spindle assembly checkpoint		Mitochondrion, cytoplasm
MALSU1	Negative regulation of translation, ribosomal large subunit biogenesis, negative regulation of mitochondrial translation, negative regulation of ribosome biogenesis	Ribosomal large subunit binding	Lysosome
MAN2B1	Mannose metabolic process, protein deglycosylation, learning or memory, carbohydrate metabolic process, metabolic process	Alpha-mannosidase activity, zinc ion binding, carbohydrate binding, catalytic activity, hydrolase activity, hydrolyzing o-glycosyl compounds, mannosidase activity, hydrolase activity, hydrolase activity, acting on glycosyl bonds, metal ion binding	Nucleoplasm, cytoplasm, plasma membrane, nucleus
MCC	Negative regulation of epithelial cell migration, establishment of protein localization, negative regulation of epithelial cell proliferation, negative regulation of canonical Wnt signaling pathway		Mediator complex
MED13L	Regulation of transcription from RNA polymerase II promoter	RNA polymerase II transcription cofactor activity	Nucleus
METTL21C	Protein methylation, peptidyl-lysine methylation, methylation	Protein-lysine n-methyltransferase activity, methyltransferase activity, transferase activity	
METTL21E	Methylation	Methyltransferase activity, transferase activity	Integral component of plasma membrane, membrane, integral component of membrane
MGC138914	L-ornithine transmembrane transport, L-lysine transmembrane transport, arginine transmembrane transport, amino acid transmembrane transport, amino acid transport	L-ornithine transmembrane transporter activity, arginine transmembrane transporter activity, L-lysine transmembrane transporter activity, antiporter activity, amino acid transmembrane transporter activity	Centriole, membrane, TCTN-B9D complex, ciliary basal body, cytoplasm, centrosome, ciliary transition zone
MKS1	Neural tube closure, determination of left/right symmetry, epithelial structure maintenance, nonmotile primary cilium assembly, embryonic digit morphogenesis, embryonic skeletal system development, branching morphogenesis of an epithelial tube, inner ear receptor stereocilium organization, head development, regulation of canonical Wnt signaling pathway, common bile duct development, regulation of smoothened signaling pathway involved in dorsal/ventral neural tube patterning, motile primary cilium assembly, embryonic brain development, regulation of Wnt signaling pathway, planar cell polarity pathway, regulation of smoothened signaling pathway, cilium assembly, cilium morphogenesis		Mitochondrion, mitochondrial inner membrane, mitochondrial large ribosomal subunit, ribosome, large ribosomal subunit, intracellular ribonucleoprotein complex
MRPL22	Mitochondrial translational initiation, mitochondrial translational	Structural constituent of ribosome, poly(a) rna binding	Cytosol, extracellular exosome,

GEN	MF	BP	CC
	elongation, translation		cytoplasm, peroxisome
MVK	Cholesterol biosynthetic process, isoprenoid biosynthetic process, phosphorylation, negative regulation of inflammatory response, metabolic process, isopentenyl diphosphate biosynthetic process, mevalonate pathway, lipid metabolic process, steroid biosynthetic process, steroid metabolic process, cholesterol metabolic process, sterol biosynthetic process	Mevalonate kinase activity, atp binding, identical protein binding, nucleotide binding, kinase activity, transferase activity, phosphotransferase activity, alcohol group as acceptor	Actin cytoskeleton, z disc, cytoplasm
MYOZ2		Actin binding, telethonin binding	Nuclear chromatin, nurd complex, nucleus
NACC2	Negative regulation of cell proliferation, posttranscriptional regulation of gene expression, histone deacetylation, cellular protein complex localization, protein homooligomerization, negative regulation of G1/S transition of mitotic cell cycle by negative regulation of transcription from RNA polymerase II promoter, positive regulation of intrinsic apoptotic signaling pathway in response to DNA damage, negative regulation of transcription, DNA-templated	RNA polymerase II core promoter proximal region sequence-specific DNA binding, transcription factor activity, RNA polymerase II transcription factor binding, transcriptional repressor activity, RNA polymerase II core promoter proximal region sequence-specific binding, histone deacetylase activity, protein homodimerization activity, histone deacetylase binding	Primary cilium, ciliary inversin compartment, ciliary base, cilium
NEK8	Protein phosphorylation, determination of left/right symmetry, heart development, organ morphogenesis, regulation of hippo signaling	Protein serine/threonine kinase activity, atp binding, nucleotide binding, protein kinase activity	Integral component of membrane, nonmotile primary cilium, membrane, integral component of plasma membrane, plasma membrane
NPY2R	Outflow tract morphogenesis, cardiac left ventricle morphogenesis, adenylate cyclase-inhibiting G-protein coupled receptor signaling pathway, neuropeptide signaling pathway, signal transduction, G-protein coupled receptor signaling pathway, cell surface receptor signaling pathway, synaptic transmission, feeding behavior	Peptide YY receptor activity, signal transducer activity, G-protein coupled receptor activity, neuropeptide Y receptor activity	Nucleoplasm, cytoplasm, extracellular exosome
NQO2			
NUDT6	Metabolic process	Hydrolase activity	Nucleus, nucleoplasm, cytoplasm
NUPL2	Protein export from nucleus	Nuclear export signal receptor activity, poly(a) rna binding, metal ion binding	Plasma membrane, integral component of membrane, membrane
OR12D2	G-protein coupled receptor signaling pathway, detection of chemical stimulus involved in sensory perception, detection of	Transmembrane signaling receptor activity, G-protein coupled receptor activity, olfactory receptor activity, signal	Apical dendrite

GEN	MF	BP	CC
	chemical stimulus involved in sensory perception of smell, signal transduction, sensory perception of smell, response to stimulus	transducer activity	
OSBP2	Lipid transport, spermatid development, transport	Cholesterol binding	Integral component of membrane, membrane
OTOP2			Integral component of nuclear inner membrane, cytosol, integral component of plasma membrane, cell-cell junction, external side of plasma membrane, neuromuscular junction, neuronal cell body, presynapse, plasma membrane, membrane, integral component of membrane, synapse
P2RX7	Activation of MAPK activity, cell morphogenesis, phagolysosome assembly, positive regulation of T cell mediated cytotoxicity, protein phosphorylation, membrane protein ectodomain proteolysis, phospholipid transfer to membrane, membrane budding, inflammatory response, mitochondrion organization, response to mechanical stimulus, response to zinc ion, positive regulation of calcium ion transport into cytosol, positive regulation of gene expression, positive regulation of glutamate secretion, positive regulation of gamma-aminobutyric acid secretion, synaptic vesicle exocytosis, protein processing, phospholipid scrambling, sensory perception of pain, cytolysis, positive regulation of bone mineralization, cellular response to extracellular stimulus, bleb assembly, positive regulation of prostaglandin secretion, response to lipopolysaccharide, positive regulation of interleukin-6 production, collagen metabolic process, response to ATP, response to fluid shear stress, positive regulation of ion transmembrane transport, purinergic nucleotide receptor signaling pathway, T cell proliferation, response to drug, T cell homeostasis, NAD transport, negative regulation of MAPK cascade, multicellular organismal protein catabolic process, phospholipid translocation, negative regulation of bone resorption, negative regulation of cell volume, positive regulation	Lipopolysaccharide binding, purinergic nucleotide receptor activity, extracellular ATP-gated cation channel activity, ATP binding, receptor activity, ion channel activity, channel activity	Cellular_component, cytoplasm

GEN	MF	BP	CC
	of glycolytic process, ceramide biosynthetic process, pore complex assembly, skeletal system morphogenesis, homeostasis of number of cells within a tissue, positive regulation of interleukin-1 alpha secretion, positive regulation of interleukin-1 beta secretion, defense response to Gram-positive bacterium, release of sequestered calcium ion into cytosol, protein oligomerization, response to calcium ion, response to electrical stimulus, membrane depolarization, positive regulation of mitochondrial depolarization, positive regulation of lymphocyte apoptotic process, cellular response to dsrna, reactive oxygen species metabolic process, extrinsic apoptotic signaling pathway, positive regulation of bleb assembly, positive regulation of protein phosphorylation, transport, ion transport, cation transport, calcium ion transport, cell volume homeostasis, plasma membrane organization, response to bacterium, response to organic substance, gene expression, programmed cell death, response to organic cyclic compound, positive regulation of interleukin-1 beta production, positive regulation of apoptotic process, positive regulation of catalytic activity, positive regulation of MAPK cascade, positive regulation of ossification, positive regulation of protein secretion, positive regulation of cytokine secretion, cellular response to organic cyclic compound, cation transmembrane transport		
P33MONOX	Biological_process, oxidation-reduction process	Molecular_function, oxidoreductase activity	
PARP12	Metabolic process	Nad+ adp-ribosyltransferase activity, poly(a) rna binding, metal ion binding	Mitochondrion
PCCA	Metabolic process	Biotin carboxylase activity, propionyl-coa carboxylase activity, atp binding, enzyme binding, metal ion binding, nucleotide binding, catalytic activity, ligase activity	Plasma membrane, integral component of membrane, membrane
PCDH10	Homophilic cell adhesion via plasma membrane adhesion molecules, cell adhesion	Calcium ion binding	
PDE5A	Signal transduction, positive regulation of cardiac muscle hypertrophy, regulation of cgmp metabolic process, negative	Cyclic-nucleotide phosphodiesterase activity, cgmp binding, metal ion binding, 3',5'-cyclic-GMP phosphodiesterase	Condensed chromosome, nucleoplasm, cytoplasm,

GEN	MF	BP	CC
	regulation of T cell proliferation, positive regulation of MAP kinase activity, cgmp catabolic process, negative regulation of cardiac muscle contraction, relaxation of cardiac muscle, positive regulation of oocyte development, metabolic process, cgmp metabolic process	activity, nucleotide binding, catalytic activity, 3',5'-cyclic-nucleotide phosphodiesterase activity, phosphoric diester hydrolase activity, hydrolase activity	membrane, preribosome, large subunit precursor, pebow complex, nucleus, chromosome, nucleolus
PES1	Maturation of LSU-rrna from tricistronic rrna transcript (SSU-rrna, 5.8S rrna, LSU-rrna), maturation of 5.8S rrna from tricistronic rrna transcript (SSU-rrna, 5.8S rrna, LSU-rrna), nucleolus organization, cell proliferation, protein localization to organelle, regulation of cell cycle, rrna processing, ribosome biogenesis, ribosomal large subunit biogenesis	RNA binding, ribonucleoprotein complex binding, poly(A) RNA binding	CAAX-protein geranylgeranyltransferase complex
PGGT1B	Protein geranylgeranylation, metabolic process	Protein geranylgeranyltransferase activity, catalytic activity, caax-protein geranylgeranyltransferase activity, zinc ion binding, prenyltransferase activity, transferase activity, metal ion binding	Histone acetyltransferase complex, cytosol, plasma membrane, nucleus, cytoplasm
PHF17	Histone H3 acetylation, histone H4-K5 acetylation, histone H4-K8 acetylation, histone H4-K12 acetylation, positive regulation of transcription from RNA polymerase II promoter, negative regulation of G1/S transition of mitotic cell cycle	RNA polymerase II transcription coactivator activity, zinc ion binding, metal ion binding	Autophagosome, extracellular exosome, cytoplasm, membrane
PIP4K2C	Phosphatidylinositol phosphorylation, positive regulation of autophagosome assembly, regulation of autophagy, phosphatidylinositol metabolic process, phosphorylation	Phosphatidylinositol phosphate kinase activity, identical protein binding, atp binding, 1-phosphatidylinositol-5-phosphate 4-kinase activity, nucleotide binding, kinase activity, transferase activity	Golgi apparatus, extracellular exosome, intracellular, cytoplasm
PITPNB	Transport	Lipid binding	Intracellular membrane-bounded organelle, intracellular
PITPNM2	Transport	Metal ion binding	Extracellular space, extracellular exosome
PON3	Response to toxic substance, aromatic compound catabolic process, negative regulation of superoxide anion generation, carboxylic acid catabolic process	Arylesterase activity, protein homodimerization activity	Cytoplasm, extracellular exosome
PRDX2	Removal of superoxide radicals, regulation of apoptotic process, oxidation-reduction process, response to oxidative stress, cellular response to oxidative stress	Thioredoxin peroxidase activity, peroxidase activity, antioxidant activity, oxidoreductase activity, peroxiredoxin activity	Nucleus, nucleotide-activated protein kinase complex
PRKAB1	Protein phosphorylation, fatty acid biosynthetic process, signal	Protein kinase activity, kinase activity	Extracellular space, endoplasmic

GEN	MF	BP	CC
	transduction, positive regulation of gene expression, regulation of protein kinase activity, lipid metabolic process, fatty acid metabolic process, phosphorylation		reticulum lumen, endoplasmic reticulum membrane, cytosol, extracellular region
PRL	Blastocyst formation, lactation, biosynthetic process, response to mechanical stimulus, positive regulation of gene expression, negative regulation of gene expression, negative regulation of nitric oxide mediated signal transduction, signal transduction involved in regulation of gene expression, peptide hormone secretion, response to food, positive regulation of NF-kappab import into nucleus, negative regulation of apoptotic process, response to external biotic stimulus, positive regulation of nitric oxide biosynthetic process, positive regulation of fatty acid biosynthetic process, positive regulation of endocytosis, long-day photoperiodism, positive regulation of NF-kappab transcription factor activity, positive regulation of lactation, regulation of meiotic cell cycle process involved in oocyte maturation, response to L-arginine	Prolactin receptor binding, hormone activity	
PROCA1			Membrane
PRRC2C	Hematopoietic progenitor cell differentiation	Protein c-terminus binding, poly(a) rna binding	Cytoplasm, plasma membrane, ciliary rootlet, ciliary basal body, intracellular, cytoskeleton, membrane, cell projection
RAB28	Intracellular protein transport, nucleocytoplasmic transport, small gtpase mediated signal transduction, metabolic process, toxin transport, signal transduction, protein transport	Gtpase activity, GTP binding, GDP binding, nucleotide binding	Intracellular, membrane
RAB34	Small gtpase mediated signal transduction, protein transport, signal transduction	GTP binding	Lipid particle, cytosol, plasma membrane, cell-cell junction, membrane, extracellular exosome, intracellular, cytoplasm, cell junction
RAP1B	Intracellular protein transport, nucleocytoplasmic transport, metabolic process, cell proliferation, Rap protein signal transduction, establishment of endothelial barrier, positive	Gtpase activity, GTP binding, GDP binding, protein complex binding, nucleotide binding	Nucleoplasm, nucleolus, cytoplasm, membrane, nucleus

GEN	MF	BP	CC
	regulation of ERK1 and ERK2 cascade, cellular response to camp, regulation of cell junction assembly, regulation of establishment of cell polarity, negative regulation of synaptic vesicle exocytosis, signal transduction, small gtpase mediated signal transduction, protein transport, negative regulation of calcium ion-dependent exocytosis		
RBM19	Positive regulation of embryonic development	Nucleotide binding, poly(a) rna binding, nucleic acid binding	Nucleoplasm, cytoplasm, ribonuclease h2 complex, nucleus
RNASEH2A	Mismatch repair, RNA catabolic process, DNA replication, removal of RNA primer, RNA phosphodiester bond hydrolysis, endonucleolytic, RNA metabolic process	RNA binding, RNA-DNA hybrid ribonuclease activity, metal ion binding, nucleic acid binding, nuclease activity, endonuclease activity, hydrolase activity	Integral component of membrane, membrane
RNFT2		Zinc ion binding, metal ion binding	
RPL23A		Nucleotide binding, molecular_function, structural constituent of ribosome, rrna binding, RNA binding	Nucleus, cytosolic large ribosomal subunit, TORC2 complex, intracellular, ribosome, intracellular ribonucleoprotein complex
RPL23A	Ribosomal large subunit assembly, translation, biological_process		
RPL23A			Nucleolus, focal adhesion, membrane, cytosolic small ribosomal subunit, extracellular exosome, intracellular, cytoplasm, cytosol, ribosome, intracellular ribonucleoprotein complex
RPS19	Ribosomal small subunit assembly, maturation of SSU-rrna from tricistronic rrna transcript (SSU-rrna, 5.8S rrna, LSU-rrna), monocyte chemotaxis, translation, nucleolus organization, Notch signaling pathway, erythrocyte differentiation, protein tetramerization, positive regulation of respiratory burst involved in inflammatory response, negative regulation of respiratory burst involved in inflammatory response, rrna processing, maturation of SSU-rrna, ribosomal small subunit biogenesis	Structural constituent of ribosome, fibroblast growth factor binding, protein kinase binding, protein homodimerization activity, poly(A) RNA binding	
RTDR1			Centrosome, ciliary basal body
RTTN	Determination of left/right symmetry, cilium organization		Extracellular region, transport

GEN	MF	BP	CC
			vesicle membrane, cytoplasmic vesicle, membrane, cytoplasmic, membrane-bounded vesicle
SCG3		Poly(A) RNA binding	Endosome, trans-Golgi network
SCOC	Positive regulation of macroautophagy, regulation of protein complex stability		Extracellular exosome
SCPEP1	Negative regulation of blood pressure, positive regulation of vasodilation, proteolysis involved in cellular protein catabolic process, proteolysis	Serine-type carboxypeptidase activity, carboxypeptidase activity, peptidase activity, hydrolase activity	Extracellular region, dolichyl-phosphate-mannose-protein mannosyltransferase complex, membrane
SDF2	Cell wall mannoprotein biosynthetic process, multicellular organism development, protein O-linked mannosylation, chain elongation of O-linked mannose residue, regulation of endoplasmic reticulum unfolded protein response	Dolichyl-phosphate-mannose-protein mannosyltransferase activity	Extracellular space, cytoplasm
SERPINB6	Negative regulation of endopeptidase activity, negative regulation of peptidase activity	Serine-type endopeptidase inhibitor activity, peptidase inhibitor activity	Mitochondrion, mitochondrial inner membrane, mitochondrial matrix
SIRT4	Regulation of glutamine family amino acid metabolic process, protein ADP-ribosylation, glutamine metabolic process, cellular response to DNA damage stimulus, peptidyl-lysine deacetylation, negative regulation of fatty acid oxidation, negative regulation of insulin secretion, positive regulation of lipid biosynthetic process, tricarboxylic acid metabolic process, regulation of pyruvate dehydrogenase activity	NAD+ ADP-ribosyltransferase activity, NAD-dependent protein deacetylase activity, metal ion binding, biotinidase activity, lipoamidase activity, NAD+ binding, transferase activity, hydrolase activity	Integral component of plasma membrane, membrane, integral component of membrane
SLC12A6	Synaptic transmission, potassium ion import, rubidium ion transport, cellular hypotonic salinity response, chloride transmembrane transport, transport, ion transport, transmembrane transport, cellular hypotonic response	Potassium:chloride symporter activity, protein kinase binding, potassium ion symporter activity, rubidium ion transmembrane transporter activity, transporter activity, potassium ion transmembrane transporter activity, cation:chloride symporter activity	Integral component of membrane, membrane
SLC15A4	Oligopeptide transport, protein transport, transmembrane transport, transport, peptide transport	Symporter activity, transporter activity	Integral component of plasma membrane, membrane, integral component of membrane
SLC26A10	Bicarbonate transport, oxalate transport, regulation of membrane potential, regulation of intracellular ph, sulfate transmembrane transport, chloride transmembrane transport, sulfate transport,	Chloride channel activity, secondary active sulfate transmembrane transporter activity, bicarbonate transmembrane transporter activity, sulfate transmembrane	Integral component of membrane, membrane

GEN	MF	BP	CC
	transmembrane transport	transporter activity, anion:anion antiporter activity, oxalate transmembrane transporter activity	
SLC35E4			Integral component of plasma membrane, extracellular exosome, plasma membrane, membrane, integral component of membrane
SLCO4C1	Sodium-independent organic anion transport, transport, ion transport	Sodium-independent organic anion transmembrane transporter activity, transporter activity	
SMTN			Kinetochore, cytoplasm, spindle microtubule, microtubule plus-end, mitotic spindle
SPAG5	Mitotic sister chromatid segregation, spindle organization, establishment of spindle orientation, regulation of attachment of spindle microtubules to kinetochore		Integral component of membrane, membrane
SPATA9	Multicellular organism development, spermatogenesis, cell differentiation		Cytoplasm, membrane, nucleoplasm, golgi apparatus
SPRY1	Organ induction, multicellular organism development, negative regulation of cell proliferation, negative regulation of Ras protein signal transduction, negative regulation of ERK1 and ERK2 cascade, regulation of signal transduction, establishment of mitotic spindle orientation, metanephros development, ureteric bud development, negative regulation of gtpase activity, negative regulation of fibroblast growth factor receptor signaling pathway, negative regulation of MAP kinase activity, negative regulation of neurotrophin TRK receptor signaling pathway, bud elongation involved in lung branching		
SUPT6H	Nucleobase-containing compound metabolic process, regulation of mrna export from nucleus, positive regulation of transcription elongation from RNA polymerase II promoter, regulation of isotype switching, regulation of mrna processing, regulation of muscle cell differentiation, negative regulation of histone H3-K27 methylation, regulation of transcription from RNA polymerase II promoter, regulation of DNA-templated transcription, elongation	DNA binding, histone binding, poly(A) RNA binding, nucleic acid binding	Integral component of membrane, membrane, cell junction, synaptic vesicle membrane, cytoplasmic vesicle, synapse
SVOP	Transmembrane transport, transport	Transmembrane transporter activity, substrate-specific transmembrane transporter activity	Focal adhesion, actin cytoskeleton
SYNPO2			Extracellular region, extracellular space, cytosol, axon, neuronal cell body
TAC1	Response to yeast, inflammatory response, positive regulation of	Receptor binding	RNA polymerase I transcription

GEN	MF	BP	CC
	cytosolic calcium ion concentration, tachykinin receptor signaling pathway, neuropeptide signaling pathway, cell-cell signaling, synaptic transmission, sensory perception of pain, antibacterial humoral response, antifungal humoral response, innate immune response, response to pain, defense response to Gram-negative bacterium, defense response to Gram-positive bacterium, cellular response to nerve growth factor stimulus		factor complex, nucleoplasm, intracellular membrane-bounded organelle
TAF1C	Transcription from RNA polymerase I promoter	RNA polymerase I CORE element sequence-specific DNA binding	Cytoplasm
TAOK3	MAPK cascade, mitotic G2 DNA damage checkpoint, positive regulation of JUN kinase activity, negative regulation of JNK cascade, protein autophosphorylation, protein phosphorylation, cellular response to DNA damage stimulus, positive regulation of stress-activated MAPK cascade, positive regulation of JNK cascade	Receptor signaling protein serine/threonine kinase activity, protein binding, ATP binding, nucleotide binding, protein kinase activity, protein serine/threonine kinase activity, transferase activity	Intracellular, endomembrane system
TBC1D8	Intracellular protein transport, regulation of vesicle fusion, activation of gtpase activity	Gtpase activator activity, calcium ion binding, Rab gtpase binding	Endoplasmic reticulum membrane, integral component of membrane, endoplasmic reticulum, membrane
TBXAS1	Oxidation-reduction process, prostaglandin biosynthetic process, lipid metabolic process, fatty acid metabolic process, fatty acid biosynthetic process, prostaglandin metabolic process	Monooxygenase activity, thromboxane-a synthase activity, iron ion binding, oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, heme binding, oxidoreductase activity, metal ion binding, isomerase activity	Extracellular exosome, extracellular region
TCN2	Cobalamin transport, cobalt ion transport, transport, ion transport	Cobalamin binding, metal ion binding	Nucleus
TEX30	Metabolic process	Hydrolase activity	Fibrinogen complex, external side of plasma membrane, extracellular matrix, platelet alpha granule, extracellular exosome, extracellular region, extracellular space, cell surface, secretory granule, endoplasmic reticulum, sarcoplasmic reticulum
THBS1	Activation of MAPK activity, negative regulation of endothelial cell proliferation, negative regulation of cell-matrix adhesion,	Phosphatidylserine binding, fibronectin binding, integrin binding, calcium ion binding, heparin binding, fibroblast	Integral component of membrane, membrane

GEN	MF	BP	CC
	sprouting angiogenesis, chronic inflammatory response, negative regulation of antigen processing and presentation of peptide or polysaccharide antigen via MHC class II, negative regulation of dendritic cell antigen processing and presentation, immune response, cell cycle arrest, cell adhesion, response to glucose, negative regulation of plasma membrane long-chain fatty acid transport, negative regulation of nitric oxide mediated signal transduction, negative regulation of cgmp-mediated signaling, negative regulation of plasminogen activation, positive regulation of fibroblast migration, cell migration, negative regulation of angiogenesis, peptide cross-linking, positive regulation of transforming growth factor beta receptor signaling pathway, regulation of cgmp metabolic process, response to magnesium ion, negative regulation of interleukin-12 production, negative regulation of fibroblast growth factor receptor signaling pathway, response to drug, positive regulation of tumor necrosis factor biosynthetic process, positive regulation of macrophage activation, negative regulation of cysteine-type endopeptidase activity involved in apoptotic process, positive regulation of blood vessel endothelial cell migration, negative regulation of blood vessel endothelial cell migration, engulfment of apoptotic cell, positive regulation of translation, positive regulation of angiogenesis, positive regulation of chemotaxis, response to calcium ion, positive regulation of protein kinase B signaling, negative regulation of fibrinolysis, positive regulation of extrinsic apoptotic signaling pathway via death domain receptors, positive regulation of endothelial cell apoptotic process, positive regulation of reactive oxygen species metabolic process, negative regulation of endothelial cell chemotaxis, inflammatory response, positive regulation of endothelial cell migration, negative regulation of endothelial cell migration, positive regulation of blood coagulation, positive regulation of cell migration, positive regulation of phosphorylation, negative regulation of apoptotic process, response to unfolded protein, response to endoplasmic reticulum stress, behavioral response to pain	growth factor binding, low-density lipoprotein particle binding, laminin binding, fibrinogen binding, collagen V binding, extracellular matrix binding	

GEN	MF	BP	CC
TLCD1			Golgi membrane, integral component of membrane, Golgi apparatus, membrane
TMEM167A			Integral component of membrane, transport vesicle, membrane
TMEM168			Integral component of membrane, ciliary transition zone, TCTN-B9D complex, ciliary membrane, plasma membrane, cilium, membrane, cell projection
TMEM17	Smoothened signaling pathway, cilium assembly, cell projection organization		Endoplasmic reticulum, integral component of membrane, membrane
TMEM98			Integral component of membrane, membrane
TMTC4			Cytoplasm
TNFAIP8L1	Negative regulation of TOR signaling		Cytoplasm, nuclear membrane, nuclear periphery, intracellular
TNPO2	Protein import into nucleus, docking, protein import into nucleus, translocation, NLS-bearing protein import into nucleus, ribosomal protein import into nucleus, intracellular protein transport	Nuclear localization sequence binding, Ran gtpase binding, protein transporter activity	Nucleus, cytoplasm
TPP2	Proteolysis	Aminopeptidase activity, serine-type endopeptidase activity, peptidase activity, serine-type peptidase activity, hydrolase activity	Nucleus, cytoplasm, intracellular
TRAF4	Activation of NF-kappab-inducing kinase activity, respiratory gaseous exchange, protein ubiquitination, respiratory tube development, regulation of apoptotic process, positive regulation of JNK cascade, positive regulation of protein homodimerization activity, signal transduction, positive regulation of protein kinase activity	Ubiquitin-protein transferase activity, tumor necrosis factor receptor binding, zinc ion binding, protein kinase binding, ubiquitin protein ligase binding, thioesterase binding, WW domain binding, metal ion binding	Integral component of membrane, extracellular exosome, membrane, integral component of plasma membrane
TSPAN8	Cell surface receptor signaling pathway		Cytoplasm, nucleus
UBC			Nucleus, nucleolus
UBLCP1	Protein dephosphorylation	Protein serine/threonine phosphatase activity, phosphoprotein phosphatase activity, hydrolase activity	Intracellular

GEN	MF	BP	CC
UNC13C	Synaptic transmission, intracellular signal transduction	Diacylglycerol binding, metal ion binding	Mitochondrial inner membrane, myelin sheath, respiratory chain, mitochondrion, membrane, integral component of membrane
UQCRFS1	Oxidation-reduction process, hydrogen ion transmembrane transport	Ubiquinol-cytochrome-c reductase activity, metal ion binding, 2 iron, 2 sulfur cluster binding, oxidoreductase activity, iron-sulfur cluster binding, oxidoreductase activity, acting on diphenols and related substances as donors	Photoreceptor inner segment, cytoplasm, actin cytoskeleton, photoreceptor connecting cilium, ciliary basal body
USH1G	Sensory perception of sound, inner ear morphogenesis, photoreceptor cell maintenance, sensory perception of light stimulus, equilibrioception, inner ear receptor stereocilium organization, inner ear receptor cell differentiation	Spectrin binding, protein homodimerization activity	Endosome, cytoplasmic vesicle membrane, integral component of organelle membrane, membrane, integral component of membrane, cytoplasmic vesicle
VOPP1	Transcription, DNA-templated, regulation of transcription, DNA-templated, signal transduction	Signal transducer activity	
VSTM2A			Integral component of membrane, membrane
WDR83OS			Transcription elongation factor complex, Cajal body, transcriptionally active chromatin, histone locus body, nucleus
ZC3H8	Negative regulation of transcription from RNA polymerase II promoter, negative regulation of T cell differentiation in thymus, snrna transcription from RNA polymerase II promoter, snrna transcription from RNA polymerase III promoter, T cell homeostasis, positive regulation of transcription from RNA polymerase III promoter, response to antibiotic, positive regulation of thymocyte apoptotic process, negative regulation of transcription, DNA-templated	RNA polymerase II intronic transcription regulatory region sequence-specific DNA binding, transcriptional repressor activity, RNA polymerase II transcription regulatory region sequence-specific binding, poly(A) RNA binding, metal ion binding, transcription factor activity, sequence-specific DNA binding, sequence-specific DNA binding	Endoplasmic reticulum, Golgi apparatus, integral component of membrane, membrane
ZDHHC22	Protein palmitoylation, protein localization to plasma membrane	Zinc ion binding, protein-cysteine s-palmitoyltransferase activity, transferase activity, transferase activity, transferring acyl groups, metal ion binding	Intracellular
ZIM2	Regulation of transcription, DNA-templated	Nucleic acid binding, metal ion binding	Nucleus
ZNF423	Notch signaling pathway, positive regulation of BMP signaling	Nucleic acid binding, metal ion binding	Intracellular

GEN	MF	BP	CC
	pathway, negative regulation of transcription, DNA-templated, positive regulation of transcription, DNA-templated		
ZNF45	Regulation of transcription, DNA-templated	Nucleic acid binding, transcription factor activity, sequence-specific DNA binding, metal ion binding	

Additional file 5.4 Respective copy numbers (CN) and number (Num GEN) of CNVR genes (GEN) identified in South African Nguni cattle.

CN	Num GEN	GEN
Deletion Duplication	30	*SCG3, SLC15A4, MGC134093, GLT1D1, ZNF45, KLHL2, OTOP2, MIR129-1, UQCRFS1, PDE5A, HHIP, GRXCR2, KCTD15, FABP2, PARP12, CNOT8, OR12D2, MYOZ2, NACC2, CDH20, TBXAS1, SCOC, MCC, SPRY1, ANAPC10, ZC3H8, LOC780876, VSTM2A, MAD2L1, PRRC2C*
Deletion	149	*AACS, SUPT6H, LRRC1, TMEM98, SPATA9, WBSCR17, STK24, CDCA7L, RNFT2, LSM14A, ERCC5, P33MONOX, LOC515976, SERPINB6, C5H12orf50, DDR1, TNFAIP8L1, KIAA0100, FAM155A, C17H5orf52, TSPAN8, CYP19A1, CRYBA1, CAP2, MIR2293, HSPB8, , HEBP1, UBLCP1, HAUS3, MKS1, TEX30, RTDR1, JUND, GAL3ST1, OSBP2, CA10, TRAF4, SLC25A2, PHF17, IGF2BP3, CACNA2D1, MDK, DCK, GNL1, P2RX7, PON3, PITPNM2, NEK8, ADCK1, EIF4E, GPR182, LALBA, ZIM2, TMEM167A, MALSU1, MIR135A-2, ANKRD50, THBS1, SYNPO2, LOC503858, FAM71D, COPS2, SPAG5, MED13L, TCN2, TPP2, PCDH10, NUDT6, MIR181D, CTRB1, SIRT4, LSM6, CALN1, ZNF423, GPNMB, TAOK3, C22H3orf18, ZNF582, SCPEP1, UNC13C, CCDC112, GABRB2, FHIT, MVK, SLCO4C1, GRIN2B, METTL21E, TAF1C, KDELC1, GRIN2C, NPY2R, MIR181C, SVOP, METTL21C, CHEK2, FOXP1, USH1G, FADS6, NQO2, E2F6, RPS19, INSIG1, PITPNB, CNOT2, INTU, FDXR, RBM19, GALNT14, KLHL7, PRKAB1, RAB34, ATP5I, NUPL2, MGC138914, DPP6, BIVM, SDF2, DGKB, IL27RA, PROCA1, MIR2444, FBXW7, SLC35E4, DUSP18, TLCD1, ETAA1, MGC157082, MRPL22, DNAH2, MGC152010, PES1, UBC, RPL23A, CD79A, CC2D1A, CRISP1, ADRBK2, DCAF15, IL12B, TMEM17, RTTN, IL15, EMP2, CLDN10, TAC1, PGGT1B, RAB28, SMTN*
Duplication	40	*PCCA, CARHSP1, IGFBP3, JUNB, RNASEH2A, HOOK2, ASNA1, C7H19orf43, GGACT, DCTN2, SLC12A6, WDR83, ATP5A1, DMRTC2, DTX3, MIR2420, TNPO2, ADRA1B, CUX2, LOC529425, MAN2B1, B4GALNT1, TMEM168, TBC1D8, SLC26A10, PRL, RAP1B, ZDHHC22, KIF5A, LYPD4, TMTC4, ITGBL1, PIP4K2C, ADCY1, DHPS, FBXW9, WDR83OS, VOPP1, PRDX2, ARHGEF25*