The Relationship between Copy Number Variations and Tick Resistance in South African Nguni Cattle

by
Leanne Hayley Pickering

Thesis presented in fulfilment of the requirements for the degree of
Master of Science (Agriculture)
in the Faculty of Agrisciences at
Stellenbosch University

Supervisor: Prof. Kennedy Dzama
Co-supervisor: Dr Magretha Diane Wang and Dr Farai Catherine Muchadeyi

December 2017

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.

_____

Date: December 2017

Abstract

The impact of ticks and tick-borne diseases on global livestock industries is an area of major concern. The ability of an animal to resist these arthropods varies within and between breeds. Tick resistance is a heritable trait, which can be exploited by using resilient breeds and incorporating them into selective breeding programmes. One such breed, known for its resilience to ticks and tick-borne diseases, is the South African Nguni. The Nguni is a locally adapted cattle breed, which has undergone minimal synthetic breeding, and is well adapted to harsh environmental conditions. Copy number variations (CNVs), present within the bovine genome, are attributable to the differences observed in adaptive and disease resistance traits in cattle. These variations comprise of deletions, duplications and insertions greater than 1kb in size. Copy number variable regions (CNVRs) overlap or lie within close proximity of genes responsible for multiple biological and molecular functions and could explain the underlying mechanisms of resistance. This study investigated the non-genetic effects of tick count and the association of CNVRs with tick resistance in South African Nguni cattle. In the first experiment, tick counts were recorded over a two-year period on 347 Nguni cattle across three different provinces in South Africa. Using SAS (Version 7.1) a general linear model was run on log transformed tick counts to determine the non-genetic effects of tick resistance. The effects of location, season, year of tick count, sex and age of the animal on tick count were tested. Factors which significantly affected tick resistance included location, season, year of tick count and the animal's age. In the second experiment, summary statistics of tick count per location were used to classify 347 Nguni cattle as susceptible (0) or resistant (1) across two levels of resistance (L1 and L2). Deoxyribonucleic acid extracted from hair and blood samples was genotyped using the Illumina BovineSNP 50 assay. After quality control and sample pruning using PLINK, 41 193 SNPs remained for further analyses. *PennCNV* identified 1 501 CNVs which were merged into 344 unique CNVRs. An association analyses using STATISTICA 64 was run which identified CNVRs associated with tick count. Seventeen CNVRs located on chromosomes 1, 2, 6, 7, 8, 9, 12, 15, 17, 20, 21, 22, 24 and 29 demonstrated a significant ($p<0.05$) association with tick resistance. Seventeen genes overlapped or lay in close proximity to these CNVRs and played a vital role in various molecular and biological processes. These processes all play an integral role in determining various cellular, immune, metabolic and reproductive responses.

Opsomming

Die globale impak van bosluise en bosluis-oordraagbare siektes op veebedrywe word as ’n belangrike kwessie beskou. Die vermoë van 'n dier om weerstand teen hierdie geleedpotiges te bied, wissel tussen rasse maar ook binne ’n ras. Bosluisweerstand is 'n oorerflike eienskap wat gebruik kan word om vir weerstandbiedende rasse te selekteer, wat op hulle beurt dan in selektiewe teelprogramme ingesluit kan word. Die Suid-Afrikaanse Nguni beesras is bekend vir sy vermoë om weerstand te bied teen bosluise en bosluisoordraagbare siektes. Die Nguni is 'n plaaslik aangepaste beesras wat minimale sintetiese teling en seleksie ondergaan het en is veral goed aangepas by moeilike omgewingstoestande. Kopie getalvariasies (CNVs) wat binne die beesgenoom voorkom, kan toegeskryf word aan die verskille in aanpassings en siekteweerstandseienskappe wat by beeste waargeneem word. Hierdie variasies bestaan uit verwyderde gene, duplikasies en invoegings groter as 1kb. Kopiegetal veranderlike streke (CNVRs) oorvleuel of lê naby aan gene wat verantwoordelik is vir verskeie biologiese en molekulêre funksies en kan moontlik ’n verklaring bied vir onderliggende weerstandsmeganismes. Hierdie studie het die verwantskap van CNVR's met bosluisweerstand in Suid-Afrikaanse Nguni-beeste ondersoek. In die eerste eksperiment is die aantal bosluise op 347 Nguni beeste, wat in drie verskillende provinsies in Suidd-Afrika voorgekom het, oor 'n tydperk van twee jaar aangeteken. Deur gebruik te maak van SAS (weergawe 7.1) is 'n algemene lineêre model analise op loggetransformeerde bosluisgetalle uitgevoer om die nie-genetiese effekte van bosluisweerstand te bepaal. Die effekte van die plek, seisoen, jaar van bosluistelling, geslag en ouderdom van die dier is in die analise ingesluit. Faktore wat ’n beduidende invloed op mate van bosluisweerstand gehad het, het plek, seisoen, jaar van bosluistelling en die dier ouderdom, ingesluit. In die tweede eksperiment is opsomming statistieke van bosluistelling per plek gebruik om die 347 Nguni-beeste as vatbaar (0) of weerstandig (1) oor twee vlakke van weerstand (L1 en L2) te klassifiseer. Deoksiribonukleïensuur wat uit hare en bloedmonsters verkry is, is aan genotipe analise met behulp van die Illumina BovineSNP 50-toets onderwerp. Na gehaltebeheer en verwydering van verdagte data met PLINK, het 41 193 SNPs vir verdere ontledings behoue gebly. PennCNV het 1 501 CNVs geïdentifiseer wat in 344 unieke CNVR's saamgevoeg is. 'n Assosiasie-ontleding met behulp van STATISTICA 64 het 17 beduidende CNVRs geïdentifiseer wat met bosluistelling geassosieer kon word. Sewentien gene oorvleuel of lê naby aan hierdie CNVRs en speel 'n belangrike rol in verskeie molekulêre en biologiese prosesse. Hierdie prosesse speel ’n belangrike in verskeie sel-, immuun-, metaboliese en reproduksie response.

Acknowledgements

First and foremost, I would like to express my sincerest gratitude to my supervisor, Kennedy Dzama. Thank you for your continued guidance, support and encouragement throughout this project.

I would also like to thank, Magretha Wang – my primary co-supervisor and mentor. The words 'thank you' will never be enough. Thank you for always challenging me and pushing me to think outside the box and excel at a level I never thought possible. Throughout this entire project you always encouraged and supported me, while still allowing me to work in my own way. I am immensely grateful for the time, patience and knowledge you have put into me and this project.

I would also like to acknowledge Farai Muchadeyi. Thank you for your valuable time, contributions, knowledge and effort into this project.

I would like to thank the National Research Foundation for their financial aid throughout the duration of this project.

This study would not have been possible without the help of Ntanganedzeni Olivia Mapholi. Thank you for assisting and providing me with the tick count data which was utilised in this study.

To my fellow students in the Animal Science department. Thank you for all being there to motivate and help one another when needed.

Timothy, Granny, Grandpa, Uncle Pick and Aunty Girlfriend. You have each contributed towards my success in your own individual ways. I am beyond grateful for your words of encouragement and motivation. Thank you.

Finally, I must express my profound gratitude and thanks to my parents. You have always provided me with unfailing support and encouragement. Thank you both for giving up so much to allow me to pursue my dreams. This accomplishment would never have been possible without you. I hope I have made you proud.

Table of Contents

List of Abbreviations

1st Qu – First Quartile

3rd Qu – Third Quartile

ANMLS – Number of animals

ARC – Agricultural Research Council

ArrayCGH - Array Comparative Genomic Hybridization

AS – De Novo Assembly

AvL – Average length

BAF – B Allele Frequency

BoLA – Bovine Lymphocyte Antigens

BP – Biological Process

CC – Cellular Component

CN – Copy number

CNV – Copy Number Variation

CNVRs – Copy Number Variable Regions

CV – Coefficient of Variation

DEL - Deletion

DNA – Deoxyribonucleic Acid

DUP - Duplication

EBVs – Estimated breeding values

EPG – Eggs per gram

FISH – Florescent In-situ Hybridization

GEN - Genes

GWA – Genome-wide Association

HMM – Hidden Markov Model

L1 – Resistance Level 1

L2 – Resistance Level 2

LRR – Log R Ratio

Max - Maximum

MaxL – Maximum length

MHC – Major Histocompatibility Complex

Min – Minimum

MinL – Minimum length

MF – Molecular Function

NGEN – Number of Genes

NGS – Next Generation Sequencing

QTL – Quantitative Trait Loci

RD – Read Depth

ResC – Resistance Category

ResL – Resistance Level

RP – Read Pair

RNA – Ribonucleic Acid

SAS – Statistical Analysis System

SCS – Somatic Cell Score

SD – Standard deviation

SNP – Single Nucleotide Polymorphism

SR – Split Read

WES – Whole Exome Sequencing

WGS – Whole Genome Sequencing

List of Genes

AIFM3 – Apoptosis inducing factor, mitochondria associated 3

ATP4B – ATPase H+/K+ transporting beta subunit

C14ORF79 – Uncharacterized protein

C1QA – Complement 1q subcomponent

CHCHD6 – Coiled-coil-helix-coiled-coil-helix domain containing 6

CTSB – Cathepsin B

GABRB2 – Gamma-aminobutyric acid A receptor, subunit beta 2

GRK1 – G protein-coupled receptor kinase 1

KYNU – Kynureninase

LSP1 – Lymphocyte-specific protein 1

LZTR1 – Leucine zipper like transcription regulator 1

MZT2 – Mitotic spindle organizing protein 2

PRODH – Proline dehydrogenase (oxidase) 1

SLC7A4 – Solute carrier family 7 member 4

TFDP1 – Transcription factor dp-1

THAP7 – THAP domain containing 7

TMCO3 – Transmembrane and coiled-coil domains 3

TNNT3 – Troponin T3, fast skeletal type

TUBA3E - Tubulin alpha 3-e

List of Tables

List of Figures

Chapter 1
General Introduction

1.1 Background

Beef and dairy products from cattle, are a key nutritional component to human diets (Canavez *et al.*, 2012). With the ever increasing human population, there is now a greater demand for meat and dairy products worldwide (Mapholi *et al.*, 2014). However, increased cattle production is hampered by the methods of control and effects of ticks and tick-borne diseases. Global economic losses due to ticks and tick-borne diseases have been estimated at US$13-18 billion annually (de Castro, 1997), with annual losses in South Africa estimated at US$92 million (Mapholi *et al.*, 2014).

Globally, there are approximately 800 breeds of cattle (Canavez *et al.*, 2012), with Africa being home to 150 of them (Rewe *et al.*, 2009). These breeds have been classified into two main species, *Bos taurus* (taurine) and *Bos indicus* (indicine), with their cross being defined as the Sanga (Ibeagha-Awemu *et al.*, 2004). Taurine and indicine breeds differ, with the latter, having a cervico-thoracic hump (Magee *et al.*, 2014) and greater physiological individualities (Canavez *et al.*, 2012).

Africa comprises predominantly of a tropical environment which presents a magnitude of harsh conditions to which cattle need to adapt in order to survive (Hoffmann, 2010; Mirkena *et al.*, 2010). The ability of local breeds to withstand severe climates, feed and water shortages and parasites and diseases, is much greater than that of exotic breeds (Mirkena *et al.*, 2010). The South African Nguni is one such breed known for its ability to withstand harsh environmental conditions and, in particular, its resilience to ticks and tick-borne diseases (Schoeman, 1989; Mirkena *et al.*, 2010). Ticks and the subsequent prevalence of tick-borne diseases, comprise a major constraint within the livestock industry causing anaemia, stress, irritation, decreased immune function, poor productivity and damage to hides (Rajput *et al.*, 2006; Abbas *et al.*, 2014). While a number of management control methods can be implemented, exploiting the innate resistance of locally adapted cattle breeds, can provide a long term solution to tick control (Mattioli *et al.*, 2000). The variation of fitness and adaptation evident within and between bovine breeds, could be due to discrepancies in genetic variation (Hoffmann, 2010; Valsesia *et al.*, 2013).

Genetic variation is the alteration of a DNA sequence in the genome (Feuk *et al.*, 2006). This can range from single base pair changes, known as single nucleotide polymorphisms (SNPs), to larger chromosomal anomalies (Freeman *et al.*, 2006; Liu *et al.*, 2010; Valsesia *et al.*, 2013). Deletion, duplication and insertion events within the genome, which are larger than 1 kb in size, are known as copy number variations (CNVs) (Tuzun *et al.*, 2005; Freeman *et al.*, 2006). When compared with SNPs, copy number (CN) changes can exert a much greater effect on the phenotype, through gene

structure and dosage modifications and/or the alteration of functional and regulatory processes (Zhang *et al.*, 2009; Gamazon & Stranger, 2015).

Copy number variations have been reported in a variety of cattle breeds (Bae *et al.*, 2010; Hou *et al.*, 2011; Liu & Bickhart, 2012). Genes detected within or near these copy number variable regions (CNVRs) are reported to play a role in a magnitude of functional and physiological processes (Kijas *et al.*, 2011). Differences in the frequency and distribution of CNVRs in these breeds, could explain the disparity of adaptive mechanisms and biological responses reported in cattle (Kijas *et al.*, 2011; Xu *et al.*, 2016). The recent discovery of CNVRs in Nguni cattle, and the influence of genes within these CNVRs, could provide evidence of the adaptive traits reported in this breed (Wang *et al.*, 2015).

1.2 Problem Statement

The economic losses experienced by farmers due to ticks and tick-borne diseases have become a grave concern. Current control methods have compromising side effects and impact on human and animal health and the environment. There is thus a need to develop alternative tick control strategies. Exploiting genetically resistant breeds and utilizing host resistance to alleviate ticks and tick-borne diseases is an alternative, permanent and cost effective solution for tick control. The South African Nguni is an indigenous Sanga type breed that is known for its resilience to ticks and tick-borne diseases. Genetic variations and their impact on phenotypic traits within cattle is not fully understood. Studying these variations within the bovine genome can provide some understanding regarding the differences in disease susceptibility and other adaptive traits in cattle. Originally, single nucleotide polymorphisms (SNPs) were considered to be the largest source of genetic variation, however copy number variations (CNVs) impact a larger percentage of the genome and could have greater effects. Copy number variations are segments of DNA that are larger or equal to 1 kb in size and consist of insertions, duplications and deletions across the genome. A previous study has reported CNVs within the Nguni (Wang *et al.*, 2015). These regions have been detected in and around genes responsible for multiple biological processes. The association of copy number variable regions (CNVRs) with tick resistance in South African Nguni cattle is unknown. Understanding and determining the role of CNVs in tick resistance, can provide valuable information for breeding and selection programmes in future.

1.3 Justification

Identifying genomic regions which impact disease susceptibility in cattle can have important implications for selection programmes. The impact of CNVRs on the resistance of Nguni cattle to ticks and tick-borne diseases has not yet been fully investigated. Understanding CNVRs and their

influence on economically important traits, can provide valuable information for genetic improvement practices. Nguni cattle show resilience to ticks and tick-borne diseases. With the presence of CNVs being reported in Nguni cattle (Wang *et al.*, 2015), this provides a valuable breed with which to investigate the association of CNVRs and tick resistance.

1.4 Objectives

This study aimed to identify CNVRs associated with tick resistance in South African Nguni cattle. The non-genetic effects on tick count in Nguni cattle across three different provinces in South Africa were investigated. The prevalence of CNVRs within the genome of the Nguni was examined as well as the biological processes, molecular functions and cellular components as to which these regions are involved. The association of CNVRs and tick resistance in Nguni cattle was investigated.

1.5 Thesis Overview and Layout

This dissertation comprises of five chapters including a general introduction chapter, a literature review, two experimental chapters and a general discussion and conclusion. The literature review (chapter 2) discusses tick resistance in cattle in light of CNVs. Chapter 3 presents the first experimental chapter where the non-genetic effects on tick count in Nguni cattle in South Africa were assessed. Tick counts of 347 Nguni cattle from three locations were recorded. Fixed effects were assessed using a general linear model. The association of CNVs with tick resistance in South African Nguni cattle was investigated in chapter 4. Copy number variations were identified using *PennCNV* software. Copy number variable regions were detected by merging overlapping CNVs. The maximum likelihood test assessed CNVRs and their association with tick resistance. Genes covered or lying within 10Mb of CNVRs identified were assessed using the *PANTHER* database to determine their biological processes, cellular components and molecular functions. A general discussion, conclusion, recommendations and ideas for future research form the final chapter. Here a critical discussion regarding the study is presented as well as implications and ideas for future research. Chapter 5 also provides details of conference presentations.

1.6 References

Abbas, R. Z., Zaman, M. A., Colwell, D. D., Gilleard, J., & Iqbal, Z. 2014. Acaricide resistance in cattle ticks and approaches to its management: The state of play. Vet. Parasitol. 203, 6–20.

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

Canavez, F. C., Luche, D. D., Stothard, P., Leite, K. R. M., Sousa-Canavez, J. M., Plastow, G., Meidanis, J., Souza, M. A., Feijao, P., Moore, S. S., & Camara-Lopes, L. H. 2012. Genome sequence and assembly of bos indicus. J. Hered. 103, 342–348.

de Castro, J. J. 1997. Sustainable tick and tickborne disease control in livestock improvement in developing countries. Vet. Parasitol. 71, 77–97.

Feuk, L., Carson, A. R., & Scherer, S. W. 2006. Structural variation in the human genome. Nat. Rev. Genet. 7, 85–97.

Freeman, J., Perry, G., & Feuk, L. 2006. Copy number variation: new insights in genome diversity. Genome Res., 16, 949–961.

Gamazon, E. R., & Stranger, B. E. 2015. The impact of human copy number variation on gene expression. Brief. Funct. Genomics. 14, 352–357.

Hoffmann, I. 2010. Climate change and the characterization, breeding and conservation of animal genetic resources. Anim. Genet. 41, 32–46.

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

Ibeagha-Awemu, E. M., Jann, O. C., Weimann, C., & Erhardt, G. 2004. Genetic diversity, introgression and relationships among West/Central African cattle breeds. Genet. Sel. Evol. 36, 673–690.

Kijas, J. W., Barendse, W., Barris, W., Harrison, B., McCulloch, R., McWilliam, S., & Whan, V. 2011. Analysis of copy number variants in the cattle genome. Gene. 482, 73–77.

Liu, G. E., Hou, Y., Zhu, B., Liu, G. E., Hou, Y., Zhu, B., Cardone, M. F., Jiang, L., Cellamare, A., Mitra, A., Alexander, L. J., Coutinho, L. L., Elena, M., Aquila, D., Gasbarre, L. C., Lacalandra, G., Li, R. W., Matukumalli, L. K., Nonneman, D., Regitano, L. C. D. A., Smith, T. P. L., Song, J., Sonstegard, T. S., Tassell, C. P. Van, Ventura, M., Eichler, E. E., Mcdaneld, T. G., & Keele, J. W. 2010. Analysis of copy number variations among diverse cattle breeds. Genome Res. 20, 693–703.

Liu, G. E., & Bickhart, D. M. 2012. Copy number variation in the cattle genome. Funct. Integr. Genomics. 12, 609–624.

Magee, D. A., MacHugh, D. E., & Edwards, C. J. 2014. Interrogation of modern and ancient genomes reveals the complex domestic history of cattle. Anim. Front. 4, 7–22.

Mapholi, N. O., Marufu, M. C., Maiwashe, A., Banga, C. B., Muchenje, V., MacNeil, M. D., Chimonyo, M., & Dzama, K. 2014. Towards a genomics approach to tick (Acari: Ixodidae) control in cattle: A review. Ticks Tick. Borne. Dis. 5, 475–483.

Mattioli, R. C., Pandey, V. S., Murray, M., & Fitzpatrick, J. L. 2000. Immunogenetic influences on tick resistance in African cattle with particular reference to trypanotolerant N'Dama (*Bos taurus*) and trypanosusceptible Gobra zebu (*Bos indicus*) cattle. Acta Trop. 75, 263–277.

Mirkena, T., Duguma, G., Haile, A., Tibbo, M., Okeyo, A. M., Wurzinger, M., & Sölkner, J. 2010. Genetics of adaptation in domestic farm animals: A review. Livest. Sci. 132, 1–12.

Rajput, Z. I., Hu, S., Chen, W., Arijo, A. G., & Xiao, C. 2006. Importance of ticks and their chemical and immunological control in livestock. J. Zhejiang Univ. Sci. B. 7, 912–21.

Rewe, T. O., Herold, P., Kahi, A. K., & Valle Zárate, A. 2009. Breeding indigenous cattle genetic resources for beef production in Sub-Saharan Africa. Outlook Agric. 38, 317–326.

Schoeman, S. J. 1989. Recent research into the production potential of indigenous cattle with special reference to the Sanga. S. Afr. J. Anim. Sci. 19, 55–61.

Tuzun, E., Sharp, A. J., Bailey, J. A., Kaul, R., Morrison, V. A., Pertz, L. M., Haugen, E., Hayden, H., Albertson, D., Pinkel, D., Olson, M. V, & Eichler, E. E. 2005. Fine-scale structural variation of the human genome. Nat. Genet. 37, 727–732.

Valsesia, A., Macé, A., Jacquemont, S., Beckmann, J. S., & Kutalik, Z. 2013. The growing importance of CNVs: New insights for detection and clinical interpretation. Front. Genet. 4, 1–19.

Wang, M. D., Dzama, K., Hefer, C. A., & Muchadeyi, F. C. 2015. Genomic population structure and prevalence of copy number variations in South African Nguni cattle. BMC Genomics. 16, 894.

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

Zhang, F., Gu, W., Hurles, M. E., & Lupski, J. R. 2009. Copy number variation in human health, disease, and evolution. Annu. Rev. Genomics. Hum. Genet. 10, 451–481.

Chapter 2
Literature Review

2.1 Introduction

Global livestock industries are affected by ticks and tick-borne diseases daily. Current control methods have been rendered ineffective and have compromising side effects; thus there is a need to develop alternative strategies. Utilizing genetically resistant breeds is one such method (Mattioli *et al.*, 2000). Disease susceptibility and resistance varies within and across breeds and can be affected by biological, morphological and physiological characteristics (Rajput *et al.*, 2006; Morris, 2007; Mapholi *et al.*, 2014). Understanding the genetic mechanisms associated with resistance, can provide valuable information for future selection and breeding programs (Ibeagha-Awemu *et al.*, 2008; Shyma *et al.*, 2013).

DNA variants in livestock, which directly affect an animal's level of resistance, are an important area of research (Kijas *et al.*, 2011). Genetic variation includes single nucleotide polymorphisms (SNPs) and structural variants which comprise of inversions, translocations and copy number variations (CNVs) (Feuk *et al.*, 2006; Liu *et al.*, 2010). Copy number variations are segments of DNA which are larger than 1 kb and consist of duplications, deletions and insertions (Feuk *et al.*, 2006; Bae *et al.*, 2010; Liu *et al.*, 2010). These variations can have an impact on the function and/or expression of genes and/or proteins which could directly influence an animal's resistance to disease (Ibeagha-Awemu *et al.*, 2008).

Identification of CNVs can be achieved through cytogenetic technologies, array hybridization approaches and next generation sequencing (NGS) technologies (Xu *et al.*, 2013; Zhao *et al.*, 2013). Following CNV detection, genome wide association (GWA) analyses can be conducted to identify CNVs and their association with particular genes or mechanisms involved in tick resistance (Lohmueller *et al.*, 2003; Bush & Moore, 2012). This review discusses tick resistance in light of copy number variations. The methods of CNV detection as well as models for CNV association are discussed.

2.2 Ticks and Tick-borne Diseases

Domestic animals, humans and livestock are all affected by the diseases transmitted by ticks (Jongejan & Uilenberg, 2004). Ticks are blood sucking ectoparasitic arthropods that are distributed worldwide, and have many animal species as hosts (Porto Neto *et al.*, 2011). With approximately 900 species of ticks that are endemic to most continents (Nava *et al.*, 2009), many different tick-borne diseases are carried by these arthropods. Of all the groups of arthropods, ticks transmit the

widest variety of pathogenic agents (Rajput *et al.*, 2006). This has led to farmers ranking these parasites as one of the most important constraints in the cattle industry (Mapiye *et al.*, 2009).

Argasidae, Ixodidae and Nuttalliellidae are the three families into which tick species are classified (Nava *et al.*, 2009). Ixodidae ticks comprise seven genera with a total of 692 species, with Argaside ticks comprising four genera with a total of 186 species (Jongejan & Uilenberg, 2004; Nava *et al.*, 2009). The genera into which Argaside ticks are divided into include *Argas, Carios, Ornithodoras* and *Otobius* with the genera of Ixodidae ticks consisting of *Amblyomma*, *Dermacentor*, *Haemaphysalis*, *Hyalomma*, *Ixodes* and *Rhipicephalus* (which includes the new subgenus – *Boophilus*) (Jongejan & Uilenberg, 2004). The Nuttalliedae family only consists of one species, *Nutalliella namaqua* (Jongejan & Uilenberg, 2004).

In the African bovine industry, Ixodidae tick species are the most prevalent and cause severe impediments, with the *Amblyomma*, *Hyalomma* and *Rhipicephalus* generas being of utmost relevance (Mapholi *et al.*, 2014). One species of tick which falls into the *Rhipicephalus* genus, is *R. microplus.* This species is the most studied (Frisch, 1999) and is globally known as the most significant and detrimental blood feeding arthropod in the livestock industry *(*Abbas *et al., 2014).*

Ticks transmit a number of diseases including babesiosis, anaplasmosis, heartwater and theileriosis (Mtshali *et al.*, 2004). In addition, tick infestations cause anaemia, blood loss, decreased immune function, hide and skin damage, irritation, paralysis, stress, toxicosis, weight loss and even death (de Castro, 1997; Rajput *et al.*, 2006; Abbas *et al.*, 2014). These direct losses, as well as the cost of the methods to control infestations, hamper livestock production, which in turn leads to large economic losses.

2.3 Factors Affecting Tick Load

Tick load is affected by species, climatic conditions and the susceptibility of the animal to tick infestation (Jongejan & Uilenberg, 2004). In tropical and subtropical regions, the occurrence of tick infestations and tick diversity is greater than in the temperate regions (Jongejan & Uilenberg, 2004). In general, areas with high rainfall and good vegetation, will have greater tick loads (Mapholi *et al.*, 2014). Large parts of South Africa have a subtropical climate with either sweet or sour rangelands (Masika *et al.*, 1997). Sweet rangeland is palatable throughout the year, while sour rangelands are only palatable during the rainy season (Marufu *et al.*, 2011a). Tick loads tend to be highest in sour rangelands and during the warmer, wetter months (Muchenje *et al.*, 2008; Mapiye *et al.*, 2009; Marufu *et al.*, 2011a). Elevated rainfall promotes tick growth particularly in semi-arid coastal climatic conditions and may explain the higher tick burden of the wetter months (Muchenje *et al.*, 2008).

Herd size and the age of the animal also affect tick burden. Diseases and parasites are more problematic in communal areas rather than small scale areas (Mapiye *et al.*, 2009). Communal farming areas are those in which many farmers share the same land for multiple farming practices (Mmbengwa, 2015). Small-scale farming areas are farms smaller than 12 hectares and are run primarily by the owners and/or family living on that land. Livestock are managed under extensive systems and have access to natural pastures (Moyo *et al.*, 2008; Mapiye *et al.*, 2011). Marufu *et al.* (2010) observed age of an animal to have an effect on tick load with younger cattle having lower tick counts than those aged four years and older. Wikel & Bergman (1997) stated that younger animals could have lower tick counts due to an innate form of protection that declines with age. Older animals are larger and thus have a greater surface area which could also attribute to the fact that older animals have higher tick loads (Marufu *et al.*, 2011a).

2.4 Methods of Tick Control

Multiple methods of tick control exist, each with its own advantages and disadvantages (Rajput *et al.*, 2006). Tick control methods include the use of acaricides (chemical control), biological control, environmental management, vaccines and breeding strategies (Frisch, 1999). The utilisation of only one control method has been deemed ineffective and therefore it is recommended that more than one strategy should be implemented (de Castro, 1997; Kamidi & Kamidi, 2005).

The predominant approach to tick control is the use of chemical acaricides (Frisch, 1999; Abbas *et al.*, 2014). These substances are quick and cost effective, however they have shown limited success due to the development of resistance and environmental contamination (Frisch, 1999; Kamidi & Kamidi, 2005; Abbas *et al.*, 2014). Consequently, to try and curb the problems encountered by chemical control, vaccines were developed (Willadsen, 1997). Vaccines are non-chemical disease and pest control agents, which allow an organism to develop antibodies against a specific number of diseases (Willadsen, 1997). The advantages of vaccines is that they reduce environmental contamination and prevent drug resistant tick species (de la Fuente *et al.*, 2007) however, they have their own concerns. The development of vaccines is complicated by antigenic variation and strain diversity and thus the efficacy of vaccines varies from region to region (Rajput *et al.*, 2006; Abbas *et al.*, 2014). With multiple species of ticks, and infestation not being a one species issue, some vaccines may not offer protection against all species (Mapholi *et al.*, 2014). Therefore, the development of multi-strain or strain specific vaccines, that are cheap, robust and effective would be beneficial (Frisch, 1999).

Environmental management entails a variety of techniques including the use of grasses which do not favour the development of ticks, burning, heavy grazing and/or rotation of pastures (de Castro, 1997; Abbas *et al.*, 2014). These methods prevent chemical residues in livestock products and are

economically viable; however they can be harmful to the environment and are not always recommended (de Castro, 1997).

Current tick control methods have compromising side effects and there is thus a need to develop alternate control strategies (Porto Neto *et al.*, 2011). The integration and exploitation of genetically resistant cattle breeds into breeding programmes is one such method. Utilizing tick resistant breeds can provide a long term solution to tick control as well as decrease the use of chemicals, improve drug efficiency and minimise production costs (Mattioli *et al.*, 2000).

2.5 Tick Resistant Breeds

One hundred and fifty different cattle breeds inhabit Africa (Rewe *et al.*, 2009). These breeds predominantly comprise of two main subspecies, *Bos taurus indicus* (indicine) and *Bos taurus taurus* (taurine), but also include respective crosses of these subspecies (Canavez *et al.*, 2012). Sanga breeds are a cross between indicine and taurine cattle, while Sanga zebu type breeds ("zenga") are a cross between Sanga and indicine breeds (*Bos indicus)* (Rege, 1999; Ibeagha-Awemu *et al.*, 2004). Indicine cattle not only differ from taurine cattle by the presence of a cervico-thoracic hump, but are also known for their advantageous physiological individualities (Canavez *et al.*, 2012; Magee *et al.*, 2014).

Most of Africa comprises a tropical environment with stressful climatic conditions, feed and water scarcity and a magnitude of diseases and pathogens (Hoffmann, 2010; Mirkena *et al.*, 2010). The adaptability of animals to such conditions varies between breeds, with indicine breeds being better suited when compared with taurine breeds (Muchenje *et al.*, 2008; Marufu *et al.*, 2011b). Adaptability is the capability of a breed or species, to survive or reproduce in a magnitude of environments due to physiological or genetic factors (Prayaga & Henshall, 2005; Barker., 2009), and is characterised by health, survival and reproductive traits (Makina *et al.*, 2014). Locally adapted African breeds, such as the Nguni (*Bos taurus africanus)*, are better suited to severe climatic conditions and display greater levels of adaptation and resistance to diseases and parasites (Mirkena *et al.*, 2010) due to biological and functional characteristics (Mattioli *et al.*, 2000).

The South African Nguni is an indigenous Sanga cattle breed that has undergone minimal synthetic breeding (Makina *et al.*, 2014). These small to medium framed animals are easily recognized by their multi-coloured patterned coats and black tipped noses. Nguni cattle are adapted to harsh environmental conditions, specifically high temperatures, and are well known for their resilience to ticks and tick-borne diseases (Schoeman, 1989). The mechanism for this resistance has not yet been established (Marufu *et al.*, 2011b), but could be due to favourable genotypic and/or phenotypic attributes (Marufu *et al.*, 2011b). The ability of Nguni cattle to be reared on natural pastures in South

Africa without the use of tick control methods is largely beneficial (Schoeman, 1989; Muchenje *et al.*, 2008). It is recommended that this indigenous breed be used in an integrated tick control strategy in South Africa, as they are more resistant to tick infestations than indigenous exotic crossbreeds suggesting that they could have a higher innate and/or acquired resistance (Marufu *et al.*, 2011a). Resistance is an acquired characteristic that varies from high to low, depending on the animal's response to infestation (Rajput *et al.*, 2006; Morris, 2007). *Bos indicus* breeds are more resistant to ticks and tick-borne diseases than *Bos taurus*, with their crosses losses being proportional to the amount of *Bos indicus* genes (Utech *et al.*, 1978; de Castro & Newson, 1993; Jonsson *et al.*, 2008). Host resistance is genetically based (Gasparin *et al.*, 2007) and can be applied by crossbreeding with *Bos indicus* or other tick resistant breeds (de Castro, 1997). By exploiting genetically resistant breeds and utilising host resistance, an effective and permanent method of tick control can be developed.

2.6 Host Resistance

Variation in disease susceptibility and resistance occurs within breeds (Rajput *et al.*, 2006; Morris, 2007). The genetic basis of parasite resistance is important for human health, animal welfare and animal production, however it is not well defined (Xu *et al.*, 2014b). Resistance to disease is a polygenic trait (Shyma *et al.*, 2013) and it is possible that multiple genes contribute to important disease characteristics (Morris, 2007). Behavioural, morphological and physiological factors vastly affect a host's resistance to ticks (Mapholi *et al.*, 2014), however this genetic variation within and between breeds needs to be explored in order for it to be efficiently utilised (Frisch, 1999).

Breed, age, sex, body size and coat characteristics of the animal affect levels of resistance in cattle. Martinez *et al.* (2006) found bulls to have higher tick loads than cows, which could be explained by the fact that testosterone can reduce acquired and innate resistance (Hughes & Randolph, 2001). However, Kabir *et al*. (2011) reported higher tick loads in cows and hypothesized that this could be due to hormonal influences. There are also differences in the level of resistance between pregnant and non-pregnant cows as immunosuppressive effects of gestational hormones cause pregnant cows to be more susceptible to ticks than those that are not pregnant (Regitano & Prayaga, 2010; Mapholi *et al.*, 2014).

The size of an animal and its coat characteristics all affect levels of tick infestation (Martinez *et al.*, 2006; Machado *et al.*, 2010; Mapholi *et al.*, 2014). Highly heritable traits including coat colour, hair type, hair length and thickness all affect tick load in cattle (Regitano & Prayaga, 2010). Skin thickness plays a role in susceptibility as animals with thicker skins are less susceptible to tick infestation than those with thinner skins (Marufu *et al.*, 2011b). Animals with lighter, shorter, smoother and straighter coats have lower tick infestations (Gasparin *et al.*, 2007; Machado *et al.*, 2010; Marufu *et al.*, 2011b).

Animals with darker coats could have more ticks due to the fact that ticks are dark coloured and thus are camouflaged from predators, such as birds (Martinez *et al.*, 2006). There is a high correlation between hair length and tick infestation suggesting that this particular trait may be involved in host resistance (Verissimo *et al.*, 2002).

Host resistance is a genetically determined survival mechanism utilized by the host as well as parasites (Jongejan & Uilenberg, 2004; Rajput *et al.*, 2006). Parasites develop their own mechanisms of resistance through the suppression of some immune response pathways (Wikel & Bergman, 1997). The utilization of host resistance is a significant factor in the economics of tick control (Frisch, 1999). Tick control using host resistance is a cost effective, permanent solution that is heritable and indefinite (de Castro, 1997), however no single breed is totally resistant (Shyma *et al.*, 2013). Improving breed resistance is the ultimate goal, which can be achieved by using resistance genes from breeds of high resistance and introgressing them into breeds with low resistance (Shyma *et al.*, 2013).

2.7 Heritability of Tick Resistance

The pattern of tick infestations and factors influencing tick load in cattle have led authors to conclude that tick count, and hence tick resistance, is heritable (Rajput *et al.*, 2006; Morris, 2007; Porto Neto *et al.*, 2011). Studies reporting resistance levels of multiple breeds provide the basis for breeding and selection strategies (Utech *et al.*, 1978). Using estimated breeding values (EBVs) to select for tick resistance is feasible, however genetic progress may be slow (Budeli *et al.*, 2009).

Studies from 1993 to 2016 have found heritability estimates of tick resistance to range from 0.15 to 0.44 in multiple cattle breeds (Table 2.1). Porto Neto *et al.* (2011), report the average heritability for tick resistance of multiple cattle breeds to be ~0.30 (Porto Neto *et al.*, 2011). A study conducted by Mapholi *et al*. (2016) on Nguni cattle found the heritability of tick resistance to be 0.12. This estimate is low compared to studies conducted by Davis (1993) and Burrow (2001) who investigated heritability estimates of Zebu cross cattle. However, Davis (1993) report on *Bos indicus* and Zebu cross breeds while Burrow (2001) report on Belmont Red, which are cross breeds of Hereford, Shorthorn, Afrikaner and Brahman. Differences in heritability estimates may be due to the breed of cattle, evaluation methods (natural vs artificial) or environmental and immunological factors (Porto Neto *et al.*, 2011; Shyma *et al.*, 2013). The heritability estimates generally range between moderate to high with a value usually greater than 0.20 (Burrow, 2001; Budeli *et al.*, 2009; Porto Neto *et al.*, 2011). The heritability of tick resistance is higher in females than in males (Burrow, 2001) and therefore it is important to identify and select tick-resistant females with which to breed (Marufu *et al.*, 2011b).

Production traits in cattle, have been reported to be correlated with tick resistance. Burrow (2001) found that selecting to increase final weights and average daily gain during wet seasons and post weaning could decrease tick counts. It has also been reported that when selecting for tick resistance, resistance against other factors such as worms and heat stress, increased (Burrow, 2001).

Table 2.1 Heritability estimates (h^2) of tick resistance in a number of cattle breeds from different locations reported between 1993 and 2016.

Breed	Location	h^2	Reference
Bos indicus and Zebu cross breeds (Beef breeds)	Australia	0.34	Davis (1993)
Hereford x Shorthorn	Australia	0.44	Henshall (2004)
Synthetic breed Belmont Red (Hereford, Shorthorn, Afrikaner and Brahman crosses)	Australia	0.42	Burrow (2001)
Bonsmara (*Bos taurus)*	South Africa	0.17	Budeli *et al.* (2009)
Brahman	Australia	0.15	Prayaga *et al.* (2009)
Bos taurus (dairy breeds)	Australia	0.37	Turner *et al.* (2010)
Gir x Holstein	Brazil	0.21	Machado *et al.* (2010)
Nguni	South Africa	0.12	Mapholi *et al*. (2016)

2.8 Genomics of Tick Resistance

Tick resistance is a complex phenotype comprising a number of different physiological processes that are controlled by a magnitude of genes (Machado *et al.*, 2010). The challenge is to identify and understand the biological and physiological mechanisms of these genes so that they can be integrated into future breeding programmes (Ibeagha-Awemu *et al.*, 2008; Shyma *et al.*, 2013). A number of studies have detected candidate genes that could be responsible for and associated with the differing levels of tick resistance between breeds (Wang *et al.*, 2007b; Carvalho *et al.*, 2008; Ibeagha-Awemu *et al.*, 2008).

An early study by Ashton *et al*. (1968) reports an association between serum amylase C and the level of tick infestation. More recently, the response of acute-phase proteins in *Bos taurus* and *Bos indicus* cattle to natural tick infestation was evaluated (Carvalho *et al.*, 2008). This study identified an increase in acute-phase proteins with greater tick infestation. Acute-phase proteins are responsible for acquired and innate immune responses (Carvalho *et al.*, 2008).The serum concentration levels of different proteins, such as haptoglobulin and transferrin, have been confirmed

and could be used as potential biomarkers to monitor levels of infestation (Porto Neto *et al.*, 2011; Shyma *et al.*, 2013).

Gene expression profiles of susceptible and resistant cattle have provided candidate genes for tick resistance (Wang *et al.*, 2007b). Type I, III and V collagen genes were expressed at higher levels in resistant animals in response to tick infestation, suggesting the role of skin structure against ticks (Wang *et al.*, 2007b). Other genes expressed at greater levels in resistant animals included *complement 1q subcomponent* (*C1QA*) and *cathepsin B* (*CTSB*), which are both related to immune response (Wang *et al.*, 2007b).

Gene variations of the major histocompatibility complex (MHC) are thought to be associated with the susceptibility of cattle to diseases (Acosta-Rodríguez *et al.*, 2005; Ibeagha-Awemu *et al.*, 2008). Bovine lymphocyte antigens (BoLA) are antigens found within the MHC on chromosome 23 and comprise of three classes, class I, II and III (Ibeagha-Awemu *et al.*, 2008). Class I and II molecules signal infection and are significant for regulation of immune functions (Ellis, 2004; Ibeagha-Awemu *et al.*, 2008). Class III molecules are involved in numerous cellular activities including cell cytotoxicity and enzymatic activity (Sharif *et al.*, 1999). It has been observed that there are some MHC BoLA class II alleles that can be associated with tick resistance or predisposition in cattle (Acosta-Rodríguez *et al.*, 2005), however further research using larger sample groups and/or different animal populations has been recommended. Although the MHC locus does significantly enact a role in host parasite resistance, the specific allele or gene variation is unknown (Porto Neto *et al.*, 2011; Shyma *et al.*, 2013).

Quantitative trait loci (QTL) are segments of a chromosome showing Mendelian inheritance and have an effect on a specific phenotype (Machado *et al.*, 2010). The discovery of specific QTLs in cattle could be used to further investigate and isolate specific genes involved in the mechanisms of resistance in cattle (Gasparin *et al.*, 2007; Coppieters *et al.*, 2009; Machado *et al.*, 2010). However, it is difficult to identify a particular gene as a potential marker for tick resistance as QTLs cover large percentages of the chromosome (Shyma *et al.*, 2013). Machado *et al.* (2010) conducted a whole genome scan, which led to the identification of six QTL regions associated with tick resistance in cattle on chromosomes 2, 5, 10, 11, 23 and 27. Most of the identified QTLs role in tick resistance were season specific, however a QTL discovered on BTA23 was prevalent in tick resistance in both the dry and rainy seasons. This genomic region on BTA23 has previously been related with the bovine histocompatibility complex (Stear *et al.*, 1990). Previously, Gasparin *et al.* (2007), had done a similar study on the same population and discovered QTL controlling tick resistance on BTA 5, 7 and 14. When identifying QTL markers for tick resistance, it is important to recognise the causal variant for future breeding and selection (Wu *et al.*, 2015).

2.9 Genetic Variation

Identifying DNA variants which directly influence an individual's phenotype is an important area of research (Kijas *et al.*, 2011). To gauge the genetic basis of the phenotypic differences within and between species, it is imperative to understand all forms of genetic variation (Liu & Bickhart, 2012). Multiple forms of genetic variation exist, ranging from single nucleotide polymorphisms (SNPs) to large chromosomal rearrangements (Feuk *et al.*, 2006; Ibeagha-Awemu *et al.*, 2008; Liu *et al.*, 2010). Chromosomal rearrangements include copy number variations (CNVs), inversions and translocations, which all fall under the umbrella term, structural variation (Feuk *et al.*, 2006; Liu *et al.*, 2010). Copy number variations comprise deletions, duplications and insertions across the genome (Feuk *et al.*, 2006; Bae *et al.*, 2010; Liu *et al.*, 2010) and can be found within the coding and regulatory regions of genes (Ibeagha-Awemu *et al.*, 2008). This type of structural variant can modify the expression and function of genes and/or proteins, which could have detrimental or beneficial consequences on an animal's health, productivity and/or disease susceptibility (Ibeagha-Awemu *et al.*, 2008). Studying the bovine genome and the genetic variations within can provide valuable information (Feuk *et al.*, 2006). By combining phenotypic data with genotypic variation, significant genotype-phenotype associations can be discovered (Zhan *et al.*, 2011).

2.9.1 Single nucleotide polymorphisms

Single nucleotide polymorphisms are single base-pair changes that occur at a frequency of greater than 1% (Zhang *et al.*, 2009) and were originally thought to be the largest source of genetic variation (Feuk *et al.*, 2006). Using genome-wide association (GWA) analysis, SNPs can be used to identify and map genetic variants associated with complex traits in cattle (Mapholi *et al.*, 2016). A few studies have linked SNP markers to tick resistance in cattle (Sollero *et al.*, 2014; Mapholi *et al.*, 2016). Sollero *et al*., (2014) found SNP markers on BTA 5, 11 and 15 to be associated with tick resistance in Brazilian Hereford and Braford cattle, while Mapholi *et al*., (2016) found associated regions on BTA 7, 10 and 19 in Nguni cattle. These discrepancies, provide evidence of different genomic regions being associated with tick resistance in different cattle breeds (Mapholi *et al.*, 2016). Although SNPs have been associated with complex traits in cattle, they only describe a small proportion of genetic variance (Manolio *et al.*, 2009). Research into the nature and pattern of SNPs in cattle has made significant progress, plus the discovery of more complex forms of genetic variation has expanded the current areas of research (Liu & Bickhart, 2012). Other forms of variation may affect disease resistance and phenotypic characteristics on a greater scale (Kijas *et al.*, 2011). Complementary to SNPs, variation in the form of CNVs has attracted recent attention (Gao *et al.*, 2017).

2.9.2 Copy number variations

Copy number variations are segments of DNA larger than or equal to 1 kb in size and consist of insertions, duplications and deletions across the genome (Bae *et al.*, 2010; Mills *et al.*, 2011; Wang *et al.*, 2016).These effects can interfere with a single gene or a magnitude of genes and cause functional losses within the phenotype (Zhang *et al.*, 2009; Liu & Bickhart, 2012).

Copy number variations have been discovered within the bovine genome (Liu *et al.*, 2010; Kijas *et al.*, 2011; Bickhart *et al.*, 2012). The publication of two cattle reference genomes, Btau_4.0 and UMD3.1 (The Bovine Consortium Hapmap, 2009), has enriched bovine genomic research (Bickhart *et al.*, 2012). These reference genomes and the advances in genome-wide technologies has provided a means to detect genomic regions impacting phenotypic variation in cattle (Fadista *et al.*, 2010; Wang *et al.*, 2015). Copy number variation detection studies have been conducted across a variety of cattle breeds (Liu *et al.*, 2010; Kijas *et al.*, 2011; Bickhart *et al.*, 2012; Wang *et al.*, 2015). Genes detected within cattle CNVs are responsible for multiple biological and molecular processes and could play a role in a variety of adaptive mechanisms (Wang *et al.*, 2015).

The role CNVs play in breed development, adaptation, health, and production traits contributes to varying frequencies of CNVs within and between cattle breeds (Liu *et al.*, 2010; Liu & Bickhart, 2012). The association of CNVs with complex traits in cattle lacks research and understanding (Bae *et al.*, 2010; Cicconardi *et al.*, 2013; Wang *et al.*, 2016). Some studies have associated CNVs with complex traits in cattle, including milk production (Xu *et al.*, 2014a), meat tenderness (Da Silva *et al.*, 2016), fertility (Glick *et al.*, 2011) and nematode resistance (Xu *et al.*, 2014b), but no study has associated CNVs with tick resistance.

The South African Nguni is a locally adapted breed that is known for its resilience to ticks and tick-borne diseases (Schoeman, 1989). The mechanisms underlying this are, however, not fully understood. The presence of CNVs within Nguni cattle has been reported (Wang *et al.*, 2015). Genes identified within these CNVs are responsible for multiple biological processes including immunity, cell communication and toxic substance responses. The presence of copy number variable regions (CNVRs) in areas associated with genes involved in these processes, suggests that CNVs in Nguni cattle could play an important role in the adaptation of the Nguni breed (Wang *et al.*, 2015).

2.10 Copy Number Variation Detection and Analysis Tools

Current CNV detection methods include cytogenetic technologies, hybridization-based microarray approaches and next generation sequencing (NGS) techniques (Xu *et al.*, 2013; Zhao *et al.*, 2013). Cytogenetic technologies include karyotyping and fluorescence in situ hybridization (FISH), however

these methods have low throughput and resolution (Alkan *et al.*, 2011). Microarray approaches, including array comparative genomic hybridization (arrayCGH) and SNP arrays were the primary platform of CNV detection (Xu *et al.*, 2013), yet the recent advances in NGS technologies have provided enhanced resolution and accuracy in bovine CNV detection (Choi *et al.*, 2016). Strategies for NGS include whole genome sequencing (WGS) and whole exome sequencing (WES) (Zare *et al.*, 2017). Although the advantages of NGS are evident, the high costs are still a diminishing factor (Choi *et al.*, 2016).

2.10.1 Hybridization based array approaches

Array comparative genomic hybridization and SNP arrays are two hybridization based array approaches utilised for SNP genotyping and CNV detection. Both methods utilise probes predefined for specific genomic regions (Zhao *et al.*, 2013) and compare these to a reference or sample population thereby deducing copy number deletions or duplications (Alkan *et al.*, 2011). The low cost of these approaches and their ability to assay large data sets are beneficial, however they are limiting as they do not detect duplications as easily as deletions (Alkan *et al.*, 2011). Other limitations include poor precision and sensitivity, low resolution, minimal coverage of the genome and difficulty in detecting rare mutations (Snijders *et al.*, 2001; Shendure & Ji, 2008; Zare *et al.*, 2017).

Array comparative genomic hybridization examines the genome by measuring hybridization intensities against a reference DNA sample and detecting gains or losses (Liu & Bickhart, 2012). Compared to SNP arrays, CGH arrays show a better signal-to-noise ratio (Pinto *et al.*, 2011), but only report the relative signal intensities (Hou *et al.*, 2011). Algorithms to detect CNVs from CGH array data have been developed, however these detection tools do not take into account B allele frequencies (BAFs) which is concerning as BAFs are an important source of data for use in CNV detection from SNP data (Xu *et al.*, 2013).

Single nucleotide polymorphism arrays, such as the Illumina BovineSNP50 Beadchip and the Higher density SNP (770K) can be used for CNV detection (Cicconardi *et al.*, 2013; Xu *et al.*, 2013). Although they are predominantly used for SNP genotyping and have lower resolution and sensitivity when compared with CGH arrays, the advantages of SNP arrays for copy number detection is unparalleled (Alkan *et al.*, 2011; Liu & Bickhart, 2012). These arrays output normalized intensities (Log R ratio – LRR) and allelic intensity ratios (B allele frequency – BAF) (Cicconardi *et al.*, 2013). The Log R ratio is the total fluorescent intensity signal from both sets of probes or alleles at each SNP (Wang *et al.*, 2007a). The BAF is the inference of the relative ratio of the fluorescent signals between two probes or alleles at each SNP (Wang *et al.*, 2007a). To detect CNVs, several algorithms utilizing the fluorescent signal intensities (LRR and BAF) from SNP arrays have been developed (Cicconardi *et al.*, 2013; Zhu *et al.*, 2016). Some of these CNV detection algorithms, which use SNP

chip data, include *PennCNV*, QuantiSNP and cnvPartition (Zhu *et al.*, 2016). QuantiSNP and *PennCNV* utilise a hidden markov model (HMM) to detect CNVs, however *PennCNV* is more commonly used due to its user friendly design and superior advantages (Wang *et al.*, 2007a; Cicconardi *et al.*, 2013). *PennCNV* is an open-source project (Xu *et al.*, 2013) that not only incorporates the LRR and BAF at each SNP marker, but also takes into account the distance between neighbouring SNPs as well as an option to include pedigree information (Seroussi *et al.*, 2010; Hou *et al.*, 2011). Data quality control is also reported by *PennCNV* for each specific CNV dataset (Xu *et al.*, 2013).

2.10.2 Next generation sequencing based approaches

Next generation sequencing is a high-throughput parallel/deep sequencing technology used to sequence DNA and RNA (Patel & Jain, 2012; Behjati & Tarpey, 2013). Copy number variations can be detected by multiple NGS based approaches including pair end mapping or read pair (RP), read depth (RD), split read (SR), *de novo* assembly (AS) or a combination of these methods (Liu & Bickhart, 2012; Zhao *et al.*, 2013). Each method has its own advantages and disadvantages, however none of these approaches are individually comprehensive (Alkan *et al.*, 2011; Zhao *et al.*, 2013). Each approach has a variety of tools which can be used to detect CNVs (Table 2.2). Although it has been stated that NGS techniques will replace microarray approaches for CNV detection in the future, this approach still presents several computational and bioinformatical challenges (Alkan *et al.*, 2011). A disadvantage of NGS is the high cost due to the storage and analysis of the data requiring intense computational resources, however their advantages outweigh the financial strain (Alkan *et al.*, 2011). Next generation sequencing is unbiased and provides high coverage and resolution as well as enhanced accuracy and precision (Alkan *et al.*, 2011; Zare *et al.*, 2017). At a genome wide level NGS technologies also facilitate the vast discovery of CNVs and SNPs which can be utilised for future discoveries and genotyping (Liu & Bickhart, 2012).

Table 2.2 Next generation sequencing approaches for CNV detection

Approach	Computational tools	Pros	Cons	References
Read-pair (RP)	Breakdancer, PEMer, VariationHunter, commonLAW, GASV, MoDIL	Can identify insertions, deletions, translocations, inversions and interspersed and tandem duplications	Cannot accurately identify copy numbers; costly; only for paired-end reads; cannot detect insertions larger than the insert size of the genome library	Korbel *et al.* (2007); Medvedev *et al.* (2009); Zhao *et al.* (2013)
Read depth (RD)	CNVnator, SegSeq, CNV-seq, RDXplorer, BIC-seq, CNAseg, Cn.MOPS, JointSLM, ReadDepth, rSW-seq, CNVnorm, CMDS, mrCaNaVar, CNVem, cnvHMM	Accurately predicts copy numbers; Can detect large insertions and complex CNVs	Relies on the length of reads; Cannot identify inversions, translocations or small CNVs	Alkan *et al.* (2009); Zhao *et al.* (2013)
Split read (SR)	Alignment with Gap Excision (AGE), Pindel, SLOPE, Split-read identification, calibrated (SRiC)	Can detect a wide range of structural variations; Accurately detects start and end positions of deletions and/or insertions	Struggles to align shorter reads; Only for unique genomic regions; Low sensitivity	Alkan *et al.* (2011); Zhao *et al.* (2013)
De novo assembly (AS)	Magnolya, Cortex assembler, TIGRA-SV, EULER-USR, ABySS, SOAPdenovo, ALLPATHS-LG	Most versatile; Does not require a reference genome; Can identify novel mutations	Struggles to identify duplications; Requires extensive computation	Alkan *et al.* (2011); Zhao *et al.* (2013)

Combination	RP + RD - SVDetect, CNVer, Genome STRiP, GASVPro, inGAP-sv RP + SR - NovelSeq RP + AS - HYDRA	Reduces false positives Detect novel insertions	Zhao *et al.* (2013)

Abbreviations: RP + RD – Read pair and read depth; RP + SR – Read pair and split read; RP + AS – Read pair and *de novo* assembly

2.11 Copy Number Variation Association Protocols

Genome-wide association (GWA) analyses are used to identify genetic variants that impact complex traits, such as disease susceptibility (Lohmueller *et al.*, 2003). The aim of these studies is to predict the specific outcome of an individual and identify the underlying genetic mechanisms for future prevention and control strategies (Bush & Moore, 2012). Association analysis using SNPs is far more established than CNV association analysis (Kim *et al.*, 2012). Many CNV calling algorithms exist, however software for CNV association is limited (Glessner *et al.*, 2013) and the statistical challenges faced are significantly different to those of CNV discovery (McCarroll & Altshuler, 2007). The generation of CNVRs is crucial for CNV association analysis but current association tools do not provide suitable definitions of these regions (Kim *et al.*, 2012). To conduct a CNV association study, three steps need to be undertaken. These are (i) CNV calling/detection; (ii) CNVR generation by CNV merging and (iii) statistical analysis (Kim *et al.*, 2012). These three steps are vital for CNV association analyses, however, not all algorithms provide all of them (Kim *et al.*, 2012). Software currently available for association analysis includes PLINK (Purcell *et al.*, 2007), Golden Helix, Parse CNV (Glessner *et al.*, 2013), CNVRuler (Kim *et al.*, 2012) and a few packages offered in R such as CNVtools (Barnes *et al.*, 2008) and CNVassoc (Subirana *et al.*, 2011). Previous studies on cattle have used PLINK (Glick *et al.*, 2011; Da Silva *et al.*, 2016) and Golden Helix SVS (Xu *et al.*, 2014b) to detect CNVs associated with complex traits. Different algorithms use different models to associate CNVs with complex phenotypes. CNVRuler uses linear and logistic regression, Chi-square and Fishers Exact tests (Kim *et al.*, 2012), CNVtools utilises the Likelihood ratio trend test (Glessner *et al.*, 2013), PLINK a permutation based test (Purcell *et al.*, 2007) and CNVAssoc the latent class model (Subirana *et al.*, 2011).

2.12 Conclusion

The interrelationship of specific genes and their association with parasite resistance at a molecular level is not completely understood (Turner et al., 2010). The presence of CNVs at particular genes studied by Wang et al. (2015) and their relationship to the enhanced ability of Nguni cattle to handle harsh environmental conditions needs further investigation. Tick resistance is a heritable trait which comprises a magnitude of physiological and biological processes. Understanding the mechanisms and genes underlying these processes, can help to exploit this trait for future breeding and selection programmes. Identifying and associating economically important traits, on a molecular level, can help incorporate them into genomic selection programmes in the future and alleviate a vast spectrum of issues in the livestock industry (Bickhart et al., 2012).

2.13 References

Abbas, R. Z., Zaman, M. A., Colwell, D. D., Gilleard, J., & Iqbal, Z. 2014. Acaricide resistance in cattle ticks and approaches to its management: The state of play. Vet. Parasitol. 203, 6–20.

Acosta-Rodríguez, R., Alonso-Morales, R., Balladares, S., Flores-Aguilar, H., García-Vazquez, Z., & Gorodezky, C. 2005. Analysis of BoLA class II microsatellites in cattle infested with *Boophilus microplus* ticks: Class II is probably associated with susceptibility. Vet. Parasitol. 127, 313–321.

Alkan, C., Kidd, J. M., Marques-Bonet, T., Aksay, G., Antonacci, F., Hormozdiari, F., Kitzman, J. O., Baker, C., Malig, M., Mutlu, O., Sahinalp, S. C., Gibbs, R. A., & Eichler, E. E. 2009. Personalized copy number and segmental duplication maps using next-generation sequencing. Nat. Genet. 41, 1061–1067.

Alkan, C., Coe, B. P., & Eichler, E. E. 2011. Genome structural variation discovery and genotyping. Nat. Rev. Genet. 12, 363–76.

Ashton, B. G. C., Seifert, G. W., & Francis, J. 1968. An association between serum amylase phenotype. Aust. J. Biol. Sci. 21, 303–308.

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

Barker, J. S. F. 2009. Defining fitness in natural and domesticated population. In: van der Werf, J. H. J., Graser, H. U., Frankham, R., Gondoro, C. (Eds.), Adaptation and Fitness in Animal Populations. Evolutionary and Breeding Perspectives on Genetic Resource Management. Springer. 3-14.

Barnes, C., Plagnol, V., Fitzgerald, T., Redon, R., Marchini, J., Clayton, D., & Hurles, M. 2008. A robust statistical method for case-control association testing with copy number variation. Nat. Genet. 40, 1245–1252.

Behjati, S., & Tarpey, P. S. 2013. What is next generation sequencing? Arch. Dis. Child. Educ. Pract. Ed. 98, 236–8.

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

Budeli, M. A., Nephawe, K. A., Norris, D., Selapa, N. W., Bergh, L., & Maiwashe, A. 2009. Genetic parameter estimates for tick resistance in Bonsmara cattle. South African J. Anim. Sci. 39, 321–327.

Burrow, H. M. 2001. Variances and covariances between productive and adaptive traits and temperament in a composite breed of tropical beef cattle. Livest. Prod. Sci. 70, 213–233.

Bush, W. S., & Moore, J. H. 2012. Chapter 11: Genome-Wide Association Studies. PLoS Comput. Biol. 8.

Canavez, F. C., Luche, D. D., Stothard, P., Leite, K. R. M., Sousa-Canavez, J. M., Plastow, G., Meidanis, J., Souza, M. A., Feijao, P., Moore, S. S., & Camara-Lopes, L. H. 2012. Genome sequence and assembly of *Bos indicus*. J. Hered. 103, 342–348.

Carvalho, W. A., Bechara, G. H., Moré, D. D., Ferreira, B. R., da Silva, J. S., & de Miranda Santos, I. K. F. 2008. Rhipicephalus (Boophilus) microplus: Distinct acute phase proteins vary during infestations according to the genetic composition of the bovine hosts, *Bos taurus* and *Bos indicus*. Exp. Parasitol. 118, 587–591.

Choi, J. W., Chung, W. H., Lim, K. S., Lim, W. J., Choi, B. H., Lee, S. H., Kim, H. C., Lee, S. S., Cho, E. S., Lee, K. T., Kim, N., Kim, J. D., Kim, J. B., Chai, H. H., Cho, Y. M., Kim, T. H., & Lim, D. 2016. Copy number variations in Hanwoo and Yanbian cattle genomes using the massively parallel sequencing data. Gene. 589, 36–42.

Cicconardi, F., Chillemi, G., Tramontano, A., Marchitelli, C., Valentini, A., Ajmone-Marsan, P., & Nardone, A. 2013. Massive screening of copy number population-scale variation in Bos taurus genome. BMC Genomics. 14, 124.

Coppieters, W., Mes, T. H. M., Druet, T., Farnir, F., Tamma, N., Schrooten, C., Cornelissen, A. W. C. A, Georges, M., & Ploeger, H. W. 2009. Mapping QTL influencing gastrointestinal nematode burden in Dutch Holstein-Friesian dairy cattle. BMC Genomics. 10, 96.

Davis, G. P. 1993. Genetic Parameters for Tropical Beef Cattle in Northern Australia: a Review. Aust. J. Agric. Res. 44, 179–98.

de Castro, J. J., & Newson, R. M. 1993. Host resistance in cattle tick control. Parasitol. Today. 9, 13–17.

de Castro, J. J. 1997. Sustainable tick and tickborne disease control in livestock improvement in developing countries. Vet. Parasitol. 71, 77–97.

de la Fuente, J., Almazán, C., Canales, M., Pérez de la Lastra, J. M., Kocan, K. M., & Willadsen, P. 2007. A ten-year review of commercial vaccine performance for control of tick infestations on cattle. Anim. Health Res. Rev. 8, 23–28.

Da Silva, V. H., De Almeida Regitano, L. C., Geistlinger, L., Pértille, F., Giachetto, P. F., Brassaloti, R. A., Morosini, N. S., Zimmer, R., & Coutinho, L. L. 2016. Genome-wide detection of CNVs and their association with meat tenderness in Nelore cattle. PLoS One. 11, 1–22.

Ellis, S. 2004. The cattle major histocompatibility complex: Is it unique? Vet. Immunol. Immunopathol. 102, 1–8.

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

Feuk, L., Carson, A. R., & Scherer, S. W. 2006. Structural variation in the human genome. Nat. Rev. Genet. 7, 85–97.

Frisch, J. E. 1999. Towards a permanent solution for controlling cattle ticks. Int. J. Parasitol. 29, 57–71.

Gao, Y., Jiang, J., Yang, S., Hou, Y., Liu, G. E., Zhang, S., Zhang, Q., & Sun, D. 2017. CNV discovery for milk composition traits in dairy cattle using whole genome resequencing. BMC Genomics. 18, 265.

Gasparin, G., Miyata, M., Coutinho, L. L., Martinez, M. L., Teodoro, R. L., Furlong, J., Machado, M. A., Silva, M. V. G. B., Sonstegard, T. S., & Regitano, L. C. A. 2007. Mapping of quantitative trait loci controlling tick [*Riphicephalus (Boophilus) microplus*] resistance on bovine chromosomes 5, 7 and 14. Anim. Genet. 38, 453–459.

Glessner, J. T., Li, J., & Hakonarson, H. 2013. ParseCNV integrative copy number variation association software with quality tracking. Nucleic Acids Res. 41.

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

Henshall, J. M. 2004. A genetic analysis of parasite resistance traits in a tropically adapted line of *Bos taurus*. Aust. J. Agric. Res. 55, 1109–1116.

Hoffmann, I. 2010. Climate change and the characterization, breeding and conservation of animal genetic resources. Anim. Genet. 41, 32–46.

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

Hughes, V. L., & Randolph, S. E. 2001. Testosterone depresses innate and acquired resistance to ticks in natural rodent hosts: a force for aggregated distributions of parasites. J. Parasitol. 87, 49–54.

Ibeagha-Awemu, E. M., Jann, O. C., Weimann, C., & Erhardt, G. 2004. Genetic diversity, introgression and relationships among West/Central African cattle breeds. Genet. Sel. Evol. 36, 673–690.

Ibeagha-Awemu, E. M., Kgwatalala, P., Ibeagha, A. E., & Zhao, X. 2008. A critical analysis of disease-associated DNA polymorphisms in the genes of cattle, goat, sheep, and pig. Mamm. Genome. 19, 226–245.

Jongejan, F., & Uilenberg, G. 2004. The global importance of ticks. Parasitology. 129, S3–S14.

Jonsson, N. N., Bock, R. E., & Jorgensen, W. K. 2008. Productivity and health effects of anaplasmosis and babesiosis on *Bos indicus* cattle and their crosses, and the effects of differing intensity of tick control in Australia. Vet. Parasitol. 155, 1–9.

Kabir, M. H. B., Mondal, M. M. H., Eliyas, M., Mannan, M. A., Hashem, M. A., & Debnath, N. C. 2011. An epidemiological survey on investigation of tick infestation in cattle at Chittagong District, Bangladesh. African J. Microbiol. Res. 5, 346–352.

Kamidi, R. E., & Kamidi, M. K. 2005. Effects of a Novel Pesticide Resistance Management Strategy on Tick Control in a Smallholding Exotic-breed Dairy Herd in Kenya. Trop. Anim. Health Prod. 37, 469–478.

Kijas, J. W., Barendse, W., Barris, W., Harrison, B., McCulloch, R., McWilliam, S., & Whan, V. 2011. Analysis of copy number variants in the cattle genome. Gene. 482, 73–77.

Kim, J. H., Hu, H. J., Yim, S. H., Bae, J. S., Kim, S. Y., & Chung, Y. J. 2012. CNVRuler: a copy number variation-based case-control association analysis tool. Bioinformatics. 28, 1790–1792.

Korbel, J. O., Urban, A. E., Affourtit, J. P., Godwin, B., Grubert, F., Simons, J. F., Kim, P. M., Palejev, D., Nicholas, J., Du, L., Taillon, B. E., Chen, Z., Tanzer, A., Eugenia, A. C., Chi, J., Yang, F., Carter, N. P., Hurles, M. E., Weissman, S. M., Harkins, T. T., Gerstein, M. B., Egholm, M., & Snyder, M. 2009. Paired-End Mapping Reveals Extensive Structural Variation in the Human Genome. Science. 318, 420–426.

Liu, G. E., Hou, Y., Zhu, B., Liu, G. E., Hou, Y., Zhu, B., Cardone, M. F., Jiang, L., Cellamare, A., Mitra, A., Alexander, L. J., Coutinho, L. L., Elena, M., Aquila, D., Gasbarre, L. C., Lacalandra, G., Li, R. W., Matukumalli, L. K., Nonneman, D., Regitano, L. C. D. A., Smith, T. P. L., Song, J., Sonstegard, T. S., Tassell, C. P. Van, Ventura, M., Eichler, E. E., Mcdaneld, T. G., & Keele, J. W. 2010. Analysis of copy number variations among diverse cattle breeds. Genome Res. 20, 693–703.

Liu, G. E., & Bickhart, D. M. 2012. Copy number variation in the cattle genome. Funct. Integr. Genomics. 12, 609–624.

Lohmueller, K. E., Pearce, C. L., Pike, M., Lander, E. S., & Hirschhorn, J. N. 2003. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. Nat. Genet. 33, 177–182.

Machado, M. A., Azevedo, A. L. S., Teodoro, R. L., Pires, M. A., Peixoto, M. G. C. D., de Freitas, C., Prata, M. C. A., Furlong, J., da Silva, M. V. G. B., Guimarães, S. E. F., Regitano, L. C. A., Coutinho, L. L., Gasparin, G., & Verneque, R. S. 2010. Genome wide scan for quantitative trait loci affecting tick resistance in cattle (*Bos taurus* x *Bos indicus*). BMC Genomics. 11, 280.

Magee, D. A., MacHugh, D. E., & Edwards, C. J. 2014. Interrogation of modern and ancient genomes reveals the complex domestic history of cattle. Anim. Front. 4, 7–22.

Makina, S. O., Muchadeyi, F. C., van Marle-Kőster, E., MacNeil, M. D., & Maiwashe, A. 2014. Genetic diversity and population structure among six cattle breeds in South Africa using a whole genome SNP panel. Front. Genet. 5, 1–7.

Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., McCarthy, M. I., Ramos, E. M., Cardon, L. R., Chakravarti, A., Cho, J. H., Guttmacher, A. E., Kong, A., Kruglyak, L., Mardis, E., Rotimi, C. N., Slatkin, M., Valle, D., Whittemore, A. S., Boehnke, M., Clark, A. G., Eichler, E. E., Gibson, G., Haines, J. L., Mackay, T. F. C., McCarroll, S. A., & Visscher, P. M. 2009. Finding the missing heritability of complex diseases. Nature. 461, 747–753.

Mapholi, N. O., Marufu, M. C., Maiwashe, A., Banga, C. B., Muchenje, V., MacNeil, M. D., Chimonyo, M., & Dzama, K. 2014. Towards a genomics approach to tick (Acari: Ixodidae) control in cattle: A review. Ticks Tick. Borne. Dis. 5, 475–483.

Mapholi, N. O., Maiwashe, A., Matika, O., Riggio, V., Bishop, S. C., MacNeil, M. D., Banga, C., Taylor, J. F., & Dzama, K. 2016. Genome-wide association study of tick resistance in South African Nguni cattle. Ticks Tick. Borne. Dis. 7, 487–497.

Mapiye, C., Chimonyo, M., Dzama, K., Raats, J. G., & Mapekula, M. 2009. Opportunities for improving Nguni cattle production in the smallholder farming systems of South Africa. Livest. Sci. 124, 196–204.

Mapiye, C., Chimonyo, M., Dzama, K., Hugo, A., Strydom, P. E., & Muchenje, V. 2011. Fatty acid composition of beef from Nguni steers supplemented with *Acacia karroo* leaf-meal. J. Food Compos. Anal. 24, 523–528.

Martinez, M. L., Machado, M. A., Nascimento, C. S., Silva, M. V. G. B., Teodoro, R. L., Furlong, J., Prata, M. C. A., Campos, A. L., Guimarães, M. F. M., Azevedo, A. L. S., Pires, M. F. A., & Verneque, R. S. 2006. Association of BoLA-DRB3.2 alleles with tick (*Boophilus microplus*) resistance in cattle. Genet. Mol. Res. 5, 513–524.

Marufu, M. C., Chimonyo, M., Mapiye, C., & Dzama, K. 2011a. Tick loads in cattle raised on sweet and sour rangelands in the low-input farming areas of South Africa. Trop. Anim. Health Prod. 43, 307–313.

Marufu, M. C., Qokweni, L., Chimonyo, M., & Dzama, K. 2011b. Relationships between tick counts and coat characteristics in Nguni and Bonsmara cattle reared on semiarid rangelands in South Africa. Ticks Tick. Borne. Dis. 2, 172–177.

Masika, P. J., Sonandi, A., & van Averbeke, W. 1997. Tick control by small-scale cattle farmers in the central Eastern Cape Province, South Africa. J. S. Afr. Vet. Assoc. 68, 45–48.

Mattioli, R. C., Pandey, V. S., Murray, M., & Fitzpatrick, J. L. 2000. Immunogenetic influences on tick resistance in African cattle with particular reference to trypanotolerant N'Dama (*Bos taurus*) and trypanosusceptible Gobra zebu (*Bos indicus*) cattle. Acta Trop. 75, 263–277.

McCarroll, S. A., & Altshuler, D. M. 2007. Copy-number variation and association studies of human disease. Nat. Genet. 39, S37–S42.

Medvedev, P., Stanciu, M., & Brudno, M. 2009. Computational methods for discovering structural variation with next-generation sequencing. Nat. Methods 6, S13–S20.

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

Mirkena, T., Duguma, G., Haile, A., Tibbo, M., Okeyo, A. M., Wurzinger, M., & Sölkner, J. 2010. Genetics of adaptation in domestic farm animals: A review. Livest. Sci. 132, 1–12.

Mmbengwa, V. 2015. Communal livestock farming in South Africa: Does this farming system create jobs for poverty stricken rural areas? Sylwan. 159, 176–192.

Morris, C. A. 2007. A review of genetic resistance to disease in Bos *taurus* cattle. Vet. J. 174, 481–491.

Moyo, B., Dube, S., Lesoli, M., & Masika, P. J. 2008. Communal area grazing strategies: institutions and traditional practices. 42nd Annu. Congrss Grassland Society of South Africa. 25, 47–54.

Mtshali, M. S., De Waal, D. T., & Mbati, P. A. 2004. A sero-epidemiological survey of blood parasites in cattle in the north-eastern Free State, South Africa. Onderstepoort J Vet Res. 71, 67–75.

Muchenje, V., Dzama, K., Chimonyo, M., Raats, J. G., & Strydom, P. E. 2008. Tick susceptibility and its effects on growth performance and carcass characteristics of Nguni, Bonsmara and Angus steers raised on natural pasture. Animal. 2, 298–304.

Nava, S., Guglielmone, A. A., & Mangold, A. J. 2009. An overview of systematics and evolution of ticks. Front. Biosci. (Landmark Ed.). 14, 2857–77.

Patel, R. K., & Jain, M. 2012. NGS QC toolkit: A toolkit for quality control of next generation sequencing data. PLoS One. 7.

Pinto, D., Darvishi, K., Shi, X., Rajan, D., Rigler, D., Fitzgerald, T., Lionel, A. C., Thiruvahindrapuram, B., Macdonald, J. R., Mills, R., Prasad, A., Noonan, K., Gribble, S., Prigmore, E., Donahoe, P. K., Smith, R. S., Park, J. H., Hurles, M. E., Carter, N. P., Lee, C., Scherer, S. W., & Feuk, L. 2011. Comprehensive assessment of array-based platforms and calling algorithms for detection of copy number variants. Nat. Biotechnol. 29, 512–20.

Porto Neto, L. R., Jonsson, N. N., D'Occhio, M. J., & Barendse, W. 2011. Molecular genetic approaches for identifying the basis of variation in resistance to tick infestation in cattle. Vet. Parasitol. 180, 165–172.

Prayaga, K. C., & Henshall, J. M. 2005. Adaptability in tropical beef cattle: Genetic parameters of growth, adaptive and temperament traits in a crossbred population. Aust. J. Exp. Agric. 45, 971–983.

Prayaga, K. C., Corbet, N. J., Johnston, D. J., Wolcott, M. L., Fordyce, G., & Burrow, H. M. 2009. Genetics of adaptive traits in heifers and their relationship to growth, pubertal and carcass traits in two tropical beef cattle genotypes. Anim. Prod. Sci. 49, 413–425.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81, 559–575.

Rajput, Z. I., Hu, S., Chen, W., Arijo, A. G., & Xiao, C. 2006. Importance of ticks and their chemical and immunological control in livestock. J. Zhejiang Univ. Sci. B. 7, 912–21.

Rege, J. E. 1999. The state of African cattle genetic resources. I. Classification framework and identification of threatened and extinct breeds. Anim. Genet. Resour. Infomation Bull. 25, 1–25.

Regitano, L., & Prayaga, K. 2010. Ticks and tick-borne diseases in cattle. In: Bishop, Breeding for Disease Resistance in Farm Animals. Third Ed. CAB International. 295–315.

Rewe, T. O., Herold, P., Kahi, A. K., & Valle Zárate, A. 2009. Breeding indigenous cattle genetic resources for beef production in Sub-Saharan Africa. Outlook Agric. 38, 317–326.

Schoeman, S. J. 1989. Recent research into the production potential of indigenous cattle with special reference to the Sanga. S. Afr. J. Anim. Sci. 19, 55–61.

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

Sharif, S., Mallard, B. A., Wilkie, B. N., Sargeant, J. M., Scott, H. M., Dekkers, J. C. M., & Leslie, K. E. 1999. Associations of the bovine major histocompatibility complex DRB3 (BoLA-DRB3) with production traits in Canadian dairy cattle. Anim. Genet. 30, 157–160.

Shendure, J., & Ji, H. 2008. Next-generation DNA sequencing. Nat. Biotechnol. 26, 1135–1145.

Shyma, K. P., Gupta, J. P., & Singh, V. 2013. Breeding strategies for tick resistance in tropical cattle: a sustainable approach for tick control. J. Parasit. Dis. 39, 1–6.

Snijders, a M., Nowak, N., Segraves, R., Blackwood, S., Brown, N., Conroy, J., Hamilton, G., Hindle, a K., Huey, B., Kimura, K., Law, S., Myambo, K., Palmer, J., Ylstra, B., Yue, J. P., Gray, J. W., Jain, a N., Pinkel, D., & Albertson, D. G. 2001. Assembly of microarrays for genome-wide measurement of DNA copy number. Nat. Genet. 29, 263–264.

Sollero, B. ., Gomes, C. C. G., Roso, V. M., Higa, R. H., Yokoo, M. J., Cardoso, L. L., Caetano, a. R., & Cardoso, F. F. 2014. Goodness of Fit Comparisons among Five Bayesian Models in Genome-Wide Association of Tick Resistance in Brazilian Hereford and Braford Beef Cattle. Proceedings, 10th World Congr. Genet. Appl. to Livest. Prod. Vancouver, Canada, 7–9.

Stear, M. J., Hetzel, D. J. S., Brown, S. C., Gershwin, L. J., Mackinnon, M. J., & Nicholas, F. W. 1990. The relationships among ecto- and endoparasite levels, class I antigens of the bovine major histocompatibility system, immunoglobulin E levels and weight gain. Vet. Parasitol. 34, 303–321.

Subirana, I., Diaz-Uriarte, R., Lucas, G., & Gonzalez, J. R. 2011. CNVassoc: Association analysis of CNV data using R. BMC Med. Genomics. 4, 47.

The Bovine Genome sequencing and analysis consortium, Elsik, C. G., Tellam, R. L., Worley, K. C. 2009. The genome sequence of taurine cattle: a window to ruminant biology and evolution. Science. 324, 522–28.

Turner, L. B., Harrison, B. E., Bunch, R. J., Neto, L. R. P., Li, Y., & Barendse, W. 2010. A genome-wide association study of tick burden and milk composition in cattle. Anim. Prod. Sci. 50, 235–245.

Utech, K. B. W., Wharton, R. H., & Kerr, J. D. 1978. Resistance to *Boophilus microplus* (Canestrini) in different breeds of cattle. Aust. J. Agric. Res. 29, 885–895.

Verissimo, C. J., Nicolau, C. V. J., Cardoso, V. L., & Pinheiro, M. G. 2002. Haircoat characteristics and tick infestation on GYR (Zebu) and crossbred (Holstein x GYR) cattle. Arch. Zootec. 51, 389–392.

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

Wang, Y. H., Reverter, A., Kemp, D., McWilliam, S. M., Ingham, A., Davis, C. A., Moore, R. J., & Lehnert, S. A. 2007b. Gene expression profiling of Hereford Shorthorn cattle following challenge with Boophilus microplus tick larvae. Aust. J. Exp. Agric. 47, 1397–1407.

Wang, M. D., Dzama, K., Hefer, C. A., & Muchadeyi, F. C. 2015. Genomic population structure and prevalence of copy number variations in South African Nguni cattle. BMC Genomics. 16, 894.

Wang, M. D., Dzama, K., Rees, D. J. G., & Muchadeyi, F. C. 2016. Tropically adapted cattle of Africa: Perspectives on potential role of copy number variations. Anim. Genet. 47, 154–164.

Wikel, S. K., & Bergman, D. 1997. Tick-host immunology: Significant advances and challenging opportunities. Parasitol. Today. 13, 383–389.

Willadsen, P. 1997. Novel vaccines for ectoparasites. Vet. Parasitol. 71, 209–222.

Wu, Y., Fan, H., Jing, S., Xia, J., Chen, Y., Zhang, L., Gao, X., Li, J., Gao, H., & Ren, H. 2015. A genome-wide scan for copy number variations using high-density single nucleotide polymorphism array in Simmental cattle. Anim. Genet. 46, 289–298.

Xu, L., Hou, Y., Bickhart, D., Song, J., & Liu, G. 2013. Comparative Analysis of CNV Calling Algorithms: Literature Survey and a Case Study Using Bovine High-Density SNP Data. Microarrays. 2, 171–185.

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

Xu, L., Hou, Y., Bickhart, D. M., Song, J., Van Tassell, C. P., Sonstegard, T. S., & Liu, G. E. 2014b. A genome-wide survey reveals a deletion polymorphism associated with resistance to gastrointestinal nematodes in Angus cattle. Funct. Integr. Genomics. 14, 333–339.

Zare, F., Dow, M., Monteleone, N., Hosny, A., & Nabavi, S. 2017. An evaluation of copy number variation detection tools for cancer using whole exome sequencing data. BMC Bioinformatics. 18, 286.

Zhan, B., Fadista, J., Thomsen, B., Hedegaard, J., Panitz, F., & Bendixen, C. 2011. Global assessment of genomic variation in cattle by genome resequencing and high-throughput genotyping. BMC Genomics. 12, 557.

Zhang, F., Gu, W., Hurles, M. E., & Lupski, J. R. 2009. Copy number variation in human health, disease, and evolution. Annu Rev Genomics Hum Genet. 10, 451–481.

Zhao, M., Wang, Q., Wang, Q., Jia, P., & Zhao, Z. 2013. Computational tools for copy number variation (CNV) detection using next-generation sequencing data: features and perspectives. Springer. BMC Bioinformatics. 14 Suppl 1, S1.

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

Chapter 3

Non-genetic Effects Affecting Tick Resistance in Nguni Cattle Across Three Provinces in South Africa

3.1 Abstract

South African Nguni cattle comprise a cross between *Bos indicus* and *Bos taurus* cattle, known as the Sanga. The Nguni is well known for its adaptability to harsh climatic conditions and its resilience to parasites and their diseases. Ticks and tick-borne diseases pose a major threat to global livestock industries. The methods of tick control and the effect of these parasites cause large economical losses and have an impact on the environment and the health of animals and humans. Within and between breed variations in tick resistance are evident. Breeding for tick resistance is a safe, effective and economically viable method for controlling ticks. This study aimed to determine the non-genetic effects on tick resistance in 347 Nguni cattle located across three different provinces in South Africa. Tick counts were taken over a two-year period from 121, 92 and 134 Nguni cattle located at Mukhuthali Nguni Community farm in Kwa-Zulu Natal, the Agricultural Research Council (ARC) Loskop farm in Limpopo and the ARC Roodeplaat farm in Gauteng respectively. A generalized linear model was run in SAS version 7.1 to determine the effect of location, season, sex, age of the animal and year of counting on tick resistance. Factors significantly influencing tick infestation included location, season, age and year of count.

3.2 Introduction

Ticks are economically one of the most important constraints in livestock industries worldwide (Mapiye *et al.*, 2009), with more than 80% of global cattle populations demonstrating tick infestations (Asmaa *et al.*, 2014). These parasites are blood feeding arthropods and comprise of approximately 900 species subdivided into three different families, Argasidae, Ixodidae and Nuttalliellidae (Nava *et al.*, 2009). The different classification of tick families is attributable to morphological, biological and ecological characteristics (Jongejan & Uilenberg, 2004; Nava *et al.*, 2009). Argasidae species live close to their host and only feed for short intervals (Jongejan & Uilenberg, 2004). Ixodidae species are differentiated according to the number of host species they attach to (Jongejan & Uilenberg, 2004). Tick species transmit a variety of diseases including anaplasmosis, babesiosis, heartwater and theileriosis (Mtshali *et al.*, 2004), and also cause direct harm to an animal through blood loss, stress and skin irritation and lesions (Jongejan & Uilenberg, 2004; Ghosh *et al.*, 2007).

The level of tick infestation varies due to climate, tick species, agro-ecological conditions and the susceptibility of an animal to infestation (Jongejan & Uilenberg, 2004; Regitano & Prayaga, 2010;

Kabir *et al.*, 2011). The occurrence of tick infestations is greater in tropical and subtropical regions when compared with temperate regions (Jongejan & Uilenberg, 2004). In general, tick burden is greater in warmer areas with high rainfall and good vegetation (Mapholi *et al.*, 2014). Rangeland also has an influence on tick load, with sour rangelands presenting greater tick burden (Muchenje *et al.*, 2008; Mapiye *et al.*, 2009; Marufu *et al.*, 2011a). The ability of an animal to withstand tick infestation and limit the number of ticks which survive to maturity is defined as tick resistance (Utech *et al.*, 1978). Tick resistance varies within and between cattle breeds and is influenced by behavioral, morphological, and physiological characteristics (Mapholi *et al.*, 2014). Morphological individualities including coat characteristics, skin thickness and body size can enhance tick resistance, by preventing tick attachment and biting (Marufu *et al.*, 2011b; Mapholi *et al.*, 2014). Physiological characteristics include cellular and immunological skin responses (Piper *et al.*, 2010; Mapholi *et al.*, 2014), age, breed and sex (Asmaa *et al.*, 2014).

A number of breeds that have developed in areas with a high prevalence of ticks demonstrate an enhanced ability to resist these arthropods and their diseases. One such breed is the South African Nguni (Schoeman, 1989; Mapiye *et al.*, 2009; Mirkena *et al.*, 2010). The Nguni is a Sanga type cattle breed which is able to survive and produce under severe environmental conditions (Mirkena *et al.*, 2010). Sanga cattle breeds are a cross between *Bos indicus* and *Bos taurus* (Ibeagha-Awemu *et al.*, 2004) and harbor less ticks when compared to foreign breeds (Mattioli *et al.*, 2000; Gifford-Gonzalez & Hanotte, 2011). The remarkable adaptability and resilience reported in the Nguni (Schoeman, 1989), provide a valuable breed in which to investigate the non-genetic effects of tick resistance. This study aimed to determine the non-genetic effects of tick resistance in South African Nguni cattle from three different provinces in South Africa.

3.3 Methods and Materials

3.3.1 Experimental animals

Tick count data was collected from 347 South African Nguni cattle from three different herds across three different provinces in South Africa. Animals were randomly selected across ages. A total of 16 males, 329 females and two animals with their sex incorrectly recorded were subject to tick count data collection.

3.3.2 Sample locations

Nguni cattle located at the Agricultural Research Council (ARC) Loskop Research Farm (29° 23' 09" E and 25° 04' 42" S) in the Limpopo province (n=92), The Agricultural Research Council (ARC) Roodeplaat Research Farm (28°35" E and 25°59" S) located in the Gauteng province (n= 134) and

the Mukhuthali Nguni Community Farm located in the Kwa-Zulu Natal province (n=121) were sampled (Figure 3.1). Kwazulu Natal has a wet sub-tropical agro-ecological climate with an average annual summer and winter rainfall of 800mm and 200mm respectively (Gbetibouo & Hassan, 2005). The average temperatures range from 23° C in summer to 16°C in winter (Gbetibouo & Hassan, 2005). The ARC-Loskop is located at 29° 23' 09" E and 25° 04' 42" S in the Limpopo Province at an altitude of 947m above sea level. Limpopo has an arid agro-ecological climate with an average annual summer and winter rainfall of 600mm and 150mm respectively (Gbetibouo & Hassan, 2005). The average temperatures range from 25° C in summer to 18°C in winter (Gbetibouo & Hassan, 2005). The ARC-Roodeplaat is located at 28°35"E and 25°59"S in the Gauteng Province at an altitude of 1232m above sea level. Gauteng has an arid agro-ecological climate with an average annual summer and winter rainfall of 600mm and 150mm respectively (Gbetibouo & Hassan, 2005). The average temperatures range from 20° C in summer to 13°C in winter (Gbetibouo & Hassan, 2005).

One hundred and twenty-one Nguni cattle were sampled at Mukhuthali. One hundred and four animals were female, 15 were male and two were unknown due to incorrect recording. Ninety-two Nguni cattle were sampled at Loskop, all of which were female. One hundred and thirty-four Nguni cattle were sampled at Roodeplaat. One hundred and thirty-three animals were female with only one being male.

Figure 3.1 Geographical location of the Mukhuthali (1), ARC Loskop (2) and ARC Roodeplaat (3) farms where 347 South African Nguni cattle were sampled for this study.

3.3.3 Tick counts

Tick count data was collected over a two-year period from May 2012 to April 2014 as described by Mapholi *et al*. (2016). Data collected was divided into two years with year one being from May 2012 to April 2013 and year two being from May 2013 to April 2014. Animals were subjected to natural tick infestation. Two specially trained technicians conducted all counts. Each animal was counted by two people at the same time with each individual counting half the body. Adult tick counts of the back, the belly (including the udder and testicles), the inside and outside of the ears, the head, legs, neck (including the gullet), perineum and tail (including underneath the tail) were performed per animal. Each month, immediately after tick count collection, each animal was spray dipped with flumethrin pour-on formulation “Drastic Deadline®”. Tick counts were recorded per season and seasons were classified as hot wet (December, January and February), cool wet (March, April and May), cool dry (June, July and August) and hot dry (September, October and November). The data recorded for each animal included location, month, year, season, animal ID, sire ID, dam ID, date of birth, age, sex, collection date and total tick count.

3.3.4 Statistical analyses

Data was analyzed using Statistical Analysis System (SAS) Enterprise guide software (Version 7.1, 2014; SAS Institute Inc, Cary, NC, USA). To ensure normality, tick count data was transformed using base 10 logarithm. A generalized linear model (PROC GLM) was used to determine the non-genetic effects of tick resistance. The model fitted the fixed effects of season and year of tick count as well as the location, sex and age of the animal. The model was as follows:

$$Y_{ijklm} = \mu + L_i + S_j + D_k + G_l + (A_m) + \varepsilon_{ijklm}$$

Y_{ijklm} = Log transformed tick count
μ = Overall mean
L_i = Location (1, 2 or 3)
S_j = Season (1, 2, 3 or 4)
D_k = Effect of year (1 or 2)
G_l = Effect of sex (1, 2 or other)
A_m = Age (Age in years at date of tick count collection)
ε_{ijklm} = Random residual error

3.4 Results and Discussion

3.4.1 Nguni tick count summary statistics

Tick counts of the 347 Nguni cattle recorded from May 2012 to April 2014 ranged from 0 to 198 with a mean tick count of 24.62 (Table 3.1). Budeli *et al*. (2009) found tick counts to range from 18.6 to 58.7 in Bonsmara cattle located in Limpopo, North-West and the Western Cape. Corbet *et al*. (2006) report mean tick count of 37 in South African Bonsmara and Australian Belmont Red cattle raised in four regions of South Africa. Turner *et al*. (2010) on the other hand report mean tick counts ranging from 30 to 124.8 in a variety of *Bos taurus* dairy cattle breeds in Australia. The mean observed in this study is less than that of Turner *et al*. (2010). This could be due to the fact that Turner *et al*. (2010) only took mean daily tick counts of 20 or more into consideration. Loskop farm not only had the highest mean tick count of 30.12, but also the highest spread of counts (Standard deviation - 20.12). The high tick counts reported at Loskop could be attributable to environmental factors favoring high tick burden.

Table 3.1 The minimum (Min), maximum (Max), mean, standard deviation (SD) and coefficient of variation (CV) of tick count data for 347 South African Nguni cattle at Mukhuthali, Loskop and Roodeplaat over a period of two years in South Africa.

Location	Min	Max	Mean	Mean*	SD	SD*	CV (%)
Mukhuthali	0	77	18.24	1.19	11.33	0.31	62.11
Loskop	0	198	30.12	1.39	20.12	0.33	66.79
Roodeplaat	0	118	26.43	1.35	18.46	0.33	69.86

*Mean and standard deviation (SD) of transformed tick counts

3.4.2 Influence of location on tick count of Nguni cattle

Location had a significant effect ($p<0.001$) on Nguni cattle tick count. Mukhuthali Nguni community farm is located in Kwazulu Natal.

Agro-ecological conditions influence tick burden in cattle with tropical and subtropical regions being host to greater infestation and tick diversity (Jongejan & Uilenberg, 2004; Regitano & Prayaga, 2010; Katiyatiya *et al.*, 2014). In South Africa, studies have shown humid-coastal areas to demonstrate greater tick loads (Katiyatiya *et al.*, 2014) which is contradictory to that of this study. Loskop reports the greatest mean tick count followed by Roodeplaat and Mukhuthali (Figure 3.2). The variation could be attributable to temperature differences, as Loskop has a higher average temperature than the other locations and could explain the greater tick infestation.

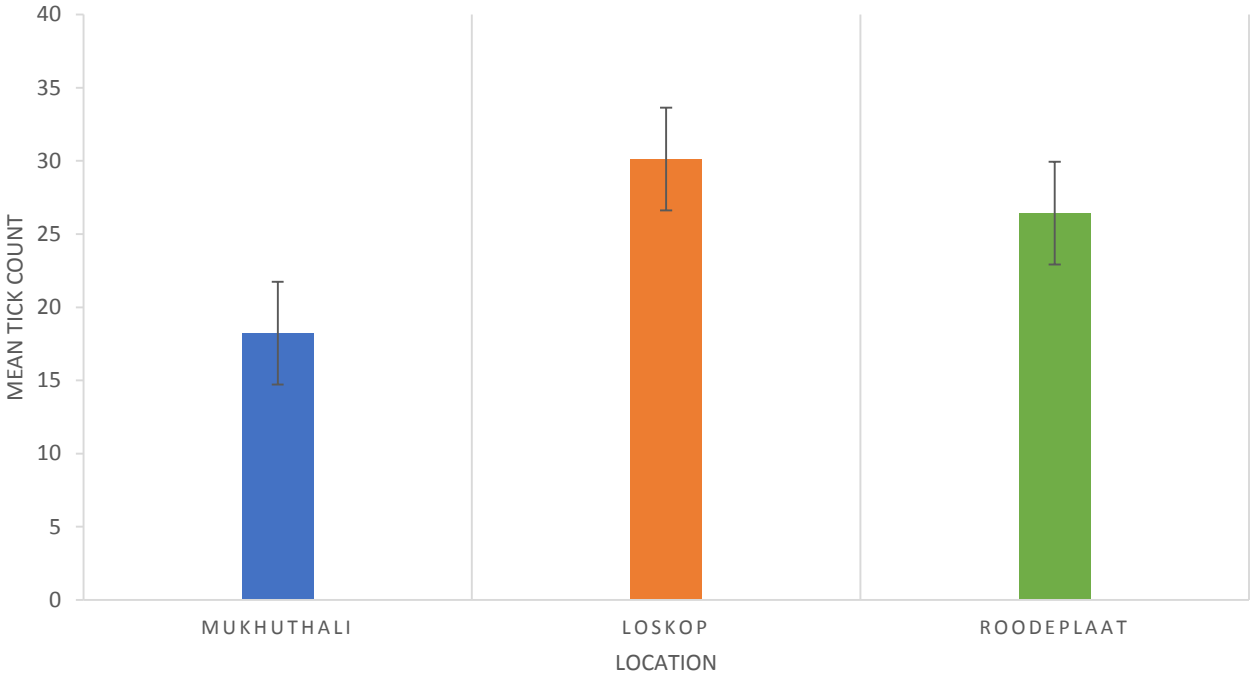

Figure 3.2 Mean tick counts of 347 Nguni cattle at Mukhuthali, Loskop and Roodeplaat

3.4.3 Influence of sex on tick count of Nguni cattle

Sex did not show a significant effect on the resistance of an animal, however a higher mean tick count was observed in female animals (Figure 3.3). These findings are in accordance with that of Kabir *et al*. (2011), Asmaa *et al*. (2014) and Rehman *et al*. (2017), who report significantly higher tick infestation in female cattle, however Martinez *et al*. (2006) report tick load to be greater in males. It has been reported that female animals are more susceptible to tick infestation due to the stressors of pregnancy and lactation (Sutherst *et al.*, 1983; Kabir *et al.*, 2011). Hormonal influences can also attribute to susceptibility of male and female animals (Lloyd, 1983; Hughes & Randolph, 2001). In females, higher levels of prolactin and progesterone can make an animal more susceptible (Lloyd, 1983), with testosterone in males reducing acquired and innate resistance (Hughes & Randolph, 2001).

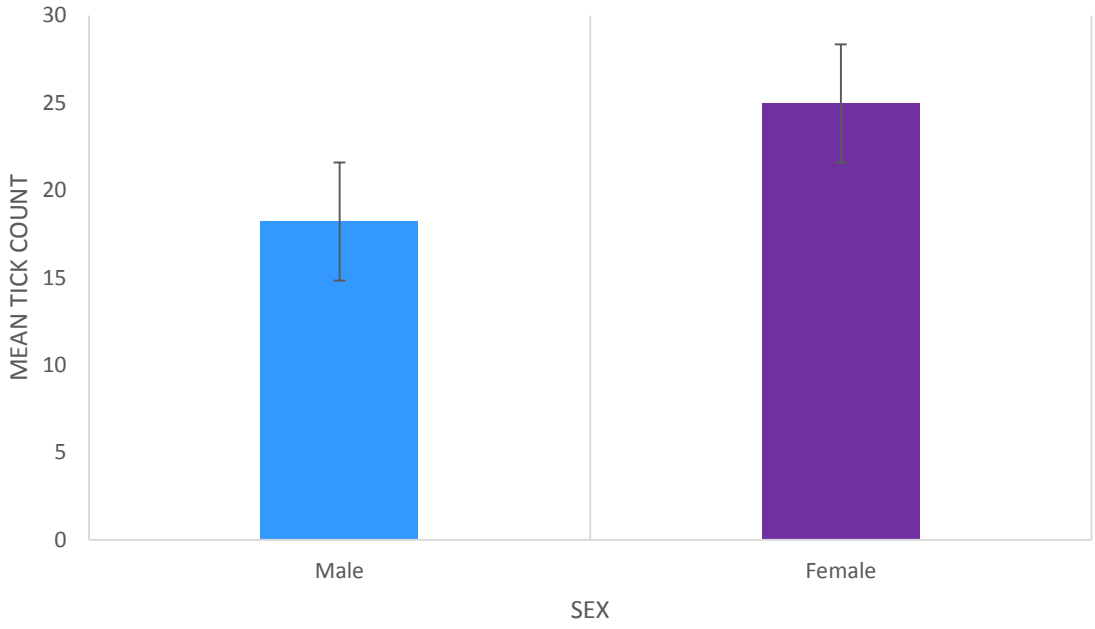

Figure 3.3 Mean tick counts of male and female Nguni cattle across three locations in South Africa

3.4.4 Influence of season on tick count of Nguni cattle

Season had a significant influence on tick count ($p<0.001$). Loskop showed the highest mean tick count for all seasons, except for the cool wet season in which Roodeplaat displayed the highest mean tick count (Figure 3.4). Mukhuthali showed the lowest mean tick counts across all seasons. The cool dry season displayed the lowest mean tick count across all locations, with the hot wet and hot dry seasons displaying the highest mean tick counts at all locations, except for Mukhuthali. These findings are in correspondence with those of Muchenje *et al*. (2008), Marufu *et al*. (2011a) and Katiyatiya *et al*. (2014), who found the prevalence of ticks to be greater in the hot wet seasons than in the cool dry seasons. The higher tick infestations during the hot wet season could be attributable to conditions more suited for breeding and survival as ticks thrive in hot, humid environments (Webb & David, 2002; Katiyatiya *et al.*, 2014).

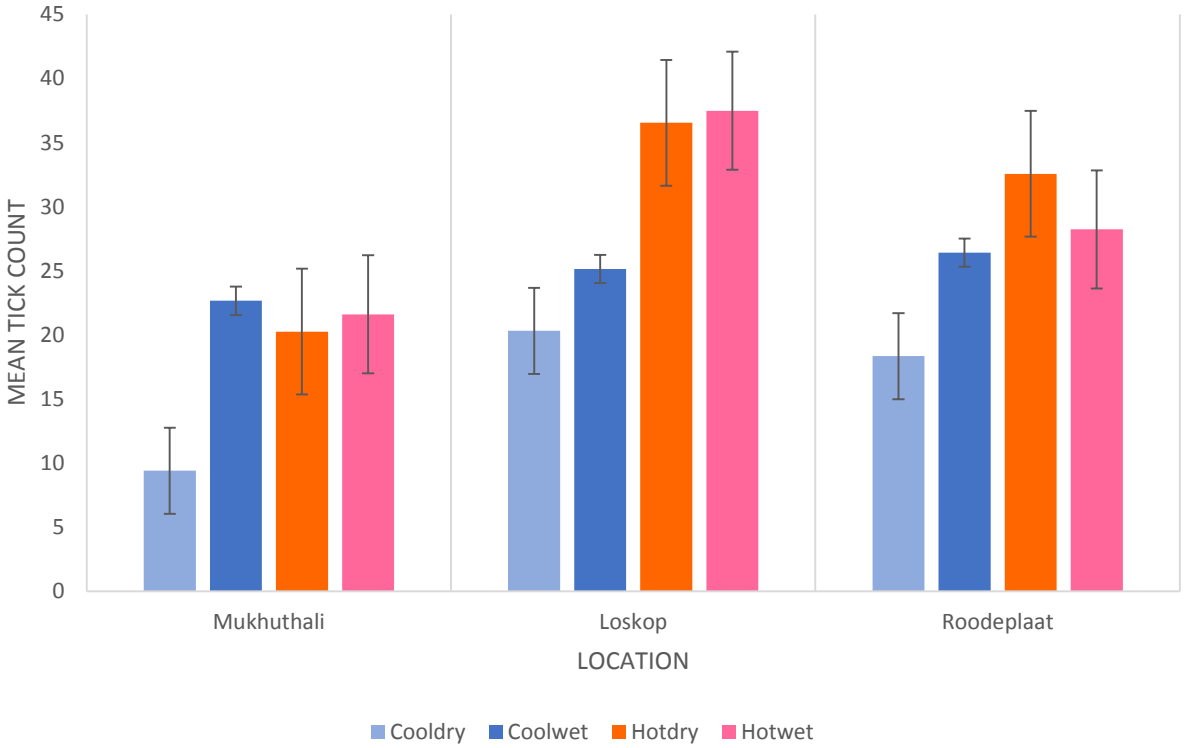

Figure 3.4 Mean tick count per season (Cool dry, cool wet, hot dry and hot wet) for 347 Nguni cattle at Mukhuthali, Loskop and Roodeplaat.

3.4.5 Influence of year on tick count of Nguni cattle

Tick counts were collected over a two-year period with year one being from May 2012 to April 2013 and year two being from May 2013 to April 2014. The year in which tick counts were collected had a significant effect on tick count ($p<0.001$). Roodeplaat had the highest mean tick count for year one, with Loskop having the highest mean tick count for year two (Figure 3.5). Mukhuthali displayed the lowest mean tick counts for both years.

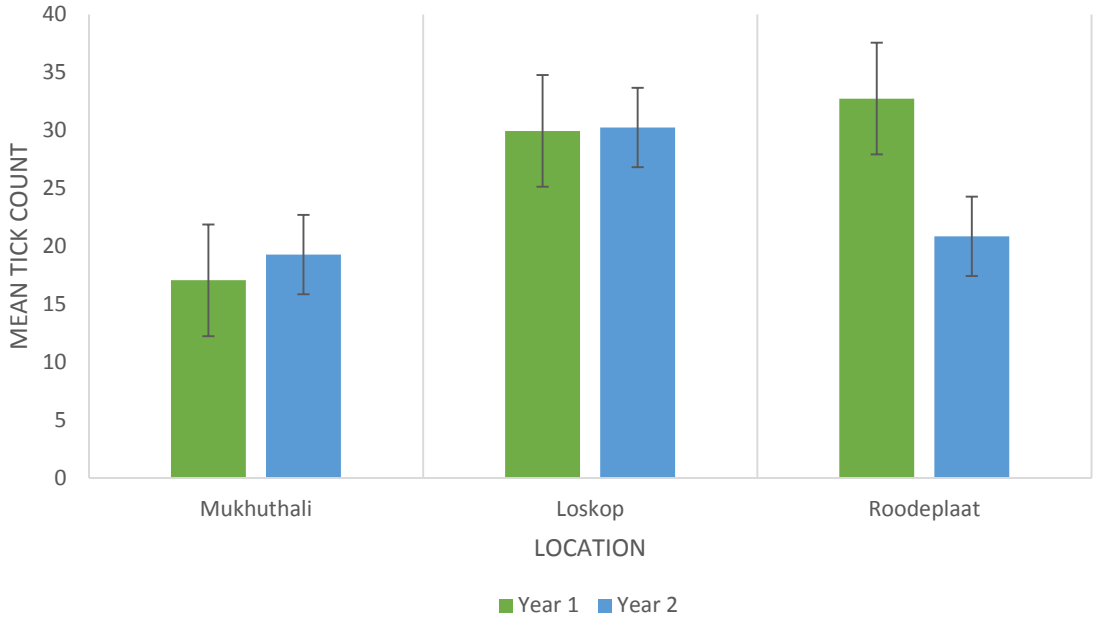

Figure 3.5 Mean tick counts per year for 347 Nguni cattle at Mukhuthali, Loskop and Roodeplaat

3.4.6 Influence of age on tick count of Nguni cattle

The age of the animal at the time of tick count collection had a significant effect ($p<0.001$) on tick count. Animals younger than five years showed the lowest mean tick counts except for Roodeplaat, which showed animals older than 5 to 10 years to have the lowest mean tick count (Table 3.3). These results are similar to that of Marufu *et al.* (2011a) and Asmaa *et al.* (2014) who report the level of tick infestation to be higher in older animals. Kabir *et al.* (2011) however, reported the prevalence of ticks to be higher in younger animals. The observation of younger animals displaying lower tick loads, could be due to an innate form of immunity which declines with age (Wikel & Bergman, 1997).

Table 3.2 Mean tick counts of Nguni cattle younger than 5 and older than 5-10, 10-15 and 15-20 years sampled from three different locations in South Africa.

Location	Age of animal in years			
	≤5	>5-10	>10-15	>15-20
Mukhuthali	17.00	19.47	20.37	-
Loskop	18.32	30.50	30.03	32.35
Roodeplaat	26.70	25.13	27.88	31.84

3.5 Conclusion

The non-genetic effects influencing tick resistance as represented by tick count in Nguni cattle from three different locations in South Africa were location, season, year of tick count and age of the animal. Loskop displayed the highest tick infestation, which could be attributable to reports that locations located further North in areas exhibiting higher temperatures showed greater tick loads. The hot wet and hot dry seasons displayed the greatest mean tick counts at Loskop and Roodeplaat, however Mukhuthali had the greatest mean tick count during the cool wet season. Animals younger than five years had lower tick counts. This could be due to an innate immunity which declines with age.

3.6 References

Asmaa, N. M., ElBably, M. A., & Shokier, K. A. 2014. Studies on prevalence, risk indicators and control options for tick infestation in ruminants. Beni-Suef Univ. J. Basic Appl. Sci. 3, 68–73.

Budeli, M. A., Nephawe, K. A., Norris, D., Selapa, N. W., Bergh, L., & Maiwashe, A. 2009. Genetic parameter estimates for tick resistance in Bonsmara cattle. South African J. Anim. Sci. 39, 321–327.

Corbet, N. J., Shepherd, R. K., Burrow, H. M., Van Der Westhuizen, J., Strydom, P. E., & Bosman, D. J. 2006. Evaluation of Bonsmara and Belmont Red cattle breeds in South Africa. 1. Productive performance. Aust. J. Exp. Agric. 46, 199–212.

Gbetibouo, G. A., & Hassan, R. M. 2005. Measuring the economic impact of climate change on major South African field crops: A Ricardian approach. Glob. Planet. Change 47, 143–152.

Ghosh, S., Azhahianambi, P., & Yadav, M. P. 2007. Upcoming and future strategies of tick control: A review. J. Vector Borne Dis. 44, 79–89.

Gifford-Gonzalez, D., & Hanotte, O. 2011. Domesticating Animals in Africa: Implications of Genetic and Archaeological Findings. J. World Prehistory 24, 1–23.

Hughes, V. L., & Randolph, S. E. 2001. Testosterone depresses innate and acquired resistance to ticks in natural rodent hosts: a force for aggregated distributions of parasites. J. Parasitol. 87, 49–54.

Ibeagha-Awemu, E. M., Jann, O. C., Weimann, C., & Erhardt, G. 2004. Genetic diversity, introgression and relationships among West/Central African cattle breeds. Genet. Sel. Evol. 36, 673–690.

Jongejan, F., & Uilenberg, G. 2004. The global importance of ticks. Parasitology 129, S3–S14.

Kabir, M. H. B., Mondal, M. M. H., Eliyas, M., Mannan, M. A., Hashem, M. A., & Debnath, N. C. 2011. An epidemiological survey on investigation of tick infestation in cattle at Chittagong District, Bangladesh. African J. Microbiol. Res. 5, 346–352.

Katiyatiya, C. L. F., Muchenje, V., & Mushunje, A. 2014. Seasonal variation in coat characteristics, tick loads, cortisol levels, some physiological parameters and temperature humidity index on Nguni cows raised in low- and high-input farms. Int. J. Biometeorol., 733–743.

Lloyd M (1983). An experiment in the organization of a minimum complex measures against haemosporidiasis in northern Tadzhikistan. J. Vet. Res., 6: 64 - 74.

Mapholi, N. O., Marufu, M. C., Maiwashe, A., Banga, C. B., Muchenje, V., MacNeil, M. D., Chimonyo, M., & Dzama, K. 2014. Towards a genomics approach to tick (Acari: Ixodidae) control in cattle: A review. Ticks Tick. Borne. Dis. 5, 475–483.

Mapholi, N. O., Maiwashe, A., Matika, O., Riggio, V., Bishop, S. C., MacNeil, M. D., Banga, C., Taylor, J. F., & Dzama, K. 2016. Genome-wide association study of tick resistance in South African Nguni cattle. Ticks Tick. Borne. Dis. 7, 487–497.

Mapiye, C., Chimonyo, M., Dzama, K., Raats, J. G., & Mapekula, M. 2009. Opportunities for improving Nguni cattle production in the smallholder farming systems of South Africa. Livest. Sci. 124, 196–204.

Martinez, M. L., Machado, M. A., Nascimento, C. S., Silva, M. V. G. B., Teodoro, R. L., Furlong, J., Prata, M. C. A., Campos, A. L., Guimarães, M. F. M., Azevedo, A. L. S., Pires, M. F. A., & Verneque, R. S. 2006. Association of BoLA-DRB3.2 alleles with tick (*Boophilus microplus*) resistance in cattle. Genet. Mol. Res. 5, 513–524.

Marufu, M. C., Chimonyo, M., Mapiye, C., & Dzama, K. 2011a. Tick loads in cattle raised on sweet and sour rangelands in the low-input farming areas of South Africa. Trop. Anim. Health Prod. 43, 307–313.

Marufu, M. C., Qokweni, L., Chimonyo, M., & Dzama, K. 2011b. Relationships between tick counts and coat characteristics in Nguni and Bonsmara cattle reared on semiarid rangelands in South Africa. Ticks Tick. Borne. Dis. 2, 172–177.

Mattioli, R. C., Pandey, V. S., Murray, M., & Fitzpatrick, J. L. 2000. Immunogenetic influences on tick resistance in African cattle with particular reference to trypanotolerant N'Dama (*Bos taurus*) and trypanosusceptible Gobra zebu (*Bos indicus*) cattle. Acta Trop. 75, 263–277.

Mirkena, T., Duguma, G., Haile, A., Tibbo, M., Okeyo, A. M., Wurzinger, M., & Sölkner, J. 2010. Genetics of adaptation in domestic farm animals: A review. Livest. Sci. 132, 1–12.

Mtshali, M. S., De Waal, D. T., & Mbati, P. A. 2004. A sero-epidemiological survey of blood parasites in cattle in the north-eastern Free State, South Africa. Onderstepoort. J. Vet. Res. 71, 67–75.

Muchenje, V., Dzama, K., Chimonyo, M., Raats, J. G., & Strydom, P. E. 2008. Tick susceptibility and its effects on growth performance and carcass characteristics of Nguni, Bonsmara and Angus steers raised on natural pasture. Animal. 2, 298–304.

Nava, S., Guglielmone, A. A., & Mangold, A. J. 2009. An overview of systematics and evolution of ticks. Front. Biosci. (Landmark Ed). 14, 2857–77.

Piper, E. K., Jackson, L. A., Bielefeldt-Ohmann, H., Gondro, C., Lew-Tabor, A. E., & Jonsson, N. N. 2010. Tick-susceptible *Bos taurus* cattle display an increased cellular response at the site of larval *Rhipicephalus* (*Boophilus*) *microplus* attachment, compared with tick-resistant *Bos indicus* cattle. Int. J. Parasitol. 40, 431–441.

Regitano, L., & Prayaga, K. 2010. Ticks and tick-borne diseases in cattle. In: Bishop, Breeding for Disease Resistance in Farm Animals. Third Ed. CAB International. 295–315.

Rehman, A., Nijhof, A. M., Sauter-Louis, C., Schauer, B., Staubach, C., & Conraths, F. J. 2017. Distribution of ticks infesting ruminants and risk factors associated with high tick prevalence in livestock farms in the semi-arid and arid agro-ecological zones of Pakistan. Parasit. Vectors. 10, 190.

Schoeman, S. J. 1989. Recent research into the production potential of indigenous cattle with special reference to the Sanga. S. Afr. J. Anim. Sci. 19, 55–61.

Sutherst, R. W., Maywald, G. F., Kerr, J. D., & Stegeman, D. A. 1983. The effect of cattle tick (*Boophilus microplus*) on the growth of *Bos indicus* x *B. taurus* steers. Aust. J. Agric. Res. 34, 317–327.

Turner, L. B., Harrison, B. E., Bunch, R. J., Neto, L. R. P., Li, Y., & Barendse, W. 2010. A genome-wide association study of tick burden and milk composition in cattle. Anim. Prod. Sci. 50, 235–245.

Utech, K. B. W., Wharton, R. H., & Kerr, J. D. 1978. Resistance to *Boophilus microplus* (Canestrini) in different breeds of cattle. Aust. J. Agric. Res. 29, 885–895.

Webb, E. C., & David, M. 2002. The efficacy of neem seed extract (*Azadirachta indica*) to control tick infestation in Tswana, Simmentaler and Brahman cattle. South African J. Anim. Sci. 32, 1–6.

Wikel, S. K., & Bergman, D. 1997. Tick-host immunology: Significant advances and challenging opportunities. Parasitol. Today. 13, 383–389.

Chapter 4

The Association of Copy Number Variations with Tick Resistance in South African Nguni Cattle

4.1 Abstract

Ticks and tick-borne diseases pose a major threat to livestock industries worldwide. The South African Nguni is a locally adapted cattle breed that is known for its resilience to ticks and tick-borne diseases. Copy number variation regions (CNVRs) comprise insertions, duplications and deletions within the genome that are larger than 1kb and play a possible role in adaptation. A preliminary investigation to determine the association between CNVRs and tick resistance in South African Nguni cattle was performed. Tick count data was collected from 347 randomly selected Nguni cattle from three different locations within South Africa over a period of two years. Data was split per location and summary statistics were used to determine quartile and interquartile ranges of tick counts. Animals which had an average tick count lower than or equal to the first quartile were classified as resistant (1) while all animals with an average tick count greater than or equal to the third quartile were classified as susceptible (0). Animals which were neither resistant nor susceptible were removed from further analyses. DNA extracted from hair and blood samples was genotyped using the Illumina BovineSNP50 assay. Quality control and sample pruning was performed using Plink (Version 1.07) leaving 41 193 high quality SNPs. LogR ratios and B allele frequency data of filtered SNPs was extracted and *PennCNV* software was utilized to identify 1 501 CNVs. A data file containing respective copy number states (0, 1, 2, 3 or 4) of CNVR loci for each animal was generated. Contingency tables testing the hypothesis that tick resistance is associated with CNVRs was run using STATISTICA 64. Seventeen CNVRs located on chromosomes 1, 2, 6, 7, 8, 9, 12, 15, 17, 20, 21, 22, 24 and 29 demonstrated a significant ($p<0.05$) association with tick resistance. Associated CNVRs covered 17 genes that play a role in multiple molecular functions and biological processes including catalytic activity, binding functions and immune, metabolic, cellular and reproduction processes respectively. This study is the first of its kind to demonstrate a significant association between tick count and copy number variations in South African Nguni cattle.

4.2 Introduction

Ticks transmit a magnitude of pathogenic diseases and are a major threat to livestock industries worldwide. Current tick control methods include acaricides, biological control, grazing management, vaccines and breeding strategies (Frisch, 1999). These approaches, do however have compromising side effects. Chemical residues in livestock products, environmental contamination

and the increasing levels of host resistance are issues faced by farmers on a daily basis (Porto Neto *et al.*, 2011). There is thus a need to develop alternative control methods. Exploiting genetically resistant breeds and utilizing host resistance is an innovative tick control strategy that is long-term, safe, effective and economically viable. The South African Nguni is an indigenous Sanga type cattle breed that is well known for its resilience to ticks and tick-borne diseases (Schoeman, 1989; Makina *et al.*, 2014). These multi-coloured, small to medium framed animals have undergone minimal synthetic breeding and are well adapted to harsh environmental conditions (Schoeman, 1989; Makina *et al.*, 2014; Wang *et al.*, 2015). The remarkable adaptability and resistance of the Nguni breed could be due to favourable phenotypic and/or genotypic characteristics however, this has not yet been fully established (Marufu *et al.*, 2011b).

Genomic variants and their impact on phenotypic variation in livestock, is an important area of research (Kijas *et al.*, 2011). Several forms of genomic variation exist and include single nucleotide polymorphisms (SNPs), gene or chromosomal rearrangements, nucleotide or gene alterations and copy number variations (CNVs) (Feuk *et al.*, 2006; Ibeagha-Awemu *et al.*, 2008; Liu *et al.*, 2010). Copy number variations are segments of chromosomal regions which are larger than 1kb in size and comprise of insertions, deletions and duplications across the genome (Sebat *et al.*, 2004; Tuzun *et al.*, 2005; Cicconardi *et al.*, 2013). These regions cover a larger percentage of the genome when compared with SNPs, and could exert an array of effects including changing gene structure and dosage, alternating gene regulation and exposing recessive alleles (Zhang *et al.*, 2009). Understanding the genetic background of functional and productive traits in cattle is of great economic importance and biological significance and therefore understanding the different forms of genetic variation, particularly CNVs, and their association with important phenotypes is an area of major interest (Cicconardi *et al.*, 2013).

Copy number variations have been detected within the bovine genome (Bae *et al.*, 2010; Hou *et al.*, 2011; Liu & Bickhart, 2012) and although their contribution towards phenotypic variation and disease susceptibility is partially known, there is still a lot more to be discovered (Cicconardi *et al.*, 2013). Following the discovery of bovine CNVs, the genetic effects of CNVs and their impact on economically important traits in cattle can be investigated (Bae *et al.*, 2010). Genes found within copy number variable regions (CNVRs) in cattle have been recognized to play a role in biological response and adaptation mechanisms (Kijas *et al.*, 2011), however it is an area that lacks research (Bae *et al.*, 2010). Recent studies in cattle, have established genes within CNVRs to be associated with multiple traits including growth (Zhou *et al.*, 2016), milk production (Xu *et al.*, 2014a), nematode resistance (Hou *et al.*, 2012), fertility (Glick *et al.*, 2011) and meat tenderness (Da Silva *et al.*, 2016). Copy number variations are evident in Nguni cattle (Wang *et al.*, 2015). Genes detected within CNVRs in this study were responsible for multiple biological processes and could play a role in the adaptive traits evident in Nguni cattle populations. Genes within CNVRs and their association with

tick resistance in Nguni cattle is an area of research still to be explored. Up to now, no study of the relationship between CNVRs and tick count in Nguni cattle has been published. The present study is the first to determine the association between CNVRs and tick count in South African Nguni cattle. Using Illumina BovineSNP50 assay data (Illumina Inc., San Diego, CA, USA) of 347 Nguni cattle across South Africa, we detected CNVs and performed an association analysis between CNVRs and tick count in order to identify genes associated with tick resistance. Genes detected within CNVRs in this study could be used in future selection and breeding programs to improve tick resistance among cattle populations.

4.3 Methods and Materials

4.3.1 Animal samples and tick counts

Tick count data was collected over a two-year period from 347 randomly selected Nguni cattle across three different provinces in South Africa. The method used to collect data has been described in chapter 3. As locality has a significant influence on tick count (Chapter 3), distribution statistics of Nguni tick counts were determined per location. These statistics were used to classify animals as susceptible (0) or resistant (1) across two levels of resistance. For resistance level one (L1), those animals demonstrating a mean tick count of less than or equal to the 1st quartile were classified as resistant, while all animals with a tick count of greater than or equal to the 3rd quartile were classified as susceptible. Those animals falling between these ranges were excluded from further analyses. For resistance level two (L2), half the standard deviation was added to the 1st quartile and subtracted from the 3rd quartile to determine the lower and upper limits for resistance and susceptibility. As per L1, those animals lying between these two limits were excluded from further analyses.

4.3.2 DNA sample collection

Blood and hair samples were taken from 347 Nguni cattle as described by Mapholi *et al*. (2016). DNA was extracted from blood and hair samples by proteinase-K digestion, followed by phenol: chloroform: isoamyl alcohol extraction and ethanol precipitation (Sambrook *et al*., 1989). The quality and quantity of extracted DNA was measured using a Thermo Scientific NanoDrop 2000 spectrophotometer. A Qubit 2.0 Fluorometer was used to verify DNA concentrations which were then normalized to 50ng/µl. Only those samples with a concentration of 45ng/µl or more were utilized for subsequent analysis and genotyping.

4.3.3 Single nucleotide polymorphism detection and quality assessment

DNA samples were genotyped using the Illumina BovineSNP50 assay (Illumina Inc., San Diego, CA, USA). The Illumina BovineSNP50 assay spans the bovine genome and consists of 54 609 highly informative markers. Quality control and sample pruning was performed using Plink (Version 1.07)

(Purcell *et al.*, 2007). Single nucleotide polymorphisms with a minor allele frequency of greater than 0.02 and/or a genotype rate of less than 0.95 were removed from the data set.

4.3.4 Generation of CNV calls and filtering

Using *GenomeStudio*, a signal intensity file was created for analysis by *PennCNV* (Wang *et al.*, 2007a). This file contained the Log R ratio (LRR), B allele frequency (BAF), G type, chromosome and position of each individual and probe. The Log R ratio is the total fluorescent intensity signal from both sets of probes or alleles at each SNP (Wang *et al.*, 2007a). The BAF is the inference of the relative ratio of the fluorescent signals between two probes or alleles at each SNP (Wang *et al.*, 2007a). *PennCNV* is a CNV detection algorithm which utilizes fluorescent signal intensities from SNP arrays (Cicconardi *et al.*, 2013; Zhu *et al.*, 2016). Other CNV detection algorithms have been developed, but *PennCNV* has outperformed many of them with low bias and enhanced accuracy and CNV calling (Wang *et al.*, 2007a; Cicconardi *et al.*, 2013; Xu *et al.*, 2013). *PennCNV* utilizes a Hidden Markov Model (HMM) to detect CNVs from SNP genotyping data (Wang *et al.*, 2007a). The HMM model presumes that the hidden copy number state of each SNP is that of the previous SNP (Wang *et al.*, 2007a). The *PennCNV* compile_pfb, detect_cnv.pl and filter_cnv.pl scripts were run. Compile_pfb was used to create the population frequency of B allele (pfb) file (Wang *et al.*, 2007a), detect_cnv.pl detected CNVs on 29 autosomes (Wang *et al.*, 2015) and filter_cnv.pl was run to perform quality control. For quality control, CNVs with a LRR standard deviation (SD) greater than 0.3, a BAF drift of greater than 0.01 and/or a waviness factor of greater than 0.05 were excluded from the data set.

4.3.5 Generation of CNVRs from CNVs for association analyses

A CNV output file was created using *PennCNV*. This was merged with L1 and L2 resistance data of the 179 and 306 animals respectively. The method of Lin *et al*. (2013) was used to create CNVRs from overlapping and adjoining CNVs identified within and across samples. Copy number variable regions were compiled per chromosome. For two CNVs with CNV1 beginning at position a and ending at position b and CNV2 running from c to d, with $a < c < b < d$, three different CNVRs would be generated (Keel *et al.*, 2016). CNVR_1 would be at position a, CNVR_2 would be from position b to c and CNVR_3 would be at position d. Details on constructing CNVRs are illustrated below (Figure 4.1). For this study copy number variable regions were considered to be greater than 1 kb in size (Bae *et al.*, 2010; Cicconardi *et al.*, 2013; Wang *et al.*, 2016) and therefore if a CNVR was detected at only one breakpoint, it was removed from the matrix. In order to reduce the number of false positive CNVs, CNVRs were removed from further analyses if they were only detected in one individual. Each animal was assigned a copy number (CN) state as per the CNV detected at that location. It was ensured that CNVRs were created according to the breakpoints and CN states as per CNV. Copy number states were classified as follows: No CNV present (0), hemizygous deletion (1), homozygous deletion (2), hemizygous duplication (3) and homozygous duplication (4). Copy

number variable region matrices, one for L1 resistance and one for L2 resistance, were created. These matrices included the sex, location and resistance data of all animals.

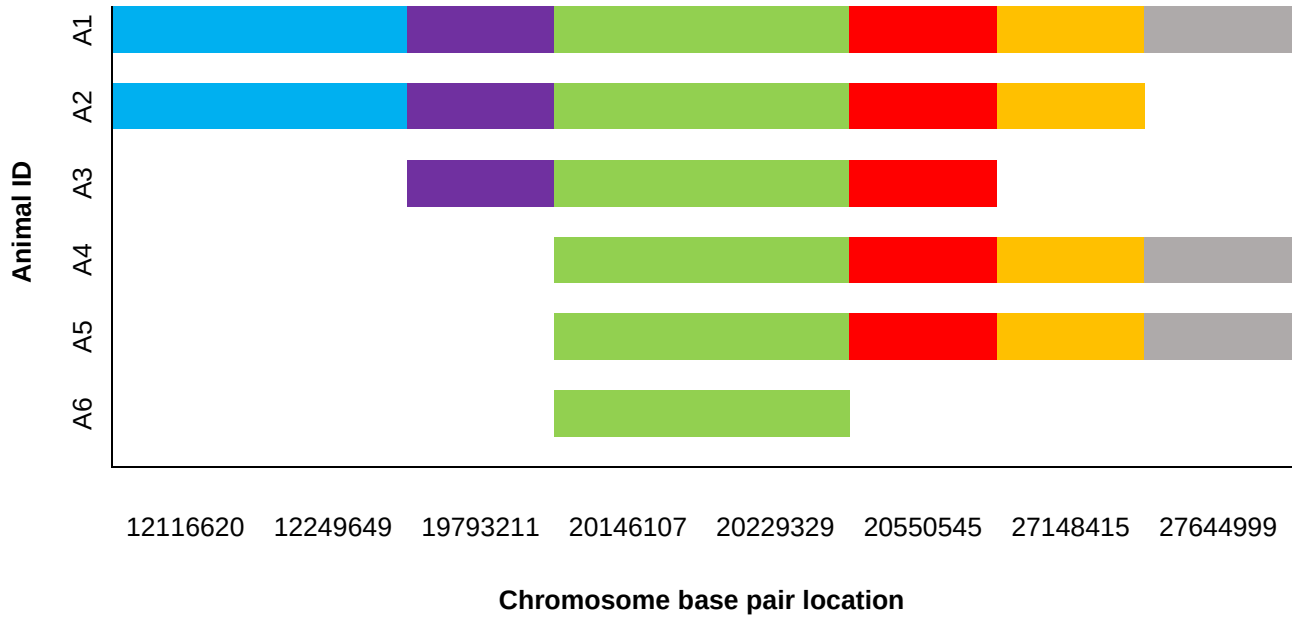

Figure 4.1 A section of chromosome one demonstrating how CNVRs were constructed from CNVs identified in South African Nguni cattle. Every CNV was deconstructed to form multiple CNVRs as indicated by different colours.

4.3.6 Association analyses

Data was analysed using STATISTICA 64 (Dell Inc. (2016). Dell Statistica (data analysis software system), version 13. https://support.software.dell.com/statistica/)

Contingency tables were used to test the following hypotheses:
H_0: There is no association between tick resistance and CNVRs in Nguni cattle
H_1: There is an association between tick resistance and CNVRs in Nguni cattle.
Both variables are nominal of nature and the maximum likelihood chi-square test was used to compare the frequencies observed due to an animals tick resistance/susceptibility [Susceptible (0) or resistant (1)] and CNVR (0, 1, 2, 3 or 4).

The test statistic G represents the maximum likelihood test and is computed as follows:

$$G = 2 \sum_{i}^{n} O_i \cdot \ln\left(\frac{O_i}{E_i}\right)$$

Where

n is the total number of observations

O_i is the observed frequency of each value (L1/L2 – 0/1; CN per CNVR – 0, 1, 2, 3, 4)

E_i is the expected frequency for each given value under the null hypothesis

The hypothesis is rejected if the value of G exceeds the 95% percentile of the chi-square distribution with (r-1)*(c-1) degrees of freedom, where r = number of levels of tick resistance and c= number of levels of CNVR as observed.

4.3.7 Gene ontology analyses

The gene content of those CNVRs found to be significantly associated with tick resistance was investigated. RefGene annotations (USCS, downloaded on htt://genome.ucsc.edu/goldenpath/gbdDescriptionsOld.html) were used to identify genes within a 10Mb region surrounding specific CNVRs. The PANTHER database was used to establish *Bos taurus* gene ontologies. Using PANTHER, the hypothesis that genes within CNVRs were over or under represented in PANTHER molecular function, pathways, biological processes and cellular components was tested using the Bonferroni corrections on the *panterdb.org* website at the p = 0.05 level.

4.4 Results and Discussion

4.4.1 Nguni tick count summary statistics

Animals were classified as susceptible (0) or resistant (1) across two levels of resistance (L1 and L2), according to cut-offs determined using distribution statistics (Table 4.1). After resistance classification 179 and 306 animals remained at L1 and L2 respectively (Table 4.2). At both levels of resistance, Loskop had the lowest number of susceptible animals, but in this study this location displayed the highest mean tick count (Chapter 3). This could be due to the expression of genes involved in immunity mechanisms of animals subject to greater tick infestation. Wang *et al*. (2007) and Piper *et al*. (2008) found animals previously subject to infestation to have enhanced levels of resistance due to up and downregulation of different genes.

Table 4.1 Distribution cut-offs for susceptible (0) and resistant (1) animals at Mukhuthali, Loskop and Roodeplaat at L1 and L2.

Resistance level	Resistance category*	Location		
		Mukhuthali	Loskop	Roodeplaat
L1	0	≥20.35	≥34.24	≥30.04
	1	≤15.82	≤25.59	≤22.96
L2	0	≥18.35	≥30.99	≥27.54
	1	≤17.82	≤28.84	≤25.46

* 0 – Susceptible; 1 – Resistant

Table 4.2 The number of susceptible (0) and resistant (1) animals at Mukhuthali, Loskop and Roodeplaat at L1 and L2

Resistance Level	Resistance Category*	Location		
		Mukhuthali	Loskop	Roodeplaat
L1	0	30	23	34
	1	31	25	36
L2	0	56	38	57
	1	55	38	62

* 0 – Susceptible; 1 – Resistant

4.4.2 Single nucleotide polymorphism quality control

The Illumina BovineSNP50 assay (Illumina Inc., San Diego, CA, USA) that spans 54 609 SNP markers was used in this study. After genotyping, 1 315 variants with a call rate of less than 95% were removed along with an additional 12 332 variants with a minor allele frequency of less than 0.02. This left 41 193 SNPs for further analyses.

4.4.3 Copy number variant detection

Using *PennCNV* a total of 1 501 CNVs in 277 of the 347 animals were detected on all autosomes. These CNVs ranged from 1 kb to 3 Mb with the highest frequency of CNVs being found on chromosome 1 (8.5%). Isolated deletion and duplication CNV events as well as CNVs comprising both deletions and duplications were identified (Table 4.3). *PennCNV* has been seen to identify more deletions than duplications (Eckel-Passow *et al.*, 2011) in discord with the current study as CNVs detected comprised of more duplications. Hou *et al.* (2011) studied CNVs in a variety of cattle breeds

and reported African breeds to have more CNV duplications than deletions when compared with composite, taurine and indicine breeds. A later study by Hou *et al*. (2012) reported more CNV duplications than deletions in Angus cattle. Selective pressures have an influence on deletions and duplications (Hou *et al.*, 2012). Deletions are known to be further away from genes and could have less of an impact, while the effect of duplications could be greater due to amplification, positive selection and conversion (Hou *et al.*, 2012).

Table 4.3 Summary statistics of CNV deletions and duplications detected in 347 South African Nguni cattle. The type (Del, dup or both), number of CNVs (CNVs), minimum length (MinL), maximum length (MaxL) and average length (AvL) of CNVs.

Type*	CNVs	MinL (bp)	MaxL (bp)	AvL (bp)
Del	705	1 023	2 193 045	222 189.03
Dup	744	2 118	3 252 826	407 660.83
Del/Dup	52	1 743	1 017 206	139 232.75

*Del – Deletion; Dup – Duplication; Del/Dup – Deletion and duplication CNV

4.4.4 Copy number variable region discovery

Copy number variable regions were generated by aggregating overlapping CNVs as described in a published protocol by Lin *et al*. (2013). A total of 344 CNVRs were discovered across 29 autosomes (Table 4.4). Sizes of CNVRs ranged from 14 to 914 kb with an average of 121 kb and a median of 91 kb. The distribution of CNVRs varied across chromosomes. Chromosome 6 displayed the greatest number of CNVRs (27) with chromosome 25 having the least CNVRs (3). Jiang *et al.* (2012) and Cicconardi *et al*. (2013) both report the greatest number of CNVRs on chromosome 6, however Jiang *et al.* (2012) detected the least amount of CNVRs (0) on chromosomes 22, 25 and 29. Cicconardi *et al*. (2013) report the least CNVRs on chromosome 29. In this study chromosome 29 presented nine CNVRs.

The number of CNVRs across chromosomes is known to differ. Larger chromosomes (Fadista *et al.*, 2010) and pericentrometric and subtelometric regions (Liu *et al.*, 2010) are seen to harbour a greater number of CNVs. Large numbers of CNVs have also been detected on chromosomes which have a greater number of segmental duplications (Liu *et al.*, 2009, 2010). Segmental duplications are DNA regions larger than 1kb whose sequences are at least 90% identical to multiple genomic loci (Bailey, 2002; Cicconardi *et al.*, 2013). These regions are catalysts and hotspots for the formation of CNV's (Sharp *et al.*, 2005; Marques-Bonet *et al.*, 2009). Chromosomes 5, 18, 27 and 29 demonstrate a large percentage of segmental duplications (Liu *et al.*, 2010). In this study chromosomes 5, 18, 27 and 29 presented 11, 6, 8 and 9 CNVRs respectively (Table 4.4).

Chromosomes displaying a large number of segmental duplications have been reported to display a small number of SNPs (Liu *et al.*, 2009). This is partially true for this study, as 1 054, 750 and 810 SNPs were detected on chromosomes 18, 27 and 29 respectively, however, 1 629 SNPs were identified on chromosome 5 which is higher than the average number of SNPs per chromosome. Estivill & Armengol (2007) state that a significant number of CNVs are not genotyped due to the fact that they fall within regions not covered by SNP arrays. In this study, many CNVRs were detected within chromosomes displaying a large number of SNPs, however we cannot rule out the possibility of more CNVRs being identified within regions not tagged by SNP arrays.

Table 4.4 The number of SNPs, the number of CNVRs and the minimum length (MinL), maximum length (MaxL) and average length (AvL) of the CNVRs identified across 29 autosomes of 347 South African Nguni cattle.

CHR	SNPs	CNVRs	MinL (bp)	MaxL (bp)	AvL (bp)
1	2 624	23	14 900	313 572	104 540.35
2	2 127	22	22 341	580 894	106523.05
3	1 956	20	22 169	714 764	187 840.65
4	1 953	24	21 493	514 764	118 281.88
5	1 629	11	20 558	443 571	120 722.91
6	2 010	27	25 877	399 918	101 747.15
7	1 777	16	23 623	375 762	110 465.38
8	1 878	11	30 741	271 288	132 265.64
9	1 536	11	28 923	301 185	95 994.18
10	1 673	9	59 446	115 398	93 385.11
11	1 688	20	26 150	534 854	132 488.60
12	1 345	20	33 437	392 714	146 818.90
13	1 344	6	49 104	195 515	125 694.33
14	1 432	9	27 947	264 011	122 138.11
15	1 319	9	52 550	277 003	125 308.56
16	1 294	5	38 759	165 774	97 461.00
17	1 255	11	37 542	914 679	170 086.36
18	1 054	6	38 700	161 641	76 220.83
19	1 062	4	21 690	92 602	48 940.00

CHR	SNPs	CNVRs	MinL (bp)	MaxL (bp)	AvL (bp)
20	1 207	15	23 865	284 950	128 003.00
21	1 118	11	25 005	408 468	116 309.36
22	1 007	4	21 522	224 514	100 427.00
23	848	4	26 699	61 754	45 677.75
24	977	8	38 738	171 046	80 942.75
25	762	3	53 250	429 046	295 852.33
26	863	10	61 992	240 104	118 450.00
27	750	8	54 342	232 768	114 140.00
28	751	8	104 015	186 333	147 767.00
29	810	9	16 685	466 283	120 210.33

Of the 344 CNVRs detected, 113 showed copy number loss, 89 showed copy number gain and 142 showed both copy number loss and gain (Additional file 4.1). These results are similar to those of other bovine studies. Jiang *et al.* (2012) report 99 CNVRs in Chinese Holstein of which 81, one and 17 were loss, gain and loss and gain events respectively. Cicconardi *et al.* (2013) report 326 CNVRs in *Bos taurus* breeds of which 192 were loss events, 31 gain events and 103 loss and gain events. These studies report more loss than gain events which is in concurrence with this study. In CNVRs, loss events are more common than gain events (Cicconardi *et al.*, 2013), which could be attributable to biological and/or technical reasons, such as the CNV detection algorithm used, the mechanism of CNV formation or selection pressures imposed (Fadista *et al.*, 2010). The *PennCNV* algorithm is reported to identify more deletions than duplications (Eckel-Passow *et al.*, 2011). More deletions are also generated through the mechanism of non-allelic homologous recombination (Turner *et al.*, 2008).

The CNVRs detected covered 1.66% of the bovine genome. This figure is slightly higher than that of Wu *et al.* (2015) who report 263 CNVRs, covering 1.41% of the genome. The CNVRs detected by Wu *et al.* (2015) comprised of 113 loss, 89 gain and 142 loss and gain events. Comparison of these results with autosomal CNVRs identified in several cattle studies showed very little overlap of exact CNVR breakpoints. The CNVRs identified in this study overlap or lie within close proximity (<1Mb) to CNVRs identified by Bae *et al.* (2010), Jiang *et al.* (2012), Cicconardi *et al.* (2013) and Wu *et al.* (2015). Many factors, including sample data, experimental method, allelic frequencies and the statistical model used, can have an influence on the exact locality of CNV breakpoints (Dellinger *et al.*, 2010).

4.4.5 Copy number variable regions associated with tick resistance

At resistance level 1 (L1), six CNVRs demonstrated a significant association ($p<0.05$) with tick count. These CNVRs were detected on chromosomes 6, 8, 12, 17, 21 and 22 (Figure 4.2). Resistance level 2 (L2) reported 14 CNVRs significantly associated ($p<0.05$) with tick count (Figure 4.3). These CNVRs were located on chromosomes 1, 2, 6, 7, 9, 12, 15, 20, 21, 22, 24 and 29. The difference in the number of significant CNVRs at the two different resistance levels could be due to stringency and the presence of more copy numbers, with the less stringent classification (L2) harbouring more significant CNVRs. Three CNVRs on three different chromosomes were significant across both levels of resistance. These CNVRs were located on chromosome 6, 12 and 21 between base pairs 10 760 779 to 10 838 635, 90 704 572 to 90 778 028 and 71 025 601 to 71 109 676 respectively. In total, 17 CNVRs were significantly associated ($p<0.05$) with tick count across L1 and L2. Significant CNVRs were detected on chromosomes 1, 2, 6, 7, 8, 9, 12, 15, 17, 20, 21, 22, 24 and 29.

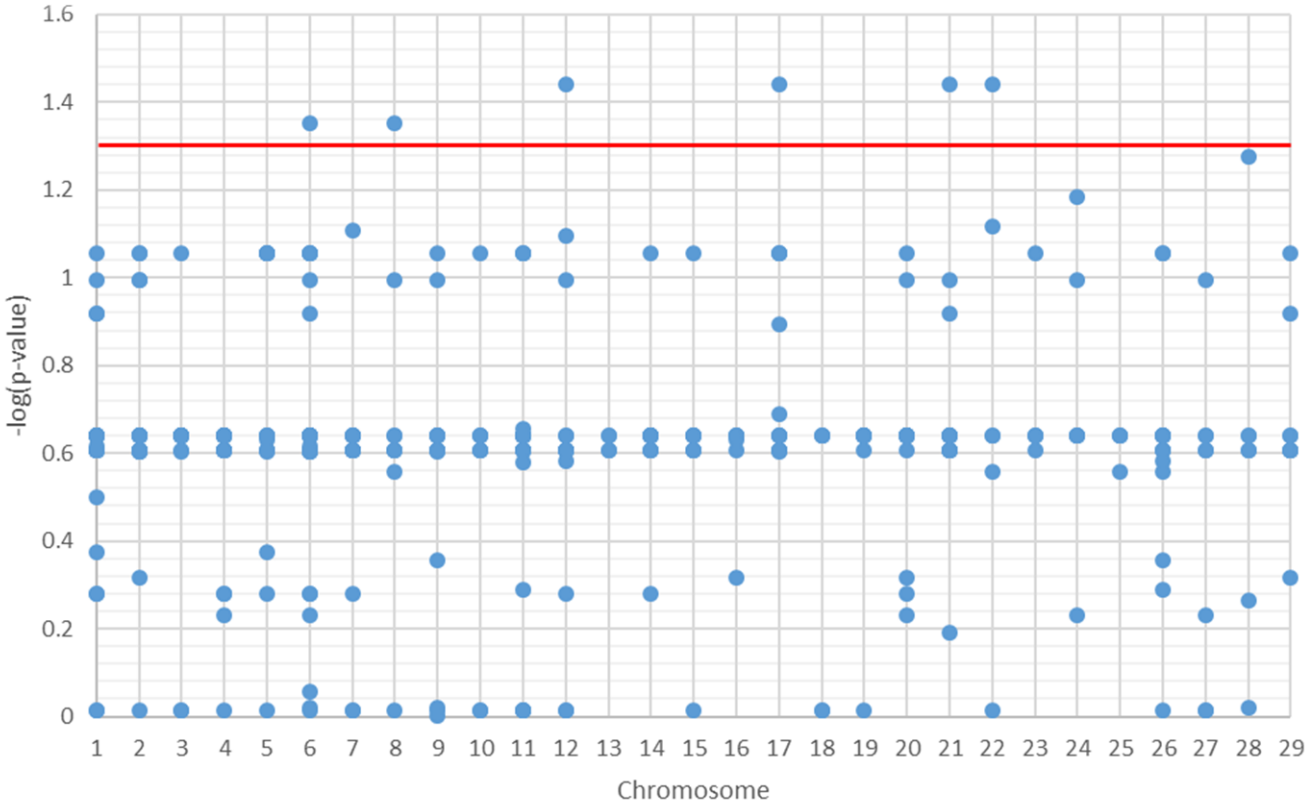

Figure 4.2 Manhattan plot for L1 CNVRs on all 29 autosomes (x-axis) and the corresponding $-\log_{10}$p-value (y-axis) indicating the association strength with tick resistance. The red line indicates the $-\log_{10}$p-value (0.05) with all CNVRs above this being significant

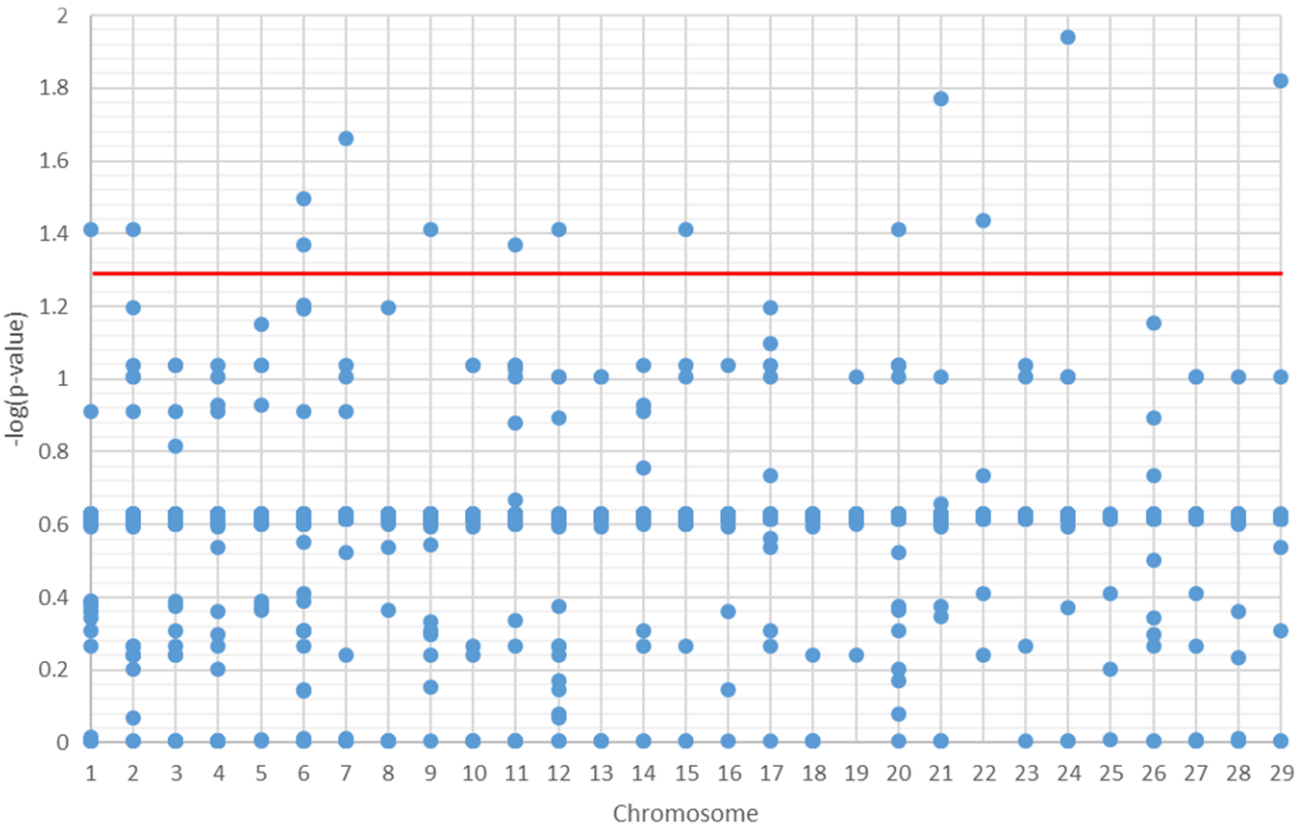

Figure 4.3 Manhattan plot for L2 CNVRs on all 29 autosomes (x-axis) and the corresponding $-\log_{10}$p-value (y-axis) indicating the association strength with tick resistance. The red line indicates the $-\log_{10}$p-value (0.05) with all CNVRs above this being significant

In total, 55 of the 347 animals harboured significant CNVRs (Table 4.5). Significant CNVRs were more prevalent in susceptible animals than in resistant animals across L1 and L2. At L1, four significant CNVRs were detected in susceptible animals, while resistant animals presented two. Eleven and seven significant CNVRs were detected in susceptible and resistant animals at L2 respectively. The average and maximum lengths of CNVRs was greater in susceptible than resistant animals across L1 and L2, however L2's average lengths were much shorter than those of L1. The minimum length of CNVRs at L1 was longer in resistant animals when compared to susceptible animals. At L2, the minimum length was the same in both resistance categories. This is due to the fact that the shortest CNVR, which was discovered on chromosome 29, was detected in both susceptible and resistant animals. Significant CNVRs overlapped a total of 12 and 9 genes in susceptible animals at L1 and L2 respectively. No genes overlapped with significant CNVRs in resistant animals at L1, however significant CNVRs in resistant animals at L2 overlapped with three genes.

Table 4.5 The number of animals (ANMLs) per resistance level (ResL), per resistance category (ResC) with their number (CNVRs), average length (AvL), minimum length (MinL) and maximum length (MaxL) of CNVRs and the number of genes (NGEN) within them.

ResL	ResC	ANMLs	CNVRs	AvL (bp)	MinL (bp)	MaxL (bp)	NGEN
L1	0	5	4	120 836.50	73 456	224 514	12
	1	6	2	90 370.00	77 856	102 888	0
L2	0	29	11	79 436.26	38 192	214 270	9
	1	23	7	64 388.31	38 192	94 535	3

Table 4.6 reports the distribution of CNVRs and SNPs (Mapholi *et al.*, 2016) significantly associated with tick resistance in South African Nguni cattle. The study by Mapholi *et al.* (2016) report the greatest number of significant SNPs (4) on chromosomes 14 and 17, however in this study zero and one significant CNVRs were detected on these chromosomes respectively. In this study chromosome 6 presented the most significant CNVRs (4), with one at L1 and three at L2. In the study by Mapholi *et al.* (2016), only one significant SNP was reported on chromosome 6. No CNVRs from this study or SNPs from the study by Mapholi *et al.* (2016) were significantly associated with tick resistance across chromosomes 4, 13, 16, 23, 25, 27 and 28. In this study, chromosomes 2, 9, 20, 21, 22, 24 and 29 presented CNVRs significantly associated with tick resistance, however Mapholi *et al.* (2016) report no significant SNPs associated with tick resistance on these chromosomes. Chromosomes which presented SNPs significantly associated with tick resistance from the study by Mapholi *et al.* (2016), but no significant CNVRs from this study, include chromosome 3, 10, 11, 14, 18, 19 and 26.

A deletion CNVR was detected on chromosome 1 between base pairs 4 957 717 and 5 034 441 in susceptible animals at L2. Cicconardi *et al.* (2013) report a deletion (CN=2) and a duplication (CN=3) in *Bos taurus* breeds that partially overlaps the region detected in this study. Duran Aguilar *et al.* (2017) report chromosome 1 to harbour two CNVs associated with somatic cell score (SCS) in Holstein cattle. These CNVs are however 89 and 142 Mb upstream from the CNV reported in the present study. Three SNP markers suggestively associated with tick resistance in Nguni cattle have also been reported on this chromosome (Mapholi *et al.*, 2016). These SNPs were located at base pairs, 96 382 717, 146 632 014 and 74 643 836 and were associated with total belly tick count and *Amblyomma hebraeum* tick tail count and total body count respectively.

Chromosome 2 encompassed a deletion CNVR between base pairs 53 894 600 and 53 951 533 at L2 in susceptible animals. The study by Hou *et al.* (2011b) report a deletion in Angus cattle, between base pairs 53 907 395 and 54 130 293 which partially overlaps this region. Machado *et al.* (2010)

report a QTL region within this chromosome to be associated with tick resistance. This QTL was located at 28.4 to 45.6 Mbp, however it did not span the region detected in this study. Eleven CNVRs were detected on chromosome 9 of which only one was significantly associated with tick resistance. This CNVR was a duplication across base pairs 94 991 477 to 95 065 382. This region has not been reported in previous CNV studies, nor have there been any QTL reported which play a role in parasite or disease resistance.

Chromosome 12 presented a deletion CNVR in susceptible animals at L1 and L2. This CNVR was located between base pairs 90 704 572 to 90 778 028. Mapholi *et al.* (2016) report a SNP marker associated with *A. hebraeum* belly tick count located at 12 141 029 base pairs.

Chromosome 15 and 21 both report CNVR deletions (CN=1) in susceptible animals (Table 4.4 and Table 4.5). Mapholi *et al.* (2016) report a SNP marker on chromosome 15 to be associated with tick resistance in Nguni cattle. This marker was located at base pairs 23 738 373 and was associated with the total tick count of *A. hebraeum* ticks on the perineum. Sollero *et al.* (2017) report three QTLs located at base pairs 37 534 610, 37 575 605 and 72 980 544 on chromosome 15 to be associated with tick resistance. Although these markers did not cover the CNV regions detected in this study there is evidence that CNVRs may exert a downstream effect on genes (Henrichsen *et al.*, 2009). These associations may be acting in isolation, but the possibility that these are linked should not be ruled out.

Table 4.6 The chromosomal distribution (CHR) of CNVR identified in this study at resistance level 1 (L1) and 2 (L2) and SNPs identified by (Mapholi *et al.*, 2016) (SNPs MAP) that demonstrate a significant association with tick resistance relative to the total number of CNVRs and SNPs on respective chromosomes detected in this study.

CHR	CNVR	CNVR L1	CNVR L2	SNPs	SNPs MAP
1	23		1	2 624	3
2	22		1	2 127	
3	20			1 956	1
4	24			1 953	
5	11			1 629	1
6	27	1	3	2 010	1
7	16		1	1 777	1*
					1
8	11	1		1 878	1
9	11		1	1 536	

CHR	CNVR	CNVR L1	CNVR L2	SNPs	SNPs MAP
10	9			1 673	1*
					2
11	20			1 688	3
12	20	1	1	1 345	1
13	6			1 344	
14	9			1 432	4
15	9		1	1 319	1
16	5			1 294	
17	11	1		1 255	4
18	6			1 054	1
19	4			1 062	1*
20	15		1	1 207	
21	11	1	1	1 118	
22	4	1	1	1 007	
23	4			848	
24	8		1	977	
25	3			762	
26	10			863	2
27	8			750	
28	8			751	
29	9		1	810	

*SNPs reaching significance at a genome-wide level ($p<0.05$), with other SNPs being significant at the suggestive level ($p<0.10$)

In this study, susceptible and resistant animals displayed CNVRs significantly associated with tick count (Table 4.7). These CNVRs comprised of hemizygous and homozygous deletions and duplications, with some regions containing both. The change in copy number can exert an effect on a phenotype through a variety of mechanisms (Gamazon & Stranger, 2015). Deletions and duplications can have an effect on genes through upregulation, reduced expression and/or the alteration of regulatory and functional processes (Gamazon & Stranger, 2015).

Three CNVRs were detected in resistant animals across chromosome 6. One of the CNVR regions was detected between base pairs 10 760 779 to 10 838 635 and comprised of a deletion and a duplication. These results concur with those of Jiang *et al.* (2012) and Cicconardi *et al.* (2013) who both report deletion, duplication CNVRs which partially and completely overlap this region respectively. An earlier study by Hou *et al.* (2011b) also report a CNVR completely overlapping this region, however it was only a deletion. Two other CNVRs detected within chromosome 6, were located between base pairs 10 105 101 to 10 199 636 and 20 297 397 to 20 350 438. Both CNVRs were deletions (CN=1) and partially overlapped with CNVRs detected by Hou *et al.* (2011b) and Boussaha *et al.* (2015) respectively. Multiple SNP markers within this chromosome have been associated with tick resistance in Braford, Hereford and Nguni cattle (Mapholi *et al.*, 2016; Sollero *et al.*, 2017). Mapholi *et al.* (2016) report a SNP marker located at 87 281 196 base pairs suggestively associated with total tick count on the belly. Three markers were reported by Sollero *et al.* (2017) and were located at base pairs 3 994 395, 4 193 024 and 49 130 874.

A single CNVR significantly associated with tick resistance was detected on chromosome 7, between base pairs 75 305 297 and 75 370 366. This CNVR was detected in one susceptible and six resistant animals, with hemizygous duplications (CN=3) corresponding to susceptibility and hemizygous and homozygous deletions (CN=1 and 2) corresponding to resistance. A duplication CNVR, located at base pairs 75 190 536 to 75 370 367, which completely spans the CNVR detected in this study, was reported by Hou *et al.* (2011b) in Angus cattle. Xu *et al.* (2014b) report a deletion CNV associated with faecal egg count in Angus cattle, located 35.4Mb downstream of the CNVR detected in this study. Single nucleotide polymorphism markers within this chromosome have been associated with tick resistance (Mapholi *et al.*, 2016; Sollero *et al.*, 2017) and SCS (Duran Aguilar *et al.*, 2017). Two markers reported by Mapholi *et al.* (2016), located at base pairs 58 781 492 and 110 608 386 were associated with total tick count on the head and *Boofilids* tick count on the perineum respectively. Sollero *et al.*, (2017) also report two SNP markers assocaited with tick resistance, however these were reported at base pairs 13 426 119 and 13 608 935. Four markers were associated with SCS in Holstein cattle as reported by Duran Aguilar *et al.* (2017).

The CNVR on chromosome 8, between base pairs 34 795 275 and 34 898 163 harboured a hemizygous deletion (CN=1) CNVR. This CNVR was detected in three resistant animals. Numerous genetic markers and variants have been located within this chromosome and associated with disease and parasite resistance. Hou *et al.* (2011a) detected a deletion CNVR in Angus cattle across base pairs 34 795 276 to 34 920 925, which overlapped with the region in this study. Quantitative trait loci spanning this region in Angus and Holstein cattle have been linked to gastrointestinal parasite resistance (Kim *et al.*, 2014) and SCS (Schnabel *et al.*, 2005) respectively. Kim *et al.* (2014) report a QTL spanning from 30.0-43.5Mbp with the QTL detected by Schnabel *et al.* (2005) located between 29.5 and 42.5Mbp. Other QTL across this chromosome have been associated with tick

resistance (Mapholi *et al.*, 2016; Sollero *et al.*, 2017).In Nguni cattle, Mapholi *et al.* (2016) report a marker associated with total *Boofilids* count on the perineum, located at base pairs 15 665 796. Sollero *et al.* (2017) report two mares assocaited with tick count, located at base pairs 50 525 859 and 107 821 440.

A hemizygous deletion (CN=1) was detected in three susceptible animals on chromosome 17, between base pairs 74 292 319 to 74 393 620. This CNVR overlapped and/or was in close proximity to a number of genes (Table 4.4). The effect of a deletion within this region could alter gene regulation and be responsible for gene regulation changes which could impact biological processes within an animal, therefore affecting its susceptibility (Zhang *et al.*, 2009; Gamazon & Stranger, 2015). Several previous studies report CNVRs to overlap with this region, however they report as complex CNVRs comprising of deletions, duplications and inversions (Hou *et al.*, 2012; Cicconardi *et al.*, 2013; Boussaha *et al.*, 2015). These CNVRs were located between base pairs 74 325 362 and 75 058 716, 74 130 891 and 76 487 767 and 74 335 993 and 74 392 322 as per Hou *et al.* (2011b), Cicconardi *et al.*, (2013) and Boussaha *et al.* (2015) respectively. Four QTL in Nguni cattle have also been associated with tick resistance within this chromosome (Mapholi *et al.*, 2016). Two of these markers were associated with the total *A. hebraeum* count on the perineum and were located at base pairs 43 990 974 and 44 000 618. The other two markers were located at base pairs 7 118 768 and 7 165 500 and associated with total *Rhipicephalus evertsi evertsi* count on the tail and whole body respectively.

At L2, Chromosome 20 presented a hemizygous deletion (CN=1) in three susceptible animals. This CNVR was located between base pairs 45 052 283 and 45 266 553. Hou *et al.*, (2011a) detected a deletion CNVR partially overlapping with this region between base pairs 45 083 602 and 45 144 742. Quantitative trait loci within chromosome 20 have been associated with tick resistance in American Braford and Hereford cattle (Sollero *et al.*, 2017), however when compared with the study of Mapholi *et al.* (2016) in Nguni cattle (Table 4.4), no SNPs were associated with tick resistance. Markers reported by Sollero *et al.* (2017) were detected at base pairs 17 837 675, 19 917 959, 56 196 291 and 71 498 820.

Chromosome 22 report two CNVRs. A hemizygous deletion (CN=1) CNVR was detected between base pairs 60 736 089 to 60 960 603 in three susceptible animals at L1. Previous studies report CNVRs overlapping with this region (Cicconardi *et al.*, 2013; Keel *et al.*, 2016). Cicconardi *et al.* (2013) report a complex CNVR overlapping with this region, located between base pairs 58 936 549 and 61 418 797. The deletion CNVR reported by Keel *et al.* (2016) was located within the region detected in this study between base pairs 60 897 501 and 60 910 000. The second CNVR significantly associated with tick resistance on chromosome 22 was located between base pairs 24 078 956 and 24 161 191. This CNVR was detected at L2 in nine susceptible and two resistant

animals. In susceptible animals, this region was complex as three animals displayed a hemizygous deletion with six animals portraying a hemizygous duplication. The CNVR in resistant animals comprised only of hemizygous duplications. No previous bovine studies report CNVRs in close proximity to this region, therefore potentially describing this chromosome as a cold spot for CNVs, however this would need further investigation. In this study this chromosome only presented four CNVRs, which is relatively low to that of CNVRs detected on other chromosomes (Table 4.4). These results are similar to that of Wang *et al.* (2015), whom report the detection of six CNVRs on chromosome 6 in Nguni cattle. No QTL have been associated with tick or parasite resistance across this chromosome either.

A CNVR detected across susceptible and resistant animals at L2 was that within chromosome 24 between base pairs 28 154 039 and 28 196 203. This CNVR was detected in three susceptible animals and five resistant animals. In susceptible animals, this region was detected as a hemizygous deletion (CN = 1), however in resistant animals this CNVR was detected as a homozygous deletion and a hemizygous duplication (CN = 2 and 3). Two previous studies in Angus cattle have reported deletion CNVRs overlapping this region located from base pairs 28 083 771 to 28 244 324 (Hou *et al.*, 2011) and 28 154 040 to 28 196 204 (Hou *et al.*, 2012). No QTL associated with tick resistance have been reported within this chromosome.

Chromosome 29 presented a CNVR at L2 in four susceptible and two resistant animals. Hemizygous deletions (CN=1) corresponded to susceptibility with hemizygous duplications (CN=3) corresponding to resistance. This CNVR was located between base pairs 50 202 589 to 50 240 781. Two previous studies report CNVRs overlapping with this region (Hou *et al.*, 2012; Cicconardi *et al.*, 2013). Hou *et al.* (2011b) report this CNVR as a duplication located between base pairs 49 922 378 and 50 797 149, while Cicconardi *et al.* (2013) report this as a complex region between base pairs 47 457 315 and 51 979 344. In the study by Mapholi *et al.* (2016), no SNP markers were significantly associated with tick resistance within this chromosome.

Table 4.7 CNVRs significantly (CNVR) associated with tick count in South African Nguni cattle at two resistant levels (ResL), their prevalence in resistant and susceptible animals (SR), the number of occurrences (TOT) with different copy numbers (CN) and the genes lying within 10 Mb of the CNVRs (GEN).

CNVR	ResL	SR	CN*				TOT	GEN
			1	2	3	4		
chr1:4957717-5034441	L2	S	3				3	-
chr2:53894600-53951533	L2	S	3				3	KYNU
chr6:10105101-10199636	L2	R	3				3	-
chr6:10760779-10838635	L1&L2	R	1		4		5	-
chr6:20297397-20350438	L2	R	3					-
chr7:75305297-75370366	L2	R	5	1			6	GABRB2
		S			1		1	
chr8:34795275-34898163	L1	R	3				3	-
chr9:94991477-95065382	L2	S			3		3	-
chr12:90704572-90778028	L1 & L2	S	3				3	TFDP1, TMCO3, ATP4B, GRK1
chr15:5474020-5537822	L2	S	3				3	-
chr17:74292319 - 74393620	L1	S	3				3	TUBA3E, AIFM3, PRODH, THAP7, MZT2, SLC7A4, LZTR1
chr20:45052283-45266553	L2	S	3				3	-
chr21:71025601-71109676	L1 & L2	S	4				4	C14orf79
chr22:24078956-24161191	L2	S	3		6		9	-
		R			2		2	

CNVR	ResL	SR	CN*				TOT	GEN
			1	2	3	4		
chr22:60736089-60960603	L1	S	3				3	CHCHD6
chr24:28154039-28196203	L2	S	3				3	-
		R		1	4		5	
chr29:50202589-50240781	L2	S	4				4	LSP1, TNNT3
		R			2		2	

*Hemizygous deletion (CN=1), homozygous deletion (CN=2), hemizygous duplication (CN=3) and homozygous duplication (CN=4).

4.4.6 Gene ontology

The ref gene annotation database was used to analyse genes of the associated CNVRs. Significant CNVRs overlapped and/or lay within 10Mb regions of 17 genes (Table 4.7 and Additional file 4.2). The molecular functions, biological processes and cellular functions of all genes within or overlapping CNVRs significantly associated with tick count are displayed in Figure 4.4 and described in Table 4.8. Copy number variations can exert effects on genes through alternating regulation, modifying their structure and dosage as well as exposing recessive alleles (Zhang *et al.*, 2009). Three genes including, *gamma-aminobutyric acid A receptor, subunit beta 2* (*GABRB2*), *lymphocyte-specific protein 1* (*LSP1*) and *troponin T3, fast skeletal type* (*TNNT3*), were detected within CNVRs across susceptible and resistant animals. *Gamma-aminobutyric acid A receptor, subunit beta 2* was identified within a CNVR on chromosome 7, with *LSP1* and *TNNT3* overlapping with a CNVR on chromosome 29. These CNVRs demonstrated a significant association with tick count. A duplication on chromosome 7 and a deletion on chromosome 29 corresponded with elevated tick counts, while the converse was true for tick resistance.

Hou *et al*. (2012) and Wang *et al*. (2015) report the *GABRB2* gene to overlap CNVRs identified in Angus and Nguni cattle populations respectively. *Gamma-aminobutyric acid A receptor, subunit beta 2* is a subunit receptor involved in the central nervous system located on chromosome 7, between base pairs 75,305,297 and 75,370,366. This CNVR was detected in seven different animals, six classified as resistant and one classified as susceptible. Copy number state of this CNVR distinguished susceptible from resistant animals with deletions corresponding to enhanced tick resistance. The *GABRB2* gene has been associated with anxiety in chickens (Johnsson *et al.*, 2016) and muscle contraction in sheep (Hui *et al.*, 2016). In the study by Hui *et al*. (2016) this gene was reported to be expressed at greater levels after infection and is hypothesized to be involved in immune homeostasis. After and during parasite infection, there is increased inflammation, which signals the response of histamines. *Gamma-aminobutyric acid A receptor, subunit beta 2* is responsible for the cellular response of histamine (Table 4.6) and muscle contraction and could explain the hypothesized role of *GABRB2* in immune response. In this study, this gene was identified in close proximity to CNVRs in multiple resistant animals and could provide the basis for a mechanism underlying immune homeostasis and in turn enhanced resistance. This is the first study to associate the *GABRB2* gene with tick resistance.

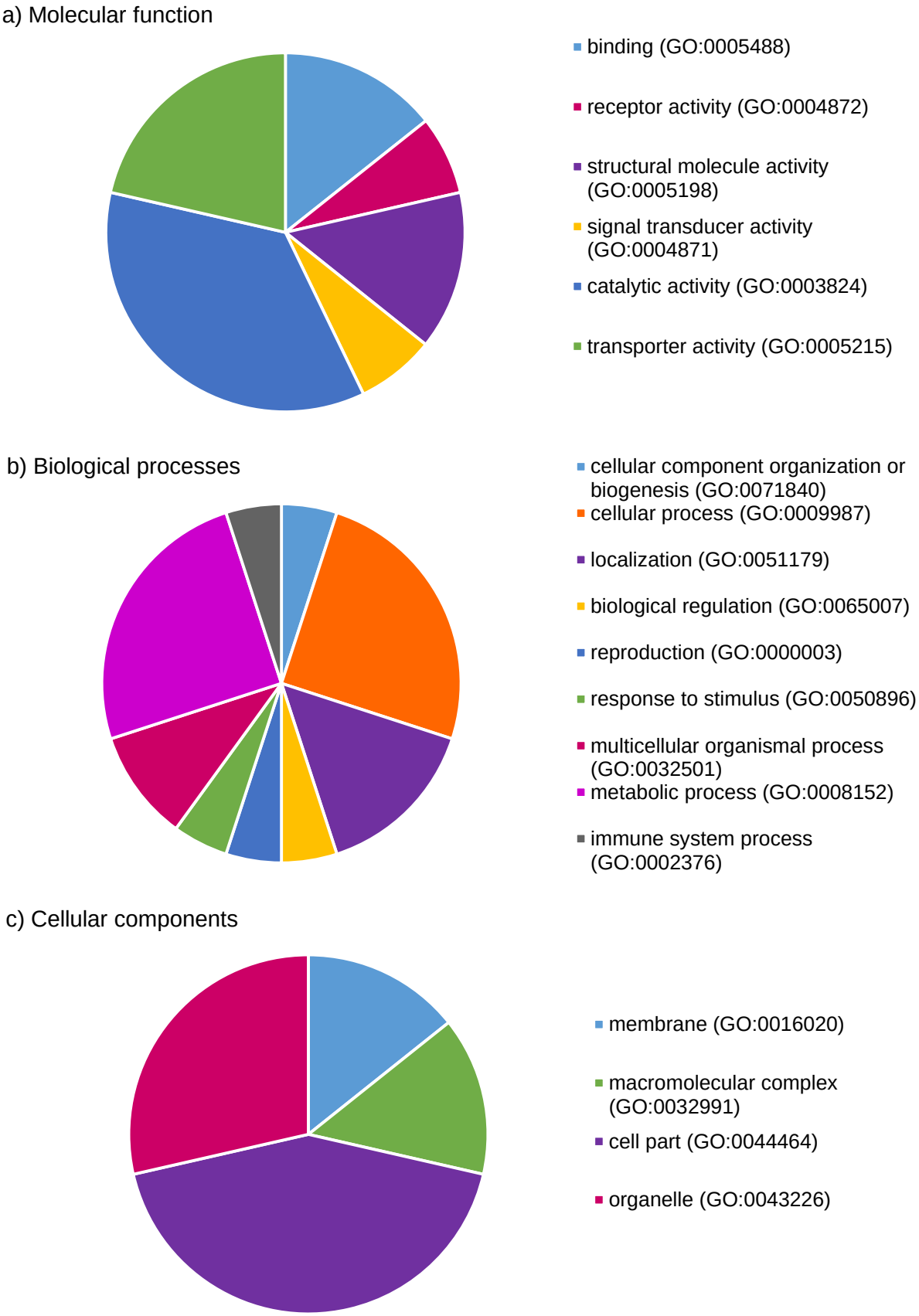

Figure 4.4 PANTHER pie charts of a) molecular functions, b) biological processes and c) cellular components of the genes detected within significant CNVRs.

Table 4.8 Molecular functions (MF), biological processes (BP) and cellular components (CC) of genes (GEN) detected within CNVRs demonstrating a significant association with tick count in South African Nguni cattle.

GEN	MF	BP	CC
AIFM3	Oxidoreductase activity, flavin adenine dinucleotide binding, 2 iron & 2 sulphur cluster binding	Oxidation-reduction process, execution phase of apoptosis	Mitochondrial inner membrane, endoplasmic reticulum, cytosol
ATP4B		Potassium ion transport, sodium ion transport	Sodium:potassium-exchanging ATPase complex
C14ORF79			
CHCHD6		Cellular response to DNA damage stimulus, cristae formation	Mitochondrion, mitochondrial inner membrane, cytosol, MICOS complex
GABRB2	GABA-A receptor activity, extracellular ligand-gated ion channel activity, chloride channel activity,	Signal transduction, sensory perception of sound, ion transmembrane transport, negative regulation of neuron apoptotic process, inner ear receptor cell development, innervation, cochlea development, cellular response to histamine, chloride transmembrane transport	Plasma membrane, integral component of membrane, cell junction, synapse, extracellular exosome, integral component of plasma membrane, cell junction, cytoplasmic vesicle membrane, chloride channel complex, postsynaptic membrane, GABA-A receptor complex
GRK1	G-protein coupled receptor kinase activity, protein binding, ATP binding, rhodopsin kinase activity	Signal transduction, visual perception, regulation of rhodopsin mediated signalling pathway, protein autophosphorylation	Photoreceptor disc membrane
KYNU	Pyridoxal phosphate binding, kynureninase activity, protein homodimerization activity	Tryptophan catabolic process to kynurenine, quinolinate biosynthetic process, response to interferon-gamma, 'de novo' NAD biosynthetic process from tryptophan, response to vitamin B6, anthranilate metabolic process, L-kynurenine catabolic process	Cytoplasm, mitochondrion, cytosol
LSP1	Signal transducer activity	Chemotaxis, defence response, signal transduction	Plasma membrane, extracellular exosome
LZTR1			Golgi apparatus

GEN	MF	BP	CC
MZT2			Centrosome, spindle, gamma-tubulin ring complex
PRODH	Proline dehydrogenase activity, FAD binding	Proline catabolic process to glutamate, oxidation-reduction process	Mitochondrion, mitochondrial matrix
SLC7A4	Amino acid transmembrane transporter activity	Amino acid transmembrane transport	Integral component of membrane
TFDP1	RNA polymerase II transcription factor activity & sequence-specific DNA binding, DNA binding, transcription coactivator activity, transcription factor binding, protein domain specific binding	Mitotic cell cycle, DNA-templated transcription, regulation of transcription from RNA polymerase II promoter, epidermis development, anoikis, positive regulation of transcription from RNA polymerase II promoter, negative regulation of fat cell proliferation, regulation of DNA biosynthetic process	Nucleus, nucleoplasm, transcription factor complex, cytoplasm, cytosol
THAP7			
TMCO3	Solute:proton antiporter activity, inorganic cation transmembrane transporter activity	Hydrogen ion transmembrane transport	Membrane, integral component of membrane
TNNT3	Actin binding, tropomyosin binding, troponin C binding, calcium-dependent ATPase activity, troponin I binding, calcium-dependent protein binding	Skeletal muscle contraction, regulation of striated muscle contraction, regulation of ATPase activity	Troponin complex
TUBA3E	GTPase activity, structural constituent of cytoskeleton, GTP binding	Cytoskeleton organization, microtubule-based process	Cytoplasm, microtubule

Two genes overlapping CNVRs within this study and that of Bae *et al*. (2010) and Hou *et al*. (2011 & 2012) are *LSP1* and *TNNT3* (Table 4.6). Both genes overlapped with a CNVR on chromosome 29 between base pairs 50,202,589 and 50,240,781. This CNVR was detected in six different animals, four being susceptible and two being resistant. These genes overlapped with deletions in susceptible animals and duplications in resistant animals. *Troponin T* is involved in the regulation of skeletal and striated muscle contraction (Verardo *et al.*, 2013; De Souza Rodrigues *et al.*, 2017). This gene is also attributable to differences in beef tenderness in Nellore and Angus breeds (De Souza Rodrigues *et al.*, 2017). *Lymphocyte-specific protein 1* is found within the endothelium, lymphocytes, macrophages and neutrophils and is a bundling cytoskeletal protein (Hou *et al.*, 2012). In our study, *LSP1* was detected as a deletion in four susceptible animals and as a duplication in two resistant animals, which is similar to the findings of Hou *et al*. (2012). In both studies, *LSP1* was detected in susceptible and resistant animals, with a higher frequency in susceptible animals, however Hou *et al*. (2012) found *LSP1* to be associated with duplication events in susceptible and resistant animals. This gene is responsible for the regulation of leukocytes to inflamed areas and, therefore, susceptible animals which express higher levels of LSP1 could have less leukocytes reaching inflamed sites, thereby increasing disease susceptibility (Hou *et al.*, 2012).

A deletion CNVR, identified in three susceptible animals, on chromosome 17 between base pairs 74,292,319 and 74,393,620 harboured seven different genes, of which two were detected in previous CNVR studies of Bae *et al*. (2010) and Hou *et al*. (2011 & 2012) (Table 4.6). The two genes detected in these previous studies include *THAP domain containing 7 (THAP7)* and *Tubulin alpha 3-e (TUBA3E)* which are both protein coding genes. The *TUBA3E* gene is reported to be downregulated in chicken cells due to heat stress (Sun *et al.*, 2015) and its protein is known to be involved in human brain malformation syndromes (Alazami *et al.*, 2015). In cattle, *THAP7* is downregulated in differential adipocytes (Yu *et al.*, 2009) and in humans could contribute to carcinogenesis (Kim *et al.*, 2010). Other genes detected within this CNVR include *apoptosis inducing factor, mitochondria associated 3 (AIFM3), proline dehydrogenase (oxidase) 1 (PRODH), mitotic spindle organizing protein 2 (MZT2), solute carrier family 7 member 4 (SLC7A4)* and *leucine zipper like transcription regulator 1 (LZTR1).* Literature relating to these genes in cattle is minimal however some research relating to these genes in other species can be found. The *AIFM3* gene is involved in the oxidation-reduction process and has been found to be differentially expressed in different pig breeds (Sodhi *et al.*, 2014). In humans, genomic variants affecting the *PRODH* gene have been associated with an increased risk of schizophrenia (Jacquet, 2002; Willis *et al.*, 2008) and mutations of the *LZTR1* gene have been linked to schwannomatosis (Kehrer-Sawatzki *et al.*, 2017). *AIFM3* and *LZTR1* were both detected within CNVRs of this study and those of Hou *et al*. (2011) and Hou *et al*. (2012).

Table 4.9 Genes detected within significant CNVRs (Number of genes – NGEN and genes – GEN) in susceptible and resistant animals in this study with those identified in other studies as reported by Bae *et al*., 2010, Hou *et al*., 2011 (a), Bickhart *et al*., 2012, Hou *et al*., 2012 (b) and Wang *et al*., 2015 (Author) which revealed 8 genes unique to the Nguni

Author	NGEN	GEN
Bae Hou(a) Hou (b) Pickering	4	*TUBA3E, THAP7, TNNT3, LSP1*
Hou (a) Hou (b) Pickering	2	*AIFM3, LZTR1*
Hou (b) Pickering	2	*CHCHD6, KYNU*
Hou (b) Wang Pickering	1	*GABRB2*
Pickering	8	*PRODH, MZT2, SLC7A4, TFDP1, TMCO3, ATP4B, GRK1, C14orf79*

The *coiled-coil-helix-coiled-coil-helix domain containing 6* (*CHCHD6)* and *kynureninase* (*KYNU)* genes were the only genes represented across CNVRs of Hou *et al*. (2012) and this study (Table 4.9). The *CHCHD6* gene is located between base pairs 60,736,089 and 60,960,603 on chromosome 22. This gene has previously been mapped to this chromosome in *Bos taurus* cattle (De Lorenzi *et al.*, 2010). In the recent study of Duran Aguilar *et al*. (2017), this gene was detected within a CNVR associated with the estimated breeding value (EBV) of somatic cell score (SCS) in Holstein cattle. The region on chromosome 2 between base pairs 53,894,600 and 53,951,533 harbours the *KYNU* gene. This gene has also been detected as an insertion on chromosome 2 within the Holstein bull genome (Kõks *et al.*, 2014) which is contradictory to this study as it was detected as a deletion. *Kynureninase* forms part of the kynurenine pathway and is the major pathway for ingested tryptophan (Keszthelyi *et al.*, 2009). After tryptophan enters this pathway, one of the major end products is quinolinic acid and this is involved in immune-regulatory processes (Moffett & Namboodiri, 2003; Keszthelyi *et al.*, 2009).

Three susceptible animals shared a deletion CNVR located on chromosome 12 between base pairs 90,704,572 and 90,778,028. *ATPase H+/K+ transporting beta subunit (ATP4B)*, *G protein-coupled receptor kinase 1 (GRK1)*, *transmembrane and coiled-coil domains 3 (TMCO3*) and *transcription factor dp-1 (TFDP1)* were four genes detected within this CNVR. The transcription factor (*TFDP1*) has been detected in a few bovine studies. One study found this gene to be upregulated in response to heat stress in Holstein calves (Srikanth *et al.*, 2017) with identifying differential expression of this gene in the liver tissue of cows due to subclinical endometriosis (Akbar *et al.*, 2014). In humans, *TFDP1* is a candidate gene for breast cancer (Melchor *et al.*, 2009). Another gene, within this CNVR,

which is a marker for cancer is *ATP4B.* DNA methylation can cause decreased expression levels of *ATP4B*, and has been reported as a potential biomarker for gastric cancer (Mahalinga Raja *et al.*, 2012). The transmembrane protein coding gene (*TMCO3*) encodes a Na^+/H^+ antiporter (Moscovich *et al.*, 2013). Mutations of *TMCO3* have been associated with anterior polar cataract and cornea guttata in humans (Chen *et al.*, 2016). An additional gene within this region also associated with an eye disorder is *GRK1.* This gene is exclusively expressed in the retina (Abraham, 2016) and its defects are known to cause Oguchi disease 2 (Melchor *et al.*, 2009). This is a recessive disorder causing morphological and functional abnormalities of the retina (Kalpana *et al.*, 2006).

4.5 Conclusion

The findings in this study reveal 344 CNVRs of which 17 are associated with Nguni cattle tick resistance. Copy number variable regions detected comprised of copy number changes in the form of deletions and duplications across the genome. These copy number changes can have an impact on gene function, regulation and expression thereby exerting an effect on the phenotype. Of the 17 significant CNVRs detected, seven overlap or lay within a 10Mb region of 17 genes. The genes identified overlapping or in close proximity to these regions are present in a number of cellular components and are responsible for many molecular and biological processes within the genome. These processes could play a role in the adaptive attributes present in Nguni populations. Three genes were detected in susceptible and resistant animals, 14 were detected in only susceptible animals and there were no genes directly associated with only resistant animals. Although none of the genes detected within or in close proximity of CNVRs of susceptible animals have a direct role in disease susceptibility, most of them are responsible for some form of disease. Two of the genes detected within or overlapping with CNVRs in susceptible and resistant animals could have a potential role in the resistance of the South African Nguni. The *LSP1* gene detected within a CNVR on chromosome 29 could be responsible for decreased resistance due to less leukocytes being able to reach inflamed areas, while GABRB2 located within a CNVR on chromosome 7 could be responsible for enhanced immunity due to cellular response of histamines and muscle contraction.

4.6 References

Abraham, G. 2016. The importance of muscarinic receptors in domestic animal diseases and therapy: Current and future perspectives. Vet. J. 208, 13–21.

Akbar, H., Cardoso, F. C., Meier, S., Burke, C., McDougall, S., Mitchell, M., Walker, C., Rodriguez-Zas, S. L., Everts, R. E., Lewin, H. A., Roche, J. R., & Loor, J. J. 2014. Postpartal subclinical endometritis alters transcriptome profiles in liver and adipose tissue of dairy cows. Bioinform. Biol. Insights. 8, 45–63.

Alazami, A. M., Patel, N., Shamseldin, H. E., Anazi, S., Al-Dosari, M. S., Alzahrani, F., Hijazi, H., Alshammari, M., Aldahmesh, M. A., Salih, M. A., Faqeih, E., Alhashem, A., Bashiri, F. A., Al-Owain, M., Kentab, A. Y., Sogaty, S., AlTala, S., Temsah, M. H., Tulbah, M., Aljelaify, R. F., Alshahwan, S. A., Seidahmed, M. Z., Alhadid, A. A., Aldhalaan, H., AlQallaf, F., Kurdi, W., Alfadhel, M., Babay, Z., Alsogheer, M., Kaya, N., Al-Hassnan, Z. N., Abdel-Salam, G. M. H., Al-Sannaa, N., AlMutairi, F., ElKhashab, H. Y., Bohlega, S., Jia, X., Nguyen, H. C., Hammami, R., Adly, N., Mohamed, J. Y., Abdulwahab, F., Ibrahim, N., Naim, E. A., Al-Younes, B., Meyer, B. F., Hashem, M., Shaheen, R., Xiong, Y., Abouelhoda, M., Aldeeri, A. A., Monies, D. M., & Alkuraya, F. S. 2015. Accelerating novel candidate gene discovery in neurogenetic disorders via whole-exome sequencing of prescreened multiplex consanguineous families. Cell Rep. 10, 148–161.

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

Bailey, J. A. 2002. Recent Segmental Duplications in the Human Genome. Science. 297, 1003–1007.

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

Boussaha, M., Esquerré, D., Barbieri, J., Djari, A., Pinton, A., Letaief, R., Salin, G., Escudié, F., Roulet, A., Fritz, S., Samson, F., Grohs, C., Bernard, M., Klopp, C., Boichard, D., & Rocha, D. 2015. Genome-wide study of structural variants in bovine Holstein, Montbéliarde and Normande dairy breeds. PLoS One. 10, 1–21.

Chen, P., Hao, X., Li, W., Zhao, X., & Huang, Y. 2016. Mutations in the TMCO3 Gene are Associated with Cornea Guttata and Anterior Polar Cataract. Sci. Rep. 6, 31021.

Cicconardi, F., Chillemi, G., Tramontano, A., Marchitelli, C., Valentini, A., Ajmone-Marsan, P., & Nardone, A. 2013. Massive screening of copy number population-scale variation in *Bos taurus* genome. BMC Genomics. 14, 124.

Da Silva, V. H., De Almeida Regitano, L. C., Geistlinger, L., Pértille, F., Giachetto, P. F., Brassaloti, R. A., Morosini, N. S., Zimmer, R., & Coutinho, L. L. 2016. Genome-wide detection of CNVs and their association with meat tenderness in Nelore cattle. PLoS One. 11, 1–22.

Dellinger, A. E., Saw, S. M., Goh, L. K., Seielstad, M., Young, T. L., & Li, Y. J. 2010. Comparative analyses of seven algorithms for copy number variant identification from single nucleotide polymorphism arrays. Nucleic Acids Res. 38, 1–14.

De Lorenzi, L., Molteni, L., & Parma, P. 2010. FISH mapping in cattle (*Bos taurus L*.) is not yet out of fashion. J. Appl. Genet. 51, 497–9.

De Souza Rodrigues, R. T., Chizzotti, M. L., Vital, C. E., Baracat-Pereira, M. C., Barros, E., Busato, K. C., Gomes, R. A., Ladeira, M. M., & Da Silva Martins, T. 2017. Differences in beef quality between Angus (*Bos taurus taurus*) and Nellore (*Bos taurus indicus*) cattle through a proteomic and phosphoproteomic approach. PLoS One. 12, 1–21.

Durán Aguilar, M., Román Ponce, S. I., Ruiz López, F. J., González Padilla, E., Vásquez Peláez, C. G., Bagnato, A., & Strillacci, M. G. 2017. Genome-wide association study for milk somatic cell score in holstein cattle using copy number variation as markers. J. Anim. Breed. Genet. 134, 49–59.

Eckel-Passow, J. E., Atkinson, E. J., Maharjan, S., Kardia, S. L., & de Andrade, M. 2011. Software comparison for evaluating genomic copy number variation for Affymetrix 6.0 SNP array platform. BMC Bioinformatics. 12, 220.

Estivill, X., & Armengol, L. 2007. Copy number variants and common disorders: Filling the gaps and exploring complexity in genome-wide association studies. PLoS Genet. 3, 1787–1799.

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

Feuk, L., Carson, A. R., & Scherer, S. W. 2006. Structural variation in the human genome. Nat. Rev. Genet. 7, 85–97.

Frisch, J. E. 1999. Towards a permanent solution for controlling cattle ticks. Int. J. Parasitol. 29, 57–71.

Gamazon, E. R., & Stranger, B. E. 2015. The impact of human copy number variation on gene expression. Brief. Funct. Genomics. 14, 352–357.

Glick, G., Shirak, A., Seroussi, E., Zeron, Y., Ezra, E., Weller, J. I., & Ron, M. 2011. Fine Mapping of a QTL for Fertility on BTA7 and Its Association with a CNV in the Israeli Holsteins. G3. Genes. Genomes. Genetics. 1, 65–74.

Henrichsen, C. N., Chaignat, E., & Reymond, A. 2009. Copy number variants, diseases and gene expression. Hum. Mol. Genet. 18, 1–8.

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

Hou, Y., Liu, G. E., Bickhart, D. M., Matukumalli, L. K., Li, C., Song, J., Gasbarre, L. C., Van Tassell, C. P., & Sonstegard, T. S. 2011b. Genomic regions showing copy number variations associate with resistance or susceptibility to gastrointestinal nematodes in Angus cattle. Funct. Integr. Genomics. 12, 81–92.

Hui, W., Jiang, S., Liu, X., Ban, Q., Chen, S., & Jia, B. 2016. Gene expression profile in the liver of sheep infected with cystic echinococcosis. PLoS One. 11, 1–14.

Ibeagha-Awemu, E. M., Kgwatalala, P., Ibeagha, A. E., & Zhao, X. 2008. A critical analysis of disease-associated DNA polymorphisms in the genes of cattle, goat, sheep, and pig. Mamm. Genome. 19, 226–245.

Jacquet, H. 2002. PRODH mutations and hyperprolinemia in a subset of schizophrenic patients. Hum. Mol. Genet. 11, 2243–2249.

Jiang, L., Jiang, J., Wang, J., Ding, X., Liu, J., & Zhang, Q. 2012. Genome-Wide Identification of Copy Number Variations in Chinese Holstein. PLoS One. 7.

Johnsson, M., Williams, M. J., Jensen, P., & Wright, D. 2016. Genetical genomics of behavior: A novel chicken genomic model for anxiety behavior. Genetics. 202, 327–340.

Kalpana, S., Muthayya, M., & Doctor, P. P. 2006. Oguchi disease. J. Postgrad. Med. 52, 143–144.

Keel, B. N., Keele, J. W., & Snelling, W. M. 2016. Genome-wide copy number variation in the bovine genome detected using low coverage sequence of popular beef breeds, Anim. Genet. 141–150.

Kehrer-Sawatzki, H., Farschtschi, S., Mautner, V. F., & Cooper, D. N. 2017. The molecular pathogenesis of schwannomatosis, a paradigm for the co-involvement of multiple tumour suppressor genes in tumorigenesis. Hum. Genet. 136, 129–148.

Keszthelyi, D., Troost, F. J., & Masclee, A. A. M. 2009. Understanding the role of tryptophan and serotonin metabolism in gastrointestinal function. Neurogastroenterol. Motil. 21, 1239–1249.

Kijas, J. W., Barendse, W., Barris, W., Harrison, B., McCulloch, R., McWilliam, S., & Whan, V. 2011. Analysis of copy number variants in the cattle genome. Gene. 482, 73–77.

Kim, Y. H., Kwei, K. A., Girard, L., Salari, K., Kao, J., Pacyna-Gengelbach, M., Wang, P., Hernandez-Boussard, T., Gazdar, A. F., Petersen, I., Minna, J. D., & Pollack, J. R. 2010. Genomic and functional analysis identifies CRKL as an oncogene amplified in lung cancer. Oncogene. 29, 1421–1430.

Kim, E. S., Sonstegard, T. S., Silva, M. V. G. B., Gasbarre, L. C., & Van Tassell, C. P. 2014. Identification of quantitative trait loci affecting gastrointestinal parasite resistance in an experimental Angus population. Anim. Genet. 45, 117–121.

Kõks, S., Reimann, E., Lilleoja, R., Lättekivi, F., Salumets, A., Reemann, P., & Jaakma, Ü. 2014. Sequencing and annotated analysis of full genome of Holstein breed bull. Mamm. Genome. 25, 363–373.

Lin, C. F., Naj, A. C., & Wang, L. S. 2013. Analyzing copy number variation using SNP array data: Protocols for calling CNV and association tests. Curr. Protoc. Hum. Genet.

Liu, G. E., Ventura, M., Cellamare, A., Chen, L., Cheng, Z., Zhu, B., Li, C., Song, J., & Eichler, E. E. 2009. Analysis of recent segmental duplications in the bovine genome. BMC Genomics. 10, 571.

Liu, G. E., Hou, Y., Zhu, B., Liu, G. E., Hou, Y., Zhu, B., Cardone, M. F., Jiang, L., Cellamare, A., Mitra, A., Alexander, L. J., Coutinho, L. L., Elena, M., Aquila, D., Gasbarre, L. C., Lacalandra, G., Li, R. W., Matukumalli, L. K., Nonneman, D., Regitano, L. C. D. A., Smith, T. P. L., Song, J., Sonstegard, T. S., Tassell, C. P. Van, Ventura, M., Eichler, E. E., Mcdaneld, T. G., & Keele, J. W. 2010. Analysis of copy number variations among diverse cattle breeds. Genome Res. 20, 693–703.

Liu, G. E., & Bickhart, D. M. 2012. Copy number variation in the cattle genome. Funct. Integr. Genomics. 12, 609–624.

Machado, M. A., Azevedo, A. L. S., Teodoro, R. L., Pires, M. A., Peixoto, M. G. C. D., de Freitas, C., Prata, M. C. A., Furlong, J., da Silva, M. V. G. B., Guimarães, S. E. F., Regitano, L. C. A., Coutinho, L. L., Gasparin, G., & Verneque, R. S. 2010. Genome wide scan for quantitative trait loci affecting tick resistance in cattle (*Bos taurus* x *Bos indicus*). BMC Genomics. 11, 280.

Mahalinga Raja, U., Gopal, G., & Rajkumar, T. 2012. Intragenic DNA methylation concomitant with repression of ATP4B and ATP4A gene expression in gastric cancer is a potential serum biomarker. Asian Pacific J. Cancer Prev. 13, 5563–5568.

Makina, S. O., Muchadeyi, F. C., van Marle-Kőster, E., MacNeil, M. D., & Maiwashe, A. 2014. Genetic diversity and population structure among six cattle breeds in South Africa using a whole genome SNP panel. Front. Genet. 5, 1–7.

Mapholi, N. O., Maiwashe, A., Matika, O., Riggio, V., Bishop, S. C., MacNeil, M. D., Banga, C., Taylor, J. F., & Dzama, K. 2016. Genome-wide association study of tick resistance in South African Nguni cattle. Ticks Tick. Borne. Dis. 7, 487–497.

Marques-Bonet, T., Girirajan, S., & Eichler, E. E. 2009. The origins and impact of primate segmental duplications. Trends Genet. 25, 443–454.

Marufu, M. C., Qokweni, L., Chimonyo, M., & Dzama, K. 2011. Relationships between tick counts and coat characteristics in Nguni and Bonsmara cattle reared on semiarid rangelands in South Africa. Ticks Tick. Borne. Dis. 2, 172–177.

Melchor, L., Saucedo-Cuevas, L. P., Muñoz-Repeto, I., Rodríguez-Pinilla, S. M., Honrado, E., Campoverde, A., Palacios, J., Nathanson, K. L., García, M. J., & Benítez, J. 2009. Comprehensive characterization of the DNA amplification at 13q34 in human breast cancer reveals TFDP1 and CUL4A as likely candidate target genes. Breast Cancer Res. 11, R86.

Moffett, J. R., & Namboodiri, M. A. 2003. Tryptophan and the immune response. Immunol. Cell Biol. 81, 247–265.

Moscovich, M., Ledoux, M. S., Xiao, J., Rampon, G. L., Vemula, S. R., Rodriguez, R. L., Foote, K. D., & Okun, M. S. 2013. Dystonia, facial dysmorphism, intellectual disability and breast cancer associated with a chromosome 13q34 duplication and overexpression of TFDP1: case report. BMC Med. Genet. 14, 70.

Porto Neto, L. R., Jonsson, N. N., D'Occhio, M. J., & Barendse, W. 2011. Molecular genetic approaches for identifying the basis of variation in resistance to tick infestation in cattle. Vet. Parasitol. 180, 165–172.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81, 559–575.

Sambrook, J., Fritsch, E. F., & Maniatis, T. 1989. Molecular cloning: A laboratory manual, 3. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Schnabel, R. D., Sonstegard, T. S., Taylor, J. F., & Ashwell, M. S. 2005. Whole-genome scan to detect QTL for milk production, conformation, fertility and functional traits in two US Holstein families. Anim. Genet. 36, 408–416.

Schoeman, S. J. 1989. Recent research into the production potential of indigenous cattle with special reference to the Sanga. S. Afr. J. Anim. Sci. 19, 55–61.

Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., Maner, S., Massa, H., Walker, M., Chi, M., Navin, N., Lucito, R., Healy, J., Hicks, J., Ye, K., Reiner, A., Gilliam, T. C., Trask,

B., Patterson, N., Zetterberg, A., & Wigler, M. 2004. Large-scale copy number polymorphism in the human genome. Science. 305, 525–528.

Sharp, A. J., Locke, D. P., McGrath, S. D., Cheng, Z., Bailey, J. A., Vallente, R. U., Pertz, L. M., Clark, R. A., Schwartz, S., Segraves, R., Oseroff, V. V., Albertson, D. G., Pinkel, D., & Eichler, E. E. 2005. Segmental Duplications and Copy-Number Variation in the Human Genome. Am. J. Hum. Genet. 77, 78–88.

Sodhi, S. S., Park, W. C., Ghosh, M., Kim, J. N., Sharma, N., Shin, K. Y., Cho, I. C., Ryu, Y. C., Oh, S. J., Kim, S. H., Song, K. D., Hong, S. P., Cho, S. A., Kim, H. B., & Jeong, D. K. 2014. Comparative transcriptomic analysis to identify differentially expressed genes in fat tissue of adult Berkshire and Jeju Native Pig using RNA-seq. Mol. Biol. Rep. 41, 6305–6315.

Sollero, B. P., Junqueira, V. S., Gomes, C. C. G., Caetano, A. R., & Cardoso, F. F. 2017. Tag SNP selection for prediction of tick resistance in Brazilian Braford and Hereford cattle breeds using Bayesian methods. Genet. Sel. Evol. 49, 49.

Srikanth, K., Kwon, A., Lee, E., & Chung, H. 2017. Characterization of genes and pathways that respond to heat stress in Holstein calves through transcriptome analysis. Cell Stress Chaperones. 22, 29–42.

Sun, L., Lamont, S. J., Cooksey, A. M., McCarthy, F., Tudor, C. O., Vijay-Shanker, K., DeRita, R. M., Rothschild, M., Ashwell, C., Persia, M. E., & Schmidt, C. J. 2015. Transcriptome response to heat stress in a chicken hepatocellular carcinoma cell line. Cell Stress Chaperones. 20, 939–95..

Turner, D. J., Miretti, M., Rajan, D., Fiegler, H., Carter, N. P., Martyn, L., Beck, S., & Hurles, M. E. 2008. The rates of de novo meiotic deletions and duplications causing several genomic disorders in the male germline. Nat Genet. 40, 90–95.

Tuzun, E., Sharp, A. J., Bailey, J. A., Kaul, R., Morrison, V. A., Pertz, L. M., Haugen, E., Hayden, H., Albertson, D., Pinkel, D., Olson, M. V, & Eichler, E. E. 2005. Fine-scale structural variation of the human genome. Nat. Genet. 37, 727–732.

Verardo, L. L., Nascimento, C. S., Silva, F. F., Gasparino, E., Martins, M. F., Toriyama, E., Faria, V. R., Botelho, M. E., Costa, K. A., Lopes, P. S., & Guimarães, S. E. F. 2013. Identification and validation of differentially expressed genes from pig skeletal muscle. J. Anim. Breed. Genet. 130, 372–381.

Wang, M. D., Dzama, K., Hefer, C. A., & Muchadeyi, F. C. 2015. Genomic population structure and prevalence of copy number variations in South African Nguni cattle. BMC Genomics. 16, 894.

Wang, M. D., Dzama, K., Rees, D. J. G., & Muchadeyi, F. C. 2016. Tropically adapted cattle of Africa: Perspectives on potential role of copy number variations. Anim. Genet. 47, 154–164.

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

Willis, A., Bender, H. U., Steel, G., & Valle, D. 2008. PRODH variants and risk for schizophrenia. Amino Acids. 35, 673–679.

Wu, Y., Fan, H., Jing, S., Xia, J., Chen, Y., Zhang, L., Gao, X., Li, J., Gao, H., & Ren, H. 2015. A genome-wide scan for copy number variations using high-density single nucleotide polymorphism array in Simmental cattle. Anim. Genet. 46, 289–298.

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

Xu, L., Hou, Y., Bickhart, D., Song, J., & Liu, G. 2013. Comparative Analysis of CNV Calling Algorithms: Literature Survey and a Case Study Using Bovine High-Density SNP Data. Microarrays. 2, 171–185.

Xu, L., Hou, Y., Bickhart, D. M., Song, J., Van Tassell, C. P., Sonstegard, T. S., & Liu, G. E. 2014b. A genome-wide survey reveals a deletion polymorphism associated with resistance to gastrointestinal nematodes in Angus cattle. Funct. Integr. Genomics. 14, 333–9.

Yu, S.-L., Lee, S.-M., Kang, M.-J., Jeong, H.-J., Sang, B.-C., Jeon, J.-T., & Lee, J.-H. 2009. Identification of Differentially Expressed Genes between Preadipocytes and Adipocytes Using Affymetrix Bovine Genome Array. J. Anim. Sci. Technol. 51, 443–452.

Zhang, F., Gu, W., Hurles, M. E., & Lupski, J. R. 2009. Copy number variation in human health, disease, and evolution. Annu Rev Genomics Hum Genet 10, 451–481.

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

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

Chapter 5
General Discussion and Conclusions

5.1 Summary of Findings

The ability of an animal to resist disease and survive and produce under severe environmental conditions is an important phenotypic characteristic. Copy number variations (CNVs) comprise of deletions, duplications and insertions and play a role in the phenotypic variation exhibited within and between cattle breeds. Locally adapted cattle are better suited to their primary environment when compared with foreign breeds. One such breed, known for its resilience to disease and harsh environmental conditions, is the South African Nguni (Schoeman, 1989; Makina *et al.*, 2014). Copy number variable regions (CNVRs) within the Nguni are involved in various molecular and biological processes (Wang *et al.*, 2015). These processes play a pivotal role in the adaptive traits evident in Nguni cattle. Understanding the association of CNVRs with tick resistance in Nguni cattle, may provide an insight into the genetic mechanisms underlying the extreme host resistance reported in this breed.

This study investigated the association of CNVRs with tick resistance in susceptible and resistant Nguni cattle. The effect of non-genetic factors on tick resistance was assessed in order to determine fixed effects that may need to be considered in the CNV association analyses. Contingency tables and the maximum likelihood chi-square test identified CNVRs demonstrating a significant association with tick resistance. Significant CNVRs were then investigated for gene content and subsequent gene ontology analyses were performed.

A general linear model assessed the non-genetic effects of location, season, year of tick count, sex and age of 347 South African Nguni cattle on tick count. Location, season, year of tick count and age of the animal had an influence on tick resistance. Warmer areas presented greater tick infestation, with the hot wet and hot dry seasons significantly influencing tick count. These findings correspond to those of Muchenje *et al.* (2008), Marufu *et al.* (2011) and Katiyatiya *et al.* (2014) whom report greater levels of tick burden during the hot wet season. Animals younger than five years presented less ticks and therefore an enhanced resistance to tick infestation. Marufu *et al.* (2011) and Asmaa *et al.* (2014) also report the low prevalence of ticks in younger animals. Tick counts ranged from 0 to 198, with an average tick count of 18.24, 30.12 and 26.43 at Mukhuthali, Loskop and Roodeplaat respectively. The highest spread of counts was reported at Loskop, which presented a standard deviation of 30.12. The standard deviations of Mukhuthali and Roodeplaat were 18.24 and 26.43 respectively.

As no specific tick count level has been determined to classify animals as resistant or susceptible, two models were determined to group animals according to tick count levels. Location presented a significant effect on tick count, and thus distribution statistics were determined per location. Using the standard deviation and quartile and inter-quartile ranges, animals were classified as susceptible (0) or resistant (1) across two levels of resistance (L1 and L2). At resistance level 1 (L1) animals having an average tick count of less than or equal to the 1st quartile were classified as resistant (1) with those being classified as resistant if their mean tick count was more than or equal to the 3rd quartile. Resistance level 2 (L2) was determined by adding half the standard deviation to the 1st quartile and subtracting it from the 3rd quartile for susceptible and resistant animals respectively. Animals which were neither susceptible nor resistant were removed from further analyses.

The BovineSNP50 beadchip, PLINK and *PennCNV* software were used to identified 1 501 CNVs, which ranged from 1 kb to 3 Mb in length, in 347 Nguni cattle. Copy number variable regions (CNVRs) were generated by aggregating overlapping CNVs. A total of 344 unique CNVRs were reported. The prevalence of CNVRs varied across chromosomes, with chromosome 6 displaying the most and chromosome 25 the least. These regions ranged from 14 to 914kb. Of these CNVRs, 17 were significantly associated with tick resistance in Nguni cattle. These regions were located on chromosomes 1, 2, 6, 7, 8, 9, 12, 15, 17, 20, 21, 22, 24 and 29. The presence of 17 genes overlapped or lay in close proximity with these regions. The biological processes, molecular functions and cellular components of these genes play a role in a variety of adaptive mechanisms. A CNVR detected on chromosome 29, which overlapped with *LSP1,* and a CNVR detected on chromosome 7 which overlapped with *GABRB2* could provide insight into the effect of CNVRs on genes and their influence on susceptibility.

5.2 General Discussion

Copy number variable regions are present within the genome of South African Nguni cattle. These regions cause phenotypic variation within and between breeds through the alteration of gene dosage, structure and regulation (Zhang *et al.*, 2009). Copy number variations can also exert effects on the genome through the exposure of recessive alleles. Three hundred and forty-four CNVRs were detected (Chapter 4). These regions ranged from 14 to 914 kb and covered 1.66% of the bovine genome. The BovineSNP50 beadchip was utilized for CNV detection and, although a high density bovine SNP chip exists, the comparison of the number of CNVRs detected between these chips does not differ significantly. Using the high density chip Wu *et al.* (2015) report 263 CNVRs, covering 1.41% of the genome in 792 Simmental cattle, while Sasaki *et al.* (2016) report 861 CNVRs, covering 1.74% of the genome in 1 481 Japanese black cattle.

The cost of CNV detection methods varies, and thus the efficient identification of CNVs in a large number of samples at a low cost is preferred. The use of the high density chip and NGS technologies is far greater than that of the 50K and, although these methods of detection can detect a varying number of CNVs, recent studies in cattle using the 50K have efficiently captured and reported a vast number of CNVs (Hou *et al.*, 2012; Jiang *et al.*, 2012). Although the use of NGS technologies for CNV detection can provide enhanced accuracy, resolution and coverage of CNVs as well as novel CNV identification, the high costs are still a diminishing factor (Alkan *et al.*, 2011).

The CNVRs detected in chapter 4 comprised different copy number states and varied across chromosomes. Copy number changes comprised of hemizygous deletions and duplications and homozygous deletions and duplications. Duplications were the predominant form of CNV. African breeds present more CNV duplications than deletions, when compared with composite, indicine and taurine breeds (Hou *et al.*, 2011). Chromosome 6 presented the most CNVs, with chromosome 25 displaying the least. The high prevalence of CNVs on chromosome 6 has been reported in Chinese Holstein (Jiang *et al.*, 2012) and *Bos taurus* breeds (Cicconardi *et al.*, 2013).

The association of CNVRs with tick resistance, presented 17 significant CNVRs. These were detected across 14 chromosomes. Copy number variable regions comprised of deletions, duplications or both. The prevalence of CNVRs in susceptible animals was greater than that of resistant animals. This could provide some indication of the mechanisms of CNVRs and their influence on the susceptibility of an animal. The comparison of CNVRs identified with SNPs identified and associated with tick resistance in Nguni cattle (Mapholi *et al.*, 2016) presented very little overlap.

The identification of the genes across or within 10Mb surrounding these CNVRs was deduced. A total of 17 genes were identified within or in close proximity to these regions. These genes are found within particular cellular components and are involved in various molecular processes and biological functions. A CNVR detected on chromosome 29 presented deletions and duplications. Deletions corresponded to susceptibility, while duplications corresponded to resistance. This CNVR identified overlapped with *LSP1* (*Lymphocyte-specific protein 1*). This gene had previously been reported within a duplication in Angus cattle, and associated with nematode resistance (Hou *et al.*, 2012). Chromosome 29 presented a deletion CNVR corresponding to resistance. This CNVR overlapped with *GABRB2 (Gamma-aminobutyric acid A receptor, subunit beta 2)*. This gene is involved in the central nervous system and is responsible for muscle contraction and the cellular response of histamine. The overlap of CNVRs with these genes could provide information for the mechanisms underlying parasite resistance.

5.3 Conclusions

The association of CNVRs and tick resistance is evident in Nguni cattle. The implications of CNVRs within the Nguni genome provides informative research on the potential mechanisms underlying tick resistance. The prevalence of CNVRs overlapping or within close proximity of genes play a pivotal role in immune related functions and processes and can have an influence on an animals resistance or susceptibility to parasites. Understanding the mechanisms and the impact of CNVs, can provide valuable information for future breeding and selection programmes. These breeding programmes can be used to curb the challenges faced through the utilisation of current tick control methods.

5.4 Summary of Contributions

5.4.1 Peer reviewed journal articles

i. Pickering, L.H., Wang, M.D., Dzama, K. and Muchadeyi, F.C. The relationship between copy number variations and tick resistance. Target journal: South African Journal of Animal Science (2017).

ii. Pickering, L.H., Wang, M.D., Dzama, K. and Muchadeyi, F.C. The association of copy number variations with tick resistance in South African Nguni cattle. Target journal: Animal Breeding and Genetics (2017).

5.4.2 Conference participation

5.4.1.1 Oral presentations

The association of copy number variations with tick resistance in South African Nguni cattle. International Symposium of Animal Genetics. Dublin, Ireland (2017).

5.4.1.2 Poster presentations

The association of copy number variations with tick resistance in South African Nguni cattle. International Symposium of Animal Genetics. Dublin, Ireland (2017).

5.5 Future research

The identification and validation of CNVRs associated with tick resistance in other breeds, could provide insight into the mechanisms of resistance in other breeds. Comparing these regions with regions in breeds of high host resistance, can provide markers for future breeding and selection programmes. Utilising alternative methods of CNV detection, such as NGS methods, and comparing CNVs detected within and between breeds, can identify further breed differences. The use of multi-omic approaches, to analyse variations and their relationship with complex traits is an upcoming area of research (Suravajhala *et al.*, 2016).

5.6 References

Alkan, C., Coe, B. P., & Eichler, E. E. 2011. Genome structural variation discovery and genotyping. Nat. Rev. Genet. 12, 363–76.

Asmaa, N. M., ElBably, M. A., & Shokier, K. A. 2014. Studies on prevalence, risk indicators and control options for tick infestation in ruminants. Beni-Suef Univ. J. Basic Appl. Sci. 3, 68–73.

Cicconardi, F., Chillemi, G., Tramontano, A., Marchitelli, C., Valentini, A., Ajmone-Marsan, P., & Nardone, A. 2013. Massive screening of copy number population-scale variation in *Bos taurus* genome. BMC Genomics. 14, 124.

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

Hou, Y., Liu, G. E., Bickhart, D. M., Matukumalli, L. K., Li, C., Song, J., Gasbarre, L. C., Van Tassell, C. P., & Sonstegard, T. S. 2012. Genomic regions showing copy number variations associate with resistance or susceptibility to gastrointestinal nematodes in Angus cattle. Funct. Integr. Genomics. 12, 81–92.

Jiang, L., Jiang, J., Wang, J., Ding, X., Liu, J., & Zhang, Q. 2012. Genome-Wide Identification of Copy Number Variations in Chinese Holstein. PLoS One. 7.

Katiyatiya, C. L. F., Muchenje, V., & Mushunje, A. 2014. Seasonal variation in coat characteristics, tick loads, cortisol levels, some physiological parameters and temperature humidity index on Nguni cows raised in low- and high-input farms. Int. J. Biometeorol. 733–743.

Liu, G. E., Ventura, M., Cellamare, A., Chen, L., Cheng, Z., Zhu, B., Li, C., Song, J., & Eichler, E. E. 2009. Analysis of recent segmental duplications in the bovine genome. BMC Genomics. 10, 571.

Liu, G. E., Hou, Y., Zhu, B., Liu, G. E., Hou, Y., Zhu, B., Cardone, M. F., Jiang, L., Cellamare, A., Mitra, A., Alexander, L. J., Coutinho, L. L., Elena, M., Aquila, D., Gasbarre, L. C., Lacalandra, G., Li, R. W., Matukumalli, L. K., Nonneman, D., Regitano, L. C. D. A., Smith, T. P. L., Song, J., Sonstegard, T. S., Tassell, C. P. Van, Ventura, M., Eichler, E. E., Mcdaneld, T. G., & Keele, J. W. 2010. Analysis of copy number variations among diverse cattle breeds. Genome Res. 20, 693–703.

Makina, S. O., Muchadeyi, F. C., van Marle-Kőster, E., MacNeil, M. D., & Maiwashe, A. 2014. Genetic diversity and population structure among six cattle breeds in South Africa using a whole genome SNP panel. Front. Genet. 5, 1–7.

Mapholi, N. O., Maiwashe, A., Matika, O., Riggio, V., Bishop, S. C., MacNeil, M. D., Banga, C., Taylor, J. F., & Dzama, K. 2016. Genome-wide association study of tick resistance in South African Nguni cattle. Ticks Tick. Borne. Dis. 7, 487–497.

Marufu, M. C., Chimonyo, M., Mapiye, C., & Dzama, K. 2011. Tick loads in cattle raised on sweet and sour rangelands in the low-input farming areas of South Africa. Trop. Anim. Health Prod. 43, 307–313.

Muchenje, V., Dzama, K., Chimonyo, M., Raats, J. G., & Strydom, P. E. 2008. Tick susceptibility and its effects on growth performance and carcass characteristics of Nguni, Bonsmara and Angus steers raised on natural pasture. Animal. 2, 298–304.

Sasaki, S., Watanabe, T., Nishimura, S., & Sugimoto, Y. 2016. Genome-wide identification of copy number variation using high-density single-nucleotide polymorphism array in Japanese Black cattle. BMC Genet. 17, 26.

Schoeman, S. J. 1989. Recent research into the production potential of indigenous cattle with special reference to the Sanga. S. Afr. J. Anim. Sci. 19, 55–61.

Suravajhala, P., Kogelman, L. J. A., & Kadarmideen, H. N. 2016. Multi-omic data integration and analysis using systems genomics approaches: methods and applications in animal production, health and welfare. Genet. Sel. Evol. 48, 38.

Wang, M. D., Dzama, K., Hefer, C. A., & Muchadeyi, F. C. 2015. Genomic population structure and prevalence of copy number variations in South African Nguni cattle. BMC Genomics. 16, 894.

Wu, Y., Fan, H., Jing, S., Xia, J., Chen, Y., Zhang, L., Gao, X., Li, J., Gao, H., & Ren, H. 2015. A genome-wide scan for copy number variations using high-density single nucleotide polymorphism array in Simmental cattle. Anim. Genet. 46, 289–298.

Zhang, F., Gu, W., Hurles, M. E., & Lupski, J. R. 2009. Copy number variation in human health, disease, and evolution. Annu Rev Genomics Hum Genet 10, 451–481.

Addendum A

Additional file 4.1 A complete list of CNVRs detected in this study, including their location, length, the number of animals presenting each specific CNVR and their respective copy number states.

CHR	CNVR_ID	CNVR	Start Position	End Position	Length (bp)	No. of animals	CN				Gain/Loss/Both
							1	2	3	4	
chr1	CNVR_1_6	chr1:5351369-5374084	5351369	5374084	22715	2	2				Loss
chr1	CNVR_1_13	chr1:9323844-9404794	9323844	9404794	80950	2	1		1		Both
chr1	CNVR_1_56	chr1:39083240-39259346	39083240	39259346	176106	2	1		1		Both
chr1	CNVR_1_66	chr1:49480638-49578738	49480638	49578738	98100	2	2				Loss
chr1	CNVR_1_68	chr1:52075648-52191701	52075648	52191701	116053	2	2				Loss
chr1	CNVR_1_77	chr1:66483743-66797315	66483743	66797315	313572	2			2		Gain
chr1	CNVR_1_85	chr1:89358747-89431170	89358747	89431170	72423	2	2				Loss
chr1	CNVR_1_92	chr1:99156076-99237377	99156076	99237377	81301	2	1		1		Both
chr1	CNVR_1_99	chr1:104152463-104344681	104152463	104344681	192218	2	2				Loss
chr1	CNVR_1_114	chr1:118714300-118815770	118714300	118815770	101470	2	1		1		Both
chr1	CNVR_1_4	chr1:4957717-5034441	4957717	5034441	76724	3	3				Loss
chr1	CNVR_1_26	chr1:20165566-20213558	20165566	20213558	47992	3		2	1		Both
chr1	CNVR_1_31	chr1:26782245-26966818	26782245	26966818	184573	3	1	1		1	Both
chr1	CNVR_1_59	chr1:39877948-40169491	39877948	40169491	291543	3	3				Loss
chr1	CNVR_1_117	chr1:120915343-120948506	120915343	120948506	33163	3	3				Loss

CHR	CNVR_ID	CNVR	Start Position	End Position	Length (bp)	No. of animals	CN				Gain/Loss/Both
							1	2	3	4	
chr1	CNVR_1_61	chr1:42142033-42265516	42142033	42265516	123483	4	3		1		Both
chr1	CNVR_1_81	chr1:77647196-77714385	77647196	77714385	67189	4	2		2		Both
chr1	CNVR_1_95	chr1:102538612-102589009	102538612	102589009	50397	4	3		1		Both
chr1	CNVR_1_96	chr1:102792723-102903978	102792723	102903978	111255	5	4		1		Both
chr1	CNVR_1_18	chr1:14693262-14708162	14693262	14708162	14900	6	5		1		Both
chr1	CNVR_1_25	chr1:17425222-17504974	17425222	17504974	79752	6	5			1	Both
chr1	CNVR_1_94	chr1:102421173-102464968	102421173	102464968	43795	6	5		1		Both
chr1	CNVR_1_102	chr1:105084197-105108951	105084197	105108951	24754	11	10		1		Both
chr10	CNVR_10_3	chr10:20146107-20229329	20146107	20229329	83222	2			2		Gain
chr10	CNVR_10_9	chr10:41790011-41849457	41790011	41849457	59446	2	2				Loss
chr10	CNVR_10_13	chr10:44040387-44143863	44040387	44143863	103476	2		1	1		Both
chr10	CNVR_10_21	chr10:58634276-58749431	58634276	58749431	115155	2			2		Gain
chr10	CNVR_10_23	chr10:59812472-59927870	59812472	59927870	115398	2	2				Loss
chr10	CNVR_10_26	chr10:71022679-71082204	71022679	71082204	59525	2	2				Loss
chr10	CNVR_10_32	chr10:90705511-90806688	90705511	90806688	101177	2			2		Gain
chr10	CNVR_10_16	chr10:45387461-45488421	45387461	45488421	100960	3			3		Gain
chr10	CNVR_10_37	chr10:103178023-103280130	103178023	103280130	102107	3	3				Loss
chr11	CNVR_11_2	chr11:2435270-2507194	2435270	2507194	71924	2	2				Loss
chr11	CNVR_11_9	chr11:6724678-6777332	6724678	6777332	52654	2	1		1		Both

CHR	CNVR_ID	CNVR	Start Position	End Position	Length (bp)	No. of animals	CN				Gain/Loss/Both
							1	2	3	4	
chr11	CNVR_11_12	chr11:13597473-13751244	13597473	13751244	153771	2	2				Loss
chr11	CNVR_11_14	chr11:16404805-16514661	16404805	16514661	109856	2	2				Loss
chr11	CNVR_11_15	chr11:17445658-17686944	17445658	17686944	241286	2	2				Loss
chr11	CNVR_11_27	chr11:25685264-25736918	25685264	25736918	51654	2			2		Gain
chr11	CNVR_11_30	chr11:27812207-27917853	27812207	27917853	105646	2	1		1		Both
chr11	CNVR_11_40	chr11:38733905-38858028	38733905	38858028	124123	2			2		Gain
chr11	CNVR_11_41	chr11:39299927-39437937	39299927	39437937	138010	2	1		1		Both
chr11	CNVR_11_44	chr11:51660353-51741896	51660353	51741896	81543	2	1		1		Both
chr11	CNVR_11_60	chr11:89680768-90215622	89680768	90215622	534854	2			2		Gain
chr11	CNVR_11_79	chr11:106682471-106804258	106682471	106804258	121787	2	2				Loss
chr11	CNVR_11_24	chr11:24892409-25036830	24892409	25036830	144421	3	1		2		Both
chr11	CNVR_11_38	chr11:38375934-38474579	38375934	38474579	98645	3	2		1		Both
chr11	CNVR_11_77	chr11:105699664-106067128	105699664	106067128	367464	3	3				Loss
chr11	CNVR_11_33	chr11:34147377-34244441	34147377	34244441	97064	4	3		1		Both
chr11	CNVR_11_73	chr11:104633267-104679644	104633267	104679644	46377	4	2	1	1		Both
chr11	CNVR_11_20	chr11:23085190-23121908	23085190	23121908	36718	7	6		1		Both
chr11	CNVR_11_70	chr11:104091220-104117370	104091220	104117370	26150	9	7		2		Both
chr11	CNVR_11_48	chr11:59622185-59668010	59622185	59668010	45825	11	5	1	4	1	Both
chr12	CNVR_12_4	chr12:4614286-4718544	4614286	4718544	104258	2	1		1		Both

CHR	CNVR_ID	CNVR	Start Position	End Position	Length (bp)	No. of animals	CN				Gain/Loss/Both
							1	2	3	4	
chr12	CNVR_12_5	chr12:4995556-5101246	4995556	5101246	105690	2	1		1		Both
chr12	CNVR_12_18	chr12:35356760-35556098	35356760	35556098	199338	2			2		Gain
chr12	CNVR_12_25	chr12:39801280-39925909	39801280	39925909	124629	2			2		Gain
chr12	CNVR_12_35	chr12:52032777-52098655	52032777	52098655	65878	2	1		1		Both
chr12	CNVR_12_40	chr12:56173146-56326750	56173146	56326750	153604	2	2				Loss
chr12	CNVR_12_44	chr12:61573515-61686477	61573515	61686477	112962	2	2				Loss
chr12	CNVR_12_45	chr12:61950490-62342403	61950490	62342403	391913	2			2		Gain
chr12	CNVR_12_47	chr12:65209818-65602532	65209818	65602532	392714	2	1		1		Both
chr12	CNVR_12_55	chr12:90915595-91017026	90915595	91017026	101431	2	1	1			Loss
chr12	CNVR_12_32	chr12:45358430-45470640	45358430	45470640	112210	3	3				Loss
chr12	CNVR_12_42	chr12:59185443-59218880	59185443	59218880	33437	3	3				Loss
chr12	CNVR_12_52	chr12:90280720-90621015	90280720	90621015	340295	3	3				Loss
chr12	CNVR_12_53	chr12:90704572-90778028	90704572	90778028	73456	3	3				Loss
chr12	CNVR_12_8	chr12:21279986-21352699	21279986	21352699	72713	4	3		1		Both
chr12	CNVR_12_2	chr12:1546821-1601317	1546821	1601317	54496	5	3		2		Both
chr12	CNVR_12_26	chr12:43551814-43638160	43551814	43638160	86346	5	3		2		Both
chr12	CNVR_12_14	chr12:31439132-31555734	31439132	31555734	116602	6	6				Loss
chr12	CNVR_12_21	chr12:36252254-36475306	36252254	36475306	223052	6			5	1	Gain
chr12	CNVR_12_29	chr12:45002070-45073424	45002070	45073424	71354	6	3		3		Both

CHR	CNVR_ID	CNVR	Start Position	End Position	Length (bp)	No. of animals	CN				Gain/Loss/Both
							1	2	3	4	
chr13	CNVR_13_2	chr13:12587622-12759014	12587622	12759014	171392	2			2		Gain
chr13	CNVR_13_4	chr13:13843476-13907892	13843476	13907892	64416	2			2		Gain
chr13	CNVR_13_9	chr13:35465554-35661069	35465554	35661069	195515	2			2		Gain
chr13	CNVR_13_11	chr13:39741913-39826900	39741913	39826900	84987	2			2		Gain
chr13	CNVR_13_23	chr13:74478509-74667261	74478509	74667261	188752	2			2		Gain
chr13	CNVR_13_24	chr13:75383374-75432478	75383374	75432478	49104	2	1		1		Both
chr14	CNVR_14_3	chr14:2319504-2468020	2319504	2468020	148516	2	2				Loss
chr14	CNVR_14_15	chr14:26685204-26949215	26685204	26949215	264011	2			2		Gain
chr14	CNVR_14_33	chr14:58203661-58342794	58203661	58342794	139133	2	1		1		Both
chr14	CNVR_14_36	chr14:75035895-75097271	75035895	75097271	61376	2	2				Loss
chr14	CNVR_14_1	chr14:1616618-1801116	1616618	1801116	184498	3	3				Loss
chr14	CNVR_14_24	chr14:34639444-34772856	34639444	34772856	133412	3	2		1		Both
chr14	CNVR_14_25	chr14:35518739-35590232	35518739	35590232	71493	3	2		1		Both
chr14	CNVR_14_30	chr14:55765964-55834821	55765964	55834821	68857	4	2		2		Both
chr14	CNVR_14_18	chr14:27380992-27408939	27380992	27408939	27947	6	5			1	Both
chr15	CNVR_15_1	chr15:3286452-3369565	3286452	3369565	83113	2	1	1			Loss
chr15	CNVR_15_8	chr15:11439502-11531149	11439502	11531149	91647	2	2				Loss
chr15	CNVR_15_13	chr15:20712027-20989030	20712027	20989030	277003	2			2		Gain
chr15	CNVR_15_21	chr15:36898323-36950873	36898323	36950873	52550	2	1		1		Both

CHR	CNVR_ID	CNVR	Start Position	End Position	Length (bp)	No. of animals	CN				Gain/Loss/Both
							1	2	3	4	
chr15	CNVR_15_28	chr15:57260972-57333896	57260972	57333896	72924	2	1		1		Loss
chr15	CNVR_15_44	chr15:76675303-76930018	76675303	76930018	254715	2			2		Gain
chr15	CNVR_15_4	chr15:5474020-5537822	5474020	5537822	63802	3	3				Loss
chr15	CNVR_15_32	chr15:62283471-62428148	62283471	62428148	144677	3	3				Loss
chr15	CNVR_15_38	chr15:70793253-70880599	70793253	70880599	87346	3	1			2	Both
chr16	CNVR_16_2	chr16:1608132-1690484	1608132	1690484	82352	2	1		1		Both
chr16	CNVR_16_11	chr16:10308240-10474014	10308240	10474014	165774	2	2				Loss
chr16	CNVR_16_13	chr16:10958964-11096101	10958964	11096101	137137	2	2				Loss
chr16	CNVR_16_6	chr16:6494072-6557355	6494072	6557355	63283	6	5			1	Both
chr16	CNVR_16_28	chr16:50759541-50798300	50759541	50798300	38759	6	3		3		Both
chr17	CNVR_17_3	chr17:3445779-3599671	3445779	3599671	153892	2			2		Gain
chr17	CNVR_17_4	chr17:6380171-6528887	6380171	6528887	148716	2			2		Gain
chr17	CNVR_17_28	chr17:39963957-40034778	39963957	40034778	70821	2	2				Loss
chr17	CNVR_17_15	chr17:24810915-24848457	24810915	24848457	37542	3	1		1	1	Both
chr17	CNVR_17_31	chr17:42661925-42711744	42661925	42711744	49819	3	2		1		Both
chr17	CNVR_17_21	chr17:29057538-29103180	29057538	29103180	45642	4	4				Loss
chr17	CNVR_17_37	chr17:73118011-74032690	73118011	74032690	914679	4	4				Loss
chr17	CNVR_17_12	chr17:23285669-23431642	23285669	23431642	145973	6	4		1	1	Both
chr17	CNVR_17_39	chr17:74292319-74393620	74292319	74393620	101301	6	5		1		Both

CHR	CNVR_ID	CNVR	Start Position	End Position	Length (bp)	No. of animals	CN				Gain/Loss/Both
							1	2	3	4	
chr17	CNVR_17_40	chr17:74481008-74627267	74481008	74627267	146259	6	4	1	1		Both
chr17	CNVR_17_1	chr17:2728598-2784904	2728598	2784904	56306	10	6	3	1		Both
chr18	CNVR_18_4	chr18:7819498-7868943	7819498	7868943	49445	2			2		Gain
chr18	CNVR_18_8	chr18:9867308-9948949	9867308	9948949	81641	2	2				Loss
chr18	CNVR_18_11	chr18:11377350-11416050	11377350	11416050	38700	2	1		1		Both
chr18	CNVR_18_25	chr18:62375495-62537136	62375495	62537136	161641	2	2				Loss
chr18	CNVR_18_26	chr18:63096692-63167945	63096692	63167945	71253	2	2				Loss
chr18	CNVR_18_21	chr18:57463530-57518175	57463530	57518175	54645	3			3		Gain
chr19	CNVR_19_21	chr19:51326750-51419352	51326750	51419352	92602	2		1	1		Both
chr19	CNVR_19_30	chr19:56401302-56426090	56401302	56426090	24788	2			2		Gain
chr19	CNVR_19_27	chr19:55235501-55257191	55235501	55257191	21690	3	1	1	1		Both
chr19	CNVR_19_18	chr19:49727374-49784054	49727374	49784054	56680	4	4				Loss
chr2	CNVR_2_2	chr2:1316010-1366643	1316010	1366643	50633	2		1	1		Both
chr2	CNVR_2_5	chr2:7720309-7772897	7720309	7772897	52588	2			2		Gain
chr2	CNVR_2_10	chr2:18834716-18885419	18834716	18885419	50703	2			2		Gain
chr2	CNVR_2_12	chr2:21696086-21904823	21696086	21904823	208737	2	1		1		Both
chr2	CNVR_2_17	chr2:29027343-29135710	29027343	29135710	108367	2		1	1		Both
chr2	CNVR_2_20	chr2:30992555-31242327	30992555	31242327	249772	2	2				Loss
chr2	CNVR_2_72	chr2:78604521-78628109	78604521	78628109	23588	2	2				Loss

CHR	CNVR_ID	CNVR	Start Position	End Position	Length (bp)	No. of animals	CN				Gain/Loss/Both
							1	2	3	4	
chr2	CNVR_2_74	chr2:84038555-84170290	84038555	84170290	131735	2	2				Loss
chr2	CNVR_2_93	chr2:136461239-136506232	136461239	136506232	44993	2	2				Loss
chr2	CNVR_2_13	chr2:22283989-22352656	22283989	22352656	68667	3	3				Loss
chr2	CNVR_2_32	chr2:53894600-53951533	53894600	53951533	56933	3	3				Loss
chr2	CNVR_2_52	chr2:60087290-60143778	60087290	60143778	56488	3	3				Loss
chr2	CNVR_2_61	chr2:71382630-71598307	71382630	71598307	215677	3			3		Gain
chr2	CNVR_2_65	chr2:77516653-77554557	77516653	77554557	37904	3	2		1		Both
chr2	CNVR_2_67	chr2:77719508-77793201	77719508	77793201	73693	3	1		2		Both
chr2	CNVR_2_70	chr2:78433731-78533652	78433731	78533652	99921	3	3				Loss
chr2	CNVR_2_84	chr2:131325971-131906865	131325971	131906865	580894	3			3		Gain
chr2	CNVR_2_89	chr2:134792816-134853068	134792816	134853068	60252	3			3		Gain
chr2	CNVR_2_42	chr2:55852340-55905779	55852340	55905779	53439	4	3		1		Both
chr2	CNVR_2_25	chr2:41786376-41829776	41786376	41829776	43400	5	5				Loss
chr2	CNVR_2_49	chr2:56583793-56636575	56583793	56636575	52782	5	3		2		Both
chr2	CNVR_2_47	chr2:56456865-56479206	56456865	56479206	22341	6	4		2		Both
chr20	CNVR_20_10	chr20:37553988-37838938	37553988	37838938	284950	2			2		Gain
chr20	CNVR_20_13	chr20:38429784-38453649	38429784	38453649	23865	2			2		Gain
chr20	CNVR_20_15	chr20:38892683-38960231	38892683	38960231	67548	2			2		Gain
chr20	CNVR_20_17	chr20:41072335-41239866	41072335	41239866	167531	2			2		Gain

CHR	CNVR_ID	CNVR	Start Position	End Position	Length (bp)	No. of animals	CN				Gain/Loss/Both
							1	2	3	4	
chr20	CNVR_20_54	chr20:54586441-54754012	54586441	54754012	167571	2	2				Loss
chr20	CNVR_20_58	chr20:60473408-60686884	60473408	60686884	213476	2	1		1		Both
chr20	CNVR_20_3	chr20:10295382-10486993	10295382	10486993	191611	3	2		1		Both
chr20	CNVR_20_21	chr20:45052283-45266553	45052283	45266553	214270	3	3				Loss
chr20	CNVR_20_35	chr20:50001881-50128719	50001881	50128719	126838	3	3				Loss
chr20	CNVR_20_51	chr20:53957730-53991925	53957730	53991925	34195	4	3		1		Both
chr20	CNVR_20_23	chr20:46035562-46080951	46035562	46080951	45389	5	4	1			Loss
chr20	CNVR_20_24	chr20:46121445-46179978	46121445	46179978	58533	5	4	1			Loss
chr20	CNVR_20_30	chr20:48749320-48903795	48749320	48903795	154475	5	5				Loss
chr20	CNVR_20_46	chr20:51940977-52011466	51940977	52011466	70489	5	3		2		Both
chr20	CNVR_20_31	chr20:49090857-49190161	49090857	49190161	99304	6	6				Loss
chr21	CNVR_21_9	chr21:19193100-19601568	19193100	19601568	408468	2			2		Gain
chr21	CNVR_21_27	chr21:60195812-60220817	60195812	60220817	25005	2			2		Gain
chr21	CNVR_21_28	chr21:60364864-60553762	60364864	60553762	188898	2	1		1		Both
chr21	CNVR_21_31	chr21:62356345-62402185	62356345	62402185	45840	2			2		Gain
chr21	CNVR_21_34	chr21:64292330-64367413	64292330	64367413	75083	2	1		1		Both
chr21	CNVR_21_36	chr21:65316438-65365440	65316438	65365440	49002	2			2		Gain
chr21	CNVR_21_45	chr21:70814197-70931906	70814197	70931906	117709	2	1		1		Both
chr21	CNVR_21_29	chr21:61310103-61370773	61310103	61370773	60670	4	3		1		Both

CHR	CNVR_ID	CNVR	Start Position	End Position	Length (bp)	No. of animals	CN				Gain/Loss/Both
							1	2	3	4	
chr21	CNVR_21_46	chr21:71025601-71109676	71025601	71109676	84075	4	4				Loss
chr21	CNVR_21_42	chr21:70272221-70373409	70272221	70373409	101188	6	5		1		Both
chr21	CNVR_21_18	chr21:51733686-51857151	51733686	51857151	123465	8	8				Loss
chr22	CNVR_22_19	chr22:59204394-59277831	59204394	59277831	73437	3	3				Loss
chr22	CNVR_22_23	chr22:60084013-60105535	60084013	60105535	21522	6	6				Loss
chr22	CNVR_22_28	chr22:60736089-60960603	60736089	60960603	224514	6	5		1		Both
chr22	CNVR_22_8	chr22:24078956-24161191	24078956	24161191	82235	12	3		9		Both
chr23	CNVR_23_4	chr23:14372603-14424516	14372603	14424516	51913	2			2		Gain
chr23	CNVR_23_9	chr23:24640422-24667121	24640422	24667121	26699	2			2		Gain
chr23	CNVR_23_18	chr23:49626490-49688244	49626490	49688244	61754	2			2		Gain
chr23	CNVR_23_16	chr23:49135538-49177883	49135538	49177883	42345	3	3				Loss
chr24	CNVR_24_10	chr24:20513331-20684377	20513331	20684377	171046	2	1		1		Both
chr24	CNVR_24_17	chr24:26080792-26152512	26080792	26152512	71720	2			2		Gain
chr24	CNVR_24_25	chr24:45471120-45539162	45471120	45539162	68042	2			2		Gain
chr24	CNVR_24_28	chr24:45973112-46027109	45973112	46027109	53997	2			2		Gain
chr24	CNVR_24_31	chr24:52047342-52086080	52047342	52086080	38738	2	1		1		Both
chr24	CNVR_24_14	chr24:24327708-24432820	24327708	24432820	105112	3	2		1		Both
chr24	CNVR_24_21	chr24:28400451-28497174	28400451	28497174	96723	7	7				Loss
chr24	CNVR_24_19	chr24:28154039-28196203	28154039	28196203	42164	8	3	1	4		Both

CHR	CNVR_ID	CNVR	Start Position	End Position	Length (bp)	No. of animals	CN				Gain/Loss/Both
							1	2	3	4	
chr25	CNVR_25_14	chr25:41191025-41596286	41191025	41596286	405261	4	4				Loss
chr25	CNVR_25_15	chr25:41935313-42364359	41935313	42364359	429046	5	5				Loss
chr25	CNVR_25_5	chr25:930509-983759	930509	983759	53250	6	6				Loss
chr26	CNVR_26_12	chr26:6620890-6692455	6620890	6692455	71565	2	1		1		Both
chr26	CNVR_26_15	chr26:12268427-12463857	12268427	12463857	195430	2			2		Gain
chr26	CNVR_26_31	chr26:45669303-45909407	45669303	45909407	240104	2			2		Gain
chr26	CNVR_26_23	chr26:19880677-19942669	19880677	19942669	61992	3	2		1		Both
chr26	CNVR_26_26	chr26:25880226-25982293	25880226	25982293	102067	3	3				Loss
chr26	CNVR_26_1	chr26:874754-1012643	874754	1012643	137889	4	3		1		Both
chr26	CNVR_26_18	chr26:17163979-17246984	17163979	17246984	83005	4	4				Loss
chr26	CNVR_26_35	chr26:51154962-51267717	51154962	51267717	112755	6	5		1		Both
chr26	CNVR_26_9	chr26:3884506-3955269	3884506	3955269	70763	7	6	1			Both
chr26	CNVR_26_3	chr26:2343667-2452597	2343667	2452597	108930	8	7		1		Both
chr27	CNVR_27_1	chr27:275256-367537	275256	367537	92281	2			2		Gain
chr27	CNVR_27_3	chr27:2298767-2378910	2298767	2378910	80143	2	2				Loss
chr27	CNVR_27_18	chr27:13767421-13847519	13767421	13847519	80098	2			2		Gain
chr27	CNVR_27_28	chr27:20559148-20791916	20559148	20791916	232768	2			2		Gain
chr27	CNVR_27_26	chr27:16453926-16571478	16453926	16571478	117552	3	3				Loss
chr27	CNVR_27_31	chr27:21615584-21725517	21615584	21725517	109933	3	3				Loss

CHR	CNVR_ID	CNVR	Start Position	End Position	Length (bp)	No. of animals	CN				Gain/Loss/Both
							1	2	3	4	
chr27	CNVR_27_8	chr27:8781446-8927449	8781446	8927449	146003	5	5				Loss
chr27	CNVR_27_22	chr27:15175109-15229451	15175109	15229451	54342	6			6		Gain
chr28	CNVR_28_5	chr28:13954510-14058525	13954510	14058525	104015	2			2		Gain
chr28	CNVR_28_18	chr28:22941657-23075883	22941657	23075883	134226	2	1		1		Both
chr28	CNVR_28_29	chr28:33262793-33449126	33262793	33449126	186333	2			2		Gain
chr28	CNVR_28_33	chr28:37799844-37962946	37799844	37962946	163102	2			2		Gain
chr28	CNVR_28_34	chr28:38004515-38136070	38004515	38136070	131555	2	1		1		Both
chr28	CNVR_28_27	chr28:25175373-25352987	25175373	25352987	177614	6	5		1		Both
chr28	CNVR_28_15	chr28:22226165-22355840	22226165	22355840	129675	9	9				Loss
chr28	CNVR_28_10	chr28:21137220-21292836	21137220	21292836	155616	11	7		1	3	Both
chr29	CNVR_29_5	chr29:13846574-13935802	13846574	13935802	89228	2	2				Loss
chr29	CNVR_29_13	chr29:27163710-27218550	27163710	27218550	54840	2	1		1		Both
chr29	CNVR_29_22	chr29:41989397-42455680	41989397	42455680	466283	2			2		Gain
chr29	CNVR_29_23	chr29:42897144-43006000	42897144	43006000	108856	2			2		Gain
chr29	CNVR_29_24	chr29:43777249-43948092	43777249	43948092	170843	2			2		Gain
chr29	CNVR_29_28	chr29:46969623-46999731	46969623	46999731	30108	2	1	1			Loss
chr29	CNVR_29_36	chr29:50569383-50586068	50569383	50586068	16685	4	4				Loss
chr29	CNVR_29_34	chr29:50202589-50240781	50202589	50240781	38192	6	4		2		Both
chr29	CNVR_29_39	chr29:51396010-51502868	51396010	51502868	106858	6	5		1		Both

CHR	CNVR_ID	CNVR	Start Position	End Position	Length (bp)	No. of animals	CN				Gain/Loss/Both
							1	2	3	4	
chr3	CNVR_3_7	chr3:18649247-19048207	18649247	19048207	398960	2			2		Gain
chr3	CNVR_3_8	chr3:21371931-21587811	21371931	21587811	215880	2			2		Gain
chr3	CNVR_3_12	chr3:32788575-32944221	32788575	32944221	155646	2			2		Gain
chr3	CNVR_3_21	chr3:40534231-40678490	40534231	40678490	144259	2			2		Gain
chr3	CNVR_3_24	chr3:41371572-41432313	41371572	41432313	60741	2	1		1		Both
chr3	CNVR_3_28	chr3:46922624-47165828	46922624	47165828	243204	2			2		Gain
chr3	CNVR_3_29	chr3:47541668-47703475	47541668	47703475	161807	2	1		1		Both
chr3	CNVR_3_32	chr3:61758600-61947686	61758600	61947686	189086	2	1		1		Both
chr3	CNVR_3_52	chr3:69716313-69832745	69716313	69832745	116432	2	1		1		Both
chr3	CNVR_3_54	chr3:71141852-71238926	71141852	71238926	97074	2			2		Gain
chr3	CNVR_3_59	chr3:84764797-84938304	84764797	84938304	173507	2			2		Gain
chr3	CNVR_3_65	chr3:95254105-95497810	95254105	95497810	243705	2	2				Loss
chr3	CNVR_3_14	chr3:33598353-33741850	33598353	33741850	143497	3	1		2		Both
chr3	CNVR_3_18	chr3:39393293-39599273	39393293	39599273	205980	3	3				Loss
chr3	CNVR_3_57	chr3:73921609-74636373	73921609	74636373	714764	3	3				Loss
chr3	CNVR_3_61	chr3:85984517-86180854	85984517	86180854	196337	3			3		Gain
chr3	CNVR_3_79	chr3:118934229-119048887	118934229	119048887	114658	3	2		1		Both
chr3	CNVR_3_38	chr3:63916846-63939015	63916846	63939015	22169	4	3		1		Both
chr3	CNVR_3_48	chr3:66296159-66381935	66296159	66381935	85776	4	3		1		Both

CHR	CNVR_ID	CNVR	Start Position	End Position	Length (bp)	No. of animals	CN				Gain/Loss/Both
							1	2	3	4	
chr3	CNVR_3_88	chr3:121025205-121098536	121025205	121098536	73331	5	5				Loss
chr4	CNVR_4_5	chr4:9615916-9748719	9615916	9748719	132803	2			2		Gain
chr4	CNVR_4_9	chr4:16800525-16899410	16800525	16899410	98885	2	2				Loss
chr4	CNVR_4_18	chr4:22026811-22227807	22026811	22227807	200996	2	1		1		Gain
chr4	CNVR_4_21	chr4:30606899-30732178	30606899	30732178	125279	2			2		Gain
chr4	CNVR_4_26	chr4:36896740-37021317	36896740	37021317	124577	2			2		Gain
chr4	CNVR_4_39	chr4:41444423-41610894	41444423	41610894	166471	2	2				Loss
chr4	CNVR_4_42	chr4:45149895-45664659	45149895	45664659	514764	2			2		Gain
chr4	CNVR_4_49	chr4:52669435-52690928	52669435	52690928	21493	2	2				Loss
chr4	CNVR_4_51	chr4:55566787-55656122	55566787	55656122	89335	2	2				Loss
chr4	CNVR_4_53	chr4:61203879-61324109	61203879	61324109	120230	2	1		1		Both
chr4	CNVR_4_66	chr4:91874324-91951169	91874324	91951169	76845	2			2		Gain
chr4	CNVR_4_67	chr4:92289023-92331760	92289023	92331760	42737	2			2		Gain
chr4	CNVR_4_74	chr4:111140851-111245497	111140851	111245497	104646	2			2		Gain
chr4	CNVR_4_76	chr4:111990062-112013902	111990062	112013902	23840	2	1		1		Both
chr4	CNVR_4_78	chr4:114975718-115112032	114975718	115112032	136314	2			2		Gain
chr4	CNVR_4_45	chr4:49717334-49824079	49717334	49824079	106745	3	1		2		Both
chr4	CNVR_4_55	chr4:67462808-67546405	67462808	67546405	83597	3			2	1	Gain
chr4	CNVR_4_64	chr4:91080419-91381987	91080419	91381987	301568	3	2		1		Both

CHR	CNVR_ID	CNVR	Start Position	End Position	Length (bp)	No. of animals	CN				Gain/Loss/Both
							1	2	3	4	
chr4	CNVR_4_37	chr4:40746706-40806862	40746706	40806862	60156	4	3			1	Both
chr4	CNVR_4_59	chr4:78440044-78523846	78440044	78523846	83802	4	4				Loss
chr4	CNVR_4_62	chr4:89568067-89648770	89568067	89648770	80703	4	3		1		Both
chr4	CNVR_4_71	chr4:108867683-108928668	108867683	108928668	60985	4	4				Loss
chr4	CNVR_4_15	chr4:21194199-21248954	21194199	21248954	54755	6	5				Loss
chr4	CNVR_4_31	chr4:38653121-38680360	38653121	38680360	27239	7	5		1	1	Both
chr5	CNVR_5_29	chr5:78206091-78352587	78206091	78352587	146496	2	1		1		Both
chr5	CNVR_5_38	chr5:110399710-110483953	110399710	110483953	84243	2			2		Gain
chr5	CNVR_5_51	chr5:118430785-118501191	118430785	118501191	70406	2	1		1		Both
chr5	CNVR_5_53	chr5:118982815-119071205	118982815	119071205	88390	2	2				Loss
chr5	CNVR_5_1	chr5:1821290-1890085	1821290	1890085	68795	3	3				Loss
chr5	CNVR_5_20	chr5:26621180-26809399	26621180	26809399	188219	3	1		2		Both
chr5	CNVR_5_10	chr5:3531860-3552418	3531860	3552418	20558	4	3			1	Both
chr5	CNVR_5_18	chr5:17995212-18075032	17995212	18075032	79820	4	4				Loss
chr5	CNVR_5_44	chr5:117133270-117194638	117133270	117194638	61368	4	3	1			Loss
chr5	CNVR_5_55	chr5:119820680-120264251	119820680	120264251	443571	4	3		1		Both
chr5	CNVR_5_48	chr5:117938671-118014757	117938671	118014757	76086	5	3	1	1		Both
chr6	CNVR_6_46	chr6:44968134-45052429	44968134	45052429	84295	2			2		Gain
chr6	CNVR_6_58	chr6:52628477-52725432	52628477	52725432	96955	2	2				Loss

CHR	CNVR_ID	CNVR	Start Position	End Position	Length (bp)	No. of animals	CN				Gain/Loss/Both
							1	2	3	4	
chr6	CNVR_6_70	chr6:56164953-56232922	56164953	56232922	67969	2	1		1		Both
chr6	CNVR_6_87	chr6:78539712-78598487	78539712	78598487	58775	2	1		1		Both
chr6	CNVR_6_98	chr6:87201599-87327708	87201599	87327708	126109	2	1		1		Both
chr6	CNVR_6_102	chr6:93137474-93297733	93137474	93297733	160259	2			2		Gain
chr6	CNVR_6_109	chr6:106981782-107045738	106981782	107045738	63956	2	2				Loss
chr6	CNVR_6_112	chr6:108891914-109022523	108891914	109022523	130609	2	1	1			Loss
chr6	CNVR_6_5	chr6:10105101-10199636	10105101	10199636	94535	3	3				Loss
chr6	CNVR_6_16	chr6:12521271-12648459	12521271	12648459	127188	3	1		2		Both
chr6	CNVR_6_25	chr6:20297397-20350438	20297397	20350438	53041	3	3				Loss
chr6	CNVR_6_31	chr6:38845992-38939012	38845992	38939012	93020	3	2		1		Both
chr6	CNVR_6_50	chr6:50046827-50255777	50046827	50255777	208950	3	2			1	Both
chr6	CNVR_6_67	chr6:55522974-55616565	55522974	55616565	93591	3	1		2		Both
chr6	CNVR_6_111	chr6:108076099-108476017	108076099	108476017	399918	3	3				Loss
chr6	CNVR_6_3	chr6:4156416-4217935	4156416	4217935	61519	4		4			Loss
chr6	CNVR_6_91	chr6:81286240-81368675	81286240	81368675	82435	4	2		2		Both
chr6	CNVR_6_1	chr6:1258166-1332167	1258166	1332167	74001	5	5				Loss
chr6	CNVR_6_20	chr6:14449728-14597357	14449728	14597357	147629	5	4		1		Both
chr6	CNVR_6_40	chr6:43037439-43089739	43037439	43089739	52300	5		4	1		Both
chr6	CNVR_6_43	chr6:44622597-44649549	44622597	44649549	26952	5	4		1		Both

CHR	CNVR_ID	CNVR	Start Position	End Position	Length (bp)	No. of animals	CN				Gain/Loss/Both
							1	2	3	4	
chr6	CNVR_6_9	chr6:10760779-10838635	10760779	10838635	77856	6	2		4		Both
chr6	CNVR_6_37	chr6:41530548-41611777	41530548	41611777	81229	6	3		3		Both
chr6	CNVR_6_61	chr6:53514737-53643730	53514737	53643730	128993	6	5		1		Both
chr6	CNVR_6_116	chr6:109610794-109682953	109610794	109682953	72159	7	7				Loss
chr6	CNVR_6_55	chr6:50981312-51007189	50981312	51007189	25877	8	8				Loss
chr6	CNVR_6_81	chr6:71994646-72051699	71994646	72051699	57053	25			25		Gain
chr7	CNVR_7_11	chr7:18910887-19286649	18910887	19286649	375762	2	1		1		Both
chr7	CNVR_7_14	chr7:21222125-21321599	21222125	21321599	99474	2	1		1		Both
chr7	CNVR_7_16	chr7:24571599-24625377	24571599	24625377	53778	2		2			Loss
chr7	CNVR_7_19	chr7:25120220-25174975	25120220	25174975	54755	2	1		1		Both
chr7	CNVR_7_21	chr7:28897553-29082004	28897553	29082004	184451	2			2		Gain
chr7	CNVR_7_23	chr7:31921125-32091641	31921125	32091641	170516	2			2		Gain
chr7	CNVR_7_42	chr7:39172213-39226471	39172213	39226471	54258	2	1		1		Both
chr7	CNVR_7_45	chr7:39944515-40136380	39944515	40136380	191865	2			2		Gain
chr7	CNVR_7_55	chr7:78584955-78672889	78584955	78672889	87934	2			2		Gain
chr7	CNVR_7_57	chr7:79012351-79096832	79012351	79096832	84481	2	2				Loss
chr7	CNVR_7_69	chr7:106170981-106274908	106170981	106274908	103927	2	2				Loss
chr7	CNVR_7_8	chr7:17913294-18048182	17913294	18048182	134888	3			3		Gain
chr7	CNVR_7_27	chr7:33722644-33746267	33722644	33746267	23623	3	1		2		Both

CHR	CNVR_ID	CNVR	Start Position	End Position	Length (bp)	No. of animals	CN				Gain/Loss/Both
							1	2	3	4	
chr7	CNVR_7_39	chr7:37150427-37180616	37150427	37180616	30189	5	4		1		Both
chr7	CNVR_7_34	chr7:34559205-34611681	34559205	34611681	52476	6	6				Loss
chr7	CNVR_7_54	chr7:75305297-75370366	75305297	75370366	65069	7	5	1	1		Both
chr8	CNVR_8_20	chr8:49981054-50148467	49981054	50148467	167413	2	1		1		Both
chr8	CNVR_8_23	chr8:50499574-50575791	50499574	50575791	76217	2	2				Loss
chr8	CNVR_8_26	chr8:54961322-55084725	54961322	55084725	123403	2			2		Gain
chr8	CNVR_8_34	chr8:83102322-83190665	83102322	83190665	88343	2			2		Gain
chr8	CNVR_8_37	chr8:88096752-88237379	88096752	88237379	140627	2			2		Gain
chr8	CNVR_8_38	chr8:88302649-88545459	88302649	88545459	242810	2			2		Gain
chr8	CNVR_8_44	chr8:94115663-94386951	94115663	94386951	271288	2	2				Loss
chr8	CNVR_8_48	chr8:101231366-101404929	101231366	101404929	173563	2			2		Gain
chr8	CNVR_8_11	chr8:34795275-34898163	34795275	34898163	102888	4	3		1		Both
chr8	CNVR_8_14	chr8:35434141-35464882	35434141	35464882	30741	4	4				Loss
chr8	CNVR_8_42	chr8:93017958-93055587	93017958	93055587	37629	4	3		1		Both
chr9	CNVR_9_30	chr9:35953813-36002535	35953813	36002535	48722	2	1		1		Both
chr9	CNVR_9_39	chr9:55771156-56072341	55771156	56072341	301185	2	1	1			Loss
chr9	CNVR_9_45	chr9:91405210-91469581	91405210	91469581	64371	2	1		1		Both
chr9	CNVR_9_51	chr9:102826828-102930778	102826828	102930778	103950	2	2				Loss
chr9	CNVR_9_14	chr9:5079903-5148301	5079903	5148301	68398	3	2		1		Both

CHR	CNVR_ID	CNVR	Start Position	End Position	Length (bp)	No. of animals	CN				Gain/Loss/Both
							1	2	3	4	
chr9	CNVR_9_35	chr9:55338203-55414103	55338203	55414103	75900	3	1		2		Both
chr9	CNVR_9_46	chr9:93685360-93884361	93685360	93884361	199001	3	3				Loss
chr9	CNVR_9_48	chr9:94991477-95065382	94991477	95065382	73905	3			3		Gain
chr9	CNVR_9_16	chr9:5901981-5949799	5901981	5949799	47818	6	2	1	3		Both
chr9	CNVR_9_8	chr9:3971122-4000045	3971122	4000045	28923	8	6		1	1	Both
chr9	CNVR_9_10	chr9:4239500-4283263	4239500	4283263	43763	10	7	2	1		Both

Additional file 4.2 The gene name, refgene annotation and gene symbol of genes overlapping or within 10Mb of CNVRs significantly associated with tick resistance in South African Nguni cattle.

CNVR_ID	CNVR	REFGENE ANNOTATION	GENE SYMBOL	GENE NAME
CNVR_1_4	chr1:4957717-5034441		NOT_FOUND	
CNVR_2_32	chr2:53894600-53951533	NM_001191308	KYNU	Kynureninase
CNVR_6_25	chr6:20297397-20350438		NOT_FOUND	
CNVR_6_5	chr6:10105101-10199636		NOT_FOUND	
CNVR_6_9	chr6:10760779-10838635		NOT_FOUND	
CNVR_7_54	chr7:75305297-75370366	NM_001192760	GABRB2	Gamma-aminobutyric acid type A receptor beta2 subunit
CNVR_8_11	chr8:34795275-34898163		NOT_FOUND	
CNVR_9_48	chr9:94991477-95065382		NOT_FOUND	
CNVR_12_53	chr12:90704572-90778028	NM_001076029	TFDP1	Transcription factor Dp-1
CNVR_12_53	chr12:90704572-90778028	NM_001098003	TMCO3	TMCO3 protein
CNVR_12_53	chr12:90704572-90778028	NM_001206140	ATP4B	Uncharacterized protein
CNVR_12_53	chr12:90704572-90778028	NM_174173	GRK1	Rhodopsin kinase
CNVR_15_4	chr15:5474020-5537822		NOT_FOUND	
CNVR_17_39	chr17:74292319-74393620	NM_001038163	TUBA3E	Tubulin alpha-3e
CNVR_17_39	chr17:74292319-74393620	NM_001046281	AIFM3	Apoptosis inducing factor, mitochondria associated 3
CNVR_17_39	chr17:74292319-74393620	NM_001075185	PRODH	Proline dehydrogenase 1, mitochondrial
CNVR_17_39	chr17:74292319-74393620	NM_001098465	THAP7	THAP domain containing 7

CNVR_17_39	chr17:74292319-74393620	NM_001099204	MZT2	Mitotic-spindle organizing protein 2
CNVR_17_39	chr17:74292319-74393620	NM_001192042	SLC7A4	Uncharacterized protein
CNVR_17_39	chr17:74292319-74393620	NM_001192111	LZTR1	Leucine zipper like transcription regulator 1
CNVR_20_21	chr20:45052283-45266553		NOT_FOUND	
CNVR_21_46	chr21:71025601-71109676	NM_001075573	C14orf79	uncharacterized protein
CNVR_22_28	chr22:60736089-60960603	NM_001076402	CHCHD6	MICOS complex subunit MIC25
CNVR_22_8	chr22:24078956-24161191		NOT_FOUND	
CNVR_24_19	chr24:28154039-28196203		NOT_FOUND	
CNVR_29_34	chr29:50202589-50240781	NM_001001441	TNNT3	Troponin T3, fast skeletal type
CNVR_29_34	chr29:50202589-50240781	NM_001075374	LSP1	Lymphocyte-specific protein 1