

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

**Genome-wide copy number variation in sheep: detection
and utility as a genetic marker for quantitative traits, with
reference to gastrointestinal nematodiasis**

Thesis presented in partial fulfilment of
the requirements for the degree of

Doctor of Philosophy

in

Animal science

At Massey University, Palmerston North,
New Zealand

Juncong Yan

2018

Abstract

Gastrointestinal nematodes are perhaps the most important parasites of domestic sheep world-wide. Genetic selection for nematode resistance in domestic sheep is being promoted in many countries including New Zealand. There are several strategies to identify genetic markers associated with quantitative traits. Single nucleotide polymorphism (SNP)-based strategies have been widely used in animal breeding. However, SNP cannot explain all the genetic variation for a particular trait. A new kind of variation, copy number variation (CNV) has been identified as contributing to genetic variation in production and disease traits. Compared with other domestic animals, CNV in sheep is poorly investigated. The primary objective of this thesis was to explore the utility of genome-wide CNV as a genetic marker for the analysis of quantitative traits in sheep. Five different studies were undertaken to fulfill the objective. The first two studies used 50 K SNP BeadChip genotype data and next generation sequencing (NGS) data to detect CNV. Extensive CNV differences were evident between breeds as well as detection algorithms. NGS-based detection resulted in better CNV resolution than that by SNP. Subsequently, a genome-wide association study (with a small sample size) using CNV detected from a high density (HD) SNP genotype data identified four CNV regions to be significantly associated with a couple of traits pertaining to gastrointestinal nematodiasis in Romney sheep, while no significant SNP associations were found. Somatic mosaicism of CNV, influenced by age (high in foetuses, compared to adults), individuals, detection algorithm and type of tissue analysed, was also evident in separate study. The final study detected CNV differences and SNP based selection signatures in two Romney lines selected for gastrointestinal nematode resistance or resilience. Several significant SNPs and line-specific CNV regions were identified. However, only one SNP overlapped to a CNV region, indicating that SNP-based selection signatures and CNV could represent different aspects of sheep immunogenetics. Overall, CNV could be a potential

genetic marker, albeit with methods for detection and validation needing to be refined. The conclusions from this thesis expand our understanding of CNV in sheep and its potential application prospects for genetic breeding of sheep in the future.

Acknowledgements

First and foremost, I would like to acknowledge my supervisors, Dr Rao Dukkipati, Prof Hugh Blair and Associate Prof Patrick Biggs. Rao, you like a father, endlessly supported and encouraged me during my PhD study. You had no complaints about my poor English and tolerated that for four years. Your passion, patience and meticulousness for science let me know what a scientist should be. Hugh, you like a grandfather, did your best to provide funds for my study. I cannot count how many times you have helped me. Your academic brilliance illuminated my academic path like a beacon and you have been my model in heart. Patrick, you opened the gate of bioinformatics for me and show me a total new interesting world.

Furthermore, I would like to acknowledge Prof Dorian Garrick, Dr Keren Dittmer, Dr Sarah Pain, Dr Andrew Greer, Mr Joseph Hamie for original data support and Dr Kristene Gedye, Rosemary Heathcott for technical support. I would also like to acknowledge the overseas laboratory group, Key Laboratory of Genetics Breeding and Reproduction of Grass Feeding Livestock; Key Laboratory of Animal Biotechnology of Xinjiang, Xinjiang Academy of Animal Science, in China. The data you supported is very important for my research. I also thank my all colleagues in room 3.06. Your support and harmonious environment in office has been crucial for my successful completion. Your help let me quickly adapt to New Zealand's life.

I would like to acknowledge research funding from different sources within Massey University as well as the Massey-Lincoln and Agricultural Industry Trust, for this project. Besides, I very much appreciate the financial support from the Massey University Doctoral scholarship, which really helped me a lot and let me focus on study. I also appreciate the financial support from the IVABS postgraduate fund which helped me to attend my first international academic conference in Australia.

I would like to acknowledge NeSi's (New Zealand eScience Infrastructure) support for high performance computing for the NGS data analysis.

Finally, I would like to thank my family. Dad, Mum and my brother, your encouragement is the power for me to climb the mountain of academic success. Your never-ending support has helped me succeed. Thank you for always believing in me.

Preface

I have undertaken this thesis in the form of publishable experimental chapters using a format of thesis by publication. The current status and publication outlet are described in the following list.

Chapter 1: Literature review

Chapter 2: Genome-wide detection of autosomal copy number variants in several sheep breeds using Illumina OvineSNP50 BeadChips.

Juncong Yan, Hugh T. Blair, Mingjun Liu, Wenrong Li, Sangang He, Lei Chen, Keren E. Dittmer, Dorian J. Garrick, Patrick J. Biggs, Venkata S.R. Dukkipati*
Published in Small Ruminant Research, 2017, 155: 24-32.
(doi:[10.1016/j.smallrumres.2017.08.022](https://doi.org/10.1016/j.smallrumres.2017.08.022))

All molecular work, data analysis, interpretation of results and manuscript write-up were completed by Juncong Yan. The original SNP data was provided by Mingjun Liu, Wenrong Li, Sangang He, Lei Chen, Keren E. Dittmer and Dorian J. Garrick. The manuscript was checked by supervisors, Venkata S.R. Dukkipati, Hugh T. Blair and Patrick J. Biggs.

Chapter 3: Detection of copy number variation in sheep by whole genome sequencing

Juncong Yan, Hugh T. Blair, Keren E. Dittmer, Patrick J. Biggs, Venkata S.R. Dukkipati
To be submitted to BMC Genomics

All molecular work, data analysis, interpretation of results and manuscript write-up were completed by Juncong Yan. The original NGS data was provided by Keren E. Dittmer. The manuscript was checked by supervisors, Venkata S.R. Dukkipati, Hugh T. Blair and Patrick J. Biggs.

Chapter 4: Genome-wide association study in sheep selectively bred for resistance or resilience to gastrointestinal nematodes

Juncong Yan, Hugh T. Blair, Andrew Greer, Joseph Hamie, Patrick Biggs, Venkata S.R. Dukkipati

To be submitted to Journal of Animal Breeding and Genetics

All molecular work, data analysis, interpretation of results and manuscript write-up were completed by Juncong Yan. The original phenotype data was provided by Andrew Greer and Joseph Hamie. Patrick J. Biggs provided bioinformatics support for gene annotation. The manuscript was checked by supervisors, Venkata S.R. Dukkipati, Hugh T. Blair and Patrick J. Biggs.

Chapter 5: Somatic mosaicism of copy number variation in sheep using Ovine Infinium® HD SNP BeadChip

Juncong Yan, Hugh T. Blair, Patrick J. Biggs, Sarah J. Pain, Venkata S.R. Dukkipati

To be submitted to PLOS ONE

All molecular work, data analysis, interpretation of results and manuscript write-up were completed by Juncong Yan. The original tissue samples were provided by Sarah J. Pain. Patrick J. Biggs provided bioinformatics support for gene annotation. The manuscript was checked by supervisors, Venkata S.R. Dukkipati, Hugh T. Blair and Patrick J. Biggs.

Chapter 6: Detection of copy number variation and genome-wide positive selection signatures using Ovine Infinium® HD SNP BeadChip in two Romney lines, selected for resistance or resilience to gastrointestinal nematodes

Juncong Yan, Hugh T. Blair, Andrew Greer, Joseph Hamie, Patrick J. Biggs, Venkata S.R. Dukkipati

To be submitted to BMC Genomics

Part of this chapter was presented as a poster at the 22nd Association for the Advancement of Animal Breeding and Genetics conference held in Townsville, QLD, Australia, 2-5 Jul 2017.

Juncong Yan, Venkata S.R. Dukkipati, Hugh T. Blair, Patrick Biggs, Joseph Hamie and Andrew Greer (2017). A genome-wide scan of positive selection signature using Ovine Infinium® HD SNP BeadChip in two Romney lines, selected for resistance or resilience to nematodes. In Proceedings of the Association for the Advancement of Animal Breeding and Genetics Vol. 22 (pp.1).

All molecular work, data analysis, interpretation of results and manuscript write-up were completed by Juncong Yan. The original phenotype data was provided by Andrew Greer and Joseph Hamie. The manuscript was checked by supervisors, Venkata S.R. Dukkipati, Hugh T. Blair and Patrick J. Biggs.

Chapter 7: General discussion

Table of Contents

ABSTRACT.....	II
ACKNOWLEDGEMENTS.....	IV
PREFACE	VI
TABLE OF CONTENTS	IX
LIST OF FIGURES	XIV
LIST OF TABLES	XVI
COMMON ABBREVIATIONS.....	XVIII
CHAPTER 1 LITERATURE REVIEW	1
1.1 Introduction	1
1.2 Genetic markers.....	2
1.2.1 Introduction.....	2
1.2.2 The history of genetic markers.....	3
1.2.3 Summary.....	6
1.3 Copy number variation (CNV).....	6
1.3.1 Introduction.....	6
1.3.2 Function of CNV.....	7
1.3.3 Molecular mechanism of formation of CNV.....	9
1.3.4 Methods for prediction of CNV	9
1.3.5 Current research on CNV.....	12
1.3.6 Summary.....	24
1.4 Details of CNV detection platforms	25
1.4.1 SNP microarray	25
1.4.2 NGS.....	30
1.4.3 Summary.....	33
1.5 Selection signatures	36
1.6 Somatic mosaicism of CNV	42
1.7 Overall summary and thesis objectives	43
CHAPTER 2 GENOME-WIDE DETECTION OF AUTOSOMAL COPY NUMBER VARIANTS IN SEVERAL SHEEP BREEDS USING ILLUMINA OVINESNP50 BEADCHIPS 45	
2.1 Abstract	46
2.2 Introduction	46

2.3	Materials and methods.....	48
2.3.1	Materials.....	48
2.3.2	Quality control	49
2.3.3	CNV detection.....	50
2.3.4	Derivation of CNVR and construction of CNVR map.....	51
2.3.5	Gene content of CNVR and functional annotation.....	52
2.3.6	CNV validation by qPCR.....	52
2.3.7	Comparison of CNV among different breeds.....	53
2.4	Results	53
2.4.1	Genome-wide CNV detection	53
2.4.2	Gene content of CNVR and functional annotation of genes	57
2.4.3	CNV validation by quantitative polymerase chain reaction (qPCR).....	58
2.4.4	Comparison of CNVR among different breeds	59
2.5	Discussion	63
2.5.1	Genome-wide CNV detection	63
2.5.2	Gene content of CNVR and functional annotation of genes	64
2.5.3	CNV validation by qPCR.....	65
2.5.4	Comparison of CNVs among different breeds	65
2.5.5	Comparison of this study with previous studies.....	66
2.6	Conclusion.....	66
2.7	Authors' contributions	68
2.8	Acknowledgements	68
2.9	Additional files	68
CHAPTER 3 DETECTION OF COPY NUMBER VARIATION IN SHEEP BY WHOLE GENOME SEQUENCING.....		69
3.1	Abstract	70
3.1.1	Background.....	70
3.1.2	Results.....	70
3.1.3	Conclusion	70
3.1.4	Keywords	71
3.2	Introduction	71
3.3	Materials and Methods	73
3.3.1	Sample collection and sequencing	73
3.3.2	Data preparation.....	73
3.3.3	CNV calling, derivation of CNV region (CNVR) and construction of CNVR map	74
3.3.4	qPCR validation	75
3.3.5	Gene annotation	76
3.3.6	Pedigree comparison	77
3.4	Results	77
3.4.1	Mapping statistics and CNV detection	77
3.4.2	qPCR validation	79
3.4.3	Gene annotation	82
3.4.4	Pedigree comparison	83
3.5	Discussion	83
3.5.1	Mapping statistics and CNV detection	83
3.5.2	qPCR validation	88
3.5.3	Gene annotation	88
3.5.4	Comparison with previous Sheep CNV studies	88
3.5.5	Pedigree comparison	91

3.6	Conclusions	92
3.7	Acknowledgments	92

CHAPTER 4 GENOME-WIDE ASSOCIATION STUDY FOR THE ASSOCIATIONS BETWEEN CNVS AND RESISTANCE OR RESILIENCE TO SHEEP GASTROINTESTINAL NEMATODES 93

4.1	Abstract	94
4.1.1	Background	94
4.1.2	Result	94
4.1.3	Conclusion	94
4.1.4	Keywords: sheep, GWAS, SNP, CNV, nematodes	95
4.2	Introduction	95
4.3	Materials and methods.....	96
4.3.1	Ethics statement	96
4.3.2	Tissue sampling, genotyping and phenotypes	97
4.3.3	Quality control and CNV detection.....	99
4.3.4	qPCR validation	100
4.3.5	Genome-wide association study (GWAS).....	101
4.3.6	Gene annotation	103
4.4	Results	103
4.4.1	Quality control	103
4.4.2	CNV detection.....	103
4.4.3	qPCR validation	103
4.4.4	Genome-wide association	104
4.4.5	Gene annotation	105
4.5	Discussion	115
4.6	Conclusion.....	116
4.7	Additional files	116

CHAPTER 5 SOMATIC MOSAICISM OF COPY NUMBER VARIATION IN SHEEP USING OVINE INFINIUM® HD SNP BEADCHIP 117

5.1	Abstract	118
5.1.1	Background	118
5.1.2	Results.....	118
5.1.3	Conclusion	118
5.1.4	Keywords: Somatic mosaicism, CNV, SNP, sheep	119
5.2	Introduction	119
5.3	Methods.....	120
5.3.1	Sample collection and genotyping	120
5.3.2	Quality control	121
5.3.3	CNV detection.....	121
5.3.4	Estimation of CNV mosaicism.....	124
5.3.5	qPCR validation	125
5.3.6	Gene annotation	127
5.4	Results	128
5.4.1	CNV detection and CNVR formation	128
5.4.2	CNVR differences between adults and fetuses	128
5.4.3	CNVR differences between tissues	132
5.4.4	Within-individual SM of CNV	134

5.4.5	qPCR validation	139
5.4.6	Gene annotation	139
5.5	Discussion	139
5.6	Conclusion.....	142
5.7	Additional files	142
CHAPTER 6 DETECTION OF COPY NUMBER VARIATION AND GENOME-WIDE POSITIVE SELECTION SIGNATURES USING OVINE INFINIUM® HD SNP BEADCHIP IN TWO ROMNEY LINES, SELECTED FOR RESISTANCE OR RESILIENCE TO GASTROINTESTINAL NEMATODES		143
6.1	Abstract	144
6.1.1	Background	144
6.1.2	Result	144
6.1.3	Conclusion	145
6.1.4	Keywords: sheep, positive selection signature, nematodes	145
6.2	Background	145
6.3	Materials and Methods	147
6.3.1	Ethics statement	147
6.3.2	Sample collection and background of lines.....	147
6.3.3	Quality control and data preparation for selective signature	147
6.3.4	CNV detection and validation	148
6.3.5	Gene annotation	150
6.3.6	Detection of selection signatures using SNP haplotypes.....	151
6.4	Results	158
6.4.1	Quality control	158
6.4.2	CNVs and CNVRs	158
6.4.3	SNP-based selection signatures.....	159
6.4.4	qPCR validation	163
6.4.5	Gene annotation	177
6.5	Discussion	188
6.6	Conclusion.....	193
6.7	Additional files	193
CHAPTER 7 GENERAL DISCUSSION		194
7.1	Thesis objective	194
7.2	Summary of results.....	194
7.3	Discussion of results.....	196
7.3.1	Genotyping platform	196
7.3.2	CNV detection algorithms.....	199
7.3.3	CNV validation by qPCR.....	201
7.3.4	CNV as a genetic marker	203
7.3.5	GWAS and selection signatures	204
7.4	Suggestions for further research	205
7.5	Overall conclusion.....	206
REFERENCES		208

APPENDIX.....	228
3.1 Code for NGS data mapping.....	228
3.2 R code for CNVR plot.....	230
3.3 Code for exploring sequencing depth in five samples	231
3.4 Code for 10kb bin.....	232
3.5 Code for violin plot.....	233
4.1 Custom written script in Perl and SQL for gene annotation	234
5.1 R code for CNVR	243
5.2 R code for CNVR in each tissue.....	244
6.1 Code for fastPHASE_v1.4.....	251
6.2 Code for selection signature regions.....	256
6.3 iHS plots in the two lines.....	259
6.4 R code for EHH and EHHS	268

List of Figures

Figure 1.1 An explanation of copy number variation.	7
Figure 1.2 Deletion of gene causes change of gene dosage (Feuk et al. 2006).	8
Figure 1.3 Deletion of upstream gene causes change of gene dosage (Feuk et al. 2006).	9
Figure 1.4 Duplication scenarios and their influence on expression (Henrichsen et al. 2009a).	10
Figure 1.5 Non-allelic homologous recombination.	11
Figure 1.6 Copy number of α -Globin (HBA) and different clinical phenotypes.	14
Figure 1.7 Illumina bead chip workflow (Illumina 2012).	26
Figure 1.8 The principle of base detection.	27
Figure 1.9 The genotyping results of one SNP point in a group of samples from the current study (Chapter 2).	28
Figure 1.10 CNV work flow.	29
Figure 1.11 Illumina flow cell (Frank-Vinken-Institute 2013).	31
Figure 1.12 Illumina NGS overview 1 (Illumina 2010).	32
Figure 1.13 Illumina NGS overview 2 (Illumina 2010).	32
Figure 1.14 Illumina NGS overview 3 (Illumina 2010)	32
Figure 1.15 Five approaches to detect CNVs from NGS short reads.	35
Figure 1.16 The formation of Selective sweep (Biswas and Akey 2006).	36
Figure 2.1 Chromosomal distribution of copy number variant regions (CNVR) detected by three algorithms.	55
Figure 2.2 Venn plot of CNVR detected by three algorithms.	56
Figure 2.3 Frequency distribution of the size range of copy number variant regions (CNVR) detected by the three algorithms.	56
Figure 2.4 Venn plot of genes found in copy number variant regions (CNVR) detected by the three algorithms.	58
Figure 2.5 Venn plot of copy number variant regions (CNVR) detected among the five breeds of sheep.	60
Figure 2.6 Venn plot of genes found in copy number variant regions (CNVR) detected among the five breeds.	61
Figure 2.7 Principal component analysis plot (PC1 and PC2) showing population stratification ...	62
Figure 3.1 Pedigree of five sheep that were subject of the study.	76
Figure 3.2 Distribution of sequencing depth-size (50 kb) bins.	78
Figure 3.3 Violin plots of sequencing depth (at whole genome level) in five individuals ...	78
Figure 3.4 Chromosomal distribution of copy number variant regions (CNVR) detected in five Romney sheep, using whole genome sequencing data.	80
Figure 3.5 Plot showing relationship between sequencing depth and number of CNVs detected in the five individuals, in 50 kb bins across the chromosomal region, ch13:46100000-5110000.	81
Figure 3.6 CNV comparison between five Romney sheep.	82
Figure 3.7 Frequency distribution of the size range of copy number variant regions (CNVR) detected in five Romney sheep, using NGS.	84
Figure 3.8 Inheritance of CNV in individual 828-05-1.	85
Figure 3.9 Inheritance of CNV in individual 828-05-3.	86
Figure 3.10 Comparison of CNVs inherited by the two half-sibs, exclusively from their sire.	87
Figure 3.11 Overlap of the CNVRs detected in the current study (based on NGS) with those from previous studies that employed SNP microarrays.	90
Figure 4.1 Principal component analysis revealing population stratification.	107
Figure 4.2 Principal component analysis - eigenvalue plot ...	108

Figure 4.3 Q-Q plot for before PCA correction (left) and after PCA correction (right) of SNP's GWAS for live weight.	109
Figure 4.4 Q-Q plot for before PCA correction (left) and after PCA correction (right) of SNP's GWAS for immunity.	110
Figure 4.5 Q-Q plot for before PCA correction (left) and after PCA correction (right) of SNP's GWAS for FEC.	111
Figure 4.6 Basic Allelic Test for association by chi-square allelic test for live weight.	112
Figure 4.7 Basic Allelic Test for association by chi-square allelic test for immunity.	113
Figure 4.8 Basic Allelic Test for association by chi-square allelic test for FEC.	114
Figure 5.1 Distribution map of CNVRs detected by both PennCNV and cnvPartition using 36 tissues from 12 sheep (adults and fetuses both).	129
Figure 5.2 UpsetR plot showing overlap of CNVs in six adult sheep.	130
Figure 5.3 UpsetR plot showing overlap of CNVs in six foetuses.	131
Figure 5.4 Venn plot of CNVR detected in tissues from adults and foetuses.	132
Figure 5.5 Distribution of difference types of CNVRs in individual chromosomes of adults and foetuses.	133
Figure 5.6 Distribution map of CNVRs in different tissues from chromosome 1 to chromosome 10.	135
Figure 5.7 Distribution map of CNVRs in different tissues from chromosome 11 to chromosome 20.	136
Figure 5.8 Distribution map of CNVRs in different tissues from chromosome 21 to chromosome 26.	137
Figure 5.9 UpsetR plot showing overlap of CNVRs across seven tissues made by UPSetR (Conway et al. 2017).	138
Figure 5.10 Relationship between CNV mosaicism and number of tissues investigated.	142
Figure 6.1 Schematic view of 11 SNPs in eight aligned chromosomes.	153
Figure 6.2 Illustration of iHH (shaded part) (Gautier et al. 2017).	154
Figure 6.3 illustrate of iES (shadow part) (Gautier et al. 2017).	157
Figure 6.4 Chromosomal distribution of copy number variant regions (CNVR) detected in gastrointestinal nematode resistant (white bars) and resilient (grey bars) lines of Romney sheep.	160
Figure 6.5 Venn plot of copy number variant regions (CNVR) detected in gastrointestinal nematode resistant (green circle) and resilient (orange) lines of Romney sheep.	161
Figure 6.6 Plot showing the differences in iHS, the within-line allele-specific extended haplotype homozygosity (EHH) test statistic, with regard to single nucleotide polymorphism (SNP) loci located on chromosome 2 between two Romney sheep lines (gastrointestinal nematode resistant and resilient).	162
Figure 6.7 the difference of positive selection signature detected between XP-EHH & Rsb.	176
Figure 6.8 Ontology and pathway analysis (molecular function) of genes harbouring the significant SNPs detected by EHH test.	178
Figure 6.9 Ontology and pathway analysis (biological process) of genes harbouring the significant SNPs detected by EHH test.	179
Figure 6.10 Ontology and pathway analysis (pathway) of genes harbouring the significant SNPs detected by EHH test.	180
Figure 6.11 Plot showing the distribution of $\log(p)$ values for Rsb and XP-EHH, between-line site-specific extended haplotype homozygosity (EHHS) test statistics, with regard to single nucleotide polymorphism (SNP) loci located on chromosome 13, in two Romney sheep lines (gastrointestinal nematode resistant and resilient).	192
Figure 6.12 Plot showing correlation between XPEHH and Rsb statistics for markers located on chromosome 13.	192

List of Tables

Table 1.1 Studies on CNV detection in domestic animals.	15
Table 1.2 Summary of popular algorithms for CNV detection using Illumina SNP microarrays.	30
Table 1.3 Summary of current algorithms for CNV detection based on NGS.	37
Table 1.4 Summary of selection tests used in published studies pertaining to selection signatures in sheep.	40
Table 1.5 Summary of studies pertaining to selection signatures in sheep.	41
Table 2.1 Details of sheep that were genotyped using ovine 50k SNP microarray.	50
Table 2.2 Number of genes and proteins found in copy number variant regions (CNVR) detected by three algorithms.	58
Table 2.3 Results of qPCR validation of copy number variants (CNV) detected by the three algorithms.	59
Table 2.4 Summary of copy number variant regions (CNVR) detected and their gene content in the five breeds of sheep.	60
Table 2.5 Pairwise population fixation index (F_{ST}) values for the five sheep breeds.	62
Table 2.6 Comparison of number and size of copy number variant regions (CNVR) detected in this study with those from previous studies.	67
Table 3.1 Identification and sex of Romney sheep and summary NGS data	74
Table 3.2 Hypothetical copy numbers of the reference and their thresholds (based on qPCR) for copy number evaluation.	76
Table 3.3 Summary of the copy number variants (CNVs) detected in five Romney sheep	79
Table 3.4 Comparison of the number and size of copy number variant regions (CNVR) detected in this study with those from previous studies in sheep.	91
Table 4.1 Hypothetical copy numbers of the reference and their thresholds (based on qPCR) for copy number evaluation.	101
Table 4.2 Results of qPCR validation of four randomly selected CNVs.	103
Table 4.3 Significant ($EMP2 < 0.05$) CNVRs detected by GWAS for live weight and FEC and gene annotation	106
Table 5.1 Details of tissue samples analysed.	123
Table 5.2 Summary of CNV mosaicism detected in foetal and adult sheep, using cnvPartition and PennCNV alone, or in combination.	126
Table 5.3 Hypothetical copy numbers in the reference sample and their thresholds (based on qPCR) for copy number evaluation	127
Table 5.4 Summary of CNVs detected by cnvPartition, PennCNV and their combination.	128
Table 5.5 CNVR differences between adults and foetuses.	132
Table 5.6 Number of CNVRs detected in individual tissues across animals	134
Table 6.1 Hypothetical copy numbers of the reference and their thresholds (based on qPCR) for copy number evaluation.	150
Table 6.2 Results of within-line allele-specific EHH test in gastrointestinal nematode resilient and resistant lines of Romney sheep: chromosome-wise number of SNPs evincing signatures of selection	163
Table 6.3 List of SNPs detected by iHS and $PiHS$, found to significant ($P < 0.0001$) positive selection signatures in the resilient line.	164
Table 6.4 List of SNPs detected by iHS and $PiHS$, found to evince significant ($P < 0.0001$) positive selection signatures in the resistant line.	167
Table 6.5 Results of qPCR validation of 4 CNVs.	169

Table 6.6 Results of between-line EHHS test (using two different algorithms, XP-EHH and Rsb) in gastrointestinal nematode resilient and resistant lines of Romney sheep: chromosome-wise number of SNPs evincing signatures of selection.	169
Table 6.7 Significant ($p < 0.0001$) SNPs detected using Rsb.	170
Table 6.8 Significant ($p < 0.0001$) SNPs detected by XPEHH.	173
Table 6.9 List of SNPs detected by both between-line EHHS algorithms, XP-EHH and Rsb, found to evince significant ($P < 0.0001$) positive selection signatures.	176
Table 6.10 Selection signature regions on chromosome 13.	177
Table 6.11 List of genes located within the unique CNVRs, those not common between the two family lines of sheep.	181
Table 6.12 List of genes located close to the significant ($P < 0.0001$) SNPs detected by EHH test in the gastrointestinal nematode resistant and resilient lines of Romney sheep.	187

Common abbreviations

aCGH	array comparative genomic hybridization
AFLP	amplified fragment length polymorphism
AMD	age-related macular degeneration
AS	de novo assembly of a genome
BAF	B allele frequency
BLUP	best linear unbiased prediction
BP	biological process
CC	cellular component
CFH	complement factor H gene
CIITA	class II Major Histocompatibility Complex transactivator
CN-LOH	mosaic copy neutral loss of heterozygosity
CNV	copy number variation
CNVR	copy number variation region
CPU	central processing unit
dH ₂ O	deionised distilled water
DLRS	derivative log ratio spread
DNA	deoxyribonucleic acid
dNTP	deoxy-ribonucleoside triphosphate
dsDNA	double-stranded DNA
EHH	extended Haplotype Homozygosity
ELISA	enzyme-Linked ImmunoSorbent Assay
EMP2	empirical p-value, corrected for all tests
EPG	eggs per gram
FDR	false discovery rate
FEC	faecal egg count
FLK	an extension of Lewontin and Krakauer (LK) test, based on population's kinship (F) matrix
F _{st}	fixation index
GO	gene ontology
GWAS	genome-wide association study
hapFLK	haplotype structure accounted FLK
HGP	human Genome Project
HIV	human immunodeficiency virus
HMMs	hidden Markov models
IBD	identity by descent
iHH	integrated allele-specific EHH
iHS	integrated haplotype Score
IQRs	inter-quartile range
ISGC	International Sheep Genomics Consortium
KEGG	Kyoto Encyclopedia of Genes and Genomes
LD	linkage disequilibrium

LRR	log R ratio
LW	live weights
MF	molecular function
MHC II	major histocompatibility complex II
MZ	monozygotic twins
NAHR	non-allelic homologous recombinations
NeSi	New Zealand eScience infrastructure
NGS	next generation sequencing
PCA	principal components analysis
PCR	polymerase chain reaction
PEM	paired-end mapping
qPCR	quantitative polymerase chain reaction
Q-Q	quantile-quantile
QTL	quantitative trait loci
RAM	random-access memory
RAPD	random Amplified Polymorphic DNA
RD	read depth
REHH	relative EHH
RFLP	restriction fragment length polymorphism
RNA	ribonucleic acid
Rsb	across Population EHH
SLE	systemic lupus erythematosus
SM	somatic mosaicism
SNP	single nucleotide polymorphism
SR	split read
SRFA	selective restriction fragment amplification
SSRs	simple sequence repeats
SVS	Golden Helix SNP & Variation Suite
TMB	tetramethyl benzidine
VNTR	variable number of tandem repeats
WF	Wave factor
XP-EHH	across Population EHH
ZHp	Z-transformed Heterozygosity Value