Genomic and transcriptomic approaches to investigating candidate genes associated with tick resistance in cattle

Kamonchanok Chooyoung

A thesis submitted for the degree of Doctor of Philosophy
School of Biodiversity, One Health & Veterinary Medicine
College of Medical, Veterinary & Life Sciences
University of Glasgow

February 2024

General Abstract

The cattle tick, *Rhipicephalus (Boophilus) microplus*, is able to transmit tick-borne diseases in cattle, which causes major production and welfare implications, particularly in tropical and subtropical areas. Problems associated with cattle ticks, such as decreased production yield, mortality, and costs of treatment are economically significant. Although resistance to ticks in cattle is moderately heritable (0.4), there are no genes or biomarkers commercially available that could be used for genomic selection. The aim of this study was to identify genes and pathways associated with tick resistance in cattle using a meta-analysis of previous studies, combined with single nucleotide polymorphism (SNP) based genotyping and long-read sequence-based transcriptomics.

Previous gene expression (GEX) and genome-wide association studies (GWAS) have identified a large number of potential genes or quantitative trait loci (QTLs) associated with tick resistance in cattle. However, different studies have found different sets of genes, using different experimental procedures and platforms. In Chapter 2, I aimed to integrate results from differential expression genes (DEGs) and QTLs from GWAS associated with host resistance to infestation with *R. microplus* in order to generate a list of genes which showed supporting evidence from multiple sources. I identified a set of 37 genes that were found in multiple studies, based on blood or skin gene expression and GWAS, including three transcription factors, 12 genes associated with immune function, nine with the extracellular matrix, six structural genes, and 13 other biological processes. This subset of genes was then used to investigate whether there was sufficient variability that they could potentially be used as targets of selection (Chapter 3).

To develop animals breed to be resilient to foreign invasion (pathogens and ectoparasites), it may be necessary to focus not only on genes associated with immune function but also on other types of pathways. The aim of Chapter 3 was to determine whether any of the genes shortlisted from Chapter 2 showed significant differences in genotypes between Scottish breed groups, including British (n=14), European (Continental) (n=10), and Hill (n=10) cattle. Focusing on breeds within Scotland provided a conservative estimate of variability of these genes. The genotyping was investigated from DNA extracted from spleen tissue samples from individuals collected from a single abattoir, by using the GeneSeek Genomic Profile (GGP) Bovine 100K SNP chip. A total of 88 SNPs were identified from the list of candidate genes, with 14 spread across six genes (HOXD1, SATB2, GIMAP7, ITGA11, PLA2G7, and PRKG1) presenting significant differences in genotype frequencies between the breed groups. Although most of the SNPs were located in

introns that were not close enough to exons to expect linkage to mutations under selection, only a single amino acid changing variant was identified in PLA2G7-1 (missense mutation), with the other 13 being either synonymous changes within an exon (ITGA11) or located in introns. This list of SNPs was then used to classify gene expression patterns from skin and spleen samples from the same individuals in Chapter 4.

All of the previous studies identifying genes showing differential expression associated with resistance to ticks in cattle have focused on skin or blood samples in Chapter 2; however, immune functions are more pronounced in tissues such as spleen or lymph nodes. Chapter 4 was a pilot study aimed at identifying genome-wide patterns of gene expression (transcriptome profiles) in both skin and spleen tissue samples using GridION from Oxford Nanopore Technology (ONT). The data were analysed using a weighted gene co-expression network analysis (WGCNA) correlated with variation in the 14 significant SNPs from Chapter 3. In skin datasets, seven modules showed significant correlations with at least one of the SNPs in the skin dataset, which contained 2,297 genes, whereas there were ten modules in the spleen dataset (3,265 genes). Overall, I found different sets of co-expressed gene modules associated with variation in the SNPs in the two tissue types, with a wider range of modules associated with the spleen but also more pathways directly related to immune function. Nevertheless, both tissues identified multiple biological pathways and interactions between pathways correlated with genotype variation at the focal genes. From the large set of genes identified in these pathways, I found three genes (FN1, ATP9A, and ECM1) from the skin dataset and five (CR2, RHOT1, SRGN, GIMAP7, and LAPTM5) from the spleen dataset that overlapped with the list of 37 candidate genes identified in Chapter 2. Only one of these (GIMAP7) also showed significant variation on the SNP genotyping panel, suggesting that it could make a useful candidate to consider further. Here I provided preliminary evidence about candidate genes connecting the genotype and phenotype (gene expression) across different tissues (skin and spleen) and developed predictions about potential biomarkers. Further analyses to identify tissue-specific novel isoforms, alternative splicing and their respective biological functions in cattle could provide additional support for the utility of the genes identified for marker-based selection.

The integration of the genes or QTLs from multiple resources, the SNP genotype variation, and the transcriptome profiling in this study offer a potential panel of biomarkers which could be evaluated in pathogen and ectoparasite resistance in cattle studies and develop improvements for the future selection and breeding programs in cattle. The understanding of the genetic basis for resistance in cattle offers preliminary evidence that

could impact beef and dairy production efficiency and product quality, especially in tropical and subtropical regions.

Table of contents

List of Tables

List of Figures

Acknowledgements

First of all, I would like to express my deepest appreciation to Professor Barbara Mable, Professor Nicholas Jonsson, and Dr. Lorenzo Viora, all of my supervisors, for their consistent encouragement, insightful direction, and valuable assistance throughout my doctoral studies. During every step of my PhD journey, their insightful suggestions, numerous comments, and constructive criticism were invaluable. I can confidently say that my supervisors consistently provided support whenever I encountered difficulties. In addition, I would like to express sincere thanks to Elizabeth Kilbride, Ryan Carter, Maria Capstick, Calum Stewart, and all staff in the OHRBID lab and roof labs for their assistance in the process of laboratory work and bioinformatics

Secondly, I express gratitude to the Thai Government for providing me with a scholarship that enabled me to pursue Ph.D. studies. Additionally, I would like to express sincere thanks to all the staff at the Office of Educational Affairs, U.K., for their assistance and support in the process of education and living in the U.K.

Thirdly, I would like to thank all the members of the School of Biodiversity, One Health, and Veterinary Medicine in the College of Medical, Veterinary, and Life Sciences at the University of Glasgow, especially my colleagues who are students of Professor Barbara Mable and Professor Nicholas Jonsson. I am also grateful to all my friends in Glasgow for enriching my life with happiness. I am thankful to everyone at my badminton club for improving my health and making my life joyful.

Finally, and of utmost significance, I would like to express my heartfelt thanks to my family members, including my parents and my siblings. The unfailing love, support, and sacrifice of my loved ones have served as my main source of motivation throughout my studies. This thesis is dedicated to the individuals mentioned.

Author's declaration

I, Kamonchanok Chooyoung, declare that the work documented in this thesis is entirely my own, except where explicitly stated otherwise. I affirm that it has not been submitted as part of a degree program elsewhere. While a substantial portion of this thesis has been collaboratively authored, I acknowledge that my individual contributions to specific chapters are enumerated below.

Chapter 2: "Integration of gene expression and genome-wide association to identify genes associated with resistance to infestation with cattle tick (*Rhipicephalus microplus*)" Principal supervisor, advisory role, contribution to experimental design and editing of the final draft were provided by Professor Barbara Mable, Professor Nicholas Jonsson. All experimental procedures, data analysis, and thesis writing and final draft were completed by Kamonchanok Chooyoung.

Chapter 3: "Genotypic variation in Scottish cattle for candidate genes associated with tick resistance from previous studies" Principal supervisor, advisory role, contribution to experimental design and editing of the final draft were provided by Professor Barbara Mable, Professor Nicholas Jonsson. Sample collections were provided by Dr. Lorenzo Viora and Kamonchanok Chooyoung. All experimental procedures (DNA extraction), data analysis, and thesis writing and final draft were completed by Kamonchanok Chooyoung.

Chapter 4: "Weighted Gene Co-expression Network Analysis (WGCNA) of long-read RNA sequencing data from skin and spleen sample of cattle" Principal supervisor, advisory role, contribution to experimental design and editing of the final draft were provided by Professor Barbara Mable, Professor Nicholas Jonsson. All experimental procedures (RNA extraction, Long-read RNA-seq) were provided by Kamonchanok Chooyoung. Data analysis (bioinformatics) were analysed by Calum Steward. Thesis writing and final draft were completed by Kamonchanok Chooyoung.

List of Abbreviations

AA	Homozygous AA
AB	Heterozygotes AB
AFFY	Affymetrix, a company that provides microarray technology
ALPL	Alkaline phosphatase
AMOS	Analysis of Moment Structures, a statistical software
AMPure	A purification kit for DNA and RNA
ATP9A	ATPase phospholipid transporting 9A
ATPase	Enzymes that catalyze the decomposition of ATP into ADP
BB	Homozygous BB
BGVD	Bovine Genome Variation Database
BLUP	Best linear unbiased prediction
BM86	A vaccine protein for tick control
BoLA	Bovine lymphocyte antigen
BoLA-DRB3	Major histocompatibility complex, class II, DRB3
BP01-BP12	Barcoding Primers 1-12
BRAF	B-Raf proto-oncogene, serine/threonine kinase
BTA	Bos taurus autosomes
BTA10	Bos taurus autosome 10
BTA2	Bos taurus autosome 2
C-C	C-C - Chemokine (C-C motif)
C-X-C	C-X-C - Chemokine (C-X-C motif)
C1QA	C1q A chain
C1QB	C1q B chain

C1QC	C1q C chain
CA	Carbonic anhydrase
CA2	Carbonic anhydrase 2
CCL2	Chemokine (C-C motif) ligand 2
CCL3	Chemokine (C-C motif) ligand 3
CCL8	Chemokine (C-C motif) ligand 8
CCR1	C-C motif chemokine receptor 1
CD14	CD14 molecule
CD19	CD19 molecule
CD25	CD25 molecule
CD84	CD84 molecule
cDNA	Complementary DNA
CFB	Complement factor B
CFD	Complement factor D
cGMP	Cyclic guanosine monophosphate
CLASP1	Cytoplasmic linker associated protein 1
COI	Cytochrome c oxidase I
COL1A1	Collagen type I alpha 1 chain
COL1A2	Collagen type I alpha 2 chain
COL3A1	Collagen type III alpha 1 chain
COL5A1	Collagen type V alpha 1 chain
COL5A2	Collagen type V alpha 2 chain
COVID19	Coronavirus Disease 2019
COX1	Cytochrome c oxidase I
CR1	Complement receptor 1

CR2	Complement receptor 2
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CSTA	Cystatin A
CXCL10	Chemokine (C-X-C motif) ligand 10
CXCL2	Chemokine (C-X-C motif) ligand 2
CXCL5	Chemokine (C-X-C motif) ligand 5
CXCL8	Chemokine (C-X-C motif) ligand 8
CXCR1	Chemokine (C-X-C motif) receptor 1
DAVID	Database for Annotation, Visualization and Integrated Discovery
DE	Differential expression
DECR1	2,4-Dienoyl-CoA reductase 1
DEG	Differential expression of genes
DHRS7	Dehydrogenase/reductase 7
DKK1	Dickkopf WNT signaling pathway inhibitor 1
DNA	Deoxyribonucleic acid
DNA-RFLP	DNA marker
DNA-SNP	Illumina BovineSNP50 BeadChip
DNAH14	Dynein axonemal heavy chain 14
dNTPs	Deoxynucleotide triphosphates
DRB3	BoLA-DRB3
DSP	Desmoplakin
EB	Elution buffer
EBV	estimated breeding value
ECM	Extracellular matrix protein
ECM1	Extracellular matrix protein 1

EdgeR	A bioconductor software package for examining differential expression of replicated count data
ENSEMBL	A genomic database project
ERP29	Endoplasmic reticulum protein 29
EV	Extracellular vesicles
F13A1	Coagulation factor XIII A chain
FAO	Food and Agriculture Organization
FB	Flush Buffer
FDR	False discovery rate
FKTN	Fukutin
FLT	Flush Tether
FN1	Fibronectin 1
FoxP3	Forkhead box P3
GC	Guanine-cytosine
GEBV	Genomic estimated breeding value
GeneID	Gene identification
GEX	Gene expression
GEXS	Gene expression studies
GGP	GeneSeek Genomic Profile
GIMAP	GTPases of immunity-associated proteins
GIMAP1	GTPases of immunity-associated proteins 1
GIMAP2	GTPases of immunity-associated proteins 2
GIMAP4	GTPases of immunity-associated proteins 4
GIMAP5	GTPases of immunity-associated proteins 5
GIMAP6	GTPases of immunity-associated proteins 6

GIMAP7	GTPases of immunity-associated proteins 7
GIMAP8	GTPases of immunity-associated proteins 8
GO	Gene ontology
GPR34	G protein-coupled receptor 34
GPS2	G-protein pathway suppressor 2
GridION	A sequencing platform from Oxford Nanopore Technologies
GS	Genomic Selection
GSTM1	Glutathione S-transferase mu 1
GTP	Guanosine triphosphate
GTPase	Enzymes that catalyze the hydrolysis of guanosine triphosphate (GTP)
GWAS	Genome-wide association studies
HMGB1	High mobility group box 1
HOX	Homeobox genes
HOXB1	Homeobox B1
HOXD	Homeobox D
HOXD1	Homeobox D 1
HSF2BP	Heat shock transcription factor 2 binding protein
HSP90B1	Heat shock protein 90 beta family member 1
HSPA5	Heat shock protein family A (Hsp70) member 5
ID	Identifier
IgG	Immunoglobulin G
IgM	Immunoglobulin M
IL-2	Interleukin 2
IL-7	Interleukin 7
IL10	Interleukin 10

IL2R	Interleukin 2 receptor
ISG15	Interferon-stimulated gene 15
ITGA11	Integrin subunit alpha 11
KCNK17	Potassium channel, subfamily K, member 17
KEGG	Kyoto Encyclopedia of Genes and Genomes
KRT1	Keratin 1
KRT10	Keratin 10
KRT14	Keratin 14
KRT17	Keratin 17
KRT2	Keratin 3
KRT3	Keratin 3
KRT5	Keratin 5
LAPTM5	Lysosomal protein transmembrane 5
LB	Loading Beads
LD	Linkage disequilibrium
LN	Liquid nitrogen
LYN	LYN proto-oncogene; Src family tyrosine kinase
MA	Minor allele
MAF	Minor allele frequency
MANF	Mesencephalic astrocyte-derived neurotrophic factor
MAPK	Mitogen-activated protein kinase
MAS	Marker-assisted Selection
MHC	Major histocompatibility complex
MHC2	Major histocompatibility complex class II
MiniKnow	A software for controlling nanopore sequencing devices

MinION	A portable sequencing device from Oxford Nanopore Technologies
MIRO1	Mitochondrial Rho GTPase 1
MM	Module membership
Module-Trait	A module-trait relationship in network analysis
mRNA	Messenger RNA
MYO5C	Myosin VC
MYOC	Myocilin
MZB1	Marginal zone B and B1 cell-specific protein
NA	Not applicable
NADP-retinol	Nicotinamide adenine dinucleotide phosphate-retinol
NCBI	National Center for Biotechnology Information
NCOR1	Nuclear receptor corepressor 1
NEB	New England Biolabs
NF-	Nuclear factor
NGS	Next-generation sequencing technology
ONT	Oxford Nanopore Technologies
ONT-RNA	RNA sequencing using Oxford Nanopore Technologies
P4-ATPPase	P4-type ATPase
P4HB	Prolyl 4-hydroxylase subunit beta
PacBio	Pacific Biosciences
PCR	Polymerase chain reaction
PCR-cDNA	PCR-amplified complementary DNA
PDGFRA	Platelet-derived growth factor receptor alpha
PDIA6	Protein disulfide isomerase family A member 6
PE	Phosphatidylethanolamine

PPP1R1B	Protein phosphatase 1 regulatory inhibitor subunit 1B
PRKG1	Protein kinase cGMP-dependent 1
PTGR1	Prostaglandin reductase 1
PTPN22	Protein tyrosine phosphatase non-receptor type 22
qRT	Quantitative real-time
qRT-PCR	Quantitative real-time polymerase chain reaction
QTL	Quantitative trait loci
RAP	Receptor-associated protein
RFLP	Restriction Fragment Length Polymorphisms
RHOT1	Ras Homolog family member T1
RNA	Ribonucleic acid
RNA-seq	RNA sequencing
rRNA	Ribosomal RNA
RT	Reverse transcription
RT-PCR	Reverse transcription polymerase chain reaction
S1P1	Sphingosine-1-phosphate receptor 1
SALL	Sal-like protein
SALL4	Sal-like protein 4 transcription factor
SATB	Special AT-rich sequence-binding protein
SATB2	Special AT-rich sequence-binding protein 2 or SATB homeobox 2
SERINC5	Serine incorporator 5
SERPINA1	Serpin family A member 1
SGS	Second-generation sequencing
SIRPA	Signal regulatory protein alpha
SMRT	Single-Molecule Real Time

SNP	single nucleotide polymorphism
SNP50K	A genotyping array with 50,000 SNPs
SNPs	Single nucleotide polymorphisms
SpotON	A sample port for nanopore sequencing devices
SQ-PCR109	A specific kit from Oxford Nanopore Technologies
SQB	Sequencing Buffer
SQK-PCB109	A specific kit from Oxford Nanopore Technologies
SRGN	Serglycin
SSP	Sequence-specific primer
SV	Structural variations
TFs	Transcription factors
TH2	T helper cell 2
TickGARD	A vaccine for tick control
TNF	Tumor necrosis factor
TOM	topology overlay matrix
TR	Tick resistance
TS	Tick susceptible
TXNDC5	Thioredoxin domain-containing protein 5
TYROBP	TYRO protein tyrosine kinase binding protein
UK	United Kingdom
US	United States
USA	United States of America
VC	Vector control
VCAN	Versican
VIPAS39	Vacuolar protein sorting-associated protein 39

VNP	VN Primers
VPS33B	Vacuolar protein sorting-associated protein 33B
WGCNA	Weighted Gene Co-expression Network Analysis
WNT	Wingless-related integration site, a family of signaling proteins

Chapter 1

General introduction

1.1 Introduction to *Rhipicephalus microplus*

1.1.1 Rhipicephalus (Boophilus) microplus

Rhipicephalus (Boophilus) microplus (R. microplus), commonly known as the cattle tick, is a hematophagous arthropod that is commonly found in tropical and subtropical regions (Figure 1.1). Cattle is the primary host of this tick, which is responsible for significant global economic losses (Jongejan and Uilenberg, 2004). Cattle ticks can transmit pathogens like *Babesia bovis*, *Babesia bigemina*, and *Anaplasma marginale*, which are causative agents of tick-borne diseases (dos Santos et al., 2022). Additionally, high infestation of *R. microplus* in cattle can result in weight loss, subsequent decreases in milk and meat production, and a decrease in the quality of leather (Corrier et al., 1979, Riek, 1962) Globally, it has been estimated that 80% of the cattle population in the world are at risk from tick infestation, causing approximate annual losses of US$ 22–30 billion (Lew-Tabor and Rodriguez Valle, 2016). Instead of chemical control (Higa et al., 2016), animal selection for tick resistance is one of the most promising strategies against *R. microplus.* This method can decrease several drawbacks of chemical use, especially the risk of contaminating the environment and cattle products (Wikel, 1988, Willadsen and Kemp, 1988). Notably, there is a consumer preference for organic products with reduced pesticide use, creating pressure on markets involved in food production (FAO, 2018).

Figure 1.1 *Rhipicephalus microplus* (cattle tick) Female (left) and male (right); photo by Daktaridudu (*Rhipicephalus microplus* ixodid ticks female and male (identify.us.com))

The subgenus of *Rhipicephalus* (*Boophilus*) includes a complex of species which includes *R. annulatus*, *R. microplus*, and *R. australis*, but it has been debated in the last decade how to identify members of this genus. Thomas et al. (2014) investigated phylogenetic analyses of mitochondrial genes based on COX1 (Cytochrome C Oxidase 1) and 16srRNA with *R. annulatus*, *R. australis*, *R. kohlsi*, *R. geigyi*, and *R. microplus* from Brazil, Cambodia and China. The results showed that there are two clades of *R. microplus* within the complex. *Rhipicephalus microplus* clade A includes *R. microplus* from Asia, South America and Africa, while *R. microplus* clade B includes *R. microplus* from Southern China and Northern India. In addition, this study suggested that *R. microplus* clade B are more closely related to *R. annulatus* than to the *R. microplus* clade A. Thomas et al. (2014) also reported that *R. microplus* from Southern China and Northern India (*R. microplus* clade B) are a cryptic species. Using five populations of *R. microplus* from Malaysia, Low et al. (2015) identified three clades of *R. microplus*, with clades A and B showing similar results with Thomas et al. (2014) and clade C identified as *R. microplus* from India and Malaysia.

1.1.2 Life cycle of *Rhipicephalus microplus* in cattle

The life cycle of *R. microplus* can be divided into parasitic and non-parasitic phases. The parasitic, blood-feeding stage can complete its life cycle in 3-4 weeks (21 days on average), involves the moulting of larvae to a nymph, and then finally into the adult tick. When the larvae come out from the egg and find a host, the rest of their life is spent on that host. After obtaining their blood meal during the blood-feeding stage, fully engorged females drop from the host, while males have been known to survive for up to 70 days, either on the host or in the vegetation (Figure 1.2) (Stewart et al., 1981). The non-parasitic stage begins as the engorged female falls to the ground and finds a suitable area to lay her eggs. This stage ends after 18 to 240 days, the larvae hatch out from the egg and jump up to the environment to find a host; they are able to live without feeding for 3-4 months in summer or around six months in winter (Snowball, 1956). The cattle tick can spend around 80% of its life in the larvae stage. Newly attached larvae, also known as 'seed ticks,' typically attach to the softer skin areas such as the inner thighs, flanks, and forelegs, and may also be found on the abdomen and brisket. Development stages, including larvae, nymph and adult, are able to feed only once for several days. After feeding, larvae and nymphs develop to the adult stage, attaching to the same host. When they change to the adult stage, male and female adult ticks become sexually mature and ready to mate. Then a female tick detaches from the host, lays down a load of eggs in the environment, and dies (Stewart et al., 1981).

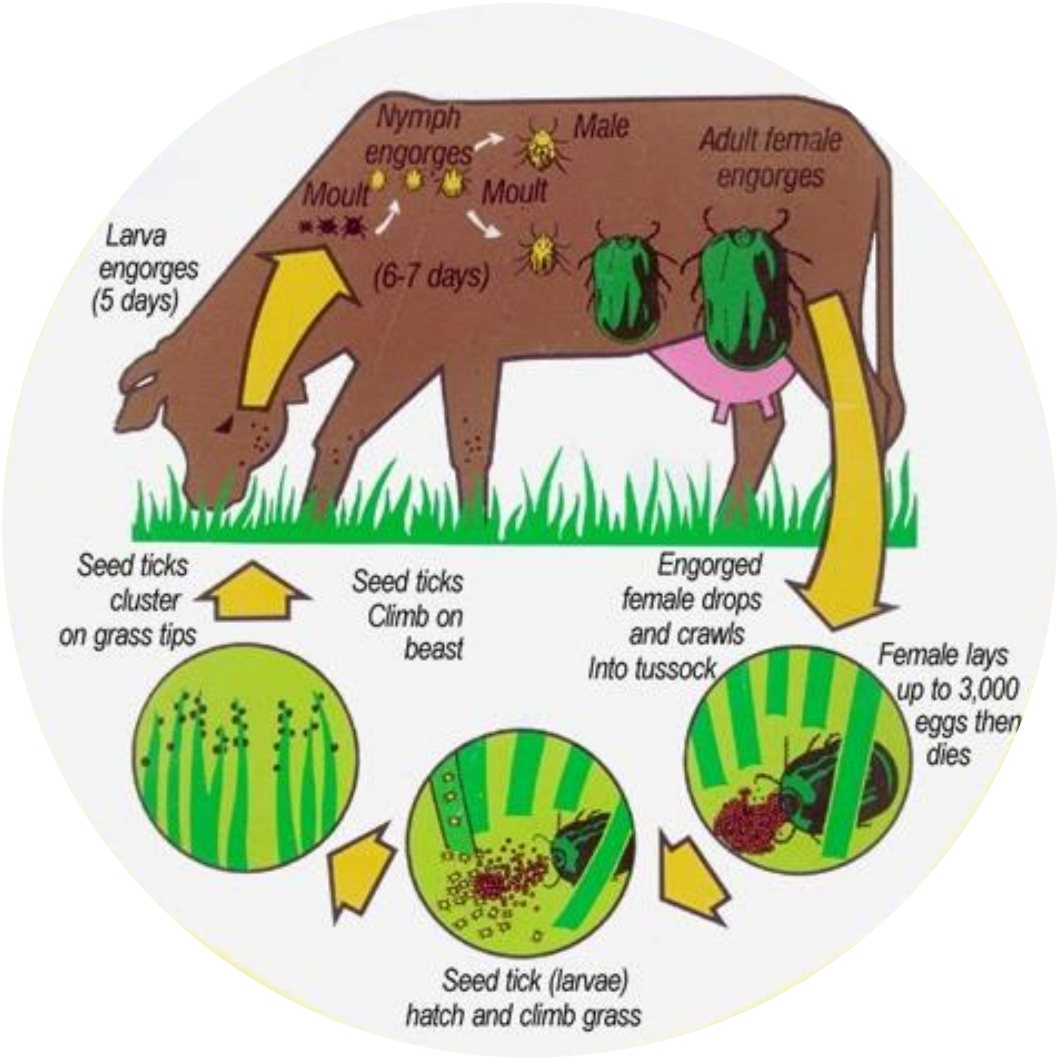


Figure 1.2 The life cycle of *Rhipicephalus (Boophilus) microplus.* An adult female ticks drop to the ground and lays around 3000 eggs before dying. They feed on a host for up to 4 weeks before detaching to lay eggs and die, often far from their original attachment point, contributing to tick spread. ("*Boophilus*" is the former genus name for *Rhipicephalus microplus*) (Queensland Government, 2016).

1.1.3 Chemical control and vaccination

Cattle tick infestation is one of the most important ectoparasite problems in tropical and subtropical regions worldwide, leading to major economic losses (Jongejan and Uilenberg, 1994). There are several types of chemicals used to control *R. microplus*, including acaricides and macrocyclic lactones, which target a broader range of parasites. Due to the excessive use of such chemical compounds, resistance has developed to all major control compounds in cattle ticks across the globe (Perez-Cogollo et al., 2010, Rodriguez-Vivas et al., 2006), including North America (USA, Mexico), Central America (Jamaica, Republica Dominicana, Cuba, Guatemala, Honduras, El Salvador, Panama, Costa Rica), South America (Colombia, Venezuela, Bolivia, Uruguay, Brazil, Argentina), Australia (Australia, New Caledonia), Asia (India, Iran), and Africa (Benin, Tanzania, South Africa, Zambia). Consequently, it is necessary to develop alternative strategies to control cattle ticks, possibly including the use of altered livestock management practices, combining different pesticide compounds, acaricide rotations, mixture formulation of pesticides, botanical and organic pesticides (plant extracts and essential oils), manual tick removal, and environmental management (Rodriguez-Vivas et al., 2018).

However, the development of vaccines has been suggested as the most effective to slow the emergence of resistance and minimize environmental impacts by reducing dependency on chemical compounds such as acaricides (Guerrero et al., 2012). Antigens secreted from the salivary gland and epidermis are referred to as exposed antigens (those naturally encountered by the immune response of hosts during feeding), while antigens secreted from the gut epithelium that are not presented to the immune response of host during tick feeding are referred to as concealed antigens. The first concealed antigen targeted for vaccine development was Bm86, which is secreted from the mid-gut epithelium of engorged *R. microplus* females. The Bm86 antigen vaccine was found to be effective in both its native and recombinant forms (Willadsen et al., 1989), inducing gut content leakage into the haemocoele of ticks, decreasing the number of engorged females, their weight and fertility. This vaccine, which was released as a commercial vaccine in Australia as TickGARD® (Willadsen et al., 1995), has been used in a huge number of cattle against tick infestation and the results showed a 90% efficacy decrease in the prevalence of tick-borne diseases from *R. microplus.* TickGARD was discontinued in Australia in 2010, as the vaccine required 3-4 boosts per year to be effective against ticks, it was not adopted by Australian beef producers as they muster their cattle only once per year (Rodríguez-Vivas et al., 2012, de la Fuente et al., 2007). Another recombinant Bm86 antigen (Gavac®) was launched in 1997

in Mexico, Argentina and Colombia (Canales et al., 1997). de la Fuente et al. (2007) investigated the study of this vaccine in Mexico, showing its efficacy in controlling infestations of *R. microplus* and *R. annulatus.* However, studies in Argentina identified the polymorphisms in the Bm86 gene that influenced the gene expression and led to the production of a soluble rather than a membrane-bound protein in ticks that were resistant to vaccination with the initial Bm86 (García-García et al., 2000).

1.1.4 Tick resistance in cattle

Genetic resistance in different breeds is one of the potential alternative strategies to control ticks (Regitano and Prayaga, 2010). Domestic cattle consist of at least three distinct lineages: *Bos taurus taurus* (European taurine), *Bos taurus indicus* (Indicine or Zebu), and *Bos taurus africanus* (Sanga and Zenga). From previous studies, taurine and indicine cattle diverged between 610,000 and 850,000 years ago (MacHugh et al., 1997). Bovine quantitative genetics studies have shown that heritability for tick resistance varies from low to high based on the cattle breed (Budeli et al., 2009, Henshall, 2004a). Taurine breeds are well-known for their high milk and beef production, high susceptibility to ticks, and optimal adaptation to the temperate climates of Europe and the Near East. In contrast, Indicine breeds are able to adapt to tropical and subtropical environments, including Asian and African socio-economic conditions, better than taurine breeds but their production yields are lower than *B. taurus* (MacHugh et al., 1997). Thus, crossbreeding between *B. indicus* and *B. taurus* has been developed to enhance production and control tick infestations through genomic selection. The achievement of genomic selection relies on appropriate genetic evaluation programs for tick resistance. Such programmes implemented in Australia and Brazil have demonstrated that the crossbred animals are productive in environments with a high prevalence of ticks (Mota et al., 2016, Shyma et al., 2015). In Africa, there are approximately 40 indigenous cattle breeds, including the Sanga *(Bos taurus africanus*: for example, Nguni, Tuli, and Tswana) and Zebu breed types (*Bos indicus*: for example, Malawi Zebu and Angoni). Additionally, a new breed category has emerged called the "Zenga," which is a crossbreed between Zebu and Sanga cattle (Frisch et al., 1997).

1.1.5 Tick counts

Wharton et al. (1970) investigated tick counting methods by counting the number of engorged ticks (between 4.5-8mm in diameter) on both sides or from one side of each individual animal following natural or artificial tick challenge in order to identify variation

in tick resistance. Tick counts are labour-intensive and necessitate skilled animal technicians and costly infrastructure to handle animals simultaneously. For natural infestation of tick, tick from environment are allowed to infest to animal randomly. The tick numbers are presented depending on the seasonal pattern of tick in the environment and requiring animals to be gathered during periods when tick numbers show the variation within the group of animals then the result showed the accuracy during natural infestation. Artificial tick infestation is another technique which can be employed; however, this requires the facilities of tick breeding and skill of technicians to accurately deliver specific quantities of tick larvae for on-farm infestations. Prayaga et al. (2009) attempted to find a simple tick counting process, this method involves single or repeated visual assessments to score the number of engorging ticks on one side of the animal after field infestations were introduced, then a range of scores of this method provided from 0 (low) to 5 (high). Tick scoring requires less infrastructure compared to tick counting because it allows for a higher throughput of animals per day. Moreover, tick scoring has significantly lower heritability than the tick count and tick scores are also affected by the same limitations that impact tick counts (Burrow, 2014). However, both tick count and tick scoring in research are still challenging, thus here is an urgent need for an easy, cost-effective method for phenotyping measurement to identify individual variation in tick resistance. This improvement is crucial for advancing genetic and management strategies aimed at enhancing host resistance to ticks (Burrow, 2014).

1.2 Genomic and transcriptomic studies in cattle – molecular approaches to selection of resistant`

During the last 20 years, researchers have investigated the mechanisms of tick resistance in various cattle breeds. Previous studies on bovine gene profiles associated with tick resistance have reported findings based on DNA, RNA, proteins, cells, and tissues. However, variations between bovine host species, tick species, and experimental designs have not been consistently considered. Additionally, differences at each tick life stage are evident, which is a limitation when comparing information. Gene expression and transcriptional studies are typically associated with and based on the relative quantification of transcripts between compared groups or under different experimental designs and conditions. RNA sequencing is one approach to comparing gene expression patterns, offering large-scale results based on absolute quantification or relative expression, which can enable the discovery of novel transcripts associated with particular phenotypes. Furthermore, proteomics and cellular level studies should be conducted to validate potentially significant findings from gene expression studies in this research field.

1.2.1 Molecular approaches for selection of resistance

The growing world population, food production and consumption requirements are increasing. Therefore, food production should be increased rapidly in the future. From the significant gap between food production and the global population, the expected productivity gap could be enhanced by improving genomic technologies. Livestock industries need to increase their products effectively; however, these will need to increase production without increasing cost and maintaining the quality of products. In the past, due to the lack of molecular knowledge, selection based on phenotypes was used in animal breeding, based on the estimated breeding value (EBV) (Garner et al., 2016). Although several approaches have been used, the best linear unbiased prediction (BLUP) uses variation in phenotypic traits in the estimate of breeding values (EBV) to improve performance predictions (Hayes and Goddard, 2010). Marker-assisted Selection (MAS) based on genetic markers has improved selective breeding compared to phenotype only based methods. The genomic estimated breeding value (GEBV) is calculated based on genome-wide dense marker data (Meuwissen et al., 2016). Mark-assisted Selection (MAS) increases reliability and is more accurate than traditional BLUP because genomic relationships can be more accurate than pedigree relationships (Fernando and Grossman, 1989).

Although the use of genetic markers has been successful, statistical power to estimate breeding values especially for complex phenotypic traits remains challenging when using individual marker-based approaches. In recent years, there have been several novel genomic technologies applied to breeding programs but one of the most popular methods in the past few decades for both animal and plant breeding programs has been Genomic Selection (GS), which has increased both technical and economic efficiency in terms of livestock production (König et al., 2009).

Figure 1.3 illustrates the number of papers based on a search for “genomic selection studies in cattle” by year of publication in the web of science, demonstrating the increasing popularity of these methods for cattle. Meuwissen et al. (2001) reported that using dense marker panels was a potential method to predict high accuracy breeding values. Information from SNPs, even with small effects, can be combined with traditional methods based on phenotypic traits Dense marker panels are required to optimally perform MAS to ensure markers are close to Quantitative Trait Locus (QTLs) which are the specific regions associated with various traits and the genome is sufficiently covered. Moreover, whole

genome sequencing has further improved and increased the accuracy of genomic selection and management of animal breeding.

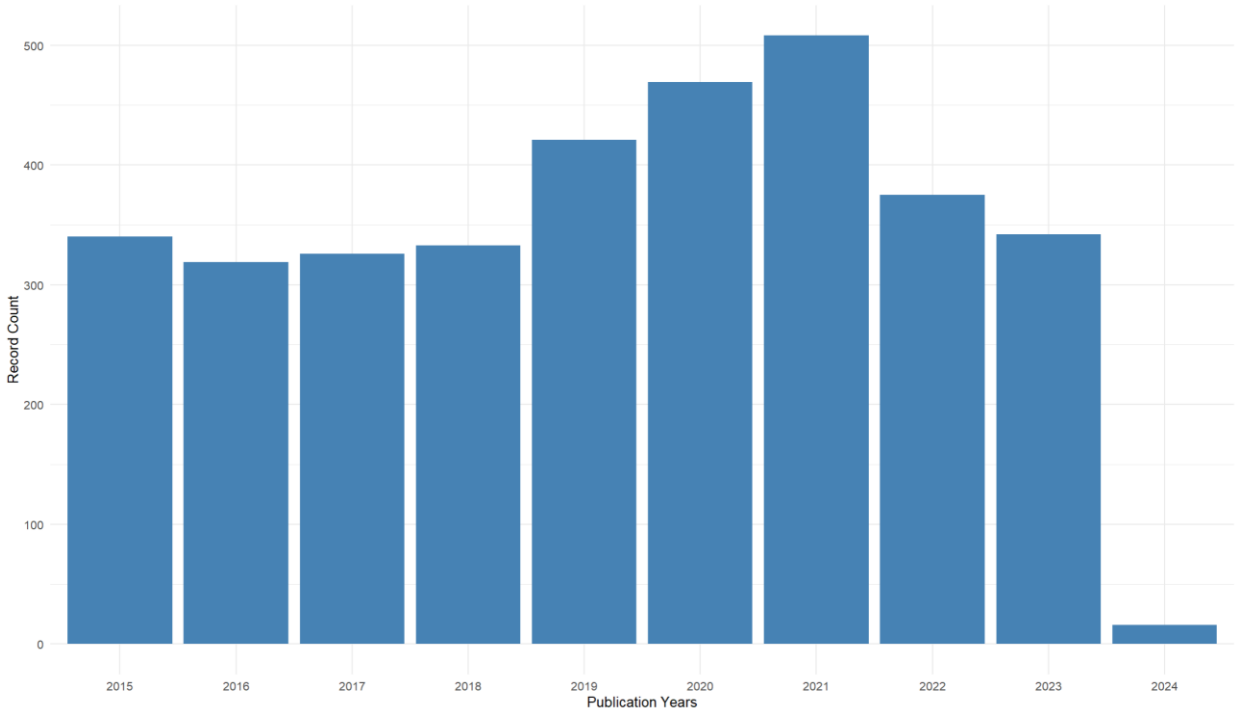


Figure 1.3 The number of genomic selection studies in cattle by year of publication in Web of Science between 2015 and 2024.

1.2.2 Genomic Selection in Cattle

In cattle, genomic selection has provided several markers associated with economically significant traits in cattle; for example, milk yield, meat quality, quantity, and cow health (Bolormaa et al., 2010). Particularly in dairy cattle, genomic selection has proven highly accurate for predicting genomic traits, offering a cost-effective approach. This has resulted in very large numbers of selection candidates genes being genotyped (Meuwissen et al., 2016). The reliability of genomic prediction in dairy cattle exceeds 0.8 for production and 0.7 for fertility and other traits (Lund et al., 2011). Although this has resulted in genetic improvements of dairy cattle in a number of countries, there is a risk that genomic selection has also increased the rate of inbreeding. In contrast, genomic prediction of beef cattle is lower than dairy cattle, ranging from 0.3-0.7 (Van Eenennaam and Young, 2014). The lower accuracy of genomic selection in beef cattle is due to the reference populations being of lower quality and smaller within breeds compared to dairy cattle but also potentially more divergent from the reference due to independent management of herds across sites (Meuwissen et al., 2016). In addition, genomic selection is more advantageous for traits that

are difficult to select for based on the traditional method. It is less advantageous in beef cattle than dairy cattle because traits that can be measured easily at a young age such as growth rate do not require progeny testing to predict future outcomes (Meuwissen et al., 2016). Combining data across countries or across breeds might possibly increase the accuracy of prediction by solving the problem of small reference populations (de Roos et al., 2009). Better collection of genotypic data and phenotyping might improve the efficiency of genomic selection in beef cattle (de Roos et al., 2009).

1.2.3 Genome-wide association studies (GWAS)

Genome-wide association studies (GWAS) use information from genetic markers or SNPs to identify associations with a trait of interest and assume that a marker is in linkage disequilibrium (LD) with it (Hayes and Goddard, 2010, Hirschhorn, 2005, Klein et al., 2005). Due to the variation of genetic architecture and polygenic nature of complex traits, different geographic regions and different pedigrees of animals, it is not easy to interpret these data sets. Fortunately, genome-wide association studies have been discovered to be an effective method to identify genes associated with various phenotypes and to elucidate complex traits on a large scale (Klein et al., 2005). In biological processes, a single gene could regulate one or more traits, leading to a high genetic association between many traits in humans and animals. In the field of livestock, GWAS has been popular in QTL mapping for important economic traits like quality and quantity of meat, milk yield fat and protein percentage, fertility traits and immune responses. Development of techniques to sequence full genomes and to resolve single nucleotide polymorphism (SNP) markers have led to improvement in QTLs mapping and GWAS. In addition, the results of GWAS can be a tool for creating breeding schemes that increase the frequency of favourable alleles (Tiezzi et al., 2015) and the accuracy of GEBV in livestock breeding (Hayes and Goddard, 2010). However, sample size is major factor in GWAS that affects statistical power and is quite hard to control (Hong and Park, 2012). In livestock breeding, measurements for complex traits are often costly and difficult to measure. One solution to deal with the issue of sample size might be to combine summaries of individual GWAS studies using meta-analysis or smaller data sets for joint GWAS (mega-analysis) (Sung et al., 2014).

GWAS allows the narrowing of genomic regions containing causal variants (AMOS, 2007, Greely, 2007), offering genetic determinant insights applicable to genetic selection programmes, potentially including tick resistance in cattle. A comprehensive review by Reed et al. (2015) presents guidelines for effective GWAS analysis, presenting a practical workflow that may be beneficial in its execution. GWAS in cattle has effectively identified

genetic factors linked to specific disease resistance or susceptibility, including tuberculosis (Bermingham et al., 2014), resistance to ticks, mastitis (Sodeland et al., 2011), and foot and mouth disease (Lee et al., 2015). Additionally, this method has proven successful in mapping genetic variants associated with meat quality (Santiago-López et al., 2018) and milk production (Chen et al., 2018). These investigations contribute valuable insights into the genetic framework of QTL, offer biological knowledge regarding the expression of economically relevant traits, and support enhancements in genomic selection. Whole genome sequencing generates massive information from a large database for biological research (Schneider and Orchard, 2011). Advances in sequencing technology have improved the processes and reduced the costs of genomic approaches over the last four decades (Koboldt et al., 2013). Databases from the National Centre of Biotechnology Information (https://www.ncbi.nlm.nih.gov/) and ENSEMBL (http://www.ensembl.org/index.html) can be used for gene annotation and to determine candidate genes associated with identified SNP markers (Zimin et al., 2009). Functional annotation databases such as DAVID and KEGG (Huang da et al., 2009a, Kanehisa et al., 2014) can also be used to identify biological functionality or pathways involved with the candidate genes. In addition, protein-protein interaction networks can be generated from open-source software such as Cytoscape (Shannon et al., 2003) or String (Szklarczyk et al., 2019). Gene networks enhance the biological understanding of interactions among genes related to a trait of interest.

The identification of common causal variants linked to tick resistance faces limitations due to several factors, including the complexity of the trait and resulting influence by more than one gene (Frazer et al., 2009, Aschard et al., 2012). Consequently, GWAS has fallen short of capturing all the genetic determinants underlying the expression of the tick resistance trait (Tam et al., 2019). The lack of a comprehensive biological understanding of how these factors impact missing heritability in cattle poses a significant obstacle to uncovering true loci associated with tick resistance. Currently, there is no perfect genomic technology capable of comprehensively capturing all the genetic information associated with the expression of complex traits (Otto et al., 2018). Despite this limitation, there is an urgent need to leverage available technologies to uncover information that can enhance traits through selection. Some of the challenges for GWAS include: sample size of animals to achieve sufficient power to detect associations, different methods used for phenotyping, and genotyping, genotyping errors, missing data, insufficient data analysis, and poor interpretation of results (Mkize et al., 2021).

1.2.4 Gene expression studies associated tick infestation in cattle

Gene expression studies based on genome-wide transcription profiles (transciptomics) have been developed over the last few decades (Martin and Wang, 2011) and are also crucial for identifying candidate genes associated with complex traits such as host response resistance to ticks in bovines. Transcriptomic approaches can not only provide quantification of differential expression of genes (DEG) in relation to particular traits but can also consider isoform variation and alternative splicing. From such results, potential vaccines and markers for selective breeding in the host response to ticks in bovines could be developed.

Gene expression studies based on microarrays in cattle skin, have identified a large number of genes that could be potential markers of tick resistance based on differences in expression between resistant and susceptible cattle (Franzin et al., 2017b, Kongsuwan et al., 2008, Moré et al., 2019, Piper et al., 2008, Wang et al., 2007). These genes show functions involved in innate, adaptive and other host responses to tick infestation in different phenotypes (Constantinoiu et al., 2018, Constantinoiu et al., 2010, Kongsuwan et al., 2010, Piper et al., 2009, Piper et al., 2017, Robbertse et al., 2018, Robbertse et al., 2020, Wang et al., 2007). For example, differential expression of the complement cascade genes associated with the immune system have been found in comparisons between susceptible and resistant cattle, with the latter showing higher levels of expression (Carvalho et al., 2014, Piper et al., 2010, Wang et al., 2007). CD14 acts as a biomarker of macrophages and monocytes and is involved in the host's immune response (Ziegler-Heitbrock and Ulevitch, 1993). It was identified as being differentially expressed in relation to tick infestation between resistant and susceptible groups of animals in skin tissue studies. However, the results showed lower levels of gene expression in resistant cattle (Piper et al., 2008, Piper et al., 2009) . The immunoglobulin genes also showed upregulation in susceptible compared to resistant individuals (Piper et al., 2010, Wang et al., 2007). Some chemokine and cytokine genes associated with the inflammation in host response (such as L-6, CXCL-8, CCL-2, HMGB1, ISG15, and PKRG), showed a delayed increase in expression in susceptible animals after tick infestation (Franzin et al., 2017b). Although some of these studies have validated differences in expression in skin using qRT-PCR (Piper et al., 2010, Piper et al., 2009, Wang et al., 2007), whole transcriptome profiling could provide more potential for discovery of differentially expressed genes and pathways, such as described in (Moré et al., 2019).

Although fewer studies have focused on blood, multiple genes and pathways have also been found to be differentially expressed in relation to tick infestation in susceptible

and resistant breeds, in both microarray (Marima et al., 2020, Piper et al., 2009) and RNA-seq studies (Mantilla Valdivieso et al., 2022). For example, IL-2, IL2Rα, TNFα, and CCR1 were significantly upregulated in resistant animals (Piper et al., 2009). In contrast, CXCL10 showed higher levels of gene expression in susceptible cattle (Piper et al., 2009)., However, (Piper et al., 2009) compared the blood samples of high and low resistant groups at the highest point of tick infestation rather than at predetermined time points (Domingues et al., 2014) used qRT-PCR to validate association of CXCL10 but they found downregulation of expression in susceptible tick-infested cattle compared to uninfested controls. They also found significant upregulation of CD25, IL10 and FoxP3 in tick-infested compared to uninfested cattle. Although CXCL10 was identified as differentially expressed in association with ticks in both studies, due to the different approaches for comparing gene expression and different experimental designs, the CXCL10 gene still requires validation in further studies.

Mantilla Valdivieso et al. (2022) presented a pathway analysis based on transcriptome profiling using RNASeq to predict the interactions between genes that were differentially expressed in tick-naïve Brangus steers (*Bos taurus* x *Bos indicu*s) at different timepoints (at 0, 3, and 12 weeks) that differed in levels of tick infestation. They reported differential expression of genes involved in chemokine-cytokine components and inflammatory host response, the IL-7 pathway, and cytokine-cytokine interactions associated with host immune responses. Because the different studies show variation in experimental designs, it is difficult to directly compare the results from different gene expression studies. For example, if the different studies provided different sampling time points or studied expression in different tissues, then the level of gene expression may show different patterns of gene expression. However, integrating across studies to look at intersections of genes common to more than one approach could help to narrow down the list of genes or pathways most associated with response to ticks. Several studies have investigated genomics and transcriptomics to identify genes or biomarkers associated with tick resistance in cattle; however, fewer studies have taken a proteomics approach, to more directly relate variation to function. A notable exception is Raza et al. (2021) who used serum samples from Santa Gertrudis cattle to quantify differences in proteins associated with resistance. Sequential window acquisition was used to identify and quantify the peptides, then ion mass spectrometry was used for the analysis. They found that 28 proteins differed significantly between resistant (TR) and susceptible (TS) groups before tick challenge, with eight of these associated with adaptive responses. In contrast, in the intragroup of animals, both groups of cattle showed similar responses, but the responses were relatively stronger in TR cattle than TS cattle. In addition, there were many significantly different proteins in the

resistant group (both before and after tick challenge) that were involved in immune function, such as complement cascades, chemotaxis, and acute immune responses (Raza et al., 2021). In 2023, Raza et al. (2023) used quantitative proteomics to analyse serum and skin protein samples from Brangus cattle (naïve tick-resistant and tick-susceptible) at two different time points after tick exposure. They found a significantly different abundance of proteins in resistant cattle after early and prolonged tick exposure (compared to resistant naïve cattle) that were involved in immune response, coagulation cascade, blood coagulation, homeostasis, and wound healing. In contrast, only some proteins of these responses after prolonged tick exposure showed significance in susceptible cattle. Further studies are needed to confirm these results in different breeds of cattle and under various tick infestation conditions (Raza et al., 2023).

1.3 Breeding cattle for tick resistance

Host resistance to tick infestation in cattle was discovered many decades ago, leading to the development of a vaccine (Bm86 antigen) in 1989 (Willadsen et al., 1989). At that time, the research focused on mechanisms to induce the immune function of the host and how to discover a novel vaccine. In the early 20th century, it was well-established that *Bos indicus* showed higher resistance to *R. microplus* than *Bos taurus* cattle (Beverley, 1996). *Bos taurus* shows high productivity but less resistance to ticks compared to *B. indicus* whereas *B. indicus* is known to have a low yield of production but is highly resistant to ticks (Wagland, 1978, Henshall, 2004b, Budeli et al., 2010). Other factors also have been shown to influence host resistance to ticks, including seasonal changes, age, sex, and nutrition. According to Wharton et al. (1970), seasonal changes affect the intensity of gene expression involved in host resistance in Australian Illawarra Shorthorn cattle. That report also was supported by other studies (Sutherst et al., 1983, Doube and Wharton, 1980). The level of resistance has been found to be significantly higher in Brahman and Brahman × British cattle than in British cattle during the spring and summer, whereas this difference decreased or disappeared in summer until early winter (Sutherst et al., 1983). The quality of nutrition also has been shown to play a key role in the maintenance of the level of tick resistance. For example, a study undertaken by Sutherst et al. (1983) demonstrated that good nutrition in late winter (a time when resistance levels decrease) expedited the recovery of resistance. There are also gender-related effects; female cattle showed higher resistance to ticks than male cattle, and the capacity of resistance in male cattle has been found to be less stable than that of females (Seifert, 1971, Utech et al., 1978b). However, in those studies, it was also

demonstrated that during pregnancy or lactation in females, the level of resistance in females was reduced. Wharton et al. (1970) discovered that responses to ticks for cows that are typically resistant were difficult to distinguish from those that were typically susceptible during lactation. However, no differences were found in the responses of lactating and nonlactating females in a subsequent study conducted by (Johnston and Haydock, 1969). Resistance levels also can be influenced by age, with advancing age contributing to shifts in the ranking and a decrease in the stability of resistance within a herd (Wharton et al., 1970). Utech et al. (1978b) corroborated this observation, although variation between age groups was not consistently significant. Individual behaviour could also affect response to tick infestation, such as self-grooming, which has been demonstrated to be an important source of tick mortality (Riek, 1962, Snowball, 1956). Bennett (1969) found that tick yields were initially increased for individuals wearing an anti-grooming harness compared to unrestricted controls but then decreased in resistant individuals, suggesting that the immune response was enhanced by the larger tick numbers. This evidence was later supported by (Koudstaal et al., 1978, Kemp et al., 1983). Other studies have assessed whether physical factors differed between high and low resistant animalsl (Wagland, 1978) found no association between the level of tick resistance and skin thickness. In addition, the coat characteristics of cattle did not significantly contribute to the different levels of resistance of animals (Doube and Wharton, 1980). These sometimes conflicting results show the complex interaction between ticks and host resistance due to differences between breeds or other characteristics of the animals, experimental designs, methods of tick counting or scoring and the environment of animals.

The above studies illustrate that various factors contribute to the level of expression of tick resistance. External factors interact in a complex way with *in vivo* biological processes which together regulate the development of resistance to ticks. Despite attempts to clarify these processes, the mechanisms underlying the development of natural resistance remains inadequately comprehended.

1.4 The variation of molecular genetic associated with tick infestation in cattle

Several studies have aimed to identify genes or biomarkers associated with host resistance to cattle ticks. These studies provided different approaches, including immunological determination, protein analysis, genotyping or candidate gene sequencing, and genotyping of SNPs or microsatellites for genomic detection of quantitative trait loci (QTL). These studies have been summarised and discussed by Porto Neto et al. (2011) and (Mapholi et al., 2014). Meta-analysis of genomics associated with tick resistance in cattle is a good example (Porto Neto et al., 2010b). Although there are many studies that present the potential information, there is insufficient data related to the development of genetic selection programs. Previous studies showed that the major histocompatibility complex (MHC), which can be referred to as bovine lymphocyte antigen (BoLA) (Acosta-Rodriguez et al., 2005, Behl et al., 2012, Machado et al., 2010, Martinez et al., 2006, Otto et al., 2018), has a key role in contributing to the variation between susceptible and resistant animals. However, there is no consistent genotype that has been associated with the immune response to ticks (Tabor et al., 2017). QTL markers also have been studied by using microsatellites and SNPs; however, the loci have tended to show only weak associations with resistances, and many of the results have been inconsistent. For example, Barendse (2007) and Turner et al. (2010) reported interesting loci associated with the resistance levels of tick infestation in cattle; however, the results showed that these loci had only influenced around 1% of this phenotypic trait. One challenge of these studies is the methods for counting ticks which is difficult, time-consuming and difficult to standardize across studies. Identifying potential biomarkers for tick resistance ideally would be based on consistently expressed phenotypes across breed groups, and various environmental and experimental conditions. However, understanding the complex mechanisms behind this resistance is challenging due to the involvement of numerous interacting factors related to adaptive and innate immune responses, pathology, and correlated responses lacking a direct functional relationship. Tick resistance in hosts is a polygenic trait that cannot be solely attributed to a few expressed genes. The integration of transcriptomics, genomics, and meta-analysis could address the multifaceted nature of this complex trait, offering valuable insights for potential inclusion in future breeding programs aimed at enhancing resistance to ticks.

1.5 Long-read RNA sequencing

In the past decade, short-read RNA sequencing, such as provided by the Illumina platform, has provided transcriptomic profiles for research. This technique produces high-accuracy readings and can analyse multiple samples in parallel. However, the short read lengths produced (up to around 600 bp) and assembly algorithms and analysis of structural variation (such as isoform quantification, isoform variation discovery, or alternative splicing) challenging (Dong et al., 2021). Long-read RNA sequencing is provided by several companies; for example, Pacific Biosciences (PacBio), Single-Molecule Real Time (SMRT) sequencing and Oxford Nanopore Technologies (ONT) nanopore sequencing (Dong et al., 2021). The advantage of these techniques is the potential to generate read lengths up to 100 kilobases (kb), which can solve the issues mentioned above when compared to short-read RNA sequencing (Figure 1.4). On the other hand, the disadvantage of long-read RNA-seq is the lower accuracy and throughput when compared with short-read RNA-seq (Hu et al., 2021). The Oxford Nanopore technique relies on a nanopore, a nanoscale protein pore that represents a detector embedded in an electrically resistant polymer membrane (MacKenzie and Argyropoulos, 2023). Single-stranded DNA or RNA molecules (negatively charged) are driven through the nanopore from the 'cis' side (negatively charged) to the 'trans' side (positively charged) in the electrolytic solution; as they do so, they disrupt the ionic currents. The time taken to move through the pore depends on the length of the strand and the specific nucleotides, which produces a unique fingerprint for that sequence (MacKenzie and Argyropoulos, 2023). The long reads produced can also be informative about processes such as alternative splicing and isoform variation.

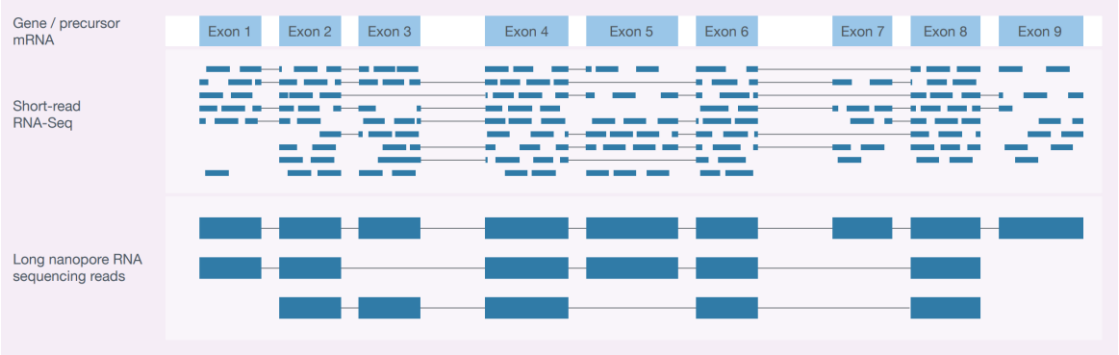


Figure 1.4 Comparison of transcriptome data between short-read RNA-seq and long read RNA-seq. Traditional short-read RNA sequencing can generate accurate sequencing but loses some positional information. However, long-read RNA-seq can be used to sequence full-length transcripts of mRNA which facilitates their identification and quantification **Gene expression | Oxford Nanopore Technologies | Oxford Nanopore Technologies**

1.6 Weighted Gene Co-expression Network Analysis (WGCNA)

Weighted Gene Co-expression Network Analysis (WGCNA) is a bioinformatics method used widely to identify biological pathways based on sets of genes showing similar patterns of expression (Zhang and Horvath, 2005). These sets of co-expressed genes (modules) can then be related to particular phenotypes, to determine if there is an association particular modules (Figure 1.5). In the first step for WGCNA networks are constructed based on the correlation pattern of gene expression across the individual samples, with hierarchical clustering then used to group sets of correlated genes into modules (Langfelder and Horvath, 2008). These co-expressed modules are then related to some grouping variable, such as phenotype groups, to identify biologically interesting modules. The relationship between genes within the modules significantly associated with the biological trait are then analysed (e.g. organisation into biological pathways). These genes might act as key roles in the biological processes involved in the modules, which could then be validated experimentally. WGCNA has been successfully used in several biological research fields; e.g., cancer studies, mouse models, and brain imaging analysis. These methods have also already been applied to analyse various previous studies on cattle, such as the feed efficiency of dairy cattle (Salleh et al., 2018) and identifying functionally enriched pathways in the lactation process (Farhadian et al., 2021). However, not only various phenotypic traits or conditions but also genotypic (genetic variation) profiles can be analysed by using WGCNA; for example, association of expression patterns with genotype variation at immune genes.

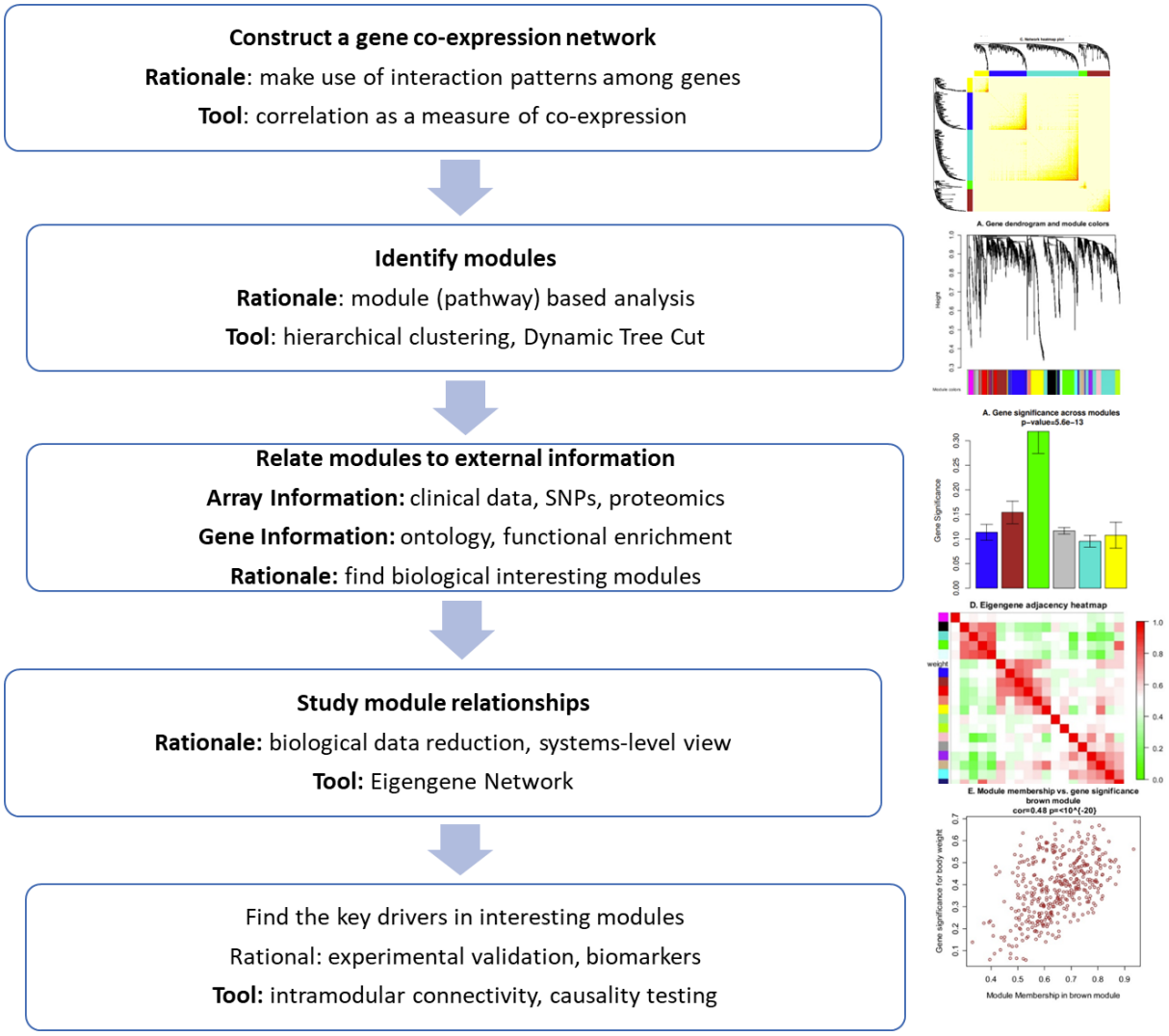


Figure 1.5 The workflow of Weighted gene co-expression network analysis (WGCNA). This flow chart presents an overview of WGCNA analysis including constructing a gene co-expression network, identifying modules, relating modules to external information, studying module relationships and identifying the key genes in interesting modules (Langfelder and Horvath, 2008)

The mechanisms underlying tick resistance within and between cattle breeds have been studied for a few decades. Previous studies on gene or biomarker profiles associated with tick resistance in cattle have discussed findings across gene expression, transcription, and proteomics studies. However, differences in experimental designs, including cattle host breeds, tick species, tick stages, tick challenges, methods of counting or scoring ticks and platforms of analysis, have made it difficult to directly compared the results from different studies. Combining information from multiple studies using a meta-analysis framework has potential to identify overlap in the types of genes or pathways that show differential expression related to tick resistance across such heterogeneous studies. Since only genes that are polymorphic across different breeds have the potential to be used to select for resistance, following this up with single nucleotide polymorphism (SNP) based genotyping and investigating differences in gene expression related to these polymorphisms using transcriptomics could further help to narrow the search and understand the drivers of tick resistance in cattle.

1.7 Aims and Objectives of the thesis

Chapter 2. Integration of gene expression and genome-wide association studies to identify genes associated with resistance to infestation with cattle ticks (*Rhipicephalus microplus*)

The aim of this study was to combine multiple information from previous gene expression studies and GWAS on cattle host resistance to infestation with *Rhipicephalus microplus*, in order to identify a list of the genes or gene products that have the potential to be used as biomarkers in the future. The specific objectives of this study were to: 1) identify genes that show differential expression related to different levels of host resistance to ticks in several previous studies based on either skin or blood tissues; 2) identify the potential SNPs or QTLs from previous GWAS associated with tick resistance in cattle; 3) describe gene functions or biological processes for the identified genes via over-representation analysis, gene ontology, and KEGG pathways; and 4) use the overlap between genes identified in different studies to identify highly confident genes associated with tick resistance in cattle to inform further studies on genetic variation (Chapter 3) and co-expression based on RNA sequencing (Chapter 4), to identify the key regulated genes of the resistance phenotype to develop effective biomarkers for future studies.

Chapter 3. Genotypic variation in Scottish cattle for candidate genes associated with tick resistance from previous studies

The aim of this study was to determine whether any of the genes identified in Chapter 2 were likely to show sufficient variation that they could potentially be used as markers for selection. The objectives of this study were to: 1) identify SNP genotypes of individual cattle in three different breed groups from a Scottish abattoir based on the list of genes associated with host resistance to tick in cattle that were previously identified from gene expression studies and GWAS in chapter 2; 2) test whether there were significant difference in genotype frequencies for each SNP between individuals classified as British, European or Hill cattle breeds; 3) characterize the genomic context of significant SNPs including the location of SNPs, exon and intron and consequence type of mutation; and 4) generate a subset of genes that could be used to interpret patterns of gene expression in Chapter 4.

Chapter 4. Weighted Gene Co-expression Network Analysis (WGCNA) of long-read cDNA sequencing data from skin and spleen samples of cattle

The aim of this study was to investigate whether there were differences in expression pattern (transcriptome profiles) related to the genes identified as showing significant differences in genotype frequency from Chapter 3, based on long-read RNA sequencing in the same individual animals and describe the biological pathways involved. Thus, the specific objectives of this study were to: 1) perform long-read cDNA sequencing (Oxford Nanopore Technology (ONT)) for skin and spleen samples from the same individual animals provided in Chapter 3; 2) use the RNA-seq data to identify sets of genes that present similarity in gene expression pattern (significant co-expression modules) related to genotype variation for each of the significant SNPs identified in Chapter 3; 3) describe the key biological processes involved in the significantly co-expressed modules; and 4) identify the relationship between the list of genes from Chapter 2 and the co-expressed gene network from the two tissue types

Chapter 5. General Discussion

This chapter provides a synthesis of the results from the three data chapters and suggests areas of further study that would be necessary to assess the potential of the proposed candidate genes as biomarkers for selection against tick infestation in cattle.

Chapter 2

Integration of gene expression and genome-wide association studies to identify genes associated with resistance to infestation with cattle ticks (*Rhipicephalus microplus*)

2.1 Abstract

The cattle tick (*Rhipicephalus microplus)* causes damage to cattle throughout the tropics and subtropics. Resistance to ticks is moderately heritable (*h2 ~ 0.4*) but it is a complex trait for which previous studies have identified different sets of genes associated with resistance. The objective of this study was to combine information from gene expression studies (GEXS) and genome-wide association studies (GWAS) on host resistance to infestation with *R. microplus*, to create a list of genes or gene products for which there were multiple sources of supporting evidence. From the literature, 16 GEXS (7 microsatellites studies, 9 NGS; 4 blood and 12 skin studies) and 12 GWAS (9 based on single nucleotide polymorphism and 3 on restriction fragment length polymorphisms) were identified that provided sufficient information to identify genes as significantly associated with tick resistance (as per authors' declarations). This yielded 10,495 differentially expressed genes (DEGs) and 288 quantitative trait loci (QTLs,) which were then filtered to only those genes for which multiple studies showed consistent results. The final list included those QTLs significant in at least two independent GWAS (n=11); DEG significant in at least 4 skin (n=6) or 2 blood (n=10) GEXS; QTLs that were also significant DEG in at 1 blood or 2 skin GEXS (n= 10). The list of genes included 2 transcription factors, 11 genes associated with immune function, 3 with the extracellular matrix, 6 structural genes, and 15 others. A total of 37 genes were identified for which multiple sources of evidence could be obtained and so provide high confidence that they are associated with resistance to tick infestation. Variability and expression of this subset of genes is further explored in chapters 3 and 4 to determine whether they have potential for use as biomarkers for genomic selection and animal breeding management.

Keywords: genome-wide association studies (GWAS), gene expression studies(GEXS), biomarkers, cattle, ticks

2.2 Introduction

The cattle tick *Rhipicephalus microplus* is the most important ectoparasite of livestock, especially in tropical and subtropical areas (Barker and Murrell, 2004), causing severe illness and adverse effects in cattle (Jongejan and Uilenberg, 2004), with annual global costs of around US$ 22–30 billion (Burrow et al., 2019, Tabor et al., 2017). In addition to its direct effects, the tick is a vector for pathogens. Resistance of cattle to ticks is moderately to highly heritable, ranging from 0.42-0.64 (Utech et al., 1978a), so it is possible to directly improve resistance of cattle to ticks by selection within breeds. *Rhipicephalus. microplus* originated in Asia and *Bos indicus* cattle (Asian) are generally more resistant to tick infestations than *Bos taurus* (Middle East, Europe, Africa) (Bradley et al., 1998). Each cattle genotype was initially domesticated in the area where it had developed natural defenses against the endemic tick parasite and the pathogens they transmit (Jonsson et al., 2014). Tick resistance is a complex trait controlled by many genes, which are involved in a variety of functions related to cellular processes or pathways that contribute to the development of host response (inflammatory response, immune response, and extracellular matrix mechanism), which are influenced by bovine genetic composition, sex and history of prior exposure to ticks (Carvalho et al., 2011, Carvalho et al., 2010, Machado et al., 2010, Piper et al., 2008, Piper et al., 2009) Other complex traits such as vascular architecture, grooming, physical characteristics of skin, composition of skin and other morphological aspects also interfere with tick feeding success and are related to tick resistance in cattle (Jonsson et al., 2014).

Several studies have attempted to quantify gene expression (GEX) associated with the resistance of cattle to tick burdens in contrasts between low and high resistance groups, with some studies comparing responses between and some within breeds. Overall, greater differences of gene expression have been found in comparisons between low-resistance taurine breeds and high-resistance indicine breeds, than between animals showing different levels of tick infestation within breeds (Hong and Park, 2012)**.** Several gene expression studies have been conducted to better understand the mechanisms of resistance and the regulation of genes influencing resistance to ticks. These studies have used different approaches, including different study designs (e.g. natural or artificial tick infestation), number of animals, timing of experiments (e.g. single or multiple timepoints), breeds of cattle, tissues (e.g. skin or blood), and history of prior exposure to ticks. Platforms for assessing differences in gene expression have included: microarrays (Carvalho et al., 2014, Franzin et al., 2017b, Marima et al., 2020, Piper et al., 2010, Piper et al., 2009, Wang et al.,

2007); targeted, quantitative real-time polymerase chain reaction (qRT-PCR) (Bagnall et al., 2009, Carvalho et al., 2014, Carvalho et al., 2010, Domingues et al., 2014, Kongsuwan et al., 2010, Marima et al., 2020, Moré et al., 2019, Nascimento et al., 2010, Piper et al., 2008); and and transcriptome profiling using deep sequencing (RNAseq) (Mantilla Valdivieso et al., 2022, Moré et al., 2019). One challenge of GEXS is the interpretation of differences depending on the time of tick challenge in relation to sampling and history of previous tick burden or other differences in the state of the experimental animals. These designs do not easily differentiate between pathological responses and adaptive responses that improve resistance. However, integration of results across different types of studies could identify genes or gene regions strongly associated with resistance to tick infestation.

Genome-wide association studies (GWAS) also have the potential to identify quantitative trait loci (QTL) for tick resistance in cattle. For example, linkage analysis based on microsatellites has been used to identify several QTL associated with tick resistance in F2 crosses between *Bos indicus* (Gyr) and *Bos taurus* (Holstein) cattle from Brazil but the gene regions identified differed between the dry and wet seasons (Gasparin et al., 2007, Machado et al., 2010, Regitano and Prayaga, 2010). Although these QTL explained only 3.3% and 5.9% of the phenotypic variance in tick burden during the dry and wet seasons, respectively, a QTL identified on BTA 23 that influenced tick burden in both seasons was located in a genomic region containing the BoLA gene complex (Machado et al., 2010). This emphasises that environment could play an important role in interpretation of GWAS results. Genotyping platforms have also been used to quantify variation in single nucleotide polymorphisms for GWAS analyses. For example, QTLs associated with tick infestation have been identified by comparing variation (Bovine 10K SNP panel) in highly resistant Brahman beef cattle and more susceptible Holstein dairy cattle (Porto Neto et al., 2011, Porto Neto et al., 2010b, Turner et al., 2010). Sollero et al. (2017) used a denser Illumina BovineSNP50 (50K) SNP assay to identify QTLs between half-sib populations of Brazilian Hereford and Braford (hybrids between Brahman and Hereford) cattle. Each of these studies identified different sets of QTLs, and the BTA 23 QTL was not identified in the SNP-based studies. However, these GWAS studies have used different phenotyping methods and genotyping strategies including the procedures of tick counts, the approach of genotyping, and the population of animals. Although SNP-based GWAS studies conducted to date have had the goal of developing genomic tools for selection for resistance to ticks, lack of consensus in the most important QTLs means that none are known to be in commercial use at the time of writing.

The overall aim of this study was to integrate information from differential gene expression and GWAS studies in order to generate a list of plausible genes or gene products that might aid in the discovery of biomarkers for host resistance to tick infestation**.** The underlying hypotheses were that: 1) genes that are identified in multiple studies addressing variation in host resistance are more likely to be truly involved in the regulation or implementation of the response than genes that are identified in single, or a few studies; 2) genes identified from multiple GWAS are likely to be involved in the regulation of the genes that are identified from multiple GEXS. The objectives were therefore to: 1) identify genes that show differential expression (DE) according to levels of host resistance in multiple GEXS in either blood or skin; 2) identify genes from multiple GWAS studies that are potential QTL for tick resistance; 3) describe gene function or biological processes for the identified genes via over-representation analysis, gene ontology (GO) and KEGG pathways; and 4) identify genes that are associated at high confidence with tick resistance in cattle, to inform further studies on levels of genetic variation (chapter 3) and co-expression networks (chapter 4) to identify important regulators of the resistance phenotype that are sufficiently variable to have potential as biomarkers.

2.3 Materials and Methods

2.3.1 Selection of Studies

A literature search of previous studies was performed (in August, 2022) to compile a set of previous studies investigating changes in gene expression and identifying SNPs or QTLs in GWAS related to tick infestation in cattle. This included unpublished data from previous PhD studies (Marima, 2022, Piper, 2010); the data from these were provided by the authors. Rather than the type of comprehensive systematic review that has been conducted previously (de Souza et al., 2018) my aim was to identify studies amenable to a synthetic analysis to specifically compare results from studies that used either GEX or GWAS to quantify differences in host responses to tick infestation, but using different tissue types, experimental designs and platforms. I then synthesized results from these studies using the workflow described in Figure 2.1.

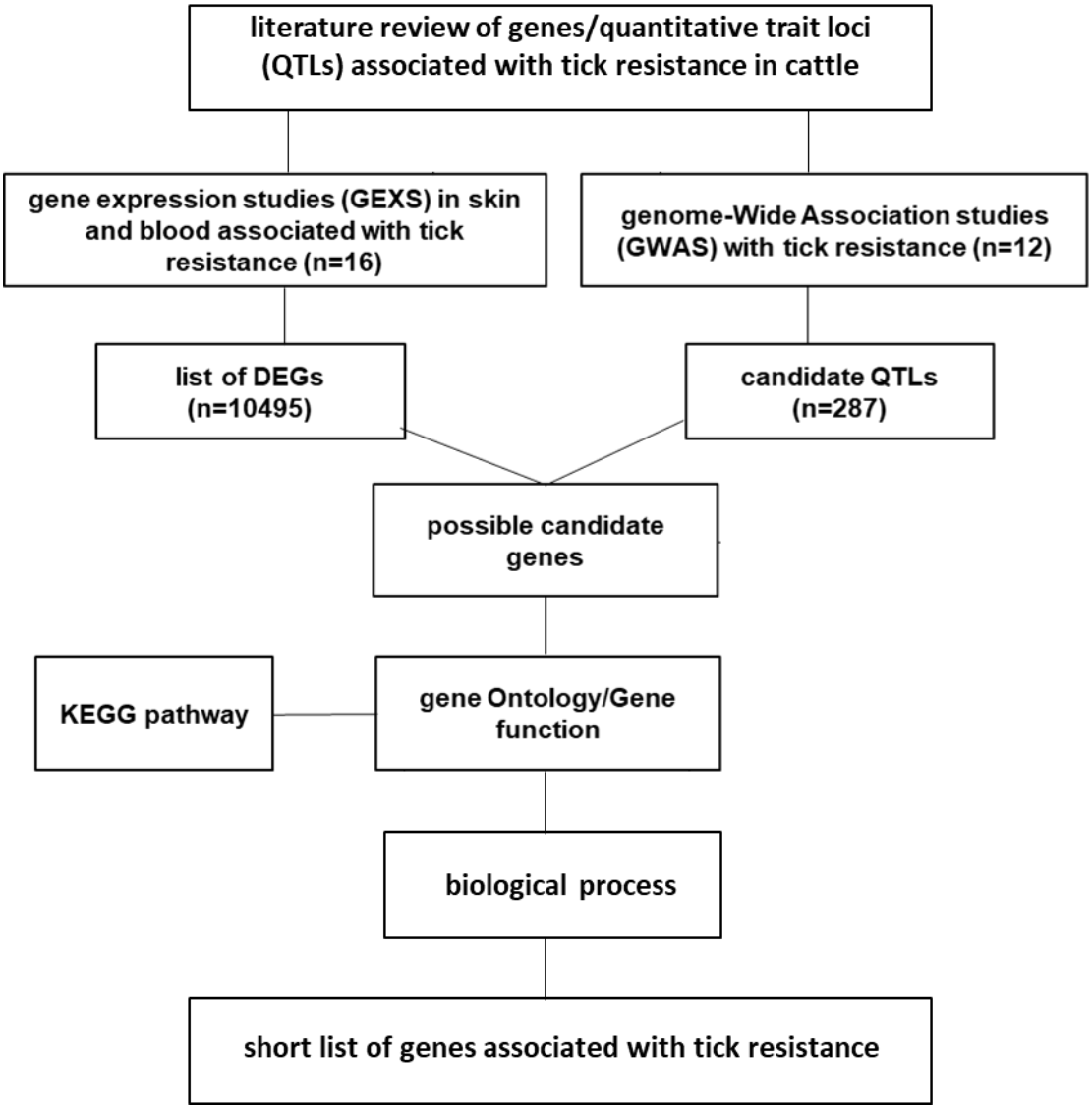


Figure 2.1 Flowchart summarizing the steps towards identification of candidate genes associated with tick resistance in cattle.

2.3.2 Synthesis of Data from Previous Studies

I recorded the following information from each of the focal GEX studies identified: 1) authors; 2) platform (microarray, qRT-PCR, or RNA sequencing); 3) tissue type(s) (skin or blood); 4) treatment comparisons (e.g. between breeds differing in resistance or between animals differing in tick infestations within breeds); 5) timing of sample collection in relation to tick infestation; 6) mode of challenge (the method of tick infestation in the experiment - natural or experimental); 7) prior exposure (animals previously exposed to ticks before joining the study or not); 8) number of animals in the study; and 9) experimental animal type (e.g. *Bos taurus*, *Bos indicus* or crossbreed) (Table 2.1). I then compiled a list of genes that

differed significantly in relation to tick infestation. Only significantly differentially expressed genes from untargeted approaches (microarray and RNA-seq) were considered. Studies that had used targeted analyses (qRT-PCR) were included to determine whether genes that were thought to be useful markers for past research studies were also found to be significantly differentiated in multiple global or untargeted studies. Some genes identified by untargeted platforms had also been validated by qRT-PCR in the same study. In all cases, the authors' declarations of significance were used. Because of the wide variation in methods used and reporting of data, no attempt was made to standardize the criteria used for declarations of significance or to undertake a quantitative analysis. I then updated names and details for genes reported as differentially expressed in each study using the ARS-UCD1.2 bovine genome annotations. Many Affymetrix Oligonucleotide Array probe identifiers that were originally reported as unannotated probes in the earlier studies had been annotated since, and these were updated using Ensembl Biomart (https://www.ensembl.org/index.html) by selecting "Attributes/Gene/External/Microarray-probes-probe sets/AFFY Bovine probe".

For the GWAS studies, QTLs that were significantly associated with the tick resistance phenotype were compiled from the studies, which had used diverse methods for genotyping: DNA markers based on Restriction Fragment Length Polymorphisms (RFLP); Bovine 10K SNP chip (Affymetrix); and Illumina Bovine SNP50K BeadChip (50 K; Illumina, San Diego, CA, USA). I recorded the following information: -1) authors; 2) genotyping platform (DNA-RFLP, SNP chips); 3) breed (*Bos taurus /Bos indicus*); 4) number of animals; 5) resistance phenotyping method (tick count or tick score); 6) mode of tick challenge (natural or experimental); and 7) number of significant QTLs (Table 2).

Table 2.1 Summary of information from the 16 previous GEX studies. The data from previous studies were classified into 11 categories including 1) reference; 2) platform (microarray, qRT-PCR, and RNA sequencing); 3) tissue sample used in the study (skin and blood); 4) timepoint (time to collect the sample after tick infestation); 5) Mode of challenge (the method of tick infestation in the experiment -natural or experimental; 6) prior exposure (animals used previously exposed to ticks before joining the study or not); 7) number of animals in the study; and 8) crossbreeding of animal.

Reference	Platform	Tissue	Timepoint	Mode of Challenge	Prior exposure	Number of Animals	Experimental animal type
Wang et al., 2007	Microarray	Skin	24 h	Artificial	Previously exposed	5	Belmont Adaptor (*B. taurus*)[a]
Piper et al., 2008	qRT PCR	Skin	NA	Artificial	Previously exposed	12	Holstein (*B. taurus*) vs. Brahman (*B. indicus*)[b]
Piper et al., 2009	Microarray	Blood	NA	Artificial	Previously exposed	12	Holstein (*B. taurus*) vs. Brahman (*B. indicus*)[b]
Piper et al., 2010	Microarray	Skin	NA	Artificial	Previously exposed	12	Holstein (*B. taurus*) vs. Brahman (*B. indicus*)[b]
Bagnall et al., 2009	qRT PCR	Skin	3 days	Artificial	Previously exposed	20	Belmont Red (*B. taurus* x *B. indicus*)[a]
Kongsuwan et al., 2010	qRT PCR	Skin	0 and 24 hours	Artificial	Previously exposed	18	Belmont Red (*B. taurus* x *B. indicus*)[a]
More et al., 2019	RNAseq	Skin	0 and 42 days	Artificial	Previously exposed	39	Braford (*B. taurus* x *B. indicus*)[a]
Marima et al., 2020	qRT PCR	Skin	0 and 12 hours	Artificial	Previously exposed	36	Angus, Brahman and Nguni[a]
Carvalho et al., 2010	qRT PCR	Skin	NA	Artificial	Previously exposed	24	Holsteins (*B. taurus*) vs Nelores (*B. indicus*)[b]
Carvalho et al., 2015	Microarray	Skin	0, 24 and 48 h	Artificial	Previously exposed	13	F2 Gir x Holstein[a]
Domingues et al., 2014	qRT PCR	Blood	0, 24 and 48 h	Artificial	Previously exposed	12	Crossbreed F2 Gir x Holstein[a]
Nascimento et al.,2010	qRT PCR	skin	NA	Artificial	Previously exposed	13	F2 (Gyr x Holstein)[a]
Piper, 2010	Microarray	Blood	NA	Artificial and Natural	naïve	30	Santa Gertrudis cattle[a] (*B. taurus x B. indicus*)
Frazin et al., 2017	Microarray	Skin	NA	Artificial	unknown	8	Holsteins (*B. taurus*) vs Nelores (*B. indicus*)[a]
Marima, 2022	Microarray	Blood	NA	Artificial	unknown	30	Angus (*B. taurus*), Brahman (*B. indicus*) and Nguni (*B. taurus africanus*)[b]

Reference	Platform	Tissue	Timepoint	Mode of Challenge	Prior exposure	Number of Animals	Experimental animal type
Mantilla Valdivieso et al. 2022	RNAseq	Blood	0,21,91 d	Artificial	naive	10	Brangus (stable composite *B. indicus × B. taurus)*

[a] indicating within breed studies

[b] indicating between breed studies

Table 2.2 Previous GWAS studies associated with tick resistance in cattle. The data from previous studies were classified 1) reference; 2) genotyping platform; 3) breed (Taurus /Indicus or hybrids); 4) number of animals; 5) phenotyping method; 6) mode of tick challenge; and 7) number of significant QTL.

Authors	Genotyping platform	Breed	No. of animals	Phenotyping Method	Mode of Challenge	No. of QTLs
Martinez et al. (2006)	DNA marker (DNA-RFLP)	F2 (Taurus x Indicus)	231	Tick count	Artificial	1
Untalan et al. (2007)	DNA marker (DNA-RFLP0	Taurus x Indicus	101	Tick count	Artificial	1
Gasparin et al. (2007)	DNA marker (DNA-RFLP)	F2 (Taurus x Indicus)	382	Tick count	Artificial	1
Turner et al. (2010)	bovine10k SNP (DNA-SNP)	Taurine	189	Tick count	Natural	16
Porto Neto et al. (2010b)	bovine10k SNP (DNA-SNP)	Taurine and Indicus	1055	Tick count and tick score	Natural	1
Porto Neto et al. (2010b)	bovine10k SNP (DNA-SNP)	Taurine and Indicus	671	Tick count and tick score	Natural	10
Porto Neto et al. (2011)	bovine10k SNP (DNA-SNP)	Taurine	189	Tick count	Natural	1
Porto Neto et al. (2012)	bovine10k SNP (DNA-SNP)	Taurine and Indicus	1122	Tick count and tick score	Natural	1
Mapholi et al. (2016)	Illumina BovineSNP50 BeadChip (DNA-SNP)	Nguni	586	Tick count	Natural	15
Sollero et al. (2017)	Illumina BovineSNP50	Taurine	3455	Tick count	Natural	103

Authors	Genotyping platform	Breed	No. of animals	Phenotyping Method	Mode of Challenge	No. of QTLs
	BeadChip (SNP_CAND_GENE)					
Otto et al. (2018)	Illumina BovineSNP50 BeadChip (SNP_CAND_GENE)	Taurus x Indicus	376	Tick count	Artificial	67
Mota et al. (2018)	Illumina BovineSNP50 BeadChip (SNP_CAND_GENE)	Taurine	4363	Tick count	Artificial	70

2.3.3 Integration of information from multiple gene expression and genome-wide association studies

I combined the lists of significant DEGs from GEXS with the genes and QTLs from GWAS using R (RStudio Team, 2022). Specifically, the following intersection categories were initially considered to identify biological pathways associated with differences in expression across multiple studies: 1) untargeted DEGs found in at least three skin studies (pairwise intersection of studies that had used microarray or RNA-seq methods in skin tissue); 2) untargeted DEGs found in at least two blood studies (pairwise intersection of studies that had used microarrays or RNA-seq on blood samples); and 3) untargeted DEGs that were identified in at least one skin and at least one blood study. In order to identify a subset of genes with high confidence to be associated with resistance to tick infestation, I also considered more stringent criteria: 4) untargeted DEGs found in at least four skin studies; 5) untargeted DEGs found in at least three blood studies; 6) QTLs identified from more than one GWAS study; and 7) untargeted DEG genes in found in two blood studies and one skin study that were also identified in at least one GWAS study. The sum of genes identified in categories 4-7 were considered as the set of confident genes to be used in further study; biological function analysis was also conducted for each. In each case, the genes were labeled according to the GeneID from NCBI or from the Affymetrix probe IDs.

2.3.4 Over-represented terms and biological pathways in response to tick infestation

To characterise the biological functions associated with the genes identified in the intersection groups described above that yielded a large number of genes (greater than 30), the over-represented biological processes and the associated pathways were identified using gene ontology (GO) terms in DAVID (DAVID Functional Annotation Bioinformatics Microarray Analysis (Huang da et al., 2009b, Huang da et al., 2009a) and KEGG (Kyoto Encyclopedia Genes and Genomes Pathways) (Kanehisa et al., 2023, Kanehisa, 2019, Kanehisa and Goto, 2000) databases, using the DAVID platform. These analyses were restricted to genes that are available in the bovine database on DAVID. Transcription factors were identified by using the information from previous study of de Souza et al. (2018)

2.4 Results

2.4.1 Integration of results from previous GEX and GWAS studies

From the 16 GEXS that met the criteria (Table 2.3), including published and unpublished studies, only nine used untargeted approaches, and these were predominantly based on microarrays (n = 2 blood; n = 5 skin) but four of these had also validated findings using qRT-PCR. Only two studies used RNA-seq (one study on blood and one on skin) and seven studies using only targeted assessment of genes already thought to be associated with tick resistance using only qRT-PCR. Across the untargeted studies, 10495 genes were identified as being differentially expressed (n = 4759 in skin; n = 5646 in blood); 47 of these had been validated using qRT-PCR in the same studies and an additional 43 genes had been tested only using qRT-PCR (Table 2.3). However, there was considerable overlap between studies, with only 4279 unique genes identified as differentially expressed based on the untargeted analyses. Twelve GWAS studies were included, from which 287 QTLs were associated with tick resistance in cattle (Table 2.2).

Table 2.3 Summary of the number of DEGs identified in the GEXS studies, According to whether they were conducted using blood or skin samples, whether they used untargeted approaches (microarrays or RNAseq) or were based on targeted assessment of genes previously associated with tick resistance (qRT PCR).

Reference	Blood	Skin	Microarray	qRT PCR	RNAseq
Wang et al. (2007)	0	258	253	5	0
Piper et al. (2008)	0	14	0	14	0
Piper et al. (2009)	502	0	494	8	0
Bagnall et al. (2009)	0	7	0	7	0
Piper et al. (2010)	0	916	901	15	0
Piper (2010)	0	996	994	2	0
Carvalho et al. (2010)	0	5	0	5	0
Nascimento et al. (2010)	0	4	0	4	0
Kongsuwan et al. (2010)	0	7	0	7	0
Carvalho et al. (2014)	6	0	0	6	0

Reference	Blood	Skin	Microarray	qRT PCR	RNAseq
Carvalho et al. (2014)	0	454	454	0	0
Franzin et al. (2017b)	0	2106	2106	0	0
Moré et al. (2019)	0	51	0	0	51
(Marima et al., 2020)	0	17	0	17	0
Marima (2022)	5074	0	5074	0	0
Mantilla Valdivieso et al. (2022)	78	0	0	0	78
Total	5660	4835	10276	90	129
Untargeted	5646	4759	20299	47	258

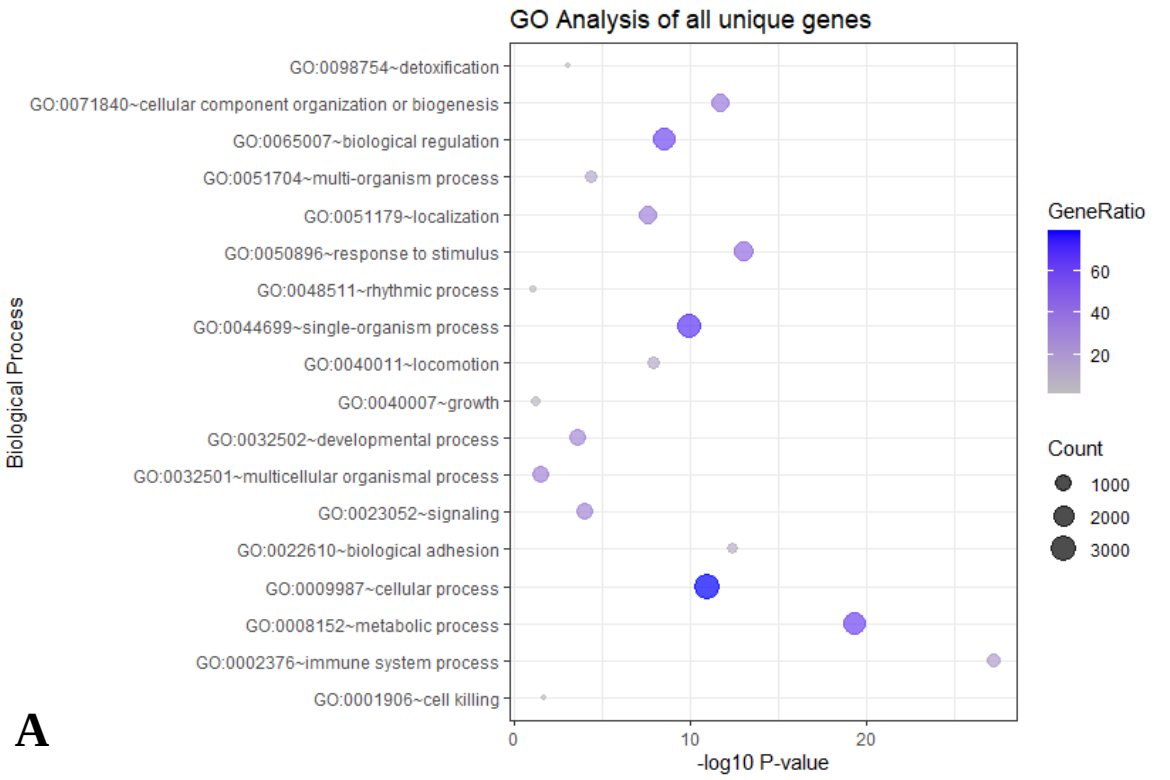
A total of 100 genes showed differential expression in at least three studies of skin GEXS (Appendix A), 114 genes showed differential expression in at least two studies of blood (Appendix B), and three genes of these overlapped between the tissue types (genes overlapped 100 genes from at least three studies of skin and 114 genes from at least two studies of blood) including DECR1, GIMAP7, and VCAN. From the more stringent analyses (Table 2.4), six genes (SRGN, CXCL8, C1QA, FN1, F13A1 and GPR34) were differentially expressed in four or more skin GEXS. Ten genes (CD84, CR2, DNAH14, ECM, HSF2BP, MYO5C, PPP1R1B, PTGR1, RHOT1, VIPAS39) were differentially expressed in at least three blood GEXS. Eleven genes were common to at least two GWAS (ATP9A, HOXD1, KCNK17, MHC2 BOLA, PRKG1, SALL4, SIRPA, U6, ITGA11, SATB2, and SERINC5). Ten genes were common to at least two skin GEXS, one blood GEXS, and one GWAS (ALPL, CA2, DHRS7, DKK1, GIMAP7, GSTM1, LAPTM5, MYOC, PLA2G7, and TYROBP) (Table 2.4). Therefore, 37 unique genes were identified from multiple GEXS and GWAS and were considered to be associated with resistance to tick infestation with high confidence (Table 2.4). All of these genes in this study showed significantly different gene expression based on P-value < 0.05.

Table 2.4 Summary of genes common to multiple studies, indicating the intersection criteria and the Gene ID.

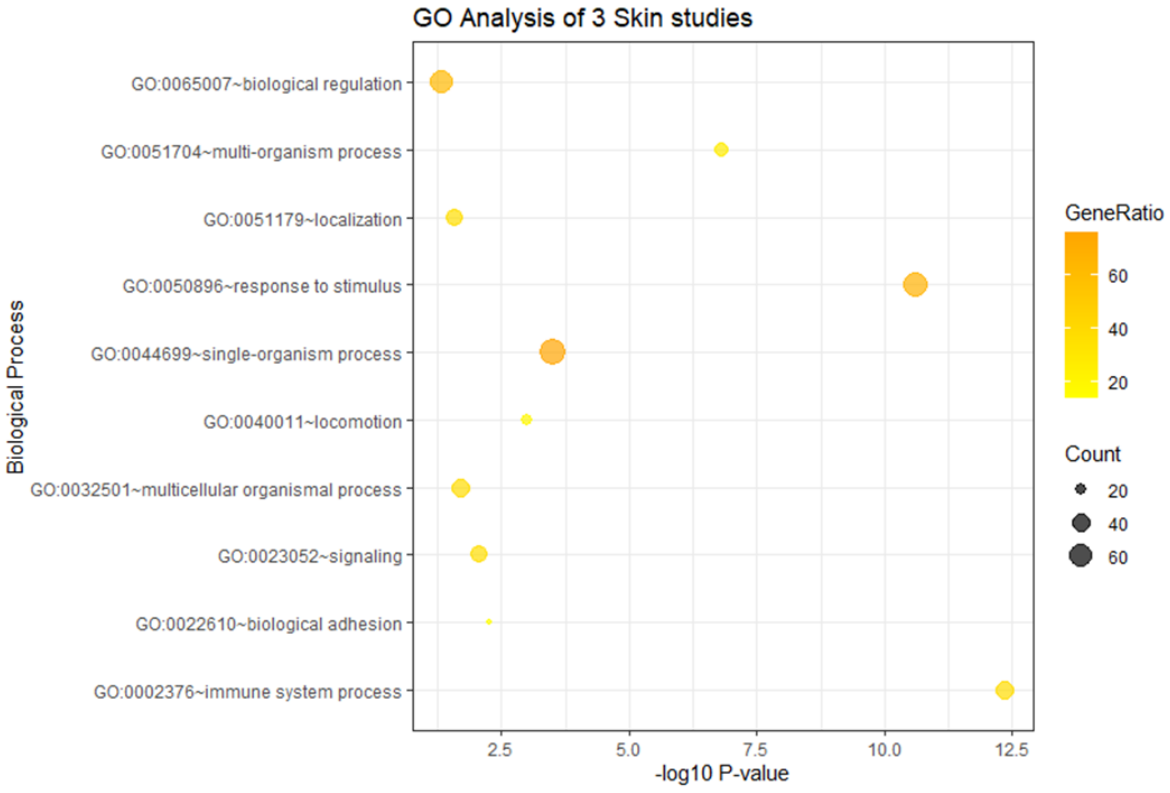
Sources	Gene ID
All three studies of gene expression in blood	CD84, CR2, DNAH14, ECM1, HSF2BP, MYO5C, PPP1R1B, PTGR1, RHOT1, VIPAS39
At least four studies of gene expression in skin	C1QA, CXCL8, F13A1, FN1, GPR34, SRGN
At least two GWAS studies	ATP9A, HOXD1, KCNK17, MHC2 BOLA, PRKG1, SALL4, SIRPA, ITGA11, SATB2, SERINC5, U6
At least two skin expression studies, one blood expression study, and one GWAS	ALPL, CA2, DHRS7, DKK1, GIMAP7, GSTM1, LAPTM5, MYOC, PLA2G7, TYROBP

2.4.2 Biological processes and pathways associated with resistance to ticks in cattle

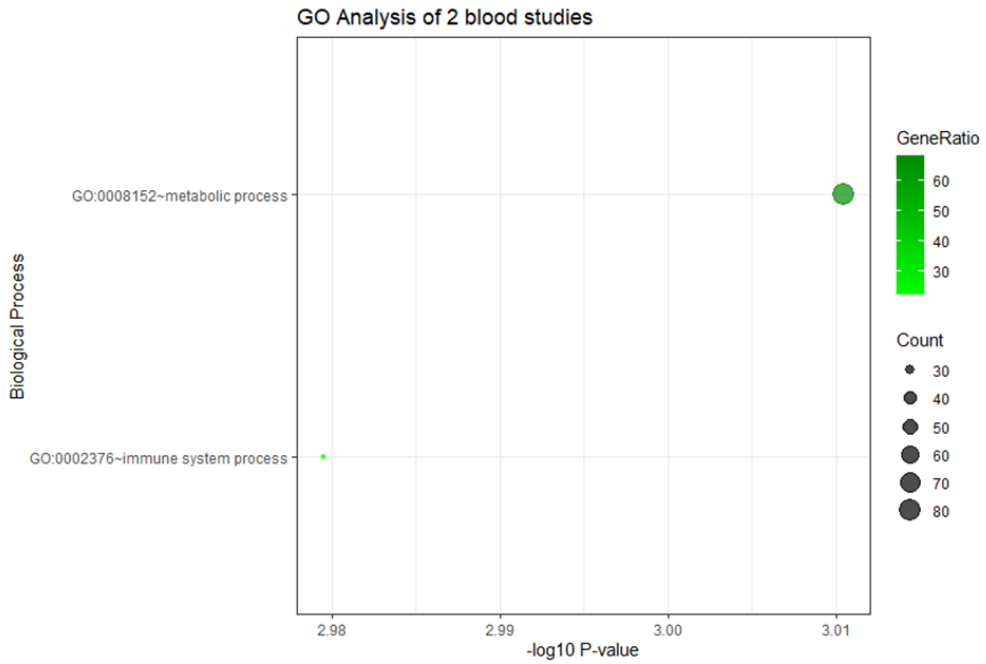
The biological process and biological pathways of the genes determined by pathway enrichment analyses based on DAVID and KEGG pathways are summarised in Figure 2.2). Overall, the unique genes identified from synthesis of the GEX studies represented 18 terms of biological process (Figure 2.2A) with the top three identified as: 1) GO:0002376 immune system process; 2) GO:0008152 metabolic process; and 3) GO:0050896 response to stimulus. For the 100 genes from at least three skin studies (Figure 2.2B), 10 terms of biological process were identified, with the topthree: 1) GO:0002376 immune system process; 2) GO:0050896 response to stimulus; and 3) GO:0050896 response to stimulus. For the 114 DEGs from at least two blood GEX studies (Figure 2.2C) only two biological processes were identified: 1) GO:0008152 metabolic process; and 2) GO:0002376 immune system process. Based on the final subset of 37 genes identified across the intersection analysis (Figure 2.2D; Table 2.5), 5 terms of the biological process were identified, with the top three: 1) GO:0014068 positive regulation of phosphatidylinositol 3 kinase signalling; 2) GO:0072378 blood coagulation, fibrin clot formation; and 3) GO: 0019725 cellular homeostasis.


GO Analysis of all unique genes
GO:0098754~detoxification
GO:0071840~cellular component organization or biogenesis
GO:0065007~biological regulation
GO:0051704~multi-organism process
GO:0051179~localization
GO:0050896~response to stimulus
GO:0048511~rhythmic process
GO:0044699~single-organism process
GO:0040011~locomotion
GO:0040007~growth
GO:0032502~developmental process
GO:0032501~multicellular organismal process
GO:0023052~signaling
GO:0022610~biological adhesion
GO:0009987~cellular process
GO:0008152~metabolic process
GO:0002376~immune system process
GO:0001906~cell killing
Biological Process
0
10
20
-log10 P-value
GeneRatio
60
40
20
Count
1000
2000
3000

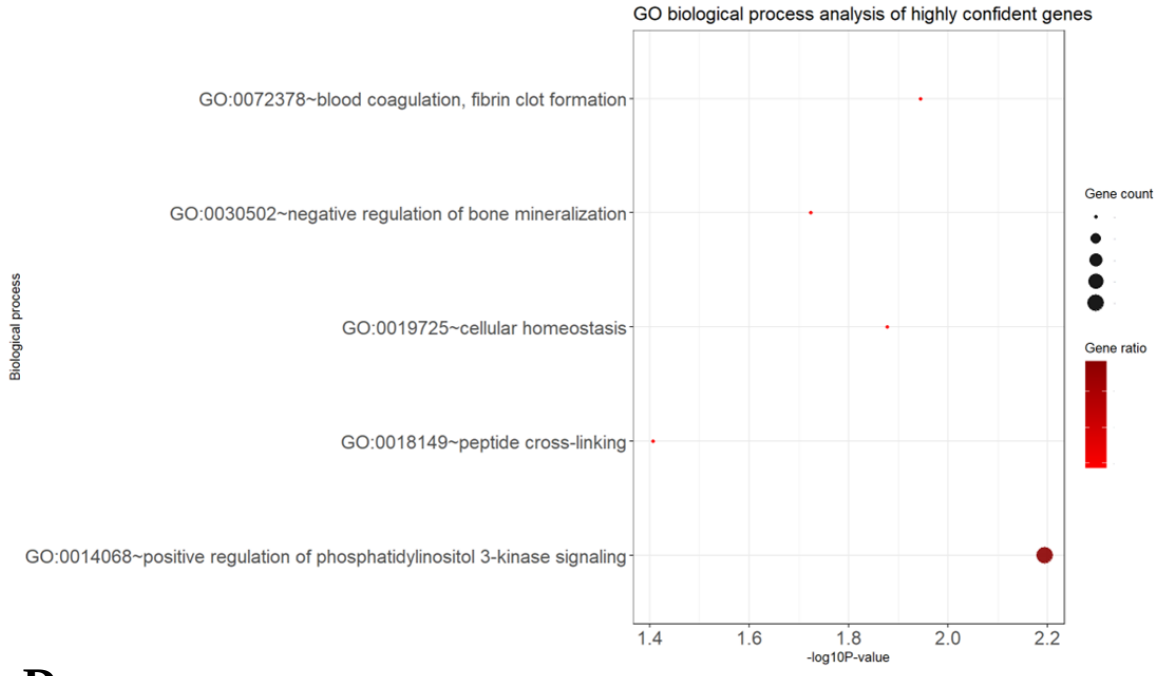
A


GO Analysis of 3 Skin studies
GO:0065007~biological regulation
GO:0051704~multi-organism process
GO:0051179~localization
GO:0050896~response to stimulus
GO:0044699~single-organism process
GO:0040011~locomotion
GO:0032501~multicellular organismal process
GO:0023052~signaling
GO:0022610~biological adhesion
GO:0002376~immune system process
Biological Process
2.5
5.0
7.5
10.0
12.5
-log10 P-value
GeneRatio
60
40
20
Count
20
40
60

B



C



D

Figure 2.2 Over-representation analysis of the list of differentially expressed genes (DEGs). The scatter plots show over-representation analysis results of three lists of genes. A) GO analysis of list of all the unique genes from previous studies (n = 4279). B) GO analysis of the list of DEGs that were significantly different according to level of tick resistance in at least three skin GEX studies (n = 100). C) GO analysis of the list of genes that were significantly different according to level of tick resistance in at least two blood

GEX studies (n = 114). D) GO analysis of the list of highly confident genes and QTLs from multiple studies. The y-axis shows the significantly ($P < 0.05$) over-represented GO terms, and the x-axis shows the -log10 P value. The number of candidate genes annotated with a GO term is mapped to the plot by point size. The gene ratio is indicated by a heat map and refers to the proportion of the genes in the list that are involved in a particular biological process.

Table 2.5 Names and brief description of the biological process/functions of the 37 genes common to multiple studies.

Gene	Source	Gene Name	Biological process/Function
C1QA	At least 4 studies of GEXS in skin	complement C1q A chain	complement activation, classical pathway
CXCL8	At least 4 studies of GEXS in skin	C-X-C motif chemokine ligand 8	chemokine-mediated signaling pathway
F13A1	At least 4 studies of GEXS in skin	coagulation factor XIII A chain	blood coagulation, fibrin clot formation
FN1	At least 4 studies of GEXS in skin	fibronectin 1	acute-phase response
GPR34	At least 4 studies of GEXS in skin	G protein-coupled receptor 34	G protein-coupled receptor signaling pathway
SRGN	At least 4 studies of GEXS in skin	serglycin	secretory granule organization
CD84	At least 3 GEXS in blood	CD84 molecule	regulation of macrophage activation
CR2	At least 3 GEXS in Blood	complement C3d receptor 2	negative regulation of complement activation, classical pathway
DNAH14	At least 3 studies of GEXS in Blood	dynein axonemal heavy chain 14	microtubule-based movement
ECM1	At least 3 studies of GEXS in Blood	extracellular matrix protein 1	inflammatory response, positive regulation of

Gene	Source	Gene Name	Biological process/Function
			endothelial cell proliferation
HSF2BP	At least 3 studies of GEXS in Blood	heat shock transcription factor 2 binding protein	double-strand break repair involved in meiotic recombination
MYO5C	At least 3 studies of GEXS in Blood	myosin VC	actin filament organisation
PPP1R1B	At least 3 studies of GEXS in Blood	protein phosphatase 1 regulatory inhibitor subunit 1B	protein phosphatase inhibitor activity
PTGR1	At least 3 studies of GEXS in Blood	prostaglandin reductase 1	prostaglandin metabolic process
RHOT1	At least 3 studies of GEXS in Blood	ras homolog family member T1	mitochondrial outer membrane permeabilization
VIPAS39	At least 3 studies of GEXS in Blood	VPS33B interacting protein, apical-basolateral polarity regulator, spe-39 homolog	protein-containing complex binding
ATP9A	At least 2 GWAS studies	ATPase phospholipid transporting 9A	regulation of retrograde transport, endosome to Golgi
HOXD1	At least 2 GWAS studies	homeobox D1	regulation of transcription by RNA polymerase II (transcription factor)
KCNK17	At least 2 GWAS studies	potassium two pore domain channel subfamily K member 17	potassium ion transmembrane transport
MHC2 BOLA	At least 2 GWAS studies	major histocompatibility complex, class II	adaptive immune response

Gene	Source	Gene Name	Biological process/Function
PRKG1	At least 2 GWAS studies	protein kinase cGMP-dependent 1	cGMP-mediated signaling
SALL4	At least 2 GWAS studies	spalt like transcription factor 4	regulation of transcription by RNA polymerase II (transcription factor)
ITGA11	At least 2 GWAS studies	integrin subunit alpha 11	integrin-mediated signaling pathway
SATB2	At least 2 GWAS studies	SATB homeobox 2	regulation of transcription by RNA polymerase II (transcription factor)
SERINC5	At least 2 GWAS studies	serine incorporator 5	innate immune response
ALPL	At least 2 GWAS studies	alkaline phosphatase, biomineralization associated	calcium ion homeostasis
U6	At least 2 GWAS studies	U6 spliceosomal RNA	Spliceosomal tri snRNP complex assembly
CA2	At least 2 skin GEXS, 1 GEXS, and 1 GWAS	carbonic anhydrase II	regulation of anion transport
DHRS7	At least 2 skin GEXS, 1 GEXS, and 1 GWAS	dehydrogenase/reductase 7	NADP-retinol dehydrogenase activity
DKK1	At least 2 skin GEXS, 1 GEXS, and 1 GWAS	dickkopf WNT signaling pathway inhibitor 1	multicellular organism development
GIMAP7	At least 2 skin GEXS, 1 GEXS, and 1 GWAS	GTPase, GIMAP family member 7	GTP binding

Gene	Source	Gene Name	Biological process/Function
GSTM1	At least 2 skin GEXS, 1 GEXS, and 1 GWAS	glutathione S-transferase M1	glutathione metabolic process
LAPTM5	At least 2 skin GEXS, 1 GEXS, and 1 GWAS	lysosomal protein transmembrane 5	activation of cysteine-type endopeptidase activity involved in apoptotic process
MYOC	At least 2 skin GEXS, 1 GEXS, and 1 GWAS	myocilin	regulation of cell-matrix adhesion
PLA2G7	At least 2 skin GEXS, 1 GEXS, and 1 GWAS	phospholipase A2 group VII	positive regulation of monocyte chemotaxis
SIRPA	At least 2 skin GEXS, 1 GEXS, and 1 GWAS	signal regulatory protein alpha	protein phosphatase binding
TYROBP	At least 2 skin GEXS, 1 GEXS, and 1 GWAS	transmembrane immune signaling adaptor TYROBP	regulation of immune response

Based on the more conservative intersection analyses (at least three skin or at least two blood studies, respectively), two pathways showed the highest number of DEGs involved and were associated with host response: 1) complement and coagulation cascades (Figure 2.3); and 2) chemokine signaling pathways (Figure 2.4). The DEGs in the complement and coagulation cascade interaction pathway (based on skin studies) were: F13A1, C1QA, C1QB, C1QC, C3, CFB, CFD and SERPINA1 (Figure 2.3). The DEGs involved in chemokine signaling pathways (based on blood studies) were: CXCL8, LYN, CCL2, CCL3, CCL8, CCR1, CXCL2, CXCL5 (Figure 2.4). I also identified 3 transcription factors (HOXD1, SALL4, and SATB2), 12 genes associated with immune function (CD84, CR2, PTGR1, C1QA, CXCL8, F13A1, FN1, BOLA, SERINC5, LAPTM5, PLA2G7, and TYROBP), 9 genes associated with the extracellular matrix and structural proteins (ECM1,

ITGA11, MYOC, DNAH14, MYO5C, RHOT1, DKK1, SRGN, and VIPAS39), and 12 associated with other biological processes (ALPL, ATP9A, CA2, DHRS7, GIMAP7, GPR34, GSTM1, HSF2BP, KCNK17, PPP1R1B, PRKG1, and SIRPA).

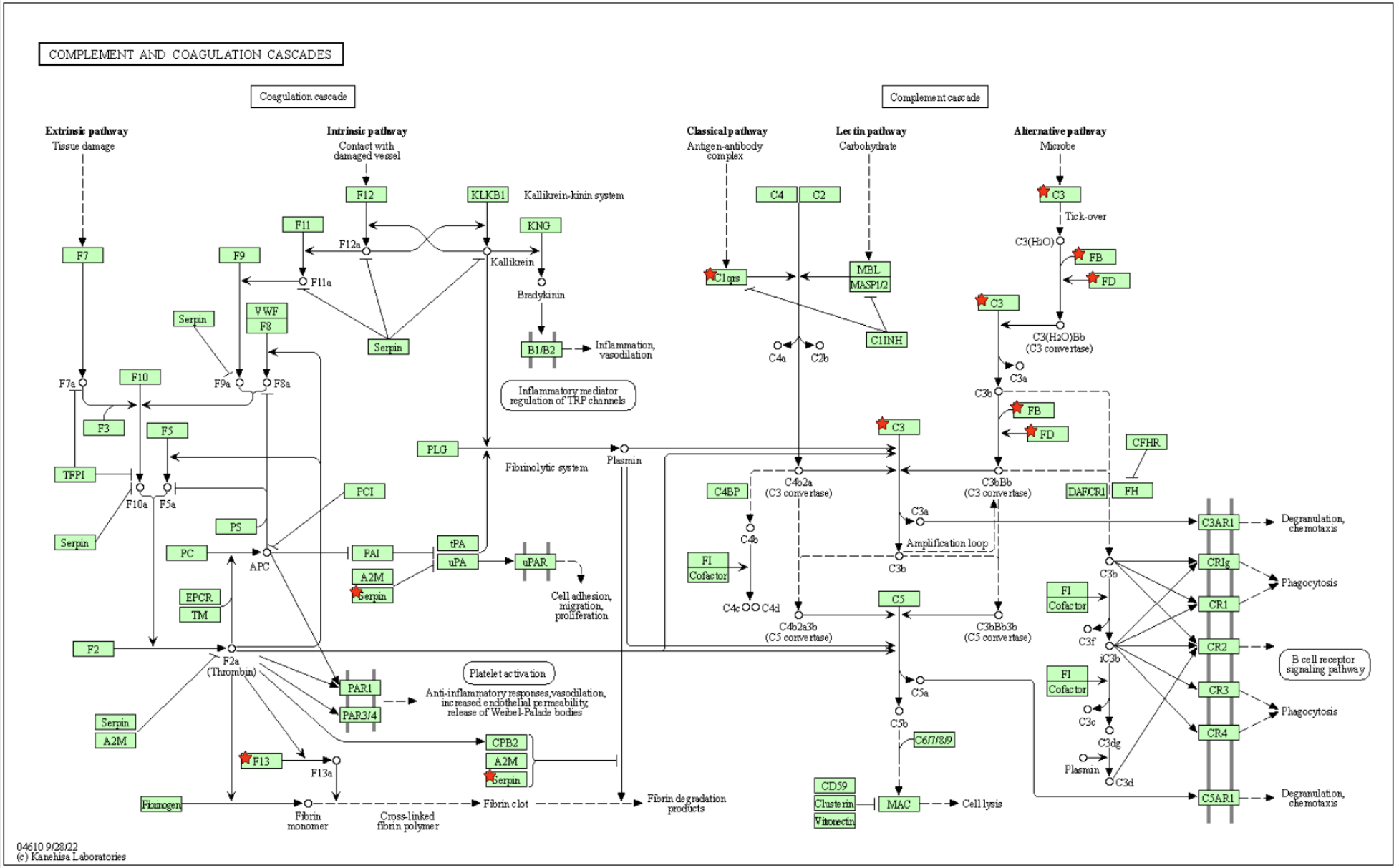

COMPLEMENT AND COAGULATION CASCADES
Coagulation cascade
Complement cascade
Extrinsic pathway
Tissue damage
Intrinsic pathway
Contact with damaged vessel
Classical pathway
Antigen-antibody complex
Lectin pathway
Carbohydrate
Alternative pathway
Microbe
F12
KLKB1
Kallikrein-kinin system
F11
KNG
F12a
Kallikrein
F7
F9
F11a
Bradykinin
Serpin
VWF
F8
B1/B2
Inflammation, vasodilation
Serpin
C4
C2
C1qrs
MBL
MASP1/2
C1INH
C4a
C2b
C3
Tick-over
C3(H2O)
FB
FD
C3(H2O)Bb
(C3 convertase)
C3a
C3b
F7a
F10
F9a
F8a
Inflammatory mediator regulation of TRP channels
F3
F5
PLG
Fibrinolytic system
Plasmin
C3
FB
FD
CFHR
TFPI
F10a
F5a
PCI
C4BP
C4b2a
(C3 convertase)
C3bBb
(C3 convertase)
DAF/CR1
FH
Serpin
PS
C3a
C3AR1
Degranulation, chemotaxis
PC
APC
PAI
tPA
uPA
uPAR
C4b
Amplification loop
FI
Cofactor
C3b
CRIg
Phagocytosis
A2M
Serpin
Cell adhesion, migration, proliferation
C4c
C4d
C5
FI
Cofactor
CR1
EPCR
TM
C4b2a3b
(C5 convertase)
C3bBb3b
(C5 convertase)
C3f
CR2
B cell receptor signaling pathway
F2
F2a
(Thrombin)
iC3b
FI
Cofactor
CR3
Phagocytosis
PAR1
Platelet activation
PAR3/4
Anti-inflammatory responses,vasodilation, increased endothelial permeability, release of Weibel-Palade bodies
C5a
C5b
C6/7/8/9
FI
Cofactor
CR4
Serpin
A2M
C3c
C3dg
CPB2
A2M
Serpin
F13
F13a
Plasmin
C3d
CD59
Clusterin
Vitronectin
MAC
Cell lysis
C5AR1
Degranulation, chemotaxis
Fibrinogen
Fibrin monomer
Cross-linked fibrin polymer
Fibrin clot
Fibrin degradation products
04610 9/28/22
(c) Kanehisa Laboratories
List genes are shown in red

Figure 2.3 Biological pathway of complement and coagulation cascades based on the KEGG database. This pathway was generated from the list of the list of DEGs that were significantly different in relation to tick infestation in at least three skin GEX studies (n = 100). The stars show the DEGs which are involved in this pathway. These are: coagulation factor XIII A chain (F13A1); complement C1q A chain (C1QA); complement C1q B chain (C1QB); complement C1q C chain (C1QC); complement C3 (C3); complement factor B (CFB); complement factor D (CFD); serpin family A member 1 (SERPINA1).

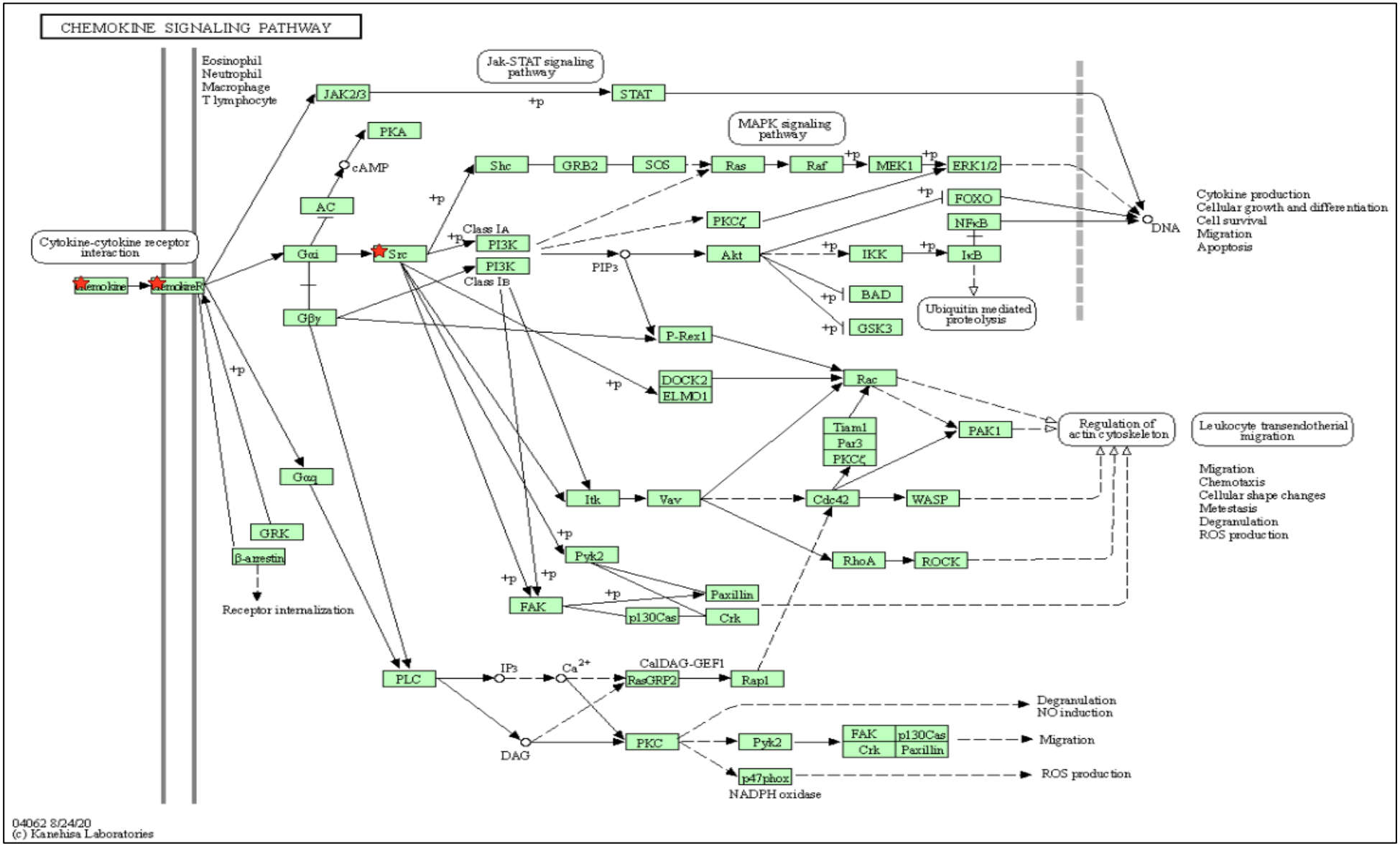

CHEMOKINE SIGNALING PATHWAY
Eosinophil
Neutrophil
Macrophage
T lymphocyte
Jak-STAT signaling pathway
JAK2/3
STAT
+p
MAPK signaling pathway
PKA
cAMP
AC
Shc
GRB2
SOS
Ras
Raf
MEK1
ERK1/2
FOXO
NFκB
IκB
IKK
BAD
GSK3
Ubiquitin mediated proteolysis
PKCζ
Class IA
PI3K
Class IB
PIP3
Akt
Cytokine-cytokine receptor interaction
chemokine
chemokineR
Gαi
Gβγ
Src
DNA
Cytokine production
Cellular growth and differentiation
Cell survival
Migration
Apoptosis
P-Rex1
DOCK2
ELMO1
Rac
Tiam1
Par3
PKCζ
PAK1
Regulation of actin cytoskeleton
Leukocyte transendothelial migration
Migration
Chemotaxis
Cellular shape changes
Metastasis
Degranulation
ROS production
Gαq
GRK
β-arrestin
Receptor internalization
Itk
Vav
Cdc42
WASP
RhoA
ROCK
Pyk2
FAK
Paxillin
p130Cas
Crk
PLC
IP3
Ca2+
CalDAG-GEF1
RasGRP2
Rap1
DAG
PKC
Degranulation
NO induction
Migration
ROS production
p47phox
NADPH oxidase
04062 8/24/20
(c) Kanehisa Laboratories

Figure 2.4 Biological pathway of chemokine signaling by KEGG. This pathway was generated from The list of genes that were significantly different according to level of tick resistance in at least two blood GEX studies (n = 114). The stars in the figure show the DEGs which occur in this pathway. These are cytokine- chemokine receptors including C-X-C motif chemokine ligand 8 (CXCL8); LYN proto-oncogene; Src family tyrosine kinase (LYN); chemokine (C-C motif) ligand 2 (CCL2); chemokine (C-C motif) ligand 3(CCL3); chemokine (C-C motif) ligand 8 (CCL8); chemokine (C-C motif) receptor 1 (CCR1); chemokine (C-X-C motif) ligand 2 (CXCL2); and chemokine (C-X-C motif) ligand 5 (CXCL5).

2.5 Discussion

2.5.1 Integration of information from multiple sources of GEXS and GWAS about host resistance of cattle to infestation with ticks

Integration of data from previous studies allowed an updated perspective on genes and biological pathways associated with resistance of cattle to tick infestation. A previous study (Porto Neto et al. 2010) had taken a similar approach but at the time there were no RNA-seq studies available and they only compared two GWAS and one GEX study; an important contribution of my study was to determine whether additional genes or pathways could be identified using more updated genomic information and integration across a larger number of studies. The final list of genes that I identified was based on DEGs significant in at least four skin GEXS (n=6) or three blood GEXS (n=10) studies; QTLs significant in at least two independent GWAS (n=11); and QTLs that were also significantly differentially expressed in at least two skin GEXS and one blood study (n= 10). Five of the genes that I identified (LAPTM5, ALPL, DHRS7, SERINC5, SIRPA) were also in the top 10 genes identified by Porto-Neto et al. (2010) but my list included a broader range of categories. Not only immune genes but genes in other biological pathways were found in my list, including immune function (n = 12), transcription factors (n = 3), extracellular matrix and structural proteins (n = 9) and genes with a range of other functions (n = 13). These findings provide a list of 37 potential genes with a high likelihood of being truly involved in variation in the resistance trait, which might be useful to identify biomarkers of tick resistance. This insight could improve cattle's resistance to ticks and increase the accuracy of genomic selection and animal breeding management.

2.5.2 Transcription factors

de Souza et al. (2018) used a manually curated compendium method to present 865 sequence-specific DNA-binding transcription factors (TFs) in bovines. These TFs were analyzed for domain family evolution, conservation, and tissue-specific expression. The authors also identified a list of putative transcription cofactors based on their interactions with the characterized TFs. This report provided general information about the regulation of gene expression and improves the understanding of regulatory mechanisms that are specific within and between species. According to de Souza et al. (2018), I identified three transcription factors (TFs) that appear to be associated with tick resistance in cattle (HOXD1, SALL4, and SATB2). The HOXD1 (homeobox D1) gene cluster belongs to the HOX gene family, which is located on chromosome 2 in cattle. Paul et al. (2011) reported that in bovine

HOX genes act as transcription factor which influence in the body axis such as limb and bone from the gastrulation stage. The results of quantitative PCR of bovine oocytes and early embryos showed the relative expression of HOXD1. In addition, HOXD1 was found in mouse oocytes and embryos. The results conclude that HOX gene in mammalian could be affected in the stage of oocyte maturation and primary processes of embryo differentiation (Paul et al., 2011) Sal-like protein 4 transcription factor (SALL4), a member of the SALL family located on chromosome 13 in cattle, this gene can express SALL4 protein. SALL4 protein is a transcription factor of stem cell that play a role in sustaining and renewing of embryonic and hematopoietic stem cells, the protein of this gene functions as an oncogene in various cancers. Despite it is also important in the biological characteristic of childhood acute lymphoblast leukaemia (Ohadi et al., 2020). In humans, SALL4 controls the progression of tumors in relation to the immune microenvironment associated with the TNF family and epigenetic mechanisms of gene expression and effects on hematopoiesis. Thus, SALL4 influences the oncogenic function in both gene transcription and the growth of tumors (Sun et al., 2022). SATB2 (SATB homeobox 2), which is located on chromosome 2 in cattle. SATB2 is a member of the SATB family, which is regulated in several processes, such as chromatin processing, exerting precise control over gene expression. SATB shows differential expression in various human tissues, especially in stem and progenitor cells. It also expresses in different type of cancers and different stages of cancers. Moreover, SATB2 is able to govern the essential gene expression for cell differentiation (Roy et al., 2020). The expression of SATB2 appears to be tissue-specific expression and regulated by various cellular signaling molecules and mediators. It also plays a crucial function in the development of both craniofacial patterns and skeletal systems (Huang et al., 2022). Among these three TFs, only the HOXD1 gene has been investigated in bovines, whereas SALL4 and SATB2 still lack information in cattle. However, the SATB2 also was reported in a study by Marima (2022), which showed the SNPs in this gene associated with tick resistance in cattle. Unfortunately, these genes, which act as transcription factors, still lack information regarding their association with high and low resistance groups of cattle during tick infestation.

2.5.3 Immune response genes

Twelve genes from the 37 genes common to multiple studies were associated with immune function: BOLA, CD84, CR2, PTGR1, C1QA, CXCL8, F13A1, FN1, SERINC5, LAPTM5, PLA2G7, TYROBP. The major histocompatibility complex (MHC), also known as bovine leukocyte antigen (BoLA) on chromosome 23. BoLA is part of a large and highly

polymorphic gene family associated with both innate and adaptive host immune responses (Behl et al., 2012). Many reports have shown the association of allelic diversity in BoLA genes in high and low tick resistance level in cattle. For example, BoLA -DRB3 (BoLa class II DRB3 locus) has been investigated in various cattle breeds and associated with infectious disease resistance, including tick infestation (Acosta-Rodriguez et al., 2005, Machado et al., 2010, Maritz-Olivier et al., 2012, Otto et al., 2018). Mantilla Valdivieso et al. (2022) showed that there was no specific pattern of BoLA-DRB3 expression with resistance phenotype in Brangus cattle; however, there was upregulation of the gene in both high and low resistance phenotypes in response to short-term tick exposure. However, the use of particular BoLA genetic variants in selection programmes might not be advisable, as diversity in the MHC is regarded as valuable for population health (Tabor et al., 2017).

Coagulation factor XIII A chain or F13A1 is the last molecule that links both intrinsic and extrinsic coagulation pathways, activates fibrin clot cross-linking, and has been found to be upregulated in resistant cattle, consistent with the view that the higher efficiency of the coagulation process (Carvalho et al., 2014). In the classic pathway (antigen-antibody complex) of the coagulation cascade, C1QA or complement C1q is the key factor of the classical complement pathway and the main link between innate immune system and adaptive immune system mediated by IgG or IgM (Ghebrehiwet et al., 2018). The chemokine signalling pathway contributes to host response and inflammatory response by recruiting leukocytes, including eosinophils, neutrophils, macrophages, and T lymphocytes to the site of inflammation (Qian et al., 2010). A previous study reported that the coagulation cascade demonstrated resistance against tick larvae by preventing their attachment, leading to early larval death (Kitsou et al., 2021). It also limited hematophagy and reduced the reproductive success of adult female ticks in tick-resistant cattle. Upregulation of expression of the CC ligand subfamily (such as CCL2, CCL3, CCL8), and the *CXC* ligand subfamily (such as CXCL2, CXCL5, and CXCL8*)*, results in upregulation of the CCR1 and CXCR1 receptors, respectively. Long term activation of this reaction can cause chronic inflammatory conditions (Gao et al., 2016). The upregulation of CXCL8 also has been shown in previous blood and skin gene expression studies in relation to tick burden in resistant cattle (Moré et al., 2019). SERINC5 encodes proteins which play a role in restriction factors against the early step of viral infection and significantly reducing viral infectivity, as demonstrated by Shi et al. (2018). This protein can be found in the plasma membrane and the surface of T-cell lymphocytes through flow cytometry (Usami et al., 2015). However, this gene still lacks information in the bovine studies. Porto Neto et al. (2010b) and Marima (2022) found SERINC5 in the list of genes associated with tick infestation in cattle. Mantilla Valdivieso

et al. (2022) investigated the transcription of leukocytes before and after tick infestation in Brangus cattle. The results showed that the leukocyte genes associated with the immune system response, cell signaling pathways, and complement and coagulation cascades were downregulated in both high-resistance and low-resistance groups during early and prolonged tick challenges. That study proposed that the downregulation of leukocyte genes (complement and coagulation cascades and cell signaling) might be due to the migration of cell populations expressing these genes out of the blood circulation. Raza et al. (2023) compared serum samples from naïve tick-resistant and tick-susceptible Brangus cattle at two-time points of tick infestation. The authors demonstrated increased levels of several proteins involved in the immune response, including complement factors, in the tick-resistant group. The results from that study also confirmed the findings of the previous transcriptomics study by Mantilla Valdivieso et al. (2022), which identified a significant increase in several complement factor proteins in the resistant cattle group. Interestingly, the list of genes from my data showed findings consistent with previous literature such as FN1, C1QA, F13A1, and chemokine receptor genes demonstrating that complement and coagulation cascades involved in the immune response play an important role in tick resistance in cattle.

2.5.4 Extracellular matrix and structural proteins

Another finding of this study was that genes associated with extracellular matrix and structural proteins were found to be associated with tick resistance in multiple studies, including: ECM1, ITGA11, MYOC, DNAH14, MYO5C, RHOT1, DKK1, SRGN, and VIPAS39. Extracellular matrix protein 1 or ECM1 encodes a variety of extracellular and structural proteins, associated with the maintenance of skin integrity and homeostasis (Oyama and Merregaert, 2017). Many reports showed the upregulation of the genes involved in ECM, including keratins and collagens after tick challenge in cattle breeds that are resistant to ticks. Both keratin and collagen proteins play roles in epidermal barrier formation to maintain skin integrity, control inflammation, and wound healing. The genes in this group were reported that showed significantly higher in resistant cattle when compared to susceptible cattle (Carvalho et al., 2010, Piper et al., 2010). The ITGA11 (Integrin alpha 11) gene shows a biological role in the control of cellular adhesion and migration and also plays a role in modulating cellular immune responses, by influencing the recruitment and adhesion of immune cells at sites of ectoparasite infestation (Kongsuwan et al., 2010). Porto Neto et al. (2010a) confirmed the location of a QTL affecting tick burden on BTA10, which is close to the ITGA11 gene. DKK1 (Dickkopf-related protein 1), is expressed in the endometrium

of cattle, sheep, and humans and has been shown to be more highly expressed in fertile animals than infertile animals.(Tríbulo et al., 2019). Ras Homolog family member T1 (RHOT1), also known as MIRO1 (Mitochondrial Rho GTPase 1), is associated with mitochondrial function and plays a role in the regulation of mitochondrial dynamics, including mitochondrial transport and distribution within cells. It interacts with the cytoskeleton and motor proteins to facilitate the movement of mitochondria along microtubules within the cell (Safiulina et al., 2019, Grossmann et al., 2019). In addition, RHOT1 also found in the early pregnancy stage in the results of transcript study in peripheral blood immune cells in heifers of *B. indicus* (Rocha et al., 2020). Previous studies have found upregulated gene expression involved in the ECM or extra cellular matrix, including collagen and keratin genes (COL1A1, KRT1, KRT3, and KRT17) (Piper et al., 2010, Marima et al., 2020, Kongsuwan et al., 2008). However, these studies compared different breeds of cattle (*Bos taurus* vs. *B. indicus*), not within the same breed. The group of ECM proteins showed a significant increase in tick resistance compared to tick susceptibility in naïve cattle. Collagen (COL1A1) plays an important role in tick resistance in cattle by activating platelet aggregation, which leads to wound healing and plays a vital role in skin integrity. Similarly, keratins are important for intracellular signaling pathways, protection from stress, and wound healing through activated keratinocytes (Zhang et al., 2019). A study on Belmont Red cattle (Kongsuwan et al., 2010) reported that KRT5 and KRT14, which represent epidermal keratins, showed upregulated gene expression in a resistant group of cattle. This finding suggests that the epidermal layer might play a crucial role in tick resistance in cattle. Therefore, ECM interactions have a remarkable impact on skin integrity, wound healing, and platelet activation as part of the host response to ticks in cattle. The proteomics study of Raza et al. (2023) found that both systemic and cutaneous levels are important in tick resistance in cattle, especially during early tick exposure (six hours after tick infestation). The results from the skin proteomics data demonstrated that proteins associated with the immune response were effectively delivered to the skin, along with ECM proteins related to skin integrity maintenance and wound healing, which were upregulated at the early stage of tick infestation. The results of this study suggested that the increased abundance of these proteins at both systemic and cutaneous levels in resistant cattle are able to develop to a robust host response to tick infestation. My finding that ECM genes might be potential genes or biomarkers associated with tick resistance in cattle is consistent with the results of these previous studies.

In the next chapters of this thesis (Chapters 3 and 4), I aimed to characterise the candidate genes identified in this chapter by using SNP genotypes and transcriptome profiles

(using Oxford Nanopore sequencing) in order to determine genetic variation of individual animal from different cattle breed, and different environment conditions. This will provide a better understanding of the genetic pathway of genes associated with host defence to pathogens in cattle. This knowledge could benefit genetic improvement programs by incorporating more biological insights into genomic prediction models, which have been shown to improve prediction accuracy across populations and breeds of cattle.

Chapter 3

Genotypic variation in Scottish cattle for candidate genes associated with tick resistance from previous studies

3.1 Abstract

The resistance of animals to pathogens is a complex developmental process that involves the regulation of hundreds to thousands of genes, leading to modifications in numerous biological signalling pathways. Breeding animals that are resilient to pathogens might thus require focus not only on genes related to immune function but on other biomarkers for resistance. The aim of this study was to: (1) obtain SNP genotypes of individual cattle obtained from a Scottish abattoir for the list of candidate genes for tick resistance that were previously identified from gene expression studies and genome-wide association studies (GWAS) in chapter 2; (2) test whether there are significant differences in genotype frequencies for each SNP between individuals classified as British, European or Hill cattle breeds; (3) summarise the genomic context of significant SNPS; and (4) classify the animals into genotype groups at each of the genes, to enable testing of whether genotypic variation relates to differences in transcriptome variation (chapter 4). Due to the impacts of COVID-19, the study was limited to collecting and analyzing Scottish *Bos taurus* breed samples, which are likely to be susceptible (low resistance) to ticks. Additionally, no tick infestation or phenotyping was conducted in these cattle because *R. microplus* is absent in Scotland. As a result, this study provided an opportunity to investigate the role that the selected genes play in various biological processes, but not their association with tick resistance. I genotyped DNA extracted from spleen samples of 34 individual cattle (obtained from an abattoir in Scotland) using the GeneSeek Genomic Profile (GGP) Bovine 100K SNP chip. The animals were grouped into three categories of breeds that might be expected to show differences in selection pressures due to differences in husbandry and environmental conditions: British (n=14), European (Continental) (n=10), and British Hill (n=10). The variants identified on the SNP chip were filtered for a panel of 37 candidate genes, which were compiled based on a meta-analysis of previous gene expression and genome wide association (GWAS) studies (Chapter 2). From the 88 SNPs identified from this list of candidate genes (Appendix C), with 14 spread across six genes (HOXD1, SATB2, GIMAP7, ITGA11, PLA2G7, and PRKG1) displaying significant differences in genotype frequencies among breeds. Even though the individuals compared were all *Bos taurus*, the finding of the loci that showed significant differences in genotype frequency among breeds based on the

subset of genes in Chapter 2 could provid into Chapter 4 that was about gene expression profiling. Additionally, the substantial differences in genotype frequencies suggest that the subset of genes identified could have potential use as biomarkers for selection. In order to assess this further, this set of SNPs was used to interpret transcriptome profiles from spleen and skin samples from the same individuals, in order to identify biological pathways associated with genotype variation at these candidate genes (Chapter 4).

Keywords: Single nucleotide polymorphism (SNP), genotyping, Scottish cattle breeds, weighted gene co-expression network analysis

3.2 Introduction

Approximately 80% of the global cattle population faces problems from ticks and diseases transmitted by ticks, leading to a substantial reduction in production. In 1996, economic losses due to ticks and tick-borne diseases were estimated to range from US$13.9 billion to US$18.7 billion annually (de Castro, 1997). More recently, estimates from 2015 suggest that the annual losses have increased to approximately US$20 – 30 billion per year (Lew-Tabor and Rodriguez Valle, 2016). However, tick control through the selection of tick-resistant cattle is currently limited due to inconsistencies in phenotyping methods and the high costs associated with identifying variations in resistance among individual animals (Kongsuwan et al., 2008). The discovery of variation associated with complex traits such as tick resistance is limited by different attributes including the nature of the trait, methodological challenges of the studies and the time and resources required to perform experiments with sufficient power. The results from previous studies in gene expression studies and GWAS have not explained all of the genetic determinates underlying the expression of these complex traits (Aschard et al., 2012, Frazer et al., 2009). Heritability estimates have also varied extensively, emphasising that environmental factors can also substantially alter interpretation of loci associated with complex traits (Otto et al., 2018).

Traditionally, the process of selective breeding of animals relied on estimated breeding values (EBV) derived from pedigree and physical characteristics (Ibeagha-Awemu et al., 2008, Hayes et al., 2009, Guerrero et al., 2012). Although genomic selection is generally suggested as a solution for improvement of complex traits that are difficult and costly to measure, it requires identification of variable gene regions that are associated with desirable traits (Mkize et al., 2021). Genotyping techniques have been developed to make genome-wide comparisons of variation based on single nucleotide polymorphisms (SNPs),

that can be tested for association with economic traits in cattle such as milk yield, meat production or resistance to pathogens and parasites (Iung et al., 2019, Liu et al., 2020, Braz et al., 2021, Johnston et al., 2020, Otto et al., 2018). The development of SNP genotyping technology in terms of costs and accessibility has led to an increase in the use of genome-wide association studies (GWAS) to develop selective markers for breed improvement. For example, using SNP markers in breeding to enhance tick resistance (resulting in fewer ticks on cattle) can provide breeders with valuable insights, aiding them in making well-informed breeding decisions (Mapholi et al., 2016, Mapholi et al., 2014). With the application of GWAS, multiple studies have been carried out to examine the genetic variation associated with tick resistance among various cattle breeds and locations. Studies that have been conducted to date have presented evidence of the association of various genomic regions with low tick load in cattle and recommended the validation of the discovered regions (Acosta-Rodriguez et al., 2005, Gasparin et al., 2007, Machado et al., 2010, Otto et al., 2018, Sollero et al., 2017). However, some of the challenges associated with only relying on GWAS include different phenotyping methods and genotyping strategies across studies, as well as obtaining enough statistical power to test for associations. It would be more powerful to combine loci identified by GWAS with gene expression studies that directly test for changes in responses to challenges with parasites in relation to particular loci.

From my previous study (Chapter 2), I integrated information about differential gene expression (DEG) from gene expression studies (GEXS) and genomic regions identified from previous genome-wide association studies (GWAS) associated with bovine resistance to infestation with the complex of cattle ticks, in order to generate a list of credible candidate biomarkers (genes or gene products) for host resistance to ticks. From the 10,495 DEGs and 288 QTLs, I identified 37 potential candidate genes that showed consistent differences in expression across studies, tissues and/or breeds. The list of genes included 3 transcription factors, 12 genes associated with immune function, 9 associated with the extracellular matrix genes, as well as 15 others. The list of genes was determined in relation to host resistance in geographic regions where exposure to ticks is high, but they could reflect more general responses to ectoparasites, which potentially could be used for genomic selection of breeds from other regions. However, in order to provide a basis for selection, it is first necessary to establish whether any of these genes show sufficient genetic variation within breeds that they could be used for selection.

As an opportunistic test of gene variation, I focused on breeds of *Bos taurus* that are farmed in the UK, which can be grouped into historically British or European (or Continental) types (Hall and Bunce, 2019). When compared to Continental European breeds in terms of characteristics, British breeds generally reach a mature size at an earlier age and are thus smaller, with less growth potential and thus yield carcasses with less saleable product, but they have relatively high fertility and calve easily, as well as producing a high quality of products (Cundiff et al., 1998). From previous studies, phylogenomic analyses indicate that the aurochs, an ancestral wild cattle species, is genetically distinct and serves as an outgroup to the domestic lineage of *B. taurus*, thereby supporting the prevalent belief that European cattle originated primarily from the Near East; in contrast, traditional British and Irish breeds show a greater number of genetic variants shared with the aurochs, compared to other European populations (Park et al., 2015). This result suggests a localised genetic exchange between aurochs and the ancestors of modern British and Irish cattle, potentially resulting from intentional restocking efforts by early herders in Britain. However, there has been some study of gene flow between British breeds and European breeds (Park et al., 2015, Todd et al., 2011). There has also been increasing interest in using more ancient and robust breeds (classified as Hill cattle in this study) as ecosystem engineers for rewilding projects or to minimise the necessity for relying on cultivated grains for feedstocks (McHugo et al., 2019). Due to the increased time spent grazing on unmanaged (or minimally managed) lands, these breeds might be expected to differ from the more intensively farmed breeds in terms of environmentally imposed selection pressures, such as exposure to pathogens or beneficial microbial communities. Although this study was limited to collecting and analysing Scottish *Bos taurus* breeds that are likely to be susceptible to ticks because *R. microplus* is a pest of tropical environments, there are likely to be differences in selection pressures due to variations in how the various breeds are maintained including differences in environmental conditions. It would thus be interesting to determine if signatures of selection can be observed at genes that have been associated with responses to ectoparasites in previous studies. The objectives of this study were to: (1) obtain SNP genotypes of individual cattle obtained from a Scottish abattoir for the list of candidate genes for tick resistance that were previously identified from gene expression studies and genome-wide association studies (GWAS) in Chapter 2; (2) test whether there are significant differences in genotype frequencies for each SNP between individuals classified as British, European or Hill cattle breeds; (3) summarise the genomic context of significant SNPS (e.g. proximity to an amino acid changing mutation); and (4) classify the animals into genotype groups at each

of the genes, to enable testing of whether genotypic variation relates to differences in transcriptome variation (chapter 4).

3.3 Materials and Methods

3.3.1 Sample collection

This study was approved by the School of Biodiversity, One Health and Veterinary Medicine Animal Ethics Committee (certificate number EA51/21, 15 December 2021). The samples were collected from animals slaughtered at an abattoir in Scotland on 29 September 2022. The animals were then classified into three groups: British - Aberdeen Angus (n=11), Shorthorn (n=3); European - Charolais (n=1), Limousin (n=8), Simmental (n=1); and Hill - Galloway (n=9), Highland (n=1) (Table 3.1). The animals were between the ages of one and two years but otherwise, selection of individuals was opportunistic. Approximately five mm of skin, spleen and lymph node samples were collected from individual animals; the tissue samples were cut using a small blade and the size was estimated by naked eye. The samples were immersed in 1.8 mL of RNAlater (ThermoFisher, USA) in 2 mL cryotubes prior to storage at 4°C for 24 h for permeation of tissues and then transferring to a -80°C freezer until further use.

Table 3.1 List of animals collected from Wishaw abattoir, This table includes their sex, the number of individuals collected, and classification into Breed groups (British, European or Hill).

Breed	Sex	Number	Breed Group
Aberdeen Angus	M	10	British
	F	1	British
Shorthorn	M	1	British
Shorthorn	F	2	British
Charolais	F	1	European
Limousin	F	8	European
Simmental	M	1	European
Galloway	M	9	Hill
Highland	M	1	Hill

3.3.2 DNA extraction

Genomic DNA was extracted from the spleen tissue samples (n= 34). The tissues were removed from storage in RNAlater, the excess RNAlater was absorbed with a Kimwipe® and samples were then transferred to 800 µL of DNA/RNA Shield reagent (Zymo research Inc, California USA) in a 2 mL tube containing a bead mix (2 mm beads). Tissues were homogenized twice in a TissueLyser II (QIAGEN Inc, Paisley, UK) at 30 Hz for 30 s, as per the manufacturer's instructions for the use of the Quick DNA/RNA Miniprep Plus kit (Zymo research Inc, California, USA). The containing tube was centrifuged at 10,000 g for 30 s. The clear supernatant was collected and transferred into a new tube, then an equal volume of DNA/RNA Lysis Buffer was added to the supernatant (1:1) and mixed well. Solutions were transferred into a Spin-Away™ Filter1 (yellow) in a collection tube and centrifuged at 10,000 x g for 30 seconds. A Spin-Away™ Filter1 was transferred to a new collection tube, 400 ul DNA/RNA Prep Buffer was added to the column and the tube was centrifuged. The flow-through was discarded. Then, 700 ul of DNA/RNA buffer was added to the column and the tube centrifuged, discarding the flow-through. Then, 400 µl of buffer was added into the column and the column was centrifuged for 2 minutes to ensure complete removal of the wash buffer, before transferring to a clean collection tube. To elute the DNA, 50 µl of nuclease free water was added directly to the column matrix and centrifuged. The eluted DNA was either used immediately or stored at -20˚C. DNA was quantified in the Nanodrop 2000 Instrument (ThermoFisher Inc, Oklahoma, USA). The absorbance ratios 260/230 nm and 260/230 nm were also recorded to check the DNA quality.

3.3.3 Genotyping

Genotyping of genomic DNA from all 34 animals was performed using the *Neogen GeneSeek® Genomic Profiler™ Bovine 100K* chip (GGP Bovine 100K). This array offers the features from the most commonly used commercial genotyping arrays, comprehensive parentage, disease, and trait-relevant SNPs by Neogen Europe Ltd. **(**Auchincruive House, Auchincruive, Ayr, KA6 5HN, Scotland, United Kingdom). The principle of SNP chips is based on nucleotide bases binding to their complementary partners (i.e. A pairs with T and C pairs with G), according to Watson-Crick base pairing. Fragmented single-stranded DNA hybridizes to arrays containing hundreds of thousands of unique nucleotide probe sequences, each designed to bind to a specific target DNA sequence. After hybridization, the specialised equipment measures the signal intensity related to each probe and its target. This signal intensity reflects the amount of target DNA in the sample and the affinity between the target

and the probe. Extensive processing and analysis of these raw intensity measures yield SNP genotype inferences (LaFramboise, 2009).The 34 samples were sent to the company, as 35 µl, at a concentration of 15-20 ng/µl of good quality and largely intact DNA (minimum 5 kb), with the DNA stored in individual tubes capped tightly by using parafilm on snap top tubes to wrap the lids to prevent accidental spillage or cross contamination. DNA samples in tubes were shipped at room temperature for overnight delivery.

3.3.4 Data processing

The BEDTools software (Quinlan and Hall, 2010) was used to find the SNPs data that were in the genes in he *Bos taurus* genome (ARS-UCD1.2) available on the Ensembl database (Figure 3.1). All subsequent analyses were performed using the R Statistical Software (v4.2.2) (RStudio Team, 2022). The packages used for analysis in RStudio environment included *tidyverse* (Hadley Wickham, 2019), *fuzzyjoin* (Robinson, 2020), *caroline* (Schruth, 2023), *data.table* , and *xlsx* (Arendt, 2022). After the data was cleaned and filtered, unique genes from the ARS-UCD1.2 genome were intersected by gene name with the list of 37 candidate genes identified in Chapter 2. This generated a list of candidate genes which had the location site in the genome and the chromosome number; this file was written in BED4 format. The SNP Map obtained from the Neogen company (Neogen Inc, Ayr, Scotland) was modified by expanding 1000 bp at the start and end site of the SNP location. Then, it was intersected with the BED4 Format of the candidate genes using the BEDtools library. From the SNP report, I excluded SNPs which showed a GC score of less than 0.6 to select for high genotyping accuracy and ordered the data by chromosome. The modified SNP report was merged with the SNP Map by SNP name to include the location of the SNP. The last step was the intersection between the list of modified candidate genes and the modified SNP report by SNP name.

In order to determine the position of the significant SNPs within their gene to enable assessment of linkage between SNPs and possible nonsynonymous mutations, the length of the gene, the location of the SNP (i.e which intron or exon), the length between the SNP and its closest exon (if it was located in an intron) were recorded. In addition, the consequences of the type of mutation of significant SNPs were predicted from the database in Ensembl. The function and biological pathways were then identified for the significant SNPs , using DAVID, as described in chapter 2.

3.3.5 Statistic analysis

Chi-square contingency (χ^2) analyses were used to test the hypothesis that there is no association between the genotypic frequencies of the different SNPs and the breed group of the cattle. The contingency chi-square test is a statistical method applied to determine the association between variation of genotypes and phenotypic traits or diseases. It aims to analyse if the distribution of genotypes differs significantly between groups of animals in the genetics association study. The Null Hypothesis (H0) is defined as there is no association between genotypes and the phenotype traits or disease. Alternative Hypothesis (H1) is defined as there is an association between the genotypes and the phenotype trait or disease. By comparing the observed and expected frequencies of different genotypes in each group of animals if there were no associations, it could interpret whether certain genotypes are linked to specific phenotypic traits or diseases (Fikret Isik, 2017).

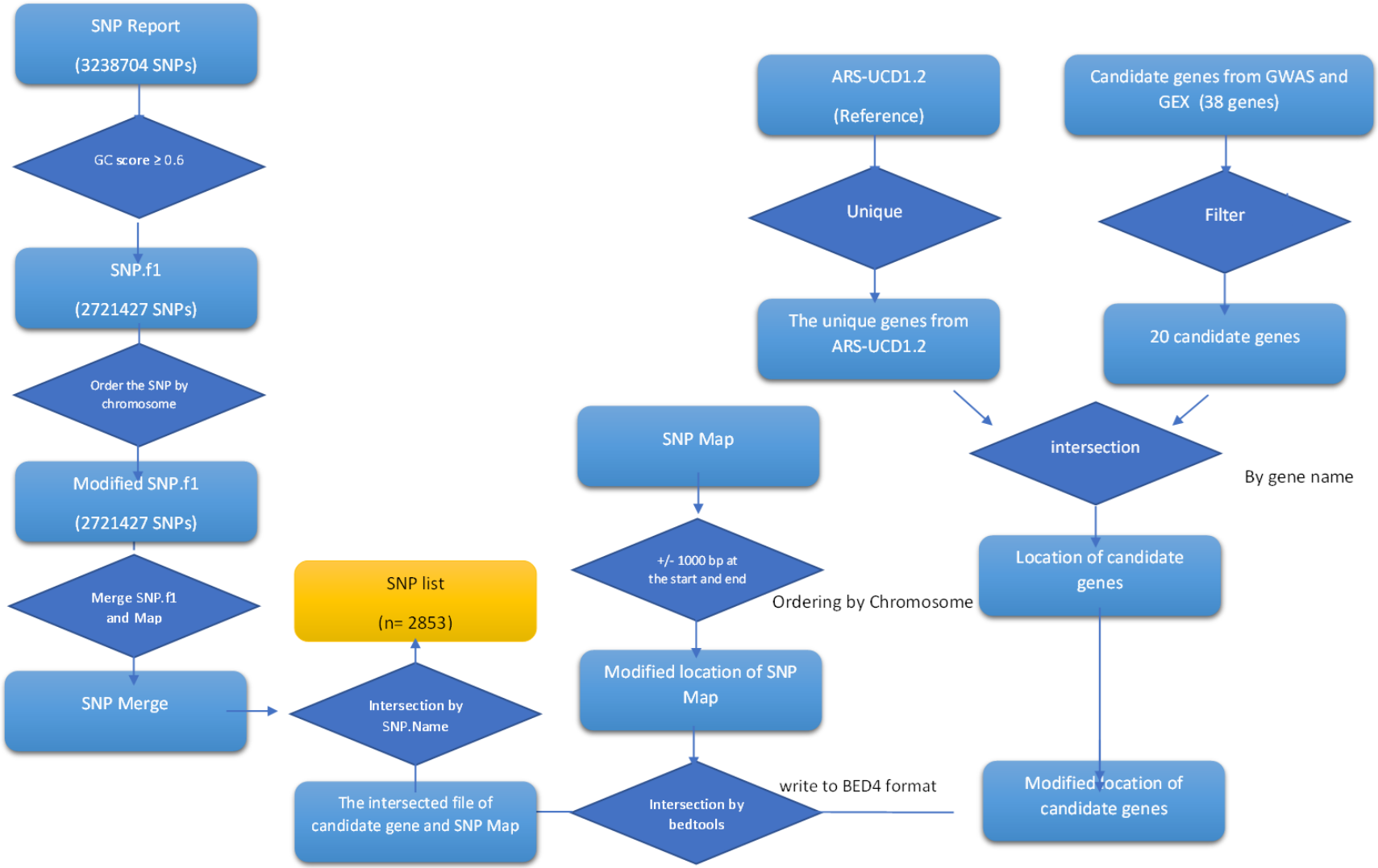

SNP Report
(3238704 SNPs)
GC score ≥ 0.6
SNP.f1
(2721427 SNPs)
Order the SNP by chromosome
Modified SNP.f1
(2721427 SNPs)
Merge SNP.f1 and Map
SNP Merge
SNP list
(n= 2853)
Intersection by SNP.Name
The intersected file of candidate gene and SNP Map
SNP Map
+/- 1000 bp at the start and end
Modified location of SNP Map
Intersection by bedtools
write to BED4 format
ARS-UCD1.2
(Reference)
Unique
The unique genes from ARS-UCD1.2
Candidate genes from GWAS and GEX (38 genes)
Filter
20 candidate genes
intersection
By gene name
Location of candidate genes
Ordering by Chromosome
Modified location of candidate genes

Figure 3.1 Flowchart summarising the steps towards identification of SNPs from the candidate genes associated with tick resistance in cattle. The SNPs report was provide by Neogen company in the xls. File, the data was analysed in R studio (RStudio Team, 2022)

3.4 Results

From the total variants identified on the SNP chip report from the GGP Bovine 150K SNP array (Neogen Inc, Ayr, Scotland), 2,721,427 were retained after filtering for quality. After filtering to only consider the 37 candidate genes identified in chapter 2, a total of 88 SNPs were identified across the samples (Supplement 3.1). There are14 SNPs from six genes showed significant differences in genotype frequencies ($p < 0.05$ based on X2 tests) among three Scottish *Bos taurus* breeds (British, European, and Hill cattle) (Table 3.2).

The list of the significant SNPs detected from the candidate genes associated with immune function and other biological processes of three Scottish *Bos taurus* breeds in Table 3.2 includes one SNP on HOXD1, two SNPs on SATB2, one on GIMAP7, one SNP on ITGA11, two SNPs on PLA2G7, and seven SNPs on PRKG1. Included SNPs (Table 2) showed a minor allele frequency (MAF) of more than 0.1 based on the Bovine Genome Variation Database (BGVD) (Chen et al., 2020). The sample size for some SNPs varied, in cases where an individual animal did not meet the quality threshold score for that SNP (Table 3.2)

Figure 3.2 presents bar plots of the 14 significant SNPs in different cattle breeds, with the X-axis showing the three Scottish *Bos taurus* breeds (British, European, and Hill cattle), and the Y-axis the observed genotype frequencies. For SATB2-1 (rs135885260), one of the homozygous genotypes (AA) was not observed for any of the breed groups but Hill cattle showed a much lower frequency of heterozygotes than the other two breeds. For SATB2-2, British cattle showed a higher frequency of one of the homozygous genotypes than heterozygotes and compared to the other breeds. For GIMAP7 (rs29011115), PLA2G7-1 (rs110518949), ITGA11 (rs110930784) one of the homozygous genotypes was observed at much lower frequency than the other and the heterozygous genotypes. For PLA2G7-1 and GIMAP7, there was a high frequency of heterozygotes (AB) for both British and Hill cattle, whereas there were no heterozygous (AB) genotypes observed in European cattle but an excess of the homozygous genotype that was rare in the former. ITGA11 showed a similar

paucity of heterozygotes in European cattle but a higher frequency of the more common homozygous genotype found in the other breeds. HOXD1 also showed an excess of one of the homozygous genotypes compared to the other breed groups. For PLA2G7-2, British cattle showed a different pattern than the other breed groups. Intriguingly, the seven SNPs found at PRKG1 showed different patterns of variation, sometimes even between adjacent SNPs (Figure 3.2).

Table 3.2 List of the significant SNPs detected from the candidate genes associated with tick resistance in three Scottish *Bos taurus* breeds. In the Gene name column, some gene showed the SNPs more that one location such as PLA2G7, PRKG1, and SATB2. SNP name showed the name of SNPs from the company.(GGK) The breed groups include British, European, and Hill cattle and the genotypes show homozygous genotype in AA and BB, heterozygous in AB. The cut-off P-value ≤ 0.05

Gene Name	Chr. no.	SNP Name	British AA	British AB	British BB	European AA	European AB	European BB	Hill AA	Hill AB	Hill BB	chi-square	df	p_value
GIMAP7	4	Hapmap59459-rs29011115	1	10	3	0	1	9	0	7	3	13.222	4	0.01024
HOXD1	2	ARS-BFGL-NGS-118009	1	7	2	7	2	0	1	4	5	16.693	4	0.002217
ITGA11	10	ARS-BFGL-NGS-119197	5	8	1	9	1	0	3	6	0	9.5654	4	0.04842
PLA2G7-1	23	ARS-BFGL-NGS-111955	1	12	1	0	0	9	1	5	4	19.802	4	0.000546
PLA2G7-2	23	BovineHD2300005156	1	10	3	5	5	0	6	3	1	9.7143	4	0.04553
PRKG1-1	26	BovineHD2600001782	1	5	8	0	1	9	8	1	1	24.216	4	7.23E-05
PRKG1-2	26	BovineHD2600001618	6	8	0	3	5	1	1	2	7	16.835	4	0.002081
PRKG1-3	26	BovineHD2600001964	6	7	1	5	4	0	2	2	6	13.315	4	0.009834
PRKG1-4	26	BTB-01078331	6	7	1	9	0	1	9	0	1	12.669	4	0.01301
PRKG1-5	26	BovineHD2600001842	0	8	6	3	3	3	1	1	8	11.028	4	0.02625
PRKG1-6	26	ARS-BFGL-NGS-73895	3	6	4	2	3	4	8	2	0	10.715	4	0.02996
PRKG1-7	26	BTA-113588-no-rs	0	3	11	2	4	4	3	5	2	9.5638	4	0.04845
SATB2-1	2	BovineHD0200025079	0	6	8	0	8	2	0	1	9	9.9529	2	0.006899
SATB2-2	2	BovineHD0200025097	8	5	1	1	4	4	1	7	2	10.842	4	0.0284

Table 3.3 The location and consequences type of the significant SNPs detected from the candidate genes associated with tick resistance in three Scottish *Bos taurus* breeds. SNP name column in this table show the universal of SNP name. The consequence type show the type of mutation of nucleotide.

Gene name	SNP Name from GGK	SNP name	Chr.	Position	Alleles	GeneID	Consequence Type
GIMAP7	Hapmap59459-rs29011115	rs29011115	4	113276469	G/A	ENSBTAG00000008550	intron variant
HOXD1	ARS-BFGL-NGS-118009	rs43293677	2	20725551	T/G	ENSBTAG00000015840	intron variant
ITGA11	ARS-BFGL-NGS-119197	rs110930784	10	15189729	A/G	ENSBTAG00000008380	synonymous variant
PLA2G7-1	ARS-BFGL-NGS-111955	rs110518949	23	19933485	A/G	ENSBTAG00000019315	missense variant
PLA2G7-2	BovineHD2300005156	rs109090865	23	19938377	G/A	ENSBTAG00000019315	intron variant
PRKG1-1	BovineHD2600001782	rs135984060	26	7529454	G/T	ENSBTAG00000018404	intron variant
PRKG1-2	BovineHD2600001618	rs136047022	26	7056812	G/T	ENSBTAG00000018404	intron variant
PRKG1-3	BovineHD2600001964	rs42235055	26	8012170	C/A	ENSBTAG00000018404	intron variant
PRKG1-4	BTB-01078331	rs42234268	26	7875261	C/G	ENSBTAG00000018404	intron variant
PRKG1-5	BovineHD2600001842	rs136992315	26	7669914	A/G	ENSBTAG00000018404	intron variant
PRKG1-6	ARS-BFGL-NGS-73895	rs109511653	26	6941707	A/G	ENSBTAG00000018404	intron variant
PRKG1-7	BTA-113588-no-rs	rs41572003	26	7907227	C/T	ENSBTAG00000018404	intron variant
SATB2-1	BovineHD0200025079	rs135885260	2	87929185	T/C	ENSBTAG00000016334	intron variant
SATB2-2	BovineHD0200025097	rs109918690	2	87992072	A/G	ENSBTAG00000016334	intron variant

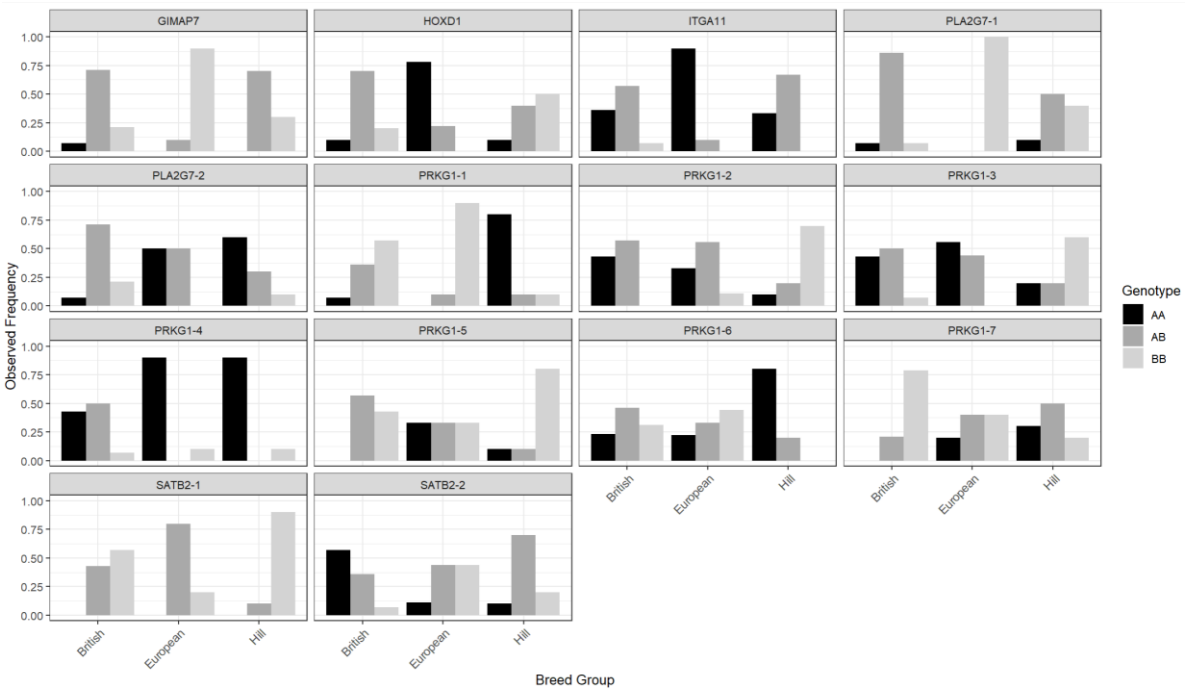


Figure 3.2 Bar plot of the significant SNPs that were identified across the panel of 37 candidate genes in British (n=14), European (n=10), and Hill (n=10) cattle. The x-axis shows the breed group, and the y-axis shows the observed frequency of the genotypes. Where more than one SNP was detected per gene, this is indicated by a dash, followed by the number, which corresponds to their sequence in Table 3.2. The numbering of the SNPs within genes was set before their location was determined. AA and BB refer to homozygous and AB refers to heterozygous genotypes.

There were 14 SNPs that showed p-values less than 0.05 from the chi-square tests; these spanned three "consequence types", including missense mutations, synonymous variants, and intron variants. ITGA11 (rs110930784) presents a synonymous variant in exon no. 5 and PLA2G7-1 (rs110518949) presents a missense mutation in exon no. 8 (Table 3.3). The rest of the significant SNPs are located in introns. GIMAP7 (Hapmap59459-rs29011115) has two exons; the intron variation is located in intron no. 1 (113,282,312), HOXD1 (rs43293677) has two exons and shows only a short length between the SNP and its closest exons (154 nt and 185 nt, respectively); the intron variant is located in intron no. 1. PLA2G7 has 16 exons; PLA2G7-2 (BovineHD2300005156) is located in intron no. 4., which is 337 nt from the nearest exon and 4,892 nt from the misssense mutation in exon 8. PRKG1 is the largest gene in this study, with 20 exons and large introns; the intron variants are located on intron numbers 3,8,2,2,3,10, and 2, respectively, but all are quite distant from exons (Table 3.4).

Table 3.4 Summary of the genomic context of the 14 significant SNPs in each gene to show the length between exons and SNPs. This table shows the location of gene on chromosome, the length of the genes, number of exon in genes, the location and length between SNP location and nearest exon (5' exon and 3' exon).

SNP name	Chr.	Genomic location of gene	Gene Length (nt)	No. of exons	5' exon location (exon:position)	Length between 5' exon and SNP (nt)	SNP location	SNP in intron or exon	3' exon location	Length between 3' exon and SNP
GIMAP7	4	113,276,469-113,287,210	10,742	2	1:113,276,515	5,797	113,282,312	intron 1	2:113,285,486	3,174
HOXD1	2	20,724,251-20,726,458	2,208	2	1:20,725,705	154	20,725,551	intron 1	2: 20,725,366	185
ITGA11	10	15,128,896-15,262,135	133,240	30		na	15,189,729	exon 5*		na
PLA2G7-1	23	19,925,024-19,963,288	38,265	16		na	19,933,485	exon 8 **		na
PLA2G7-2	23	19,925,024-19,963,288	38,265	16	4:19,940,032	1,655	19,938,377	intron 4	5:19,938,040	337
PRKG1-1	26	6,895,298-8,313,728	1,418,431	20	3:779,5001	265,547	7,529,454	intron 3	4:7,411,643	117,811
PRKG1-2	26	6,895,298-8,313,728	1,418,431	20	8:7,071,721	14,909	7,056,812	intron 8	9:7,022,540	34,272
PRKG1-3	26	6,895,298-8,313,728	1,418,431	20	2:8,132,649	120,479	8,012,170	intron 2	3:7,795,114	217,056
PRKG1-4	26	6,895,298-8,313,728	1,418,431	20	2:8,132,649	257,388	7,875,261	intron 2	3:7,795,114	80,147
PRKG1-5	26	6,895,298-8,313,728	1,418,431	20	3:7,795,001	125,087	7,669,914	intron 3	4:74,11,643	258,271
PRKG1-6	26	6,895,298-8,313,728	1,418,431	20	10:6,950,757	9,050	6,941,707	intron 10	11:6,926,186	15,521
PRKG1-7	26	6,895,298-8,313,728	1,418,431	20	2:8,132,649	225,422	7,907,227	intron 2	3:7,795,114	112,113
SATB2-1	2	87,840,935-88,044,294	203,360	12	6:87,942,802	13,617	87,929,185	intron 6	7:87,916,825	12,360
SATB2-2	2	87,840,935-88,044,294	203,360	12	6:87,942,802	49,270	87,992,072	intron 6	7:87,916,825	75,247

* synonymous mutation

** missense mutation

3.5 Discussion

From the results of this chapter, we found variation in the genotypes of six genes involved in immune function and transcription factors in three Scottish *Bos taurus* breeds. According to the study by Mallard et al. (2015) the genotype of immune response genes is affected by the environment and epigenetics, meaning animals from different environments could exhibit different immune response phenotypes to pathogens. For this reason, the goal of this chapter was to demonstrate the genotypic variation of immune function genes and other genes influenced by pathogens in British, European, and Hill cattle, although the phenotypes of individual animals were not clarified. Chapter 4 identified whether there is differential expression of these genes in skin or spleen tissues and used weighted gene co-expression network analysis to identify the range of pathways with which they are associated. This study was conducted to explore the genotypic and its potential impact on resistance and susceptibility phenotypes. SNP genotyping was applied to characterise the genetic variation. The results revealed patterns of genetic diversity among the different breeds, highlighting the role of both historical selection pressures and environmental adaptations.

Despite the fact that the study was limited to *Bos taurus* in Scotland, considerable variation was apparent in the number of individuals presenting a specific genotype in the different cattle breed groups for the 14 SNPs from six genes that were identified as being significantly differentiated. However, only a single amino acid changing variant was identified (PLA2G7-1), with the other 13 being either synonymous changes within an exon (ITGA11) or located in introns. Linkage to mutations under selection that were not included on the SNP chip could explain why the identified genes showed differential gene expression in previous studies in relation to tick exposure (Chapter 2). Intron variants also can increase transcription levels by altering the rate of transcription, nuclear export processes, and stability of transcripts, and can also increase the efficiency of mRNA translation (Shaul, 2017). However, only two of the intron variants were located close enough to an exon that tight linkage to an amino acid changing mutation would be expected. For example, the HOXD1 SNP, which was located in the first intron, was only a short distance from its flanking exons, which could mean that it is indirectly associated with expression of the gene. The function of HOXD1 was mentioned in Chapter 2. In bovines, HOXD1 transcripts have been identified in oocytes and early embryos at various frequencies depending on the stage of development (Paul et al., 2011). The variation between breed groups detected in this study thus could reflect their expected differences in rates of development and morphology

The synonymous variant in ITGA11 could also be linked to a functional mutation. However, synonymous mutations can still have effects on processes like mRNA stability, translation efficiency, and protein folding (Sauna and Kimchi-Sarfaty, 2011). The ITGA11 (integrin alpha 11) gene plays a role in regulating cellular adhesion, as well as cellular immune responses (Porto Neto et al., 2010a). In humans, ITGA11 acts as a multifunctional integrin in different pathways, including collagen binding and attachment of muscle tissue to the extracellular matrix (Zeltz and Gullberg, 2016). The significant differences detected between breed groups here, in particular the lower number of heterozygotes and excess of one of the homozygotes in European compared to the other breed groups, could again reflect differences in development and morphology. ITGA11 has also been linked to tick burden in various dairy cattle breeds, including Australian Red, Brown Swiss, Channel Isle, Holstein, composites, and Brahman beef cattle (Porto Neto et al., 2010a). It is interesting that GIMAP7, which is part of a large immune related gene family, shows a similar pattern of variation between the breeds as ITGA11which is high level of heterozygotes in British and Hill cattle.

GTPases of immunity-associated proteins (GIMAPs) are integral regulators of lymphocyte survival and homeostasis. The expression of GIMAP2 and GIMAP7 shows distinct regulation in various human T cell lymphoma lines. GIMAP4, GIMAP5, and GIMAP7 have functions associated with immune response and hematopoiesis (Schwefel et al., 2013, Chen et al., 2011). These GIMAPs play a significant role in modulating immune functions by regulating cell death and T cell activation (Ho and Tsai, 2017). However, the information about the GIMAP7 gene in cattle is limited.

The single amino acid changing mutation identified (one of the two SNPs in PLA2G7) was classified as missense which changed the codon from ATC to ACG (Methionine to Threonine); however, it could not predicted to result in loss or changing function of the gene. In humans, PLA2G7 plays a critical role in the regulation of adipose tissue metabolism related to immunity (Candels et al., 2022), and it is also a biomarker of diffuse large B cell lymphoma (Zheng et al., 2021). Interestingly, although there is no information about the function of this gene in cattle, it showed different patterns of genotype frequencies for the breed groups compared here: there was an excess of heterozygotes for British cattle but an excess of one of the homozygotes for European cattle. This could be evidence for differential selection of these genes but further study would be required to test this hypothesis.

Although the sample size used in this study was small, there were some other interesting patterns of differences in genotype frequency between the cattle breed groups. For example, SATB2 and HOXD1 are both homeobox transcription factors that have been implicated in regulating immune responses and other biological pathways (Roy et al., 2020, Dobreva et al., 2003, Li et al., 2020, Zhang et al., 2023); it is thus interesting that they both showed the absence of one of the expected homozygous genotypes in at least one of the breeds, which could suggest immune-related selection against one of the alleles. SATB2 is associated with the immune system in relation to binding the matrix attachment region of the immunoglobulin micro locus and regulation of expression of pre-B cells(Dobreva et al., 2003). In addition, SATB2 has been recognized as an oncogene involved in carcinogenesis, offering a potential target for both cancer treatment and prevention (Roy et al., 2020). SATB2 plays a crucial role in orchestrating various physiological and pathological processes by modulating gene transcription and expression through the organization of chromatin architecture (Huang et al.,2021 and Shao et al.,2020). A QTL upstream of the SATB2 gene was among the 25 identified locations significantly associated with tick burden in cattle in a previous GWAS study (Turner et al., 2010).

The largest number of significantly differentiated SNPs were found for the PRKG1 gene, which was also the largest of the genes assessed in my study, with 20 exons and very large introns. Despite the small sample sizes used in my study, finding significant variation between three opportunistically sampled British groups of cattle from a single abattoir provides preliminary evidence that the six genes identified here will show sufficient variation to be potential candidates for marker-assisted selection. The differences in genotype frequencies detected in this chapter could be due to differences in development or ecology of the breed groups as well as immune or other functions. Mota et al. (2018) found that the PRKG1 gene was identified near a SNP (ARS-BFGL-NGS-73895; BTA26) that showed a significant effect exclusively in cases of low tick resistance in cattle. This gene has been previously presented in systemic lupus erythematosus, an autoimmune disease affecting multiple organ systems, including the skin, musculoskeletal, renal, and hematologic systems in humans (Kariuki et al., 2015). Gene function of PRKG1 is not supported in other immune cells like T and B cells, it is significant in interferon-α-producing cells. Additionally, PRKG1 has been linked with terms such as Large-Conductance Calcium-Activated Potassium Channel α Subunits, GABA-B Receptors, and Thioredoxin Reductase (Mota et al., 2018).

Comparing results across different studies is challenging due to the complexity of the traits involved and the variations among populations, as well as differences in pathogen prevalence and the methodologies employed. For example, the horn fly, *Haematobia irritans* belongs to the Muscidae family, which includes various parasites and pests that particularly affect domestic animals, especially cattle (Moon, 2009). This fly is an obligatory pest of cattle and is considered the most prevalent ectoparasite affecting them. Basiel et al. (2021) identified three genes associated with horn fly resistance in organic Holstein cattle. Focusing on genes within the genomic region explaining the most variation in their analysis, Basiel identified CHIC2, PDGFRA, GSX2, and KIT, all located on *Bos taurus* autosome 6. CHIC2 is expressed in several bovine tissues involved in cellular vesicles and the plasma membrane (Papatheodorou et al., 2020, Cools et al., 2001). However, this gene still lacks sufficient supporting information for the results. GSX2 is known for its role in neural development and regeneration (López-Juárez et al., 2013). Variants in GSX2 could potentially affect horn fly avoidance behaviours, such as kicking and tail flicking, although there is currently no literature on the variation of these behaviours among individuals. PDGFRA and KIT encode tyrosine kinase cell surface receptors and are physically associated in multiple animals (Lord et al., 1996). These genes are involved in several pathways, such as RET signalling, synovial fibroblast apoptosis, G-protein-coupled receptors, the growth factor receptor-bound protein 2 associated binder 1 signalosome, and extracellular-signal-related kinase signalling (Stelzer et al., 2016). The author highlighted a significant SNP located within the KIT gene associated with low resistance to horn flies in cattle, which potentially plays a role in both coat colouration and resistance to horn flies in Holstein cattle. The Basiel et al. (2021) study did not identify overlapping SNPs or genes with the significant SNPs detected in candidate genes associated with the immune response of transcription factors in the three Scottish *Bos taurus* breed groups from my study. Nevertheless, the list of SNPs from this chapter is interesting to be candidates for investigating their association with tick resistance, horn fly resistance, or other pathogens in Scottish cattle and buffalo in the future.

Chapter 4

Weighted Gene Co-expression Network Analysis (WGCNA) of full-length cDNA sequencing data from skin and spleen samples of cattle

4.1 Abstract

In order for genes to be useful for marker-based selection, they not only have to show variability, but the variation needs to be related to phenotypic differences on which selection can act. One way to predict whether genetic variation at a locus is likely to result in functional changes is to use gene expression data as the phenotype in a model with the genotype at the locus as the potential explanatory variable. The genes I identified in Chapter 2 were based on differential gene expression or GWAS in relation to tick infestation but the gene expression studies were based on skin and blood samples. The purpose of this chapter was to test whether the subset of SNPs identified in Chapter 3 that were significantly different between Scottish cattle breed groups had functional consequences in terms of gene expression in skin and/or spleen samples from the same animals. I used the GridION from Oxford Nanopore Technology (ONT) for full-length cDNA sequencing of transcripts from RNA extracted from skin and spleen samples to identify transcript expression profiles associated with genotype variation at the 14 significant SNPs from Chapter 3. A weighted gene co-expression network analysis (WGCNA) approach was used to identify co-expressed modules that were significantly associated with each of the 14 SNPs, analysed separately for skin and spleen samples. Seven modules showed significant correlations with at least one of the SNPs in the skin dataset and ten modules in the spleen dataset. The genes in these modules represented a range of biological pathways that were affected by genetic variation at the SNPs, but the predicted functional pathways differed substantially between skin and spleen. Three genes (FN1, ATP9A, and ECM1) from the skin dataset and five (CR2, RHOT1, SRGN, LAPTM5 and GIMAP7) from the spleen dataset showed overlap with the list of 37 genes associated with tick resistance that were identified in chapter 2. Overall, the spleen samples showed more complex patterns of co-expression and involvement of a wider range of interacting pathways than the skin samples.

Keywords: cattle, transcriptome profiling, long-read sequencing, Weight gene co-expression network analysis (WGCNA).

4.2 Introduction

The previous studies used to identify candidate genes associated with tick resistance described in Chapter 2 quantified changes in gene expression in relation to ticks. To understand the genetic mechanisms of host response to pathogens or immune system in cattle, it can be helpful to elucidate the global expression profiles of different tissues. Sequencing of genome-wide RNA (RNA-seq) has proven more efficient at providing a global picture of differential expression patterns than older technologies such as microarrays (Rao et al., 2018). For example, the introduction of RNA-seq using next-generation sequencing technology (NGS) has provided a platform to dissect the complex details of tick resistance while addressing the biological heterogeneity of different cattle breeds challenged with various cattle tick species (Colgan et al., 2017). The transcriptional outputs produced by RNA-seq may be instrumental in providing insight into the complex genetic networks mediating host-tick interactions in cattle, as well as other complex biological processes.

In an era of genomics, third-generation long-read sequencing is mostly represented by Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT). Consequently, many -omics technologies are rapidly developing in these fields, significantly enhancing the scientific knowledge accumulated over the past decades with short-read sequencing methods (Jenjaroenpun et al., 2018). The fact that assembled transcripts from short reads do not cover full-length transcripts of genomes is a significant limitation of short-read sequencing. Illumina technology (short-read sequencing) is well-known and widely used in transcriptomic studies; however, the limitations of short-read sequencing can be addressed with long-read sequencing methods (Byrne et al., 2017, Depledge et al., 2019, Amarasinghe et al., 2020). Additionally, during library construction, PCR amplification bias can occur (Tilgner et al., 2013, Puglia et al., 2020). In long-read sequencing, PacBio read lengths exceed 15 kb and ONT read lengths exceed 30 kb, which is more than sufficient to cover most RNA molecules in eukaryotes (Byrne et al., 2017). In ONT long-read sequencing, sample preparation does not require PCR amplification, reducing bias but leading to decreased throughput (Jenjaroenpun et al., 2018, Byrne et al., 2017, Giolai et al., 2017). Previous studies compared the quantification of transcripts between Illumina and ONT. The results demonstrated a high level of correlation in gene expression quantification using both technologies (Cui et al., 2020, Seki et al., 2019, Byrne et al., 2017, Oikonomopoulos et al., 2016). This confirms that results from ONT RNA-seq approaches are comparable to those from short-read sequencing platforms (Stark et al., 2019). The advantages offered by ONT and PacBio make these long-read technologies highly appealing and suitable choices for

discovering isoforms and fusion transcripts, as well as for conducting *de novo* transcriptomic analysis.

PacBio introduced a DNA sequencing method with its Single Molecule Real-Time (SMRT) technology, developed by BioSciences. This technique offers longer read lengths compared to second-generation sequencing (SGS) technologies, making it highly suitable for overcoming limitations in -omics studies. PacBio sequencing can address gaps in existing reference assemblies and analyze structural variations (SV) in personal genomes. Its longer reads facilitate sequencing through extended repetitive regions and the identification of mutations, including those associated with diseases (Anthony and Kin Fai, 2015).

In recent years, long-read sequencing has become more and more applicable. Oxford Nanopore Technologies (ONT) is one of the most popular providers of long-read sequencing technology (Burgess, 2018, Pollard et al., 2018). Nanopore technology relies on a method of sequencing which uses a nanoscale protein pore in an electrically resistant polymer membrane to identify the sequence of nucleotides of DNA or RNA molecules (Wang et al., 2021). The nucleic acid molecules pass through the nanopore, and the nucleotides are detected as they pass through the pores by measuring changes in the ionic current. ONT has been used to analyse both full-length cDNA and mRNA samples from tissue and individual cells. Nanopore sequencing is able to generate reads as long as 30 kb (recently represented by the ONT GridION and PromethION instruments (Oxford Nanopore Technologies Inc.). An advantage of RNA-seq based on long-read data is that it can also be used for rare isoform discovery and isoform expression quantification (Byrne et al., 2017, Dong et al., 2021, Jain et al., 2016, Wang et al., 2021), in addition to comparing levels of gene expression. Grünberger et al. (2022) investigated ONT RNA-seq in *Escherichia coli,* comparing the capabilities and advantages of different ONT-RNA sequencing protocols such as direct RNA, direct cDNA, and PCR-cDNA. Their findings indicated that PCR-cDNA-seq offers improved yield and accuracy compared to direct RNA sequencing. Importantly, PCR-cDNA-seq is well-suited for quantitative measurements and enables simultaneous and precise detection of transcript 5' and 3' boundaries, analysis of transcriptional units, and exploration of transcriptional heterogeneity. In simple comparisons of discrete groups, differential expression analysis can be informative to discover genes or pathways associated with particular traits. However, for more complex comparisons based on multiple groups, alternative approaches could have more power. Weighted gene co-expression network analysis (WGCNA) is one of the most effective gene clustering biological methods, which can be used to associate multiple genes with multiple phenotypes or pathways (Chai et al.,

2022). It generatesthe gene-gene interactions based on similarity of expression that is used to identify set of co-expressed (modules), which can then be correlated with the levels of the categories hypothesized to influence expression (e.g. phenotypes, tissues, species) (Geng et al., 2020). The representative genes in the module are summarized by the term “eigengene,” representing the overall gene expression. This is based on the assumption that co-expressed genes, which share similar functions, suggest their involvement in the same biological pathway (Langfelder and Horvath, 2008). WGCNA is also able to link gene modules to specific metadata characteristics; for example, providing potential insights into the molecular mechanisms of a disease. These methods have been proven to be effective in several biological studies, such as in cancer research, mouse genetics, and brain imaging analysis (Langfelder and Horvath, 2008). It has also been applied to predict immune function in humans associated with variation in a particular candidate gene (Zhong et al., 2022). However, there is also potential to use this approach to look at the influence of multiple genes on global gene expression profiles.

The purpose of this chapter is to explore the expression of several genes of interest from the preceding work (Chapter 2 and Chapter 3) in skin and spleen using long-read cDNA sequencing and determine whether this varies in relation to their genotypes for the SNPs that were found to differ significantly between breed groups (Chapter 3). Specifically, I assessed the potential functional significance of variation in genotypes at these genes and whether this differed by the tissue type. Thus, the aims of this chapter were to: 1) generate full-length cDNA sequencing data from the same individuals used in chapter 3, by using Oxford Nanopore Technology (ONT; 2) use these data to identify sets of genes that show similarity in expression patterns (co-expression modules) in relation to genotypes for each of the 14 SNPs identified in chapter 3; 3) identify the main biological functions associated with the significant co-expression modules; and 4) determine what proportion of the 37 genes I identified in chapter 2 are associated with these co-expressed complexes in the two tissue types.

4.3 Materials and Methods

4.3.1 RNA extractions

Skin and spleen biopsies from the 34 sampled cattle (see Chapter 3) were removed from storage in RNAlater, and the excess RNAlater was absorbed with a kimwipe. The skin sample slices were immersed in liquid nitrogen (LN) for 1 minute. The skin sample was put in the mortar and ground 3-4 times with a pestle, then immediately immersed in LN for 30 seconds. The grinding steps were repeated three to four times until the tissue was completely disrupted. Care was taken to avoid sample thawing in between steps (Piper et al., 2008). Then, the skin tissues and spleen slices were then transferred to 800 µL of DNA/RNA Shield reagent (Zymo research, Irvine, CA,USA) in a 2mL tube containing a bead mix (2 mm beads). Tissues were homogenized in a TissueLyser II (QIAGEN, Germantown, MD, USA) at 30 Hz for 30 sec twice. Total RNA isolation was performed as per manufacturer's instructions for the use of the Quick DNA/RNA Miniprep Plus kit (Zymo research, Irvine, CA, USA).

The quality and quantity of each sample was checked by using a Nanodrop 2000 Instrument (ThermoFisher, Waltham, MA, USA). Samples showing concentrations of less than 50 ng/µl and 260/280 absorbance ratios less than 1.7 were re-extracted.

Table 4.1 Sample organization table for full-length cDNA sequencing using a GridIron, This table indicates the individual, its breed and sex, along with which flow cell it was run on. Each flow cell includes 12 samples; each number refers to both skin and spleen tissues for the six individuals in each batch. I also included technical replication in each run by including the same reference sample (Animal No. 5) in Flow cell 1,3, and 5.

Animal No.	Breed	Sex	Flow Cell 1	Flow Cell 2	Flow Cell 3	Flow Cell 4	Flow Cell 5	Flow Cell 6
5	Aberdeen Angus[a]	F	X		X		X	
21	Aberdeen Angus[a]	M	X					
22	Aberdeen Angus[a]	M		X				
23	Aberdeen Angus[a]	M		X				
24	Aberdeen Angus[a]	M			X			
25	Aberdeen Angus[a]	M				X		
26	Aberdeen Angus[a]	M				X		
29	Aberdeen Angus[a]	M					X	
32	Aberdeen Angus[a]	M						X
33	Aberdeen Angus[a]	M						X
34	Aberdeen Angus[a]	M						X
4	Beef Shorthorn[a]	F					X	
27	Beef Shorthorn[a]	F					X	
9	Belted Galloway	M					X	
7	Charolais[b]	F					X	
1	Dairy Shorthorn[a]	M						X
8	Galloway[c]	M	X					
10	Galloway[c]	M	X					
11	Galloway[c]	M		X				
12	Galloway[c]	M		X				
13	Galloway[c]	M			X			
14	Galloway[c]	M			X			
15	Galloway[c]	M				X		
16	Galloway[c]	M				X		
20	Highland[c]	M						X

Animal No.	Breed	Sex	Flow Cell 1	Flow Cell 2	Flow Cell 3	Flow Cell 4	Flow Cell 5	Flow Cell 6
2	Limousin[b]	F	X					
3	Limousin[b]	F	X					
17	Limousin[b]	F		X				
18	Limousin[b]	F		X				
19	Limousin[b]	F			X			
28	Limousin[b]	F			X			
30	Limousin[b]	F				X		
31	Limousin[b]	F				X		
6	Simmental[b]	M						X

Note: [a] refers to British cattle

[b] refers to European cattle

[c] refers to Hill cattle

4.3.2 Overview of the PCR-cDNA barcoding protocol for SQK-PCB109 Kit

The PCR-cDNA barcoding protocol for the kit SQK-PCB109 from Oxford Nanopore technology (Oxford Nanopore Technologies, 2024) was used for full-length cDNA sequencing, using R9.4.1 flow cells. The protocol provided by the Oxford nanopore community website was used, with individual components purchased separately from the various suppliers. Each flow cell of this kit could run only 12 samples, so I decided to run both skin and spleen in the same flow cell for 6 individuals at a time (Table 4.1). For this protocol, total RNA of ~50 ng was required. Using too little or too much RNA, or RNA of poor quality (e.g. fragmented or containing chemical contaminants) can affect your library preparation.

Library preparation (Reverse transcription and strand switching)

Total RNA (~50 ng) was added to a 0.2 ml PCR tubes, with 1 µl of 2 µM VN Primers (VNP) (Oxford Nanopore Technologies, 2024) and 1 µl of 10 mM dNTPs; the total volume was adjusted to 11 µl with Rnase-free water (Table 4.2). The reaction was mixed by flicking

the tube, and the tube was spun down. The reaction was incubated at 65° C for 5 minutes and then snap cooled on a pre-chilled freezer block.

Table 4.2 Preparation of reactions for reverse transcription and strand-switching

Reagent	Volume
~50 ng total RNA	x µl
VN Primers (VNP), at 2 µM	1 µl
10 mM dNTPs 1	1 µl
Rnase-free water	9-x µl
Total volume	11 µl

In a separate tube, 4 µl of 5x reverse transcription (RT) Buffer, 1 µl of RNaseOUT, 1 µl of Nuclease-free water, and 2 µl of Strand-Switching Primers (SSP, at 10 µM) (Oxford Nanopore Technologies, 2024) were mixed, in a total volume of 8 µl (Table 4.3). The reaction was mixed gently by flicking the tube, and spun down.

Table 4.3 The preparation of reactions for strand-switching buffer

Reagent	Volume
5x RT Buffer	4 µl
RNaseOUT	1 µl
Nuclease-free water	1 µl
Strand-Switching Primer (SSP, at 10 µM)	2 µl
Total	8 µl

Next, 11 µl of the preparing reaction in Table 4,2 was mixed with 8 µl of strand-switching buffer in Table 4.3 by flicking the tube and spinning down. The reaction was incubated at 42° C for 2 minutes and then 1 µl of Maxima H Minus Reverse Transcriptase (200 U/µl) was added. The total volume of 20 µl was mixed gently by flicking the tube, and spinning down. The reaction was incubated according to the protocol summarized in Table 4.4. The first step was activation between total RNA, primers, dNTPs,and RT and strand-switching enzymes at 42° C for 90 minutes. Then, reactions were inactivated by incubation at 85° C for 5 minutes and the reaction then held at 4° C.

Table 4.4 The temperature and incubation times of reverse transcription and strand-switching reactions

Cycle step	Temperature Time	Time
Reverse transcription and strand-switching	42° C	90 mins
Heat inactivation	85° C	5 mins
Hold	4° C	∞

Selecting full-length transcripts by PCR and barcoding of samples

Samples were barcoded, amplified, end-repaired and purified using the protocol provided by SQK-PCB109 (Oxford Nanopore Technologies, 2024). For each sample in the batch for individual flow cells (n=12), 5 µl of reverse-transcribed RNA sample was mixed with 1.5 µl of Barcoding Primers (BP01-BP12) (Technologies, 2008) and 18.5 µl of nuclease-free water. Then 25 µl of 2x LongAmp Taq Master Mix was added, for a total volume of 50 µl. This barcoding reaction was prepared at room temperature. The samples were amplified using the cycling conditions summarized in Table 4.5.

Table 4.5 The PCR cycling conditions for barcoding and amplification of samples

Cycle step	Temperature	Time	No. of cycles
Initial denaturation	95° C	30 secs	1
Denaturation	95° C	15 secs	18
Annealing	62° C	15 secs	18
Extension	65° C	50 secs per kb	18
Final extension	65° C	6 mins	1
Hold	4° C	∞	∞

In order to end-repair the DNA, 1 µl of New England Biolabs (NEB) Exonuclease was added directly to each barcoding reaction in PCR tubes, the reactions were mixed by pipetting, and then incubated at 37° C for 15 minutes, followed by 80° C for 15 minutes. To purify the reactions the samples were bound to AMPure XP magnetic beads, which were resuspended by vortexing, and 40 µl was added to the reactions and mixed by pipetting, then incubated on a Hula rotator mixer for 5 minutes at room temperature. The tube was spun

down for 30 sec and kept on the magnetic rack. The supernatant was then pipetted off, and the tube was kept on the magnet while washing the beads with 200 µl of freshly prepared 70% ethanol without disturbing the pellet. The ethanol was removed using a pipette and discarded. The washing step was repeated. The tube was spun down and placed back on the magnetic rack. Residual 70% ethanol was pipetted off and the tube allowed to dry. The tube was removed from the magnetic rack and the pellet was resuspended in 12 µl of Elution Buffer (EB; Oxford nanopore sequencing kit) and then incubated on a Hula mixer for 10 minutes at room temperature. Pelleted beads were kept on the magnet until the eluate was clear and colourless. The 12 µl of elute which contains the DNA library was removed and transferred to a clean 1.5 ml Eppendorf DNA LoBind tube. 1 µl of the purified DNA was analysed for concentration using a Nanodrop200 (ThermoFisher, Waltham, MA, USA). In a 1.5 ml Eppendorf DNA LoBind tube, the 12 samples were then pooled for a total of 100 fmol of the amplified cDNA barcoded samples to a final volume of 11 µl in elution Buffer (EB; Oxford nanopore sequencing kit).

4.3.3 Adapter addition for amplified cDNA

1 µl of Rapid Adapter (RAP; ONT SQK-PCB109) was added to the amplified cDNA library, mixed by pipetting and spun down. Then, the reaction was incubated for 5 minutes at room temperature, spun down briefly and the library was stored on ice until ready to load.

Priming and loading the SpotON flow cell

From the SQ-PCR109 kit, the Sequencing Buffer (SQB), Loading Beads (LB), Flush Tether (FLT) and one tube of Flush Buffer (FB) were thawed at room temperature, mixed by vortexing and spun down at room temperature. Then, the MinION Mk1B lid was opened and the flow cell was slid under the clip and the priming port cover was turned clockwise to open the priming port. Any air bubbles under the cover were removed by drawing back a small volume by setting a P1000 pipette to 200 µl, inserting the tip into the priming port and turning the wheel until the dial showed 220-230 µl, or until a small volume of buffer entered the pipette tip. The slide was then visually checked to ensure that there was a continuous buffer from the priming port across the sensor array. Next, the flow cell priming mix was prepared by adding 30 µl of thawed and mixed Flush Tether (FLT) directly to the tube (1000 µl) of thawed and mixed Flush Buffer (FB), and mixed by vortexing. Next, 800 µl of the priming mix was loaded into the flow cell via the priming port, avoiding the introduction of air bubbles and waiting for 5 minutes. The loading library was then prepared by mixing 37.5

ul of SQB with 25.5 ul of LB (thoroughly mixed by pipetting just prior to mixing) and 12 ul of the purified and barcoded DNA library (Table 4.6).

Table 4.6 Preparation of the library for loading to flow cells.

Reagent	Volume
Sequencing Buffer (SQB)	37.5 µl
Loading Beads (LB), mixed immediately before use	25.5 µl
DNA library	12 µl
Total	75 µl

The SpotON sample port cover was carefully lifted to expose the SpotON sample port. Subsequently, 200 µl of the priming mix was injected into the flow cell through the priming port (not the SpotON sample port), taking care to prevent air bubbles from entering. Following this, 75 µl of the prepared library (which was gently mixed by pipetting up and down just prior to loading) was gradually added to the flow cell via the SpotON sample port, ensuring each drop was fully absorbed before adding the next..The SpotON sample port cover was gently replaced, making sure the bung entered the SpotON port, closing the priming port and replacing the MinION Mk1B lid. Then, the machine was set via the MiniKnow software, with the programme selected as RNA sample and the SQK-PCB109 Kit, and high accuracy base-calling, with a threshold quality score more than 9.

4.3.4 Data processing workflow

FAST5 files from the GridION were converted to basecalls in FASTQ format using Guppy (V3.3.0) (GPU-accelerated-guppy-basecalling/GPU-accelerated_guppy_v2.png at main asadprodhan/GPU-accelerated-guppy-basecalling (github.com)), and porechop (V0.2.4) (rrwick/Porechop: adapter trimmer for Oxford Nanopore reads (github.com)) was used to trim adapters on the ONT reads. Only reads that passed the quality control by Filtlong (V0.2.1) (GitHub – rrwick/Filtlong: quality filtering tool for long reads) were retained for subsequent analyses the parameters of filter was exclude the sequences which have read lengths less than 25 base pairs. The ONT reads were aligned to the *Bos taurus* reference ARS-UCD1.2 using minimap2 (V2.2.14) (Li, 2018). Transcript expression levels were generated using salmon (v1.10.1) (Patro et al., 2017) and EdgeR (Robinson et al., 2010) was

used to normalize, generate counts per million (cpm) and filter transcripts by abundance level for each library. The cmp data were then selected based on the 14 significant SNPs from 6 genes from Chapter 3 (GIMAP7, HOXD1, ITGA11, SATB2-1, SATB2-2, PLA2G7-1, PLA2G7-2, PRKG1-1, PRKG1-2, PRKG1-3, PRKG1-4, PRKG1-5, PRKG1-6, and PRKG1-7) (Figure 4.1).

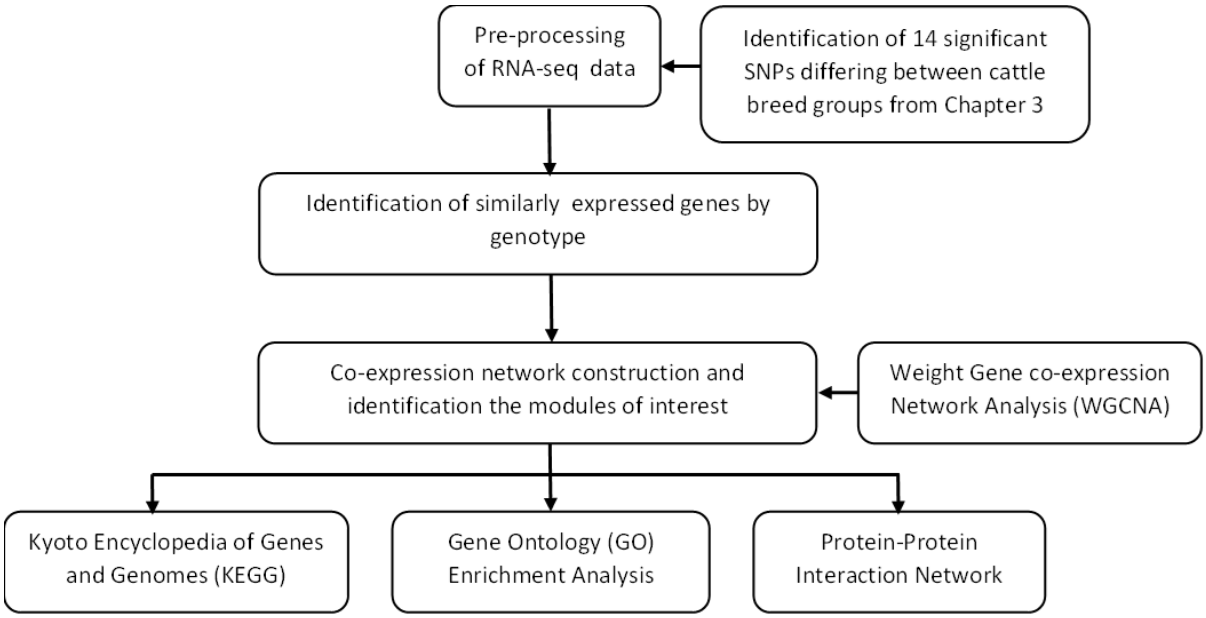


Figure 4.1 Flow chart presenting the work flow from preparation of data, data processing and data analysis

4.3.5 Weighted gene co-expression Analysis (WGCNA) and Identification of Modules of interest

A weighted gene co-expression network was constructed using the WGCNA package in R.3.4.2 (Langfelder and Horvath, 2008) to identify consensus modules for skin and spleen data, separately. The groups of similarly expressed genes were classified by the 14 significant SNPs described in Chapter 3. The concept of this analysis was used to identify co-expression of the genes in each modules. The genes that show the similarity of gene expression profiles were classified in the same modules. All genes which showed the pattern of expression were weighted by 14 significant SNPs this result was converted to a topology overlay matrix (TOM), and modules were detected by cluster analysis during module selection. Module membership (MM) for each SNP was based on the significance of the correlation coefficient between genes and modules and was used to describe the reliability

of a gene belonging to a particular module. Lastly, then, modules that showed a significance ($p < 0.05$) in each significant SNPs were be the modules of interest.

4.3.6 Gene Ontology and pathway enrichment analysis, identification of genes of interest, and protein-protein interaction analysis

I first identified genes involved in co-expression modules that were significantly associated with the focal SNPs that overlapped with the full set of 37 genes identified in chapter 3. I then selected the module that contained the largest number of significant SNPs, one module each from skin and spleen, to visualize protein-protein interaction networks. The relationship between biological pathways and the modules of interest were processed with Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses by using the Database for Annotation, Visualization and Integrated Discovery (DAVID) online tool (http://david. abcc.ncifcrf.gov).

4.4 Results

4.4.1 Data quality

Tables 4.7 and 4.8 show the amount of information obtained after the data processing step in the skin and spleen tissue samples, respectively. Across the 34 skin samples sequenced, there was an average N_{50} of 401, and an average read length of 241,737 bases. The sequencing from the 34 spleen samples showed an average N_{50} of 445 and an average read length of 242,704 bases. The average GC content in the skin sample was 41.61% and in the spleen sample 41.3%. The results of data quality was good enough to continue to the next step.

Table 4.7 Summary of the quality of full-length cDNA sequencing from 34 individual skin samples. This table shows information after the data processing step including the read length (base pair), the minimum and maximum read lengths (base pairs, bp), the mean and median read lengths (base pair), the percent G-C content, N_{50}, L_{50}, N_{90}, and L_{90}.

Sample	reads (bp)	min (bp)	max (bp)	mean (bp)	median (bp)	GC (%)	N_{50} [a]	L_{50} [b]	N_{90} [c]	L_{90} [d]
Skin01	409,646	97	6604	426.4	335	40.3	487	114855	235	325362
Skin02	360,970	94	25254	469.3	348	40.8	551	91649	243	280531
Skin03	229,463	101	72459	528.4	401	43.5	619	58257	269	176227
Skin04	215,143	103	5745	451.3	341	39.9	502	56526	244	169680
Skin05	486,576	88	14488	498.2	349	41.6	615	116460	247	374751
Skin06	293,648	89	8126	398.8	301	41.3	412	79730	228	236694
Skin07	148,650	110	6561	360.9	287	40.5	353	45376	226	122591
Skin08	167,202	89	10555	351.9	284	41.4	346	52108	222	138193
Skin09	281,952	96	13136	363.9	299	42.4	358	90086	232	233697
Skin10	259,109	111	13403	394	309	40.5	393	76650	240	212271
Skin11	242,588	99	13209	372.7	304	40.8	368	76008	234	200110
Skin12	146,141	117	9333	376.9	301	40.9	368	44444	234	120235
Skin13	224,089	107	7604	412.1	322	41.8	429	64058	243	181833
Skin14	372,629	92	8185	350	295	43	357	123414	226	309513
Skin15	277,901	94	11387	347.9	288	43.2	341	91115	228	231882
Skin16	246,887	115	8515	356.5	306	41.3	356	84062	235	206097
Skin17	473,202	109	10974	388.6	311	41.6	396	143733	238	388580
Skin18	243,827	110	7108	425.1	326	42.3	458	67612	243	196496
Skin19	132,298	108	6733	327.2	274	41.9	320	44420	219	110830
Skin20	268,905	103	7803	373.6	296	42.3	370	80556	228	220457
Skin21	210,282	103	5926	366.3	287	42.8	362	63042	227	173259
Skin22	238,863	106	4684	379	312	41	377	75440	238	197089
Skin23	144,580	106	6095	390.9	302	41.9	394	41563	232	118006
Skin24	304,654	95	5515	388.5	308	41.9	389	90062	233	248935
Skin25	129,982	112	4949	400.7	318	40.3	433	37939	234	105082
Skin26	241,165	100	26168	393.9	319	41.6	419	72401	235	195642
Skin27	272,824	98	18943	375.8	300	42.6	385	81651	227	222667

Sample	reads (bp)	min (bp)	max (bp)	mean (bp)	median (bp)	GC (%)	N_{50} [a]	L_{50} [b]	N_{90} [c]	L_{90} [d]
Skin28	249,369	110	19430	329.1	278	40.9	324	84452	220	208711
Skin29	214,261	90	50802	351.2	279	40.9	346	65615	220	176955
Skin30	651,545	99	9911	332.9	273	41.3	322	211731	218	543818
Skin31	62,239	88	25826	370.6	293	40.9	361	18930	233	51400
Skin32	45,106	111	2750	302.8	268	43.3	301	16327	211	38041
Skin33	119,840	109	5183	396.1	331	42.2	406	38765	252	98922
Skin34	54,835	136	5126	399.4	323	42	403	16440	242	44810
Skin35	129,607	97	5527	406	317	41.4	434	38028	242	105712
Skin36	152,549	107	23559	384.5	309	41.8	384	46974	239	125646
Average	241,737	103	13544	387.3	308.17	41.61	401.08	72235.5	233.8	196965

N_{50} [a] is defined as the weight of the median and mean lengths of a set of sequences. It represents the length of the shortest read within the group of longest sequences that collectively account for at least 50% of the nucleotides in the entire set.

L_{50} [b] is defined as the number of congits whose length sum product of $N_{50.}$

N_{90} [c] is defined as the length for which the collection of all contigs of that length or longer contains at least 90% of the sum of the lengths.

L_{90} [d] is defined as the number of congits whose length sum product of N_{90}

(Castro and Ng, 2017).

Table 4.8 Summary of the quality of full-length cDNA sequencing from 34 individual spleen samples. This table shows information after the data processing step, including the overall read lengths (base pair), the minimum and maximum read lengths (base pairs, bp), the mean and median read lengths (bp), the percent G-C content, N_{50}, L_{50}, N_{90}, and L_{90}

Sample	reads (bp)	min (bp)	max (bp)	mean (bp)	median (bp)	GC (%)	N_{50} [a]	L_{50} [b]	N_{90} [c]	L_{90} [d]
Spleen01	438,300	101	22877	525	400	41.5	637	112491	269	336047
Spleen02	111,909	104	7069	571	439	41.5	698	28328	286	84705
Spleen03	64,347	101	7303	503	389	42.5	587	16899	262	49834
Spleen04	218,634	96	90075	430.7	316	41.2	478	56707	236	173887
Spleen05	544,271	97	64322	500.5	371	41.7	604	137999	257	420653
Spleen06	493,294	103	7452	416.7	313	40.9	445	132630	234	395515
Spleen07	157,629	106	15108	384.8	307	40.6	390	47409	235	129103
Spleen08	154,065	109	32957	446.6	339	39.8	493	42096	250	123153
Spleen09	140,708	113	4359	388.2	305	41	392	41394	235	114895
Spleen10	85,775	99	4934	353.1	289	40.9	370	25942	206	69698
Spleen11	161,484	115	12865	448.2	326	40.9	533	41636	239	127304
Spleen12	274,697	97	6261	345.1	295	40.9	342	93396	230	229093
Spleen13	322,530	116	34479	403.3	318	42	430	94417	238	261786
Spleen14	161,425	109	8184	406.3	330	41.2	432	48815	243	131213
Spleen15	370,576	98	15500	387.6	310	42	395	111736	236	303500
Spleen16	187,544	93	4069	363.4	301	41.7	376	58868	226	153479
Spleen17	183,489	108	43822	374.6	290	41.1	387	52907	223	149483
Spleen18	516,257	105	10886	426.2	344	42.2	461	155645	254	419265
Spleen19	242,262	97	14787	396.9	316	41.2	408	72261	239	197800
Spleen20	142,590	91	44235	417.3	328	40.4	443	41427	245	115612
Spleen21	127,283	107	5273	379.4	302	41.8	382	38222	232	104317
Spleen22	128,996	110	9048	393.3	314	41.3	403	38646	238	105367
Spleen23	887,287	94	20891	491.3	387	41.9	562	249054	270	701515
Spleen24	526,189	95	23597	450	358	41.7	504	151556	255	420745
Spleen25	368,461	106	23043	413.3	330	41.4	448	108512	244	297400
Spleen26	104,447	109	5265	373.7	303	40.4	394	31826	226	85160
Spleen27	171,738	101	7000	360.1	291	41.8	368	52534	221	140528
Spleen28	159,603	92	44119	341.8	273	41.4	350	48459	207	131044
Spleen29	302,730	111	12643	400	322	41	431	89968	235	245458
Spleen30	403,472	100	15610	401.4	332	40.6	419	126375	246	330337
Spleen31	142,310	117	10748	428.9	350	41.4	465	43559	257	115701
Spleen32	79,701	109	3470	412.3	343	40.7	440	25245	255	65269
Spleen33	68,871	110	4838	382.7	319	41.9	391	21948	243	56600
Spleen34	61,291	103	5492	300.6	261	41.7	289	22151	213	51925
Spleen35	127,092	141	6063	428	341	40.9	465	37764	252	103035
Spleen36	106,087	134	43909	386.6	325	40.9	394	34681	248	87836

Sample	reads (bp)	min (bp)	max (bp)	mean (bp)	median (bp)	GC (%)	N_{50} [a]	L_{50} [b]	N_{90} [c]	L_{90} [d]
Average	242,704	105.5	19238	412	327.139	41.3	444.61	70375.08	241.3	195229.5

N_{50} [a] is defined as the weight of the median and mean lengths of a set of sequences. It represents the length of the shortest read within the group of longest sequences that collectively account for at least 50% of the nucleotides in the entire set.

L_{50} [b] is defined as the number of congits whose length sum product of $N_{50.}$

N_{90} [c] is defined as the length for which the collection of all contigs of that length or longer contains at least 90% of the sum of the lengths.

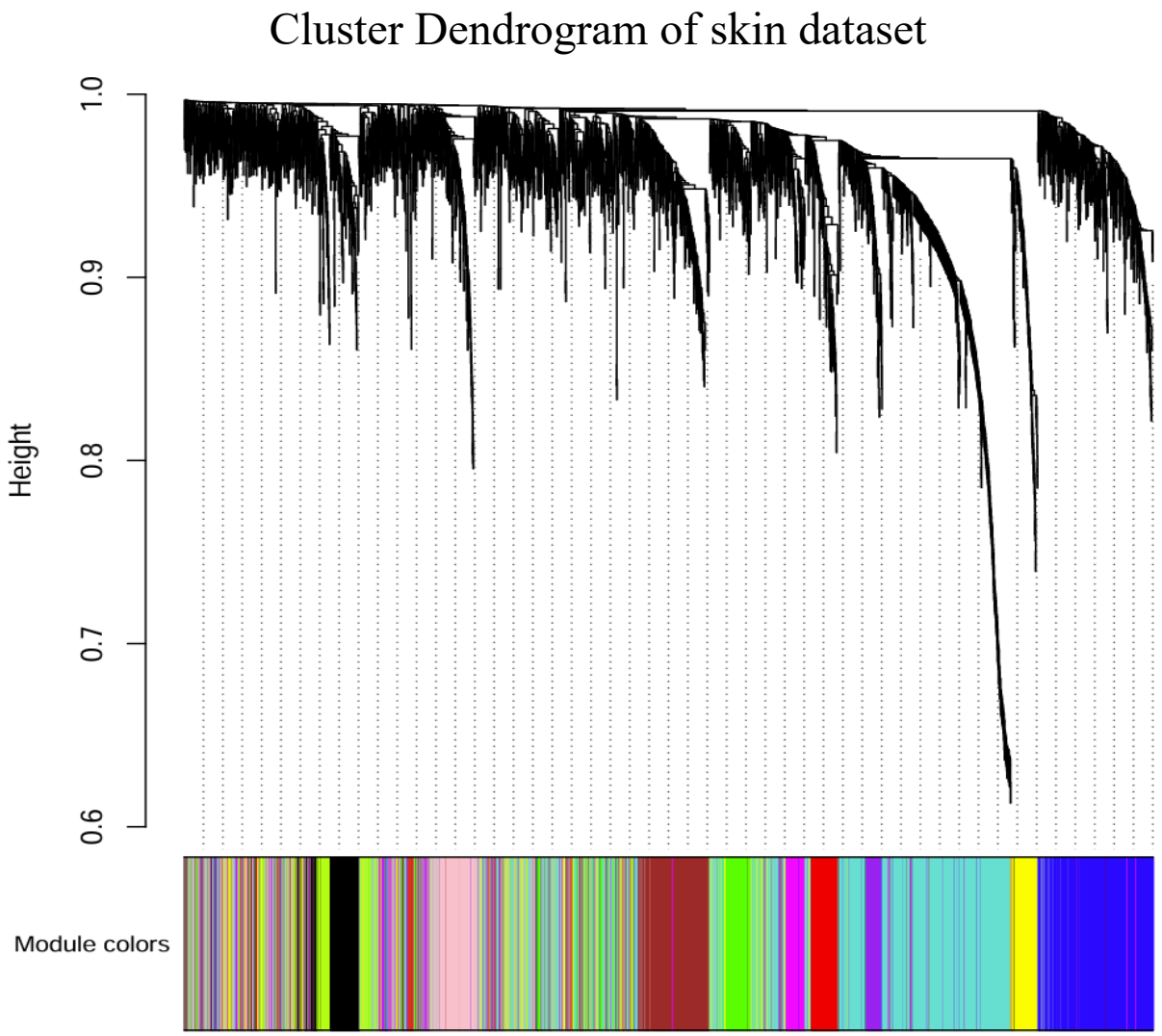
L_{90} [d] is defined as the number of congits whose length sum product of N_{90}

(Castro and Ng, 2017).

4.4.2 Weighted gene co-expression networks

Expression profiles from a total of 3,475 genes from 34 skin tissue samples and 4,138 genes from 34 spleen tissue samples were retained after quality control. Then the clustering of expression profiles of the individual samples in relation to their genotypes were showed substantial differences in expression between skin and spleen (Figures 4.2A,B) transcriptome profiles. Clustering of genes into co-expression modules also showed substantially different profiles in the skin (Figure 4.2A) compared to the spleen datasets (Figure 4.2B). Overall, there were 12 different co-expression modules identified from the skin dataset (Figure 4.3A) and there were 15 for the spleen dataset (Figure 4.3B). Note that the colours are arbitrary so they cannot be compared between the two datasets. The coloured row below the dendrogram indicates module membership identified by the dynamic tree cut method (Figure 4.2A,B). There was considerable variation in associations between individual SNPs and predicted membership in individual co-expression modules. Numbers in the table represent the correlations scores between MEMs and significant SNPs in each column, with the p-values of the correlations in parentheses. The table is color-coded by correlation according to the color legend in each row, considering the correlation p-value cut of 0.05. In WGCNA, genes with a positive correlation of co-expression tend to be expressed at similar levels across samples. When one gene shows upregulated expression, the other genes also tend to show increased expression levels. Conversely, genes with a negative correlation are inversely co-expressed; when one gene's expression increases, the other gene's expression tends to decrease. There were 7 significant modules in the skin dataset including pink (PRKG1-1: r=-0.0, p=0.02;, PRKG1-2: r=0.4, p=0.02), greenyellow (GIMAP7: r=-0.42; p=0.02; PLA2G7-2: r=-0.35, p=0.05), blue (GIMAP7, r=-0.36; p=0.04), magenta (PLA2G7-6, r=-0.45; p=0.009) brown (GIMAP7, r=0.38; p=0.03, ITG11, r=-0.43; p=0.01, PLA2G7-2, r=0.35; p=0.05), turquoise (PLA2G7-2, r=0.48; p=0.007), and yellow (PLA2G7-2, r=0.37; p=0.03) (Figure 4.3A). While the modules were 10 significantly associated with significant SNPs such as brown (GMAP7, r=0.41; p=0.01, ITGA11, r=-0.39; p=0.02), magenta (SATb2, r=-0.35; p=0.03, ITGA11, r=0.36; p=0.03, PRKG1-4, r=-0.37; p=0.03, PRKG1-7, r=0.33; p=0.05), red (PRKG1-1, r=-0.45; p=0.006), green (HOXD1, r=-0.33; p=0.05, PRKG1-3, r=-0.38; p=0.02), purple (PRKG1-4, r=-0.4; p=0.02), greenyellow (ITGA11, r=0.33; p=0.05, PRKG1-2, r=-0.37; p=0.03, PRKG1-3, r= -0.42; p=0.01, PRKG1-6, r=0.48; p=0.003), blue (ITGA11, r=0.33; p=0.05, PRKG1-3, r=-0.38; p=0.02), turquoise (PRKG1-3, r=-0.54; p=6e-04), Cyan (HOXD1, r=-0.38; p=0.02,

PRKG1-1, r=0.37; p=0.02, PRKG1-5, r=-0.38; p=0.02), and grey (ITGA11, r=0.44; p=0.007) (Figure 4.3B).



A.

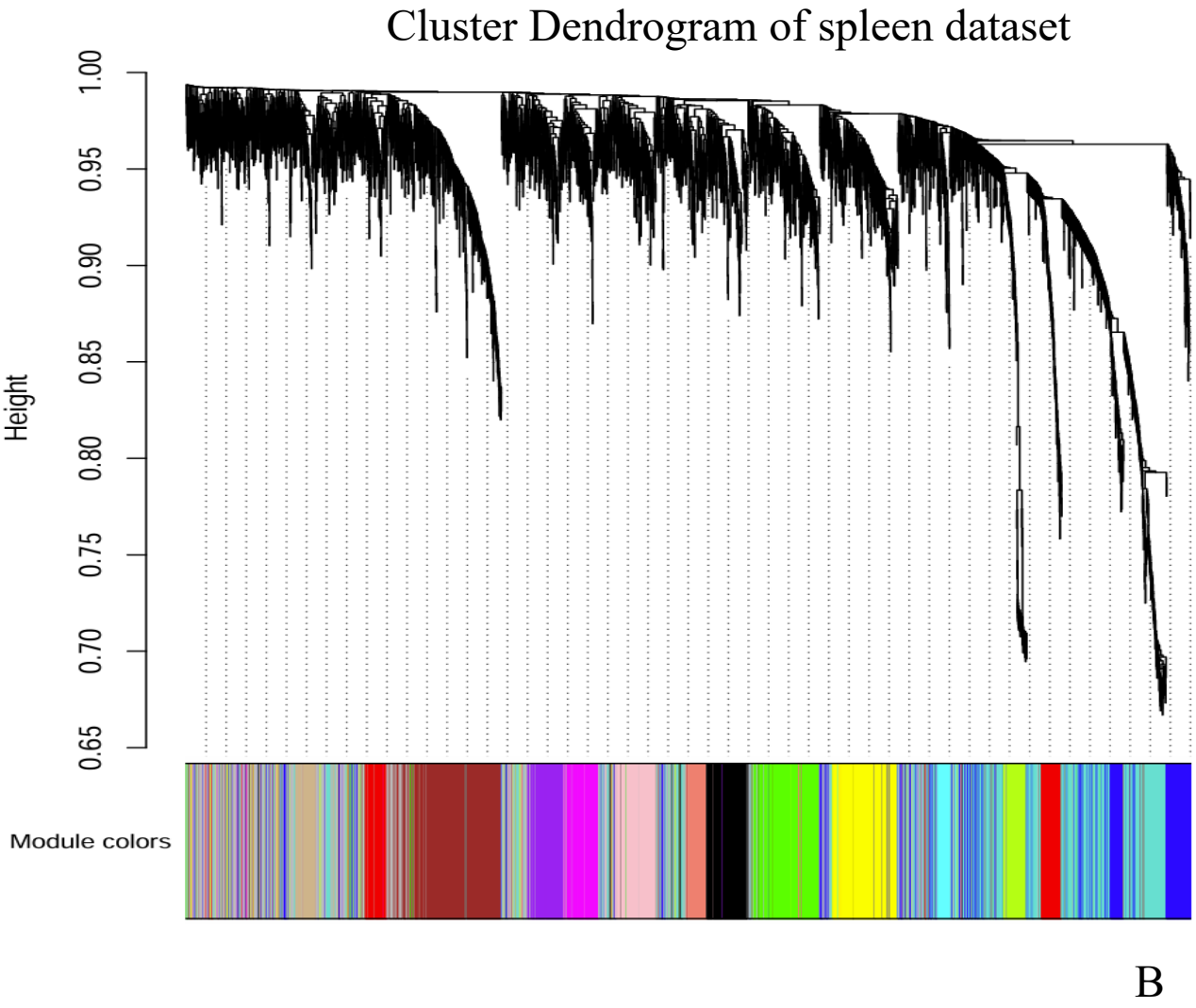
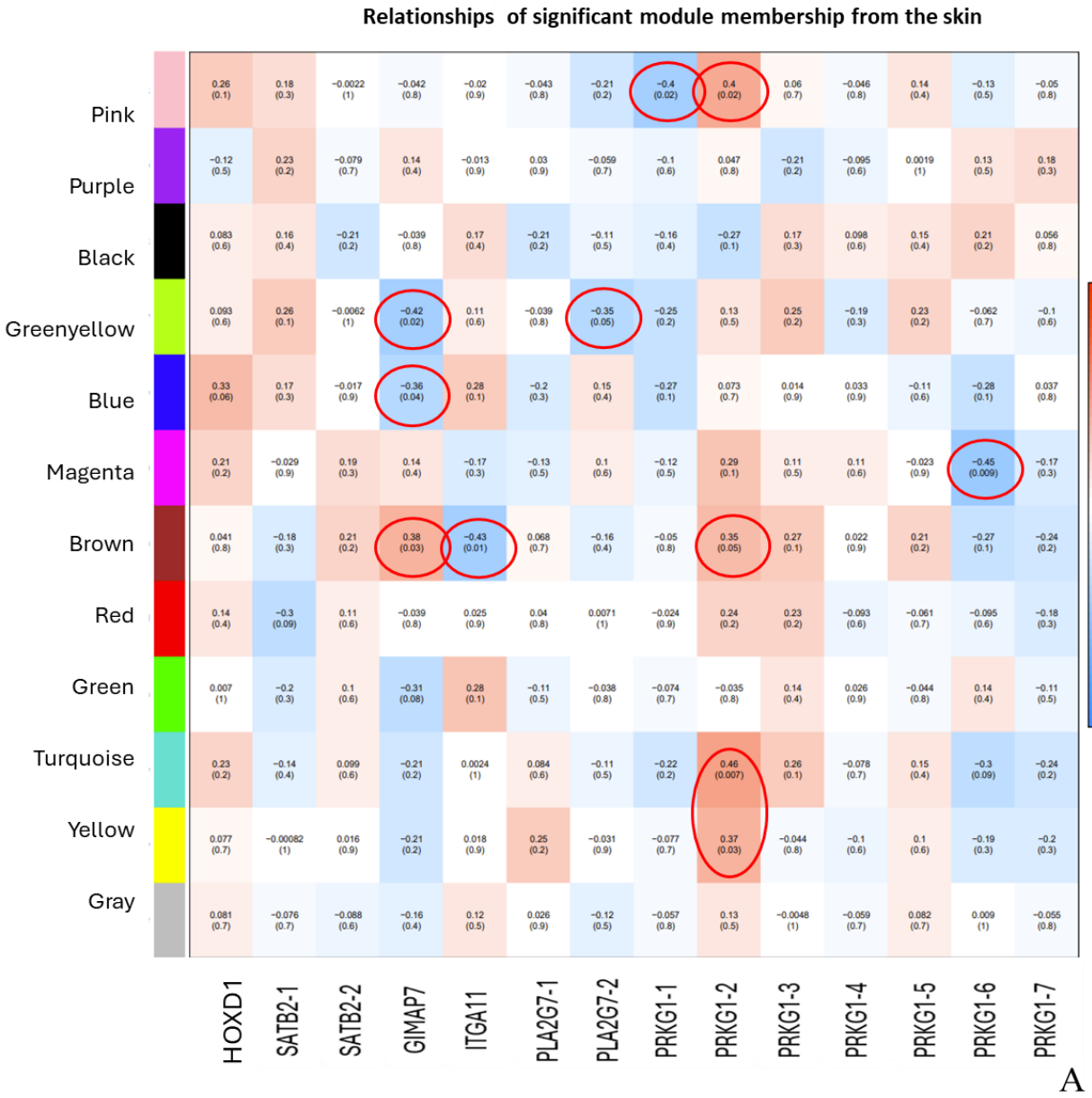


Figure 4.2 Cluster analysis based on the WGCNA pipeline to identify clusters of coexpressed genes from the full-length cDNA sequencing from skin (A) and spleen (B). Genes are divided into co-expressed modules based on the clustering of expression profiles derived from 14 SNP genotypes, across the 6 variable genes identified in Chapter 3. The cluster dendrogram illustrates the similarity between the expression profiles of individual genes, with the height indicating the relative strength of the correlation coefficients for the genes within each module. The gene dendrogram is generated using average linkage hierarchical clustering. The color row below the dendrogram indicates module assignments as determined by the Dynamic Tree Cut method.


Relationships of significant module membership from the skin
Pink
Purple
Black
Greenyellow
Blue
Magenta
Brown
Red
Green
Turquoise
Yellow
Gray
HOXD1
SATB2-1
SATB2-2
GIMAP7
ITGA11
PLA2G7-1
PLA2G7-2
PRKG1-1
PRKG1-2
PRKG1-3
PRKG1-4
PRKG1-5
PRKG1-6
PRKG1-7
1
0.5
0
-0.5
-1

A

.

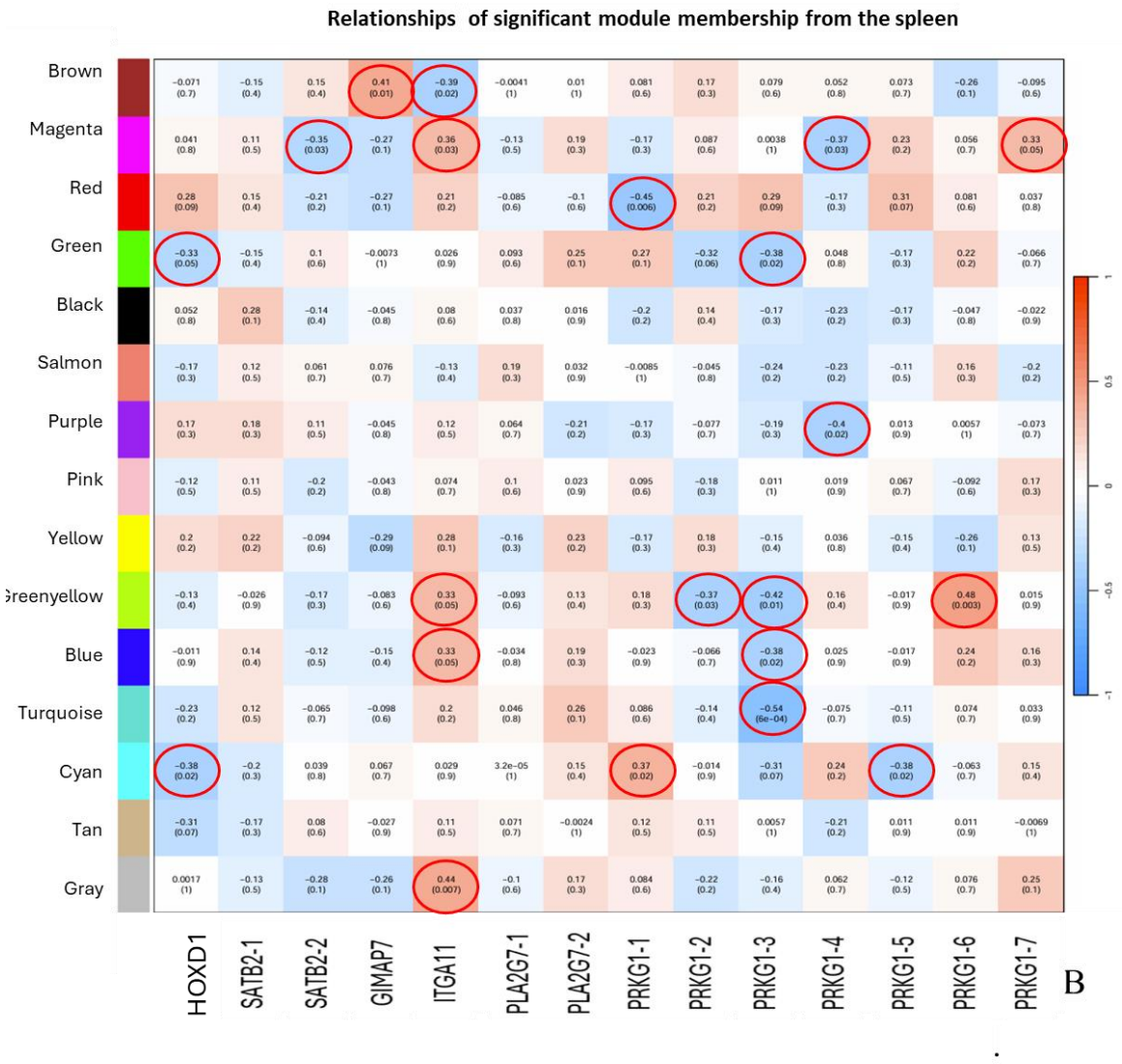


Figure 4.3 Relationships of consensus co-expression module membership (MM) from the skin (A) and spleen (B) datatsets relative to the 14 SNPs that differed significantly between breed groups. Each row in the table corresponds to a co-expression module and each column to a SNP. Numbers in the table report the correlations coefficients predicting MM for each of the co-expression modules and the SNPs (Module-Trait relationships), with the p-values of the correlations in parentheses. The table is color-coded by the strength of the correlations according to the heat map legend on the right, with darker colours indicated stronger associations; blue indicates a negative correlation while red indicateds a positive one. Red circles indicate significant associations.

4.4.3 Identification of modules of interest

In Weighted Gene Co-expression Network Analysis (WGCNA), positive and negative correlations refer to the relationships between the expression levels of genes across different samples. Positive correlations are used to identify modules (clusters) of co-expressed genes. These modules are groups of genes that have similar expression patterns, indicating they might be functionally related. In contrast, negative correlations indicate genes that are part of different pathways or regulatory networks that act in opposition to each other. Positive and negative correlations in WGCNA help to understand the relationships between genes based on their expression patterns, providing insights into gene function, regulation, and the underlying biological processes. For the skin dataset, seven modules were significantly correlated with at least one of the focal SNPs, but this included only 6/14 of the SNPs (Table 4.9). There were no significant correlations identified for HOXD1, neither of the SATB2 SNPs, one of the PLA2G7 SNPs, nor four of the PRKG1 SNPs. The pink module, which contained 170 genes, showed a significant correlation with genotypes of PRKG1-1 and PRKG1-2 but none of the other SNPs in PRKG1 or the other genes. PRKG1-2 also showed a unique significant correlation with the turquoise (853 genes) and yellow (270 genes) modules; although it shared the brown co-expression module (283 genes) with ITGA11, the association was in opposite directions (positive in the former and negative in the latter). PRKG1-6 was uniquely associated with the magenta module (168 genes). The greenyellow module (135 genes) showed a significant correlation with PLA2G7-2 and GIMAP7 but GIMAP7 also showed a unique association with the blue module(418 genes). For the spleen dataset, ten modules showed a significant correlation with 11/14 of the focal SNPs (Table 4.9). In this case, neither of the PLA2G7 SNPs was significantly correlated with a co-expression module but one of the SATB2 SNPs did show a significantly negative association (magenta; 155 genes). The magenta module was also negatively associated with PKRG1-4 and positively associated with PRKG1-7 and ITGA11. HOXB1 was negatively associated with two modules (green and cyan), both of which were also associated with SNPs in the PRKG1 gene. Interestingly, the cyan and greenyellow modules showed both positive and negative associations for different SNPs within the PRKG1 gene. In contrast to the skin dataset, all 7 SNPs for PRKG1 showed associations with at least one co-expression module. GIMAP7 and ITGA11 again shared an association in opposite directions with a module (also arbitrarily labeled brown), but it included 404 genes for the spleen data. In contrast to the skin data, ITGA11 also showed a positive correlation with four other co-expression modules, all but one (grey: 733 genes) of which were also associated with other SNPs.

Table 4.9 Significant modules were identified for each of the focal SNPs in both skin and spleen tissues. Positive correlations mean that the group of genes show similar expression patterns across samples. Negative correlations indicate that the group of genes show different expression patterns across samples Arbitrary colors are used to define the co-expressed modules, but these colors are not consistent between the skin and spleen datasets.

SNP	Skin		spleen	
	Positive correlation	Negative correlation	Positive correlation	Negative correlation
HOXD1	-	-		Green, Cyan
SATB2-1	-	-	-	-
SATB2-2	-	-		Magenta
GIMAP7	Brown	Greenyellow, Blue	Brown	
ITGA11		Brown	Magenta, Greenyellow, Blue, Grey	Brown
PLA2G7-1	-	-	-	-
PLA2G7-2		Greenyellow	-	-
PRKG1-1		Pink	Cyan	Red
PRKG1-2	Pink, Brown, Turquoise, Yellow	-		Greenyellow
PRKG1-3	-	-		Green, Greenyellow, Blue, Turquoise
PRKG1-4	-	-		Magenta, Purple
PRKG1-5	-	-		Cyan
PRKG1-6		Magenta	Greenyellow	
PRKG1-7	-	-	Magenta	

Note: (-) is defined as no significant module in positive or negative correlations associated with the focal SNPs

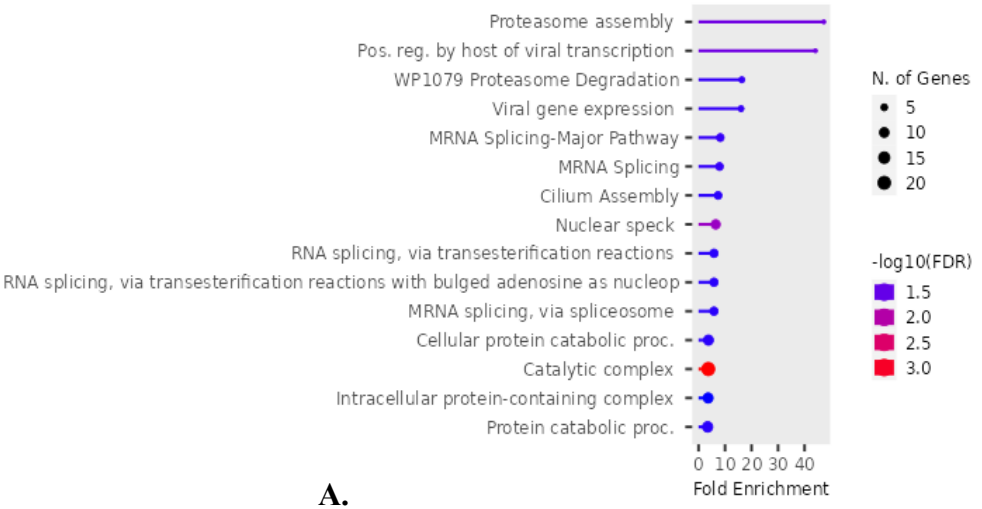
4.4.4 GO Analysis for the Module Genes

Based on GO analysis of the co-expression modules significantly associated with the focal SNPs in the skin dataset, 15 significant biological processes were identified for the pink module, with the two highest fold enrichment involved with proteasome assembly and post regulation by host of viral transcription (Figure 4.4A). The greenyellow module showed 20 biological processes, with the complex of collagen trimers (COL1A1 COL1A2 COL3A1 COL5A2 COL5A1) showing the highest fold enrichment (Figure 4.4B). There were 20 biological processes for the blue module, with the highest fold enrichment identified as amyotrophic lateral sclerosis (involving 22 genes, Figure 4.4C). In the magenta module, there were 13 biological processes, with negative regulation of macromolecule biosynthetic processes (involving 24 genes) showing the highest fold enrichment (Figure 4.4D). There

were 19 biological processes in the brown module, with the highest fold enrichment identified as WP1071 Cytoplasmic Ribosomal Proteins (20 genes; Figure 4.4E). The turquoise module presented 19 biological processes, with the highest fold enrichment associated with regulation of protein localization (24 genes; Figure 4.4F). The yellow module was associated with 20 biological processes, with the highest fold enrichment process identified as cornified envelopment, including 6 genes (CSTA, DSP, KRT10, KRT2, KRT1) (Figure 4.4G)

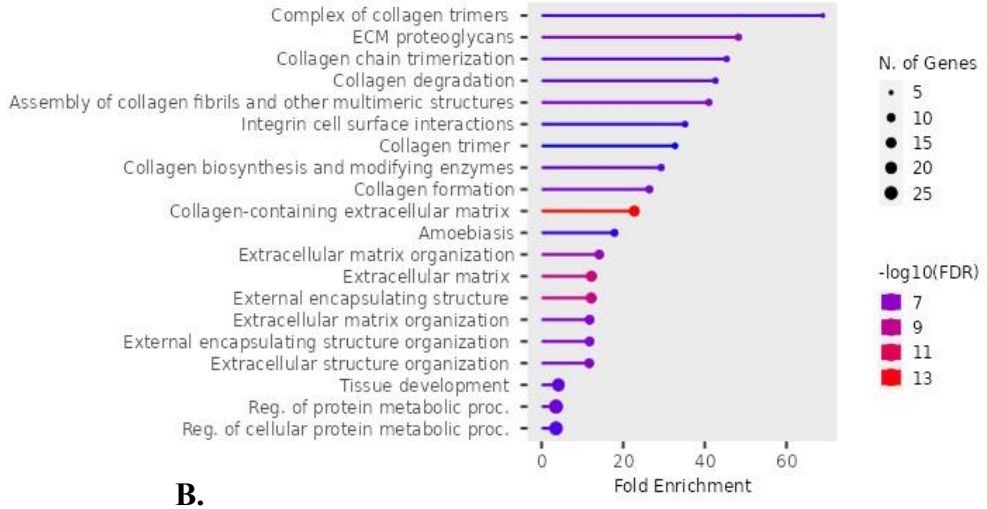
For the spleen dataset, the highest fold enrichment of biological pathways for the ten significant co-expression modules (Figure 4.5A-J) were: 1) positive regulation of transcription by RNA polymerase II (41 genes); 2) negative regulation of protein modification processes (14 genes); 3) negative regulation of stress-activated MAPK cascade (our genes: FKTN, GPS2, NCOR1, PTPN22); 4) autophagy (15 genes); 5) mRNA metabolic processes (13 genes);6) body morphogenesis (three genes: PDGFRA, CLASP1,BRAF); 7) WP1071 cytoplasmic ribosomal proteins (21 genes); 8) regulation of mRNA metabolic processes (24 genes); 9) endoplasmic reticulum lumen (8 genes: PDIA6, HSP90B1, P4HB, ERP29, HSPA5, TXNDC5, MANF, MZB1); and 10) nuclear protein-containing complex (72 genes).

Pink module



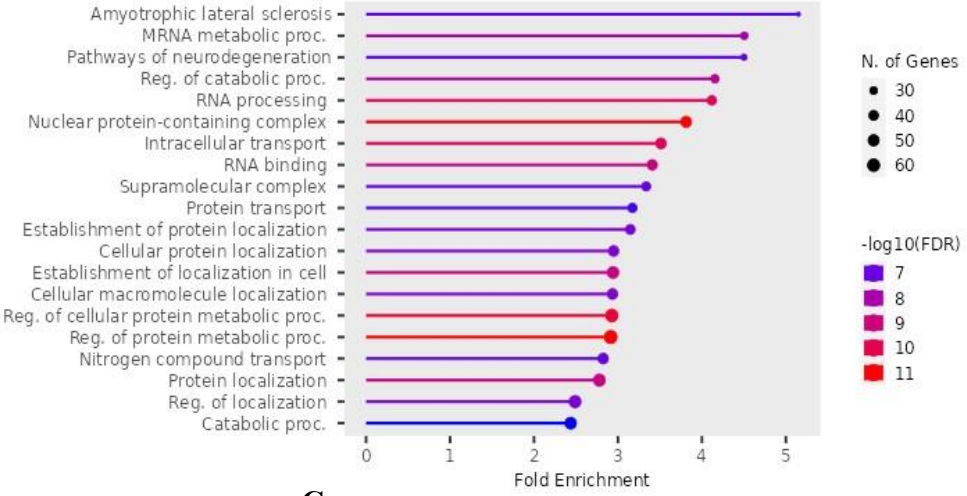
A.

Greenyellow



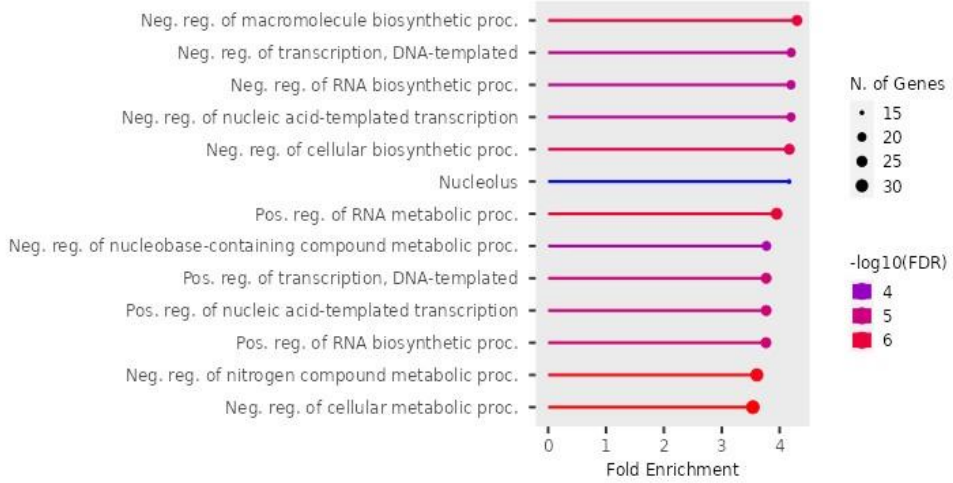
B.

Blue module



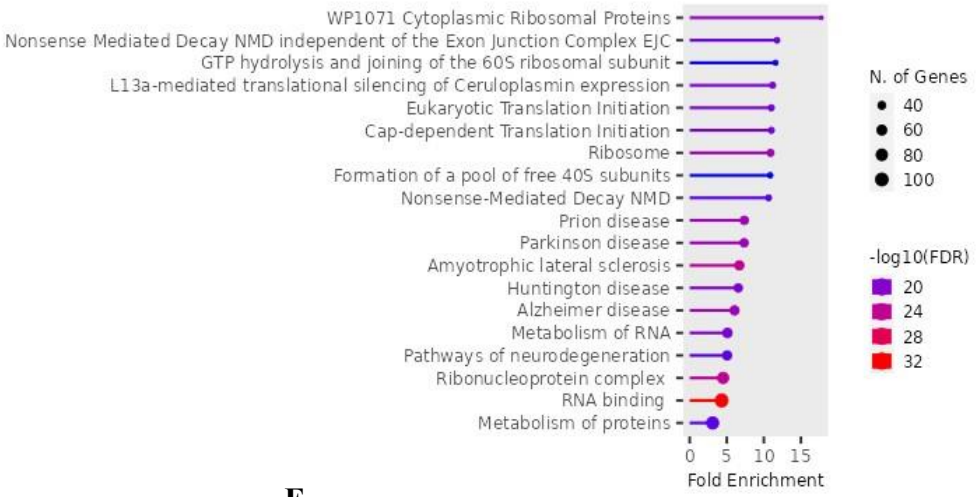
C.

Magenta



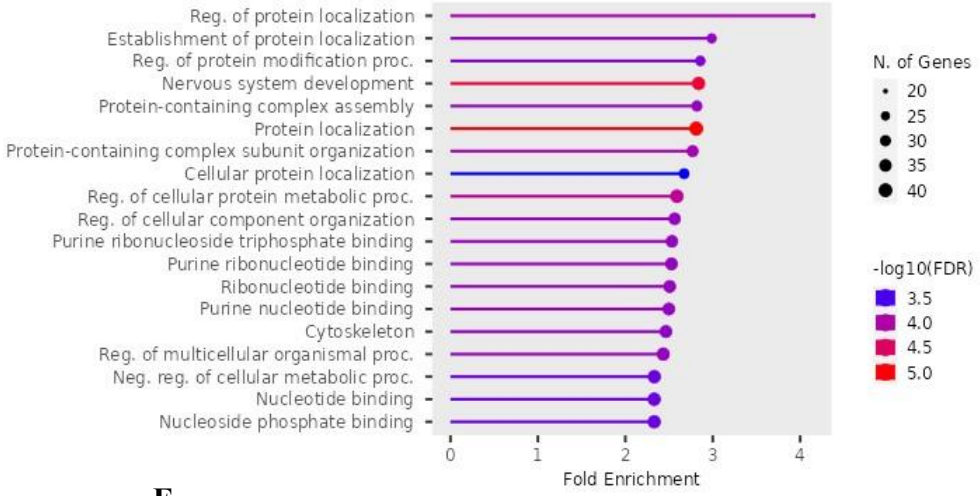
D.

Brown



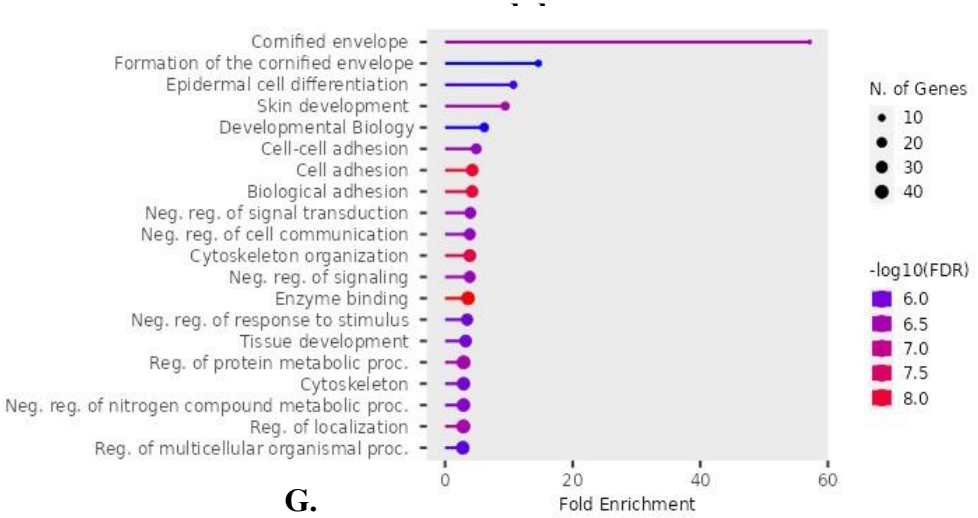
E.

Turquoise



F.

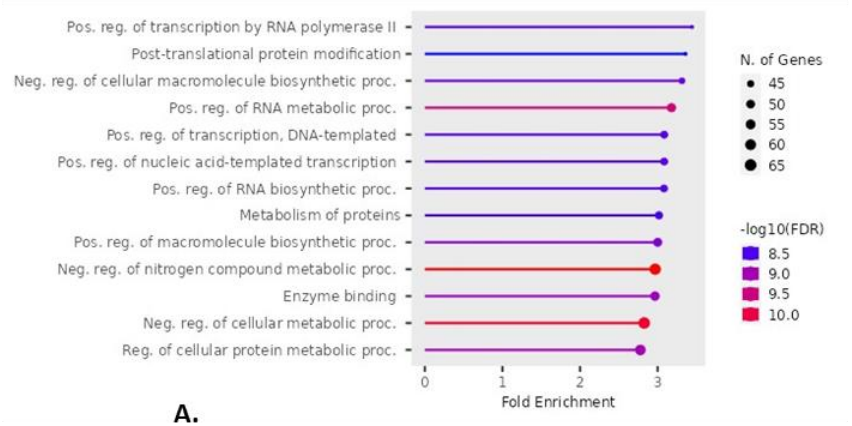
Yellow



G.

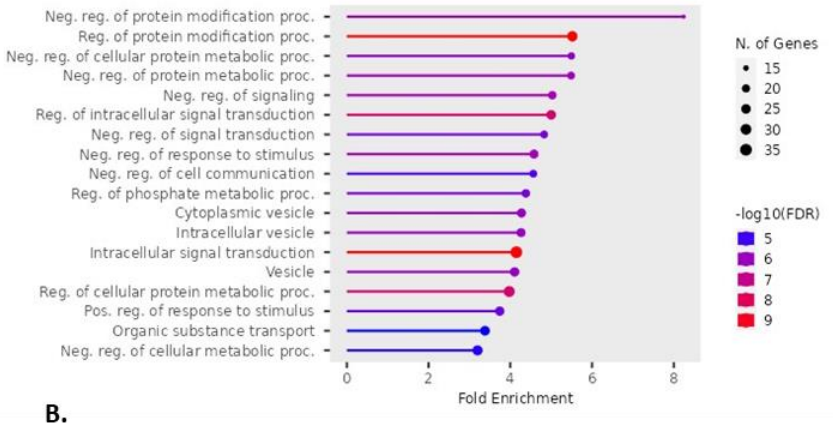
Figure 4.4 Gene ontology analysis of significant co-expression modules from the skin dataset. A) pink; B) greenyellow; C) blue; D) magenta; E) brown; F) turquoise; and G) yellow. The x-axis indicates the fold enrichment and the y-axis shows the GO biological processes. The colours Indicate the significance level of enrichment (-log10 (FDR) and the size shows the number of genes in the module associated with a particular biological process.

Brown module



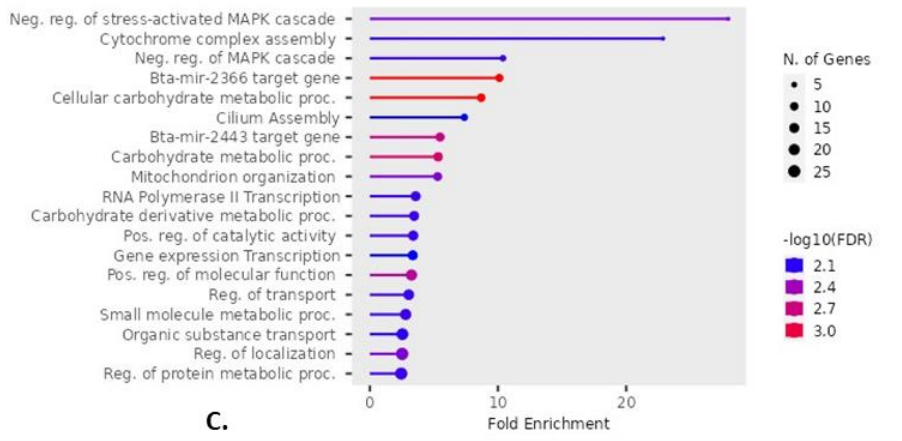
A.

Magenta module



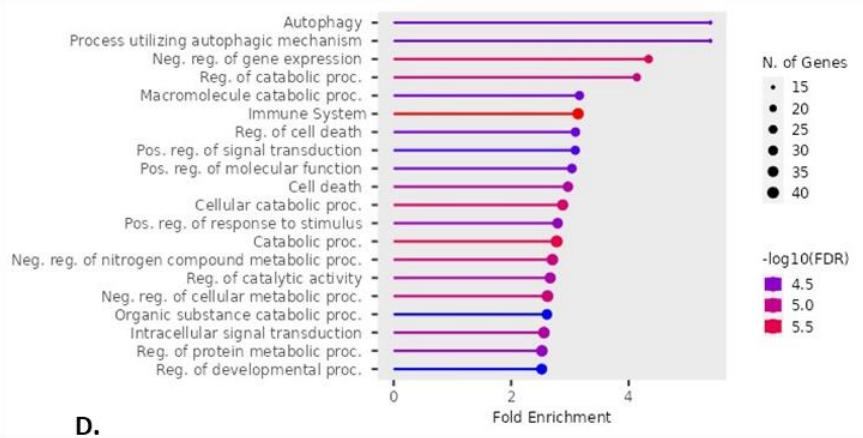
B.

Red module

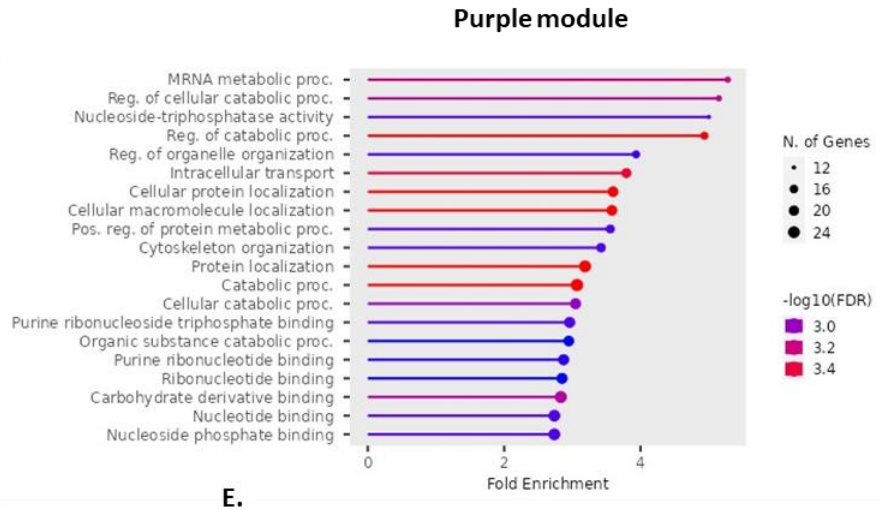


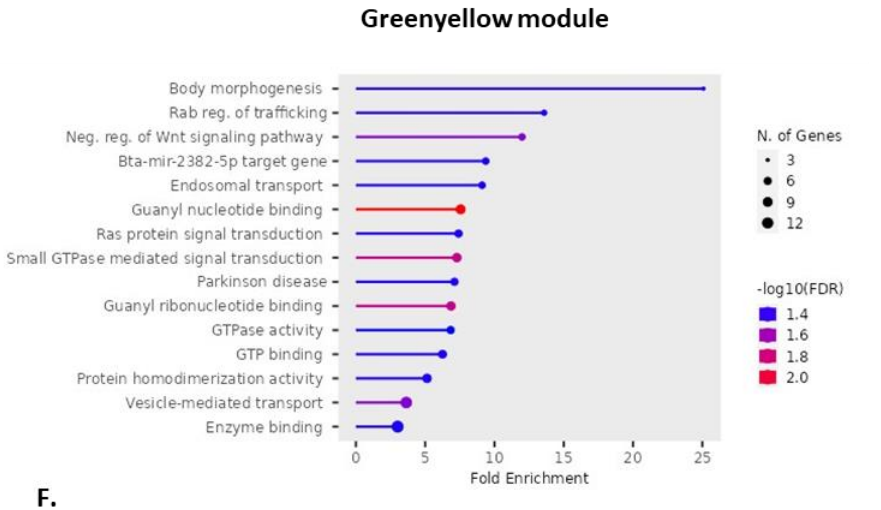
C.

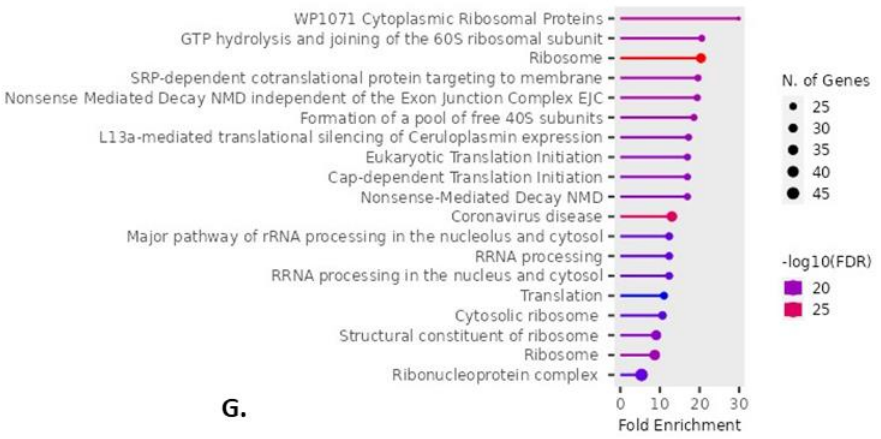
Green module

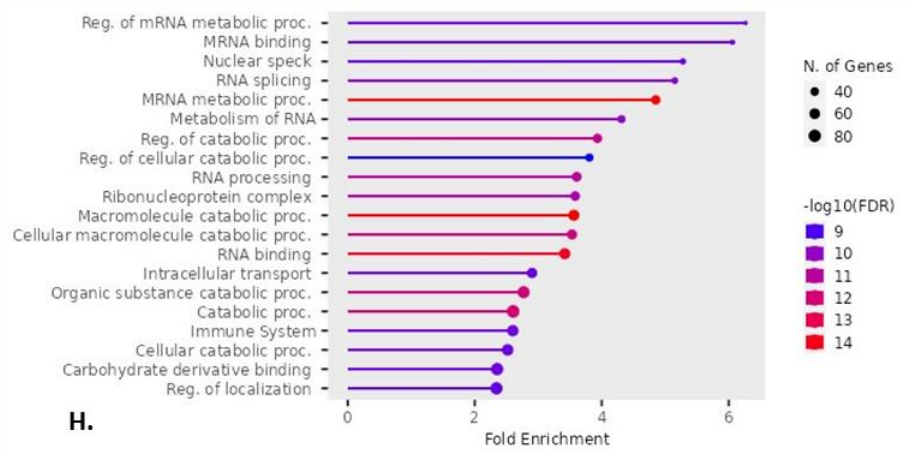


D.


Purple module
MRNA metabolic proc.
Reg. of cellular catabolic proc.
Nucleoside-triphosphatase activity
Reg. of catabolic proc.
Reg. of organelle organization
Intracellular transport
Cellular protein localization
Cellular macromolecule localization
Pos. reg. of protein metabolic proc.
Cytoskeleton organization
Protein localization
Catabolic proc.
Cellular catabolic proc.
Purine ribonucleoside triphosphate binding
Organic substance catabolic proc.
Purine ribonucleotide binding
Ribonucleotide binding
Carbohydrate derivative binding
Nucleotide binding
Nucleoside phosphate binding
0
2
4
Fold Enrichment
N. of Genes
12
16
20
24
-log10(FDR)
3.0
3.2
3.4
E.


Greenyellow module
Body morphogenesis
Rab reg. of trafficking
Neg. reg. of Wnt signaling pathway
Bta-mir-2382-5p target gene
Endosomal transport
Guanyl nucleotide binding
Ras protein signal transduction
Small GTPase mediated signal transduction
Parkinson disease
Guanyl ribonucleotide binding
GTPase activity
GTP binding
Protein homodimerization activity
Vesicle-mediated transport
Enzyme binding
0
5
10
15
20
25
Fold Enrichment
N. of Genes
3
6
9
12
-log10(FDR)
1.4
1.6
1.8
2.0
F.


Blue module
WP1071 Cytoplasmic Ribosomal Proteins
GTP hydrolysis and joining of the 60S ribosomal subunit
Ribosome
SRP-dependent cotranslational protein targeting to membrane
Nonsense Mediated Decay NMD independent of the Exon Junction Complex EJC
Formation of a pool of free 40S subunits
L13a-mediated translational silencing of Ceruloplasmin expression
Eukaryotic Translation Initiation
Cap-dependent Translation Initiation
Nonsense-Mediated Decay NMD
Coronavirus disease
Major pathway of rRNA processing in the nucleolus and cytosol
RRNA processing
RRNA processing in the nucleus and cytosol
Translation
Cytosolic ribosome
Structural constituent of ribosome
Ribosome
Ribonucleoprotein complex
0
10
20
30
Fold Enrichment
N. of Genes
25
30
35
40
45
-log10(FDR)
20
25
G.


Tauqouise module
Reg. of mRNA metabolic proc.
MRNA binding
Nuclear speck
RNA splicing
MRNA metabolic proc.
Metabolism of RNA
Reg. of catabolic proc.
Reg. of cellular catabolic proc.
RNA processing
Ribonucleoprotein complex
Macromolecule catabolic proc.
Cellular macromolecule catabolic proc.
RNA binding
Intracellular transport
Organic substance catabolic proc.
Catabolic proc.
Immune System
Cellular catabolic proc.
Carbohydrate derivative binding
Reg. of localization
0
2
4
6
Fold Enrichment
N. of Genes
40
60
80
-log10(FDR)
9
10
11
12
13
14
H.

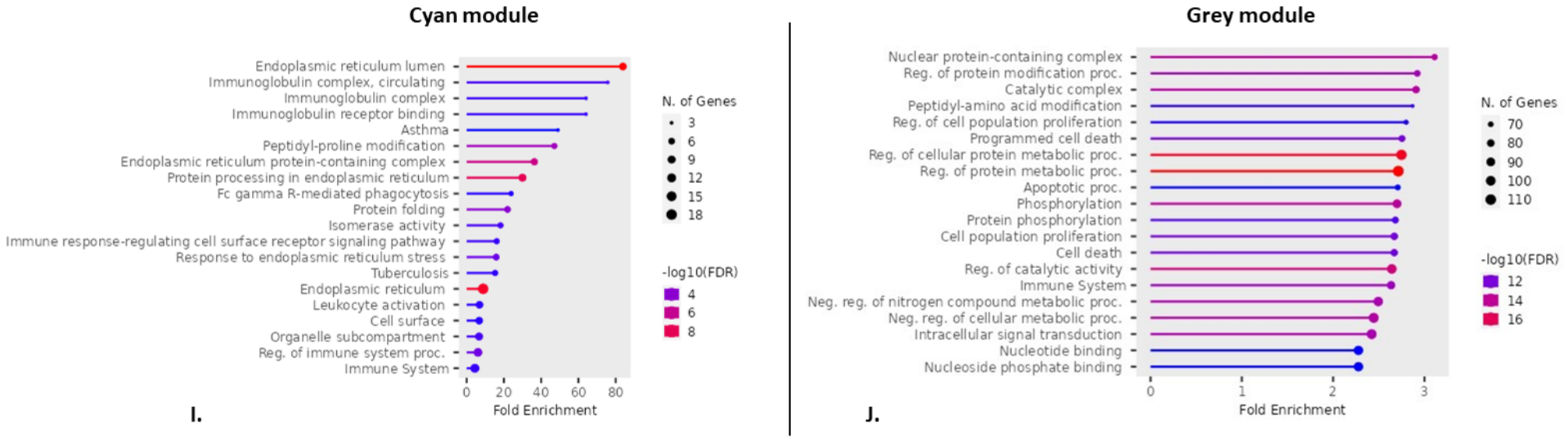


Figure 4.5 Gene ontology analysis of significant corexpression modules from the spleen dataset. A) brown;, B) magenta; C) red; D) green, E) purple; F) greenyellow; G) blue; H) turquoise; I) cyan; and J) grey. The x-axis indicates the fold enrichment and the y-axis shows the GO biological processes. The colours Indicate the significance level of enrichment (-log10 (FDR) and the size shows the number of genes in the module associated with a particular biological process.

4.4.5 Gene Ontology and pathway enrichment analysis, identification of genes of interest, and protein-protein interaction analysis

For the skin dataset, a total of 2,297 genes were identified across the seven co-expression modules significantly associated with the focal SNPs (Appendix D). Only three of these overlapped with the set of genes that I identified in chapter 2 as being associated with tick resistance at high confidence: FN1 (greenyellow module), ATP9A (turquoise module), and ECM1 (yellow module). For the spleen dataset, a total of 3,265 were identified across the ten co-expression modules (Appendix E). Only five of these overlapped with the genes identified in chapter 2 but none were the same as for the skin samples: CR2 (turquoise module), RHOT1 (purple module), SRGN (turquoise module), and LAPTM5 (turquoise module). GIMAP7 was the only one of the genes included in the SNP analysis that was also identified as a component of co-expression modules, but was included in four different modules (grey, pink, brown and turquoise).

For the skin dataset, I generated the protein-protein interaction network for the greenyellow module, which was associated with SNPs in two different genes (GIMAP7 and PRKG1) and included FN1 from the full set of 37 genes. The GO analysis had shown that biological functions within this module were associated predominantly with extracellular matrix and regulation of protein metabolic processes (Figure 4.4). However, the protein-protein network identified interactions between different processes. There are genes in yellow cluster are associated with basement membrane organization; genes in purple cluster are associated with extracellular matrix assembly; genes in orange cluster are with collagen fibril organisation; genes in pink cluster are associated with type I interferon signaling pathway: and gene in green cluster are associated with collagen metabolic process (Figure 4.6; note the interactions with the BOLA gene complex and FN1 at yellow star). Interestingly, several genes showed the role in many biological processes in one gene. For the spleen dataset, the turquoise module, which was significantly associated with one of the SNPs in the PRKG1 gene, included 4/5 of the genes that overlapped with the set identified in chapter 2. The primary biological process was associated with regulation of mRNA and metaboloic processes (Figure 4.5).The magenta module was used to generate protein-protein network (Figure 4.7). module in spleen datasets due to this module showed the largest number of significant SNPs (SATB2-2, ITGA11, PRKG1-4, and PRKG1-7) which were predominantly involved in basement membrane organization; extracellular matrix assembly; collagen fibril organization; type I interferon signaling pathway; and collagen metabolic process

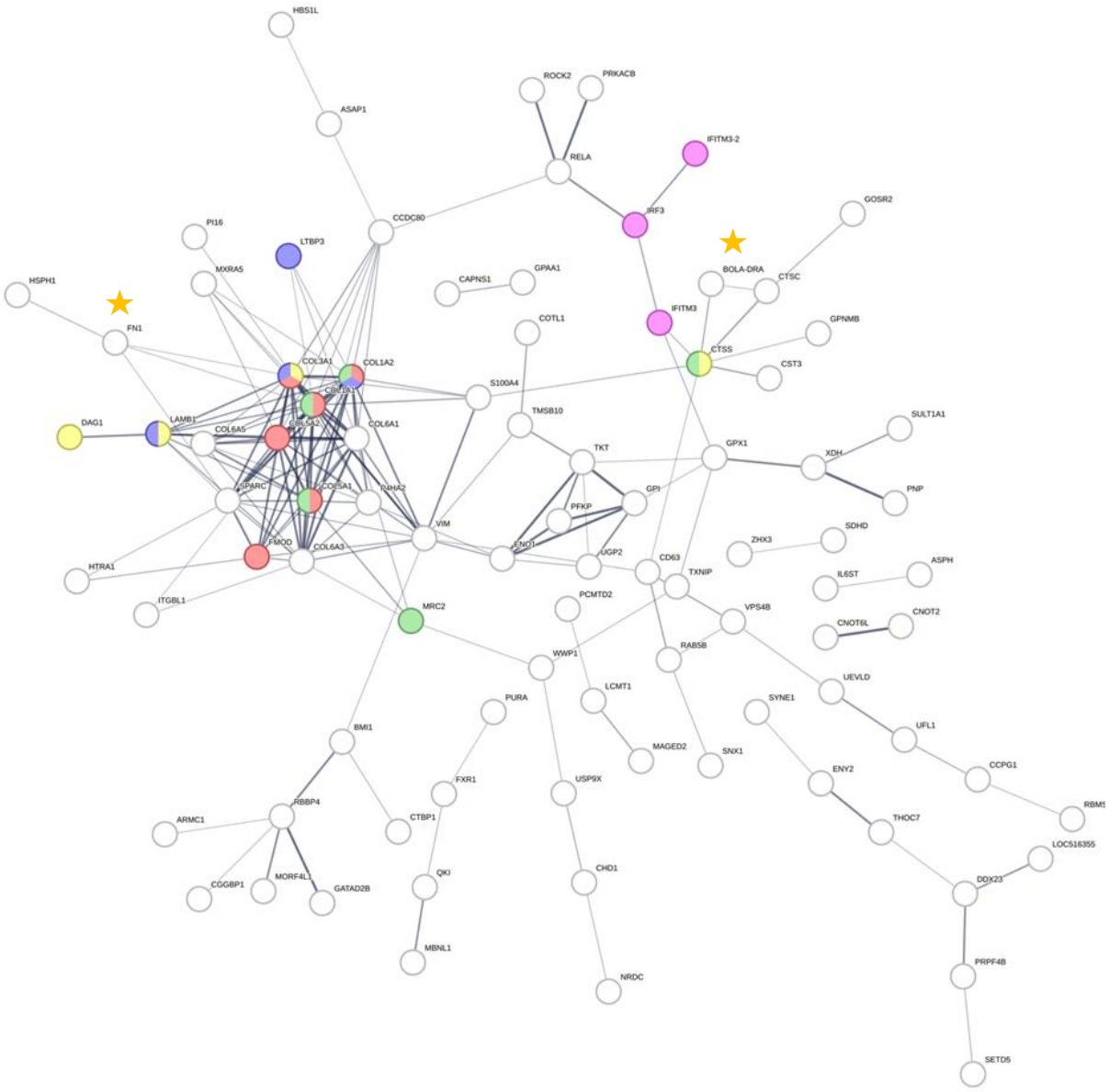


Figure 4.6 Protein-protein interaction network of the greenyellow module (132 genes) from the skin dataset. Genes in yellow are predominantly associated with Basement membrane organization; genes in purple are predominantly associated with Extracellular matrix assembly; genes in orange are predominantly associated with Collagen fibril organization; genes in pink are predominantly associated with Type I interferon signaling pathway; gene in green are predominantly associated with Collagen metabolic process and genes in white are others function. Yellow stars represent FN1 and BOLA genes with identify as candidate genes from chapter 2.

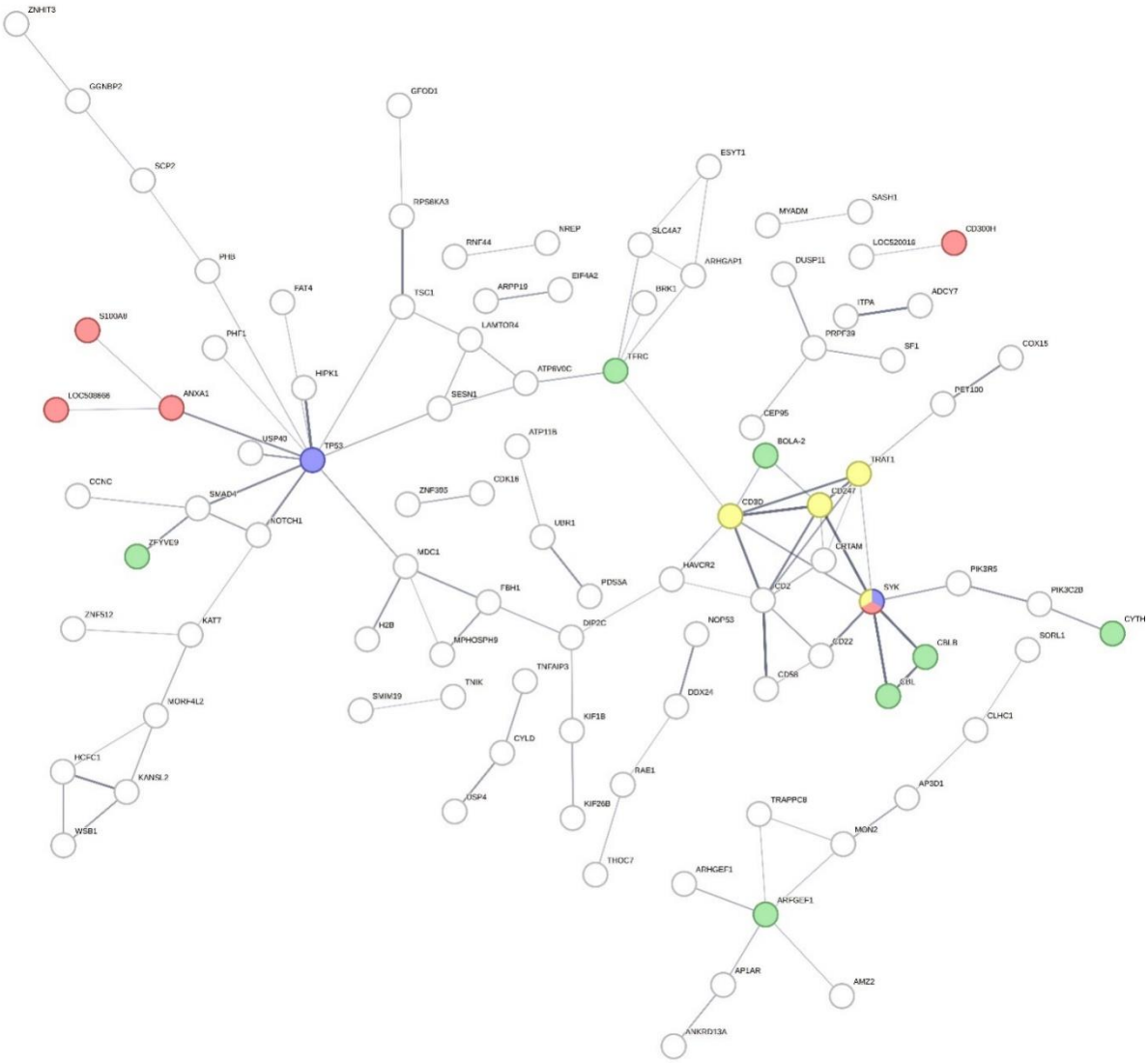


Figure 4.7 Protein-protein interaction network of the module (142 genes) from the spleen dataset. The genes in purple are predominantly associated with Positive regulation of reactive oxygen species metabolic process; the genes in red are predominantly associated with Granulocyte chemotaxis; the genes in yellow are predominantly associated in T-cell receptor complex; the genes in green are presented in Endocytosis pathway; genes in white are others function

4.5 Discussion

Infectious diseases result from the complex interactions between the host, pathogen (including both intracellular and extracellular pathogens), and environment. After encountering a pathogen, the host's immune responses deploy several mechanisms to eliminate it. However, environmental conditions and pathogen characteristics can affect the host immune system and influence the outcome of the infection (Emam et al., 2019, Mallard et al., 2015). Emam et al. (2019) reviewed the genetic and epigenetic factors that contribute to individual differences in immune protection. The author also identified the various methods researchers have used to study the genetic regulation of immunity in ruminants and examines the different sources of genetic information. The results indicate that different research groups have applied diverse strategies to unravel this complex network and understand the genetics of disease resistance. These strategies can be categorised according to the types of genetic information analysed (Emam et al., 2019). Genetic pathways identified in this chapter, such as immune response gene clusters and extracellular matrix gene clusters, suggest that it may be possible to estimate genomic breeding values for immune response to improve immune traits, health, and pathogen infection outcomes (Mallard et al., 2015).

The co-expression analysis based on the ONT RNA-seq data generated here, combined with the genotype data from chapter 3, identified a range of biological pathways showing differential expression of genes correlated with SNPs showing differences in genotype frequencies between three Scottish cattle breeds. Although co-expression analyses are normally based on associations with particular traits or phenotypes such as the lactation process or reproductive performance (Fonseca et al., 2020, Farhadian et al., 2021), the novelty here was applying the approach to individual genotypes across multiple genes. Although the results were complicated, the GO analyses allowed me to interpret the dominant pathways and interactions between pathways associated with different processes that are predicted to be influenced by the focal SNPs. For example, ITGA11 was associated with five different co-expression modules in the spleen, then this gene might be involved in several biological pathways. Interestingly, the only SNP showing an amino acid polymorphism based on the SNP genotyping (PLA2G7-1) in chapter 3 did not show significant association with any of the co-expression modules. Other interesting findings from this study are the differences detected between genes and pathways that show different patterns of co-expression in skin compared to spleen tissues. Most previous studies on tick resistance have focused on expression of genes in skin or blood samples rather than spleen.

It is well known that skin is the initial location of tick attachment (Francischetti et al., 2009) however, spleen also plays a key role in immune system which is useful for immune response studies (Goff et al., 2010) While this is likely due to the fact that spleen tissues need to be sampled destructively, my study could suggest that if immune function is the target, spleen could be more informative.

Intriguingly, only a handful of the genes associated with tick resistance based on the synthetic review in Chapter 2 showed significance in the WGCNA analysis, which could suggest that these could be useful targets to consider for future studies on potential biomarkers. For the skin, of the three genes that intersected between the analyses (FN1, ATP9A, ECM1), most would not be classified as "immune" genes so their role in tick resistance is interesting. Fibronectin 1 or FN1 is a large glycoprotein component of the extracellular matrix including two nearly identical subunits, with various binding domains of cell surface and extracellular ligands. An important function of FN1 is support for cell adhesion, but it also affects cytoskeleton organization and cell migration. This gene is involved in several biological processes such as wound healing, thrombosis and ageing (Yamada, 1991, Moursi et al., 1996). Goossens et al. (2009) reported that FN1 might play a role in early blastocyst formation specifically, alternative splicing at 3' defined regions of FN1 may be involved in bovine preimplantation embryo development. Bhattarai et al. (2024) studied in human and suggested that the vascular deposition of FN1 and loss-of-function (LOF) variants in FN1 may reduce Alzheimer's disease (AD) risk, providing novel therapeutic intervention clues targeting the extracellular matrix (ECM) to mitigate AD risk. The findings offer a cross-species design with both pathological and functional validation, demonstrating that the ECM component fibronectin is related to key pathological aspects of AD, such as toxic protein clearance, blood-brain barrier integrity, and microglial activity. The study investigated the first knockout zebrafish model for fibronectin 1 (FN1) in relation to amyloid toxicity and identified cellular changes associated with fibronectin activity (Bhattarai et al., 2024). ATPase phospholipid transporting 9A (putative) or ATP9A, is a P4-ATPPase, which acts as a regulator for endosomal recycling (Tanaka et al., 2016). It is well known that during the cell cycle, it influences cell to regulate phosphatidylethanolamine (PE) and release of extracellular vesicles (EV) (Irie et al., 2017). Naik et al. (2019) reported that ATP9A regulates exosome release from human cells. Extracellular matrix protein (ECM1) has been identified to regulate endochondral bone formation and activates endothelial cell proliferation, it also has been found in angiogenesis and some tumor stages (Wang et al., 2003, Han et al., 2001, Bhalerao et al., 1995). Li et al. (2011) found that ECM1 can control T_H2 cell trafficking when the immune response regulates through $S1P_1$

expression. In addition, ECM1 also has two and three splice variants in mouse and human, respectively (Hamada, 2002).

Although more of the genes overlapped for the spleen dataset (GIMAP7, CR2, RHOT1, SRGN, LAPTM5), it is interesting that none are the same as for the skin dataset and that only one of the genes (GIMAP7) was also one of the genes identified as differing significantly between the cattle breed groups. GIMAP7 or GTPase, IMAP family member 7 is one of seven in the gene family (GIMAP1, GIMAP2, GIMAP4, GIMAP5, GIMAP6, and GIMAP8) (Krücken et al., 2004). GIMAP mRNA expression presents in the spleen and lymph nodes, indicating a potential role in immune regulation and it has been demonstrated to have a connection to tumor development (Schnell et al., 2006). It was included in four different co-expression modules (grey, pink, brown, turquoise) and its genotype was significantly associated with the brown module, suggesting that it is involved with a wide range of biological processes. The other four genes that intersected for the spleen have also been associated with immune function and so also could provide interesting biomarkers to consider. Complement C3d receptor 2 (CR2) encodes a CR2 protein that has been found in cattle and protein expression of CR2 has been confirmed in cattle and a monoclonal antibody, it also has a commercial antibody available called CC21 (Sopp, 1996, Naessens et al., 1990). RT-PCR in cattle showed that there are four distinct complement receptors that are created from the CR2 gene by alternative splicing (Pringle et al. 2012). Two of the variants of CR2 in cattle represented homologues of murine CR1 and CR2. In addition, CR2 and CD19 can be co-signaling molecules in cattle (Pringle et al., 2012). The Ras homolog family member T1 (RHOT1) has been found to be expressed in the early pregnancy stage in immune cells of the peripheral blood in *Bos indicus* heifers (Rocha et al., 2020). In humans, serglycin or SRGN has been identified as a proteoglycan present in the intracellular secretory compartment;, it is also expressed by several cell types and plays an important role in both normal and infection conditions (Scully et al., 2012, Kolset and Pejler, 2011). Serglycin is expressed in most hematopoietic cells and acts as a key in inflammation with various inflammatory mediators (Kolset and Pejler, 2011, Korpetinou et al., 2014). LAPTM5 (Lysosomal protein transmembrane 5) encodes a protein that is expressed in immune cells (Adra et al., 1996); on one hand this protein acts as a negative regulator of T and B cell receptors in the plasma membrane (Ouchida et al., 2008, Ouchida et al., 2010), on another hand it acts as a positive regulator of NF-κB and MAPK signaling. LAPTM5 plays a crucial role in the inflammatory signaling pathways in macrophages (Glowacka et al., 2012). The overlapping genes between chapter 2and chapter 4 showed the supportive information to each others. Most of the genes that exhibited overlap were related to the extracellular matrix

and host response in the skin. In contrast, the overlapping genes in the spleen were associated with immune response processes relevant to spleen function.

This study presents a transcriptome analysis from full-length cDNA sequencing analysed by WGCNA from skin and spleen tissue sample from Scottish cattle. Robust genes that related to significant SNPs from chapter 3 showed associations with the multiple biological processes, but particularly the extracellular matrix, immune response, and inflammatory pathways. My findings present preliminary evidence of the role of these candidate genes and provide the basis of host immune response in cattle for further investigation. Further studies, alternative splicing and more extensive validating studies along with the information of transcriptome level associated with immune-related genes would provide a better understanding of the molecular mechanisms of host immune response in cattle.

Mantilla Valdivieso et al. (2022) conducted a transcriptional study on leukocytes from naïve Brangus steers at various time points following artificial tick infestation to identify genes associated with high and low tick resistance in cattle. Tick infestation in animals at 3 and 12 weeks revealed leukocyte gene expression linked to chemotaxis, cytokine secretion, and inflammatory responses, with IL-17 and cytokine-cytokine interaction pathways playing crucial roles in the host immune response. Before initial infestation in high and low resistance groups, significant genes involved in immunity, tissue remodeling, and angiogenesis were identified. The study also highlighted the importance of the bovine MHC complex in tick infestation in cattle (Mantilla Valdivieso et al., 2022). Although this study did not investigate leukocyte or blood samples directly, spleen samples indicated host immune responses (GIMAP7, RHOT1), inflammatory signaling pathways in macrophages (LAPTM5), and complement receptor (CR2).

A few years later, (Manjunathachar et al., 2014) a new study on gene expression in bovine skin exposed to *Rhipicephalus australis.* The findings identified 158 significant transcription factors (TFs), including GRHL3 and DTX1, and a list of differentially expressed genes (DEGs) associated with immune response and host response to infectious diseases. Interferon Type 1-stimulated genes (including MX2, ISG15, MX1, and OAS2) were up-regulated in low tick resistance steers. Additionally, the study performed on high and low tick resistance in tick-infested Brangus cattle showed significant downregulation of keratin and keratin-associated protein genes (KRTAPs) in both groups after 12 weeks compared to pre-infestation samples. These genes, crucial for hair shaft structure and mechanical strength, were notably affected, with KRTAP9-1, KRTAP9-2, and other

KRTAP9-like genes showing the highest downregulation, suggesting potential keratinocyte deformity in response to tick infestation (Mantilla Valdivieso et al., 2024). Interestingly, the findings in this study also highlighted the involvement of genes related to keratin and collagen in skin datasets. The expression of both keratin (KRT10, KRT2, KRT1) in yellow module and collagen (COL1A1 COL1A2 COL3A1 COL5A2 COL5A1) in greenyellow module are essential for maintaining the integrity of the epidermal barrier, suggesting that these genes may play a role in the local immune response to tick infestation.

Although the list of genes identified in this study differs from previous transcription studies, the biological processes or gene functions appear strongly similar. The gene expression patterns suggest that animals with different resistance phenotypes experience distinct stages of wound healing due to varying tick infestation levels. Several studies have explored genomics and transcriptomics to identify genes or biomarkers associated with tick resistance in cattle. However, interpreting the diverse findings remains challenging, and there is a notable absence of information on proteomics studies in this field. Raza et al. (2021) conducted a proteomics study using serum samples from Santa Gertrudis cattle. They employed sequential window acquisition for peptide quantification and ion mass spectrometry for analysis, revealing significant differences in the levels of 28 proteins between resistant (TR) and susceptible (TS) groups before tick challenge. Eight of these proteins were associated with adaptive responses including conglutinin (P23805), kinesin family member 12 (F1MMK9), kininogen-1 (A0A140T8C8), apolipoprotein C-III (V6F9A3), uncharacterised protein (F1MLW8), C8 beta chain (F1N102), clusterin (P17697) and complement factor I (Q32PI4). Intragroup comparisons indicated similar responses within each group, with stronger responses observed in TR cattle compared to TS cattle. Furthermore, many significantly different proteins identified in the resistant group, both before and after tick challenge, were involved in immune functions such as complement cascades, chemotaxis, and acute immune responses (Raza et al., 2021). In 2023, Raza et al. extended their research with quantitative proteomics analysis of serum and skin protein samples from Brangus cattle categorised as tick-resistant and tick-susceptible, at two different time points following tick exposure. They found significantly different protein abundances in resistant cattle after both early and prolonged tick exposure, compared to naïve resistant cattle, particularly involving immune response, coagulation cascade, blood coagulation, homeostasis, and wound healing pathways. However, only some proteins within these categories showed significance in susceptible cattle after prolonged tick exposure. Further studies for transcriptomics and proteomics are are necessary to validate

these findings across different cattle breeds, tissue types and under varying tick infestation conditions.

Chapter 5
Discussion

Understanding the mechanism of genomic selection for host resistance to ectoparasites, especially tick-borne diseases in the beef cattle industry, could improve breeding programs to develop more resistant cattle and animal production (Mapholi et al., 2014). Host resistance leads to a decrease in tick infestation, making it a measurable phenotype. However, tick count or scores are difficult to handle, which constrains the ability to obtain many animal phenotypes for genetic improvement (Cardoso et al., 2021). While a hereditary component influences variations in host resistance, the role of the host immune response remains essential but not thoroughly defined (Jonsson et al., 2014). The progress in high-throughput transcriptomics and the availability of genomic resources in cattle are increasingly affordable and thus before the discovery of biomarkers associated with tick resistance in cattle. It may be practical to determine the genomic differences between and within breeds initially. Indeed, determining the relationship between tick infestation and host immune responses using genomics and transcriptomics would be fascinating.

Because of the COVID-19 pandemic, this study faced limitations in collecting and analysing data exclusively from Scottish *Bos taurus* breeds, which are generally considered susceptible to ticks. Additionally, no tick challenge or phenotyping was conducted during this study, as Scotland does not typically face infestations of *Rhipicephalus (Boophilus) microplus*, the southern cattle tick. Despite these constraints, the study provided a unique opportunity to investigate the selected genes associated with host response to pathogens. The samples for this study were collected from various *Bos taurus* breeds commonly found in Scotland, including British, European, and Hill cattle. These breeds were chosen because they represent the typical cattle populations in the region, allowing for a more comprehensive analysis of genetic variations and their potential associations with host response to pathogens. By focusing on these breeds, the study aimed to uncover the novel SNPs that could be crucial for breeding programs to associate tick resistance in cattle populations susceptible to ticks. This research is particularly important as it contributes to the understanding of genetic factors influencing tick resistance, despite the absence of direct tick challenge data. The findings could provide valuable insights for future studies and practical applications in cattle breeding, especially in regions where tick infestations are a significant concern.

5.1 Integration of information from multiple gene expression and genome-wide association studies on host resistance of cattle to infestation with *Rhipicephalus microplus*

5.1.1 Summary of gene expression and GWAS association studies on host resistance of cattle to infestation with *Rhipicephalus microplus*

According to the systematic literature review results in Chapter 2, I observed 16 GEXs and 12 GWAS associated with tick resistance in cattle, which were identified as meeting the established criteria in our survey. Among them are different aspects, approaches, or study methods such as breeding, tissue type of study, tick challenging, and tick exposure to cattle. From the previous studies, I focused on two types of tissue study, including skin and spleen. The reason for choosing the skin tissue type for the study is skin is the first localisation of tick attachment and the first place that shows the host immune response system (Constantinoiu et al., 2018, Piper et al., 2008, Piper et al., 2010, Piper et al., 2009, Wang et al., 2007). Whereas the spleen sample was chosen because the spleen is one of the most important organs involved in immune function (Gasisova et al., 2017). Many studies have examined gene expression (GEXS) in cattle with high and low resistance to ticks, and several gene-wide association studies (GWAS) have been conducted; however, no genetic or phenotypic markers are in commercial use. Besides these studies, the information on the genetic determination of mechanisms and gene markers associated with tick resistance in cattle is still limited. This yielded 10,495 DEGs and 288 QTLs, which were then filtered to only those genes for which multiple studies showed consistent results. The final list included those QTLs significant in at least two independent GWAS in 11 genes. For the GEXS, DEGs significant in at least four skin studies are six genes and DEGs significant in 2 blood studies are ten genes. DEGs and QTLs that were also significant in 1 blood, two skin GEXS or 2 GWAS are ten genes. From the biological processes networks, the total list of 37 genes associated with tick resistance in cattle included two transcription factors (HOXD1, SALL4, and SATB2), 12 genes associated with immune function (BOLA, CD84, CR2, PTGR1, C1QA, CXCL8, F13A1, FN1, SERINC5, LAPTM5, PLA2G7, and TYROBP), 3 with the extracellular matrix, six structural genes including ECM1, ITGA11, MYOC, DNAH14, MYO5C, RHOT1, DKK1, SRGN, and VIPAS39. The HOXD1 is a transcription factor in the Homeobox family. GWAS has detected HOXD1 for muscularity in cattle (Doyle et al., 2020). Niciura et al. (2022) reported that HOXD1 associated with resistance to gastrointestinal nematodes has been detected in Morada Nova sheep. SATB2 has been

known as an oncogene in carcinogenesis; it can be the target for treatment and prevention of cancer. SATB2 can also be a regulator in innate immune responses and unfold protein mitochondria to mediate mitochondrial homeostasis (Huang et al., 2022). Turner et al. (2010) reported that the SATB2 gene showed an association with tick burden in cattle, and it has also reported that it can bind to the matrix attachment region of the immunoglobulin micro locus and activate expression in pre-B-cells (Dobreva et al., 2003). Among all immune system genes associated with the host response to pathogens, the major histocompatibility complex genes [MHC or bovine lymphocyte antigen (BoLA)] are the most studied in cattle. Located on BTA 23, these genes encode cell surface glycoproteins, which work as receptors on antigen-presenting cells, connecting and presenting antigenic peptides to T-lymphocytes, thus triggering the onset of immune response in the host (TIZARD). In the same position on chromosome 23, Machado et al. (2010) found highly significant QTL associated with tick counts using the same experimental population as the current work. The association between MHC class II alleles and tick resistance was also reported by Martinez et al. (2006), who found an association between a low number of ticks and gene-BoLA-DRB 3.2, alleles 18, 20, and 27, suggesting that BoLA-DRB 3.2 alleles could be used to aid in the selection of tick-resistant animals. Other studies also reported a significant relationship between MHC genes and tick counts (Untalan et al., 2007, Acosta-Rodriguez et al., 2005). The expression of CXCL8 decreased or remained the same after the tick exposure. This gene showed variation associated with the discouraging long-term supply of blood feeding to the tick and those involved with host immune responses (Franzin et al., 2017b). The CD84 is an immunoglobulin superfamily member and is regulated during dendritic cell activation (Breloer and Fleischer, 2008); it has been identified potential tick resistance gene when analysing gene expression microarrays in the first hours of tick infestation (Carvalho et al., 2014). In addition, the extracellular matrix genes (ECM1, ITGA11, MYOC, DNAH14, MYO5C, RHOT1, DKK1, SRGN, and VIPAS39.) such as collagen, keratocan, osteoglycin, and lumican were up-regulated in the high resistant level of cattle (Piper et al., 2010). Carvalho et al. (2010) studied the relating coagulation cascade in the skin of high and low resistant cattle infested with R. microplus, while susceptible hosts had an increased blood clotting time at tick haemorrhagic feeding pools in comparison to normal skin and the skin of resistant hosts. Neutrophils might favour infestation by damaging the extracellular matrix around the area of tick attached and thereby allowing access to blood feeding (Tatchell and Bennett, 1969, Tatchell and Moorhouse, 1970). In cattle, the immune system plays a crucial role in determining susceptibility or resistance to tick infestation. In my study, the integration of information from multiple gene expression and genome-wide association studies on host

resistance of cattle to infestation with *Rhipicephalus microplus* identified the candidate genes and their potential association with the immune system and extracellular matrix associated to tick resistance, these genes are strong candidates to be associated with tick resistance. Consequently, I may propose a robust analysis approach involving a systematic review and functional analysis of candidate genes with possible variants for important economical traits in livestock. Recently, the advantage of the new approaches and technologies including genomics, transcriptomics, immune-molecular characterisation, understanding naturally acquired resistance, and the development of innovative arthropod and animal models, holds the potential to enhance investigations into breeds naturally resistant to ticks and tick-borne pathogens. The new technologies may help in the discovery of new candidate biomarkers with resistance against ticks and tick-borne pathogens. By comprehending the interaction between ticks and hosts and identifying common factors in immunity, it becomes possible to manipulate the acquired immune response, enhancing the effectiveness of novel biomarkers associated with the host immune response to tick. Additionally, the knowledge gained may contribute to genomics selection to develop an immune response against ticks in cattle.

5.1.2 Future directions -omics association studies on host resistance of cattle to infestation with *Rhipicephalus microplus*

Several studies have explored genomics and transcriptomics to identify genes or biomarkers related to tick resistance in cattle, but the results have been difficult to interpret. A study on Brangus steers to investigate gene expression in leukocytes following artificial tick infestation was conducted by Mantilla Valdivieso et al. (2022), aiming to identify genes associated with different levels of tick resistance in cattle. They found that at 3 and 12 weeks after infestation, tick exposure induced gene expressions in leukocytes involved to chemotaxis, cytokine secretion, and inflammatory responses. Crucial pathways such as IL-17 and cytokine-cytokine interactions were highlighted in host immune response in steers. Before infestation, significant genes linked to immunity, tissue remodeling, and angiogenesis were identified in both high and low tick resistance groups, emphasising the role of the bovine MHC complex in tick infestation (Mantilla Valdivieso et al., 2022). Although the study did not directly examine leukocyte or blood samples, spleen samples indicated immune responses involving genes like GIMAP7, RHOT1, as well as inflammatory pathways in macrophages (LAPTM5) and the complement receptor (CR2). In their subsequent study, Mantilla Valdivieso et al. (2024) investigated gene expression in bovine skin exposed to *Rhipicephalus australis*. They presented 158 significant transcription

factors (TFs), including GRHL3 and DTX1, and differentially expressed genes (DEGs) involved with host immune response and host defense against infectious diseases. Additionally, analysis of high and low tick resistance groups after 12 weeks of tick challenging showed significant downregulation of keratin and keratin-associated protein genes (KRTAPs) compared to pre-infestation samples. KRTAP9-1, KRTAP9-2 and related genes indicated potential alterations in keratinocyte function in response to tick infestation (Mantilla Valdivieso et al., 2024). Interestingly, the study highlighted the involvement of genes related to keratin and collagen in skin datasets. The expression of keratin in the yellow module and collagen in the greenyellow module are critical for maintaining the integrity of the epidermal barrier, suggesting their role in the local immune response to tick infestation. Although the list of genes identified in this study differs from previous transcription studies, the biological processes or gene functions appear strongly similar. The gene expression patterns suggest that animals with different resistance phenotypes experience distinct stages of wound healing due to varying tick infestation levels. Several studies have explored genomics and transcriptomics to identify genes or biomarkers associated with tick resistance in cattle. However, interpreting the diverse findings remains challenging, and there is a notable absence of information on proteomics studies in this field.

For proteomics study, there are a few studies associated with tick resistance in cattle. Raza et al. (2021) conducted a proteomics study using serum samples from Santa Gertrudis cattle. They used sequential window acquisition and ion mass spectrometry to identify and quantify peptides. They found that 28 protein levels differed significantly between resistant (TR) and susceptible (TS) groups before the tick challenge, with eight proteins associated with adaptive responses. Within the intragroup analysis, both groups exhibited similar responses, but these responses were stronger in TR cattle than in TS cattle. Additionally, many significantly different proteins in the resistant group, both before and after the tick challenge, were involved in immune functions such as complement cascades, chemotaxis, and acute immune responses (Raza et al., 2021). In 2023, Raza et al. (2023) studied quantitative proteomics to analyze serum and skin protein samples from Brangus cattle (naïve tick-resistant and tick-susceptible) at two different time points after tick exposure. They found significantly different protein abundances in resistant cattle after early and prolonged tick exposure (compared to resistant naïve cattle) involved in immune response, coagulation cascade, blood coagulation, homeostasis, and wound healing. Only some of these responses showed significance in susceptible cattle after prolonged tick exposure.

Further studies are needed to confirm these results in different cattle breeds and under various tick infestation conditions (Raza et al., 2023). In the future, more research should be conducted on proteomics across different breeds, experimental designs, and tick conditions. This expanded research effort will provide a more comprehensive understanding of the mechanisms underlying tick resistance in cattle, helping to identify key biomarkers and pathways involved in the immune response to tick infestation. By investigating the proteomic profiles in diverse cattle populations and under different environmental conditions, researchers can uncover potential genetic and molecular targets for breeding programs aimed at enhancing tick resistance. Metabolomic study might be another option for -omics study associated with tick resistance in cattle. Metabolomic studies also highlight the role of gut microbiota and their metabolic products in influencing host resistance. Integrating metabolomics with other omics approaches, such as genomics and proteomics, provides an understanding of the biological mechanisms underpinning tick resistance, potentially leading to the development of targeted interventions to enhance cattle resilience against tick-borne diseases. Although a few studies have shown that host volatile organic compounds (VOCs) provide an important role in *Rhipicephalus* spp. ticks attachment (Franzin et al., 2017a, Jaires et al., 2016, Lígia Miranda Ferreira et al., 2015, Louly et al., 2010, Tabor et al., 2017), further study is required to uncover the relationship between host microbiota and VOC production associated with tick attraction or repulsion. This research could potentially uncover additional factors contributing to tick susceptibility, thereby providing novel changes to improve the tick control processes the livestock ectoparasite, *R. microplus.*

5.2 Variation in genotypic and gene expression among tick resistance in cattle.

5.2.1 Single nucleotide polymorphism (SNP) genotypes in different cattle breeds

Single nucleotide polymorphisms (SNP), another approach used to detect evidence of association with tick resistance throughout the genome. Most markers explained a small proportion (~1%) of the phenotypic variation of the trait (transformed tick counts). This suggested that if SNP were to be used in selection programs, it would have to be a panel of markers rather than a single marker with a high predictive value. Alternatively, a genomic selection approach (Meuwissen et al., 2001) could be taken. This refers to the analysis in

which many SNP are fitted together in the statistical model, potentially resulting in a greater proportion of the variance explained and improving the accuracies of the predictions. This study showed the genotype variation among the three cattle breeds, including British, European, and Hill cattle, using the GGP Bovine 100K array (Neogen). In addition, it also identified candidate genes that have been associated with resistance to ticks (R*hipicephalus microplus*) in cattle based on previous gene expression studies (Chapter 2). It identified SNP in a subset of these genes that showed significant differences in genotypes between breeds. My finding reported that of the 88 SNPs identified from this list of candidate genes associated with tick resistance, six genes (HOXD1, SATB2, GIMAP7, ITGA11, PLA2G7, and PRKG1) presented a total of 14 SNPs that displayed significant differences in genotype frequencies in the different breeds. The SNPs regions were found on BTA2,4,10,23, and 26, respectively. The consequence types were; synonymous variant (on ITGA11); missense variant (on PLA2G7-1); and intron variant (on GIMAP7, HOXD, PLA2G7, PRKG1, and SATB2). Obvious candidate mutations include a premature stop codon, a mutation at a splice site, or a non-synonymous mutation causing a difference in the amino acid sequence (Yuan et al., 2006). The ITGA11 (integrin alpha 11) gene has a role in regulating cellular adhesion and immune responses. ITGA11 genes are located on BTA10 and could potentially be implicated in governing cellular adhesion and movement during the skin infection process resulting from tick infestation (Porto Neto et al., 2010a) and have been linked to tick burden in various dairy cattle breeds, including Australian Red, Brown Swiss, Channel Isle, Holstein, composites, and Brahman beef cattle. In humans, the PLA2G71 gene is identified as a critical regulation gene related to the role of adipose tissue metabolism in regulating immune metabolic effects (Candels et al., 2022), and it is also a novel biomarker of diffuse large B cell lymphoma. Unfortunately, there is no gene function information related to this gene in cattle. Genes that showed variation are associated with immune responses. Identifying biomarkers correlating to tick resistance could improve the sensitivity of prediction tools, subsequently impacting the rational approach to developing novel biomarkers. Additional biomarkers are necessary to improve the accuracy of selection programs and predictive tests for tick-resistant cattle. However, the primary detection of the SNPs presented insight worth exploring using greater population sizes in further study. This study showed considerable variation in the number of individuals presenting a specific genotype in the different breeds, including British, European, and Hill cattle. All of these breeds are *Bos Taurus*, and they were collected from the same abattoir. However, I knew the information from farmers that they had different habitats, different livestock management, and different backgrounds of pathogen exposure so that they could show the variation of

genotype. Then, this is a simple way to generate the group of animals depending on the breed of cattle. Although the results were insufficient to support the confident speculation that the SNPs identified would affect tick resistance in cattle between different breeds, I can generate the group of animals depending on the genotype frequencies from this Chapter for further study.

5.2.2 Future directions of genetic variation and gene expression among tick resistance in cattle

Recent advancements in high-throughput genotyping technologies have significantly improved the identification of new genetic variants, particularly single nucleotide polymorphisms (SNPs), associated with economically important traits in cattle. SNPs are favoured genetic markers due to their heritability and wide distribution across the genome. Genome-wide association studies (GWAS) have become a standard approach to explore SNP markers linked to various traits in animal production. This method correlates phenotype data with genotype data to identify genetic variants causally linked to traits of interest using statistical methods (Mkize et al., 2021). For example, SNP markers for tick resistance in cattle have the potential to assist breeders in making informed decisions to enhance host resistance against ticks. Several GWAS studies have investigated genetic variants for tick resistance across different cattle breeds and regions, experimental design, and tick challenging condition (Acosta-Rodriguez et al., 2005, Gasparin et al., 2007, Machado et al., 2010, Otto et al., 2018, Sollero et al., 2017). Challenges in GWAS include variability in phenotyping methods and genotyping strategies, which could be mitigated by standardizing methods for tick count measurement (Marufu et al., 2011). Genotype imputation has been recognized as a cost-effective solution to handle missing genotypic data and enhance the power of GWAS. Furthermore, collaborative efforts are recommended to overcome limitations related to small discovery populations and the operational costs of GWAS focused on tick resistance. Despite the numerous GWAS studies on tick resistance conducted to date, data availability remains a challenge for enhancing tick resistance traits through genomic selection. Several researchers have emphasized the necessity for a comprehensive understanding of genome variation related to tick resistance in cattle (Ibeagha-Awemu et al., 2008, Mapholi et al., 2014). The generation of genomic data for host resistance to ticks is increasingly prioritized due to the critical need for knowledge on genetic factors influencing this trait.

5.3 Genetic variation and transcriptomic approaches regulate host immune response.

5.3.1 Weighted Gene Co-expression Network Analysis (WGCNA) for full-length cDNA sequencing Data of Scottish cattle to regulate host immune response

The cattle breeding development have been focused on the molecular level including genetics transcriptomics and proteomics, to improve genetics selection strategies to breed cattle with increased resistance to pathogen. Previous gene expression and QTL studies have presented the multiple genes associated with cattle resistance to tick infestation (Franzin et al., 2017b, Marima et al., 2020, Moré et al., 2019, Piper et al., 2008, Piper et al., 2010, Porto Neto et al., 2010a, Porto Neto et al., 2011, Porto Neto et al., 2012, Porto Neto et al., 2010b). In addition, DEGs between high resistance level and low resistance level cattle has been shown at the skin level (Carvalho et al., 2014, Kongsuwan et al., 2008, Wang et al., 2007),which is the initial area of tick attachment and feeding. The key factors hindering the development of cattle breeding is the lack of fully characterised mechanism of host response to pathogen then, there are still no commercial genes or biomarkers in the market. In Chapter 4 was used long-read RNA-seq from ONT to identify DEGs and the biological pathways in the different tissue samples, I also applied the WGCNA which is a well-known analysis to determine genes in the various fields (Garner et al., 2016, Li et al., 2018, Sabino et al., 2018). The significant modules represents the gene expression profiles depending on co-expression analysis of 14 significant SNPs from chapter 3. Seven significant modules from skin datasets including pink, greenyellow, blue, magenta, brown, turquoise, and yellow showed 2,297 genes, whereas there were ten significant modules from spleen datasets including brown, magenta, red, green, purple, greenyellow, blue, turquoise, cyan, and grey, these modules showed 3,265 genes. Interestingly, there were genes which overlapping with the list of 37 genes from Chapter 2 and gene expression profiles in Chapter 4. FN1, ATP9A, and ECM1 genes showed the overlapping from the list of the 37 genes from chapter 2 and gene expression profile from skin datasets, CR2, RHOT1, SRGN, GIMAP7, and LAPTM5 genes showed the overlapping from the list of the 37 genes from Chapter 2 and gene expression profile from spleen datasets. The summarised function of these genes associated in the immune response and extracellular matrix which have been proven important to host response to pathogen. In addition, the results of the key genes also correlated with gene ontology and protein-protein interaction network. Based on the ONT RNA-seq, the co-expression analysis generated here, the combination of the genotype data from Chapter 3,

described various biological pathway presenting differential expression gene related with SNPs showing differences in genotype frequencies between three Scottish cattle breeds. Here, this studies attempted to apply the approach to individual genotypes across multiple genes. Interestingly, the genes associated with tick infestation in cattle based on the synthetic review in Chapter 1 showed significant in co-expression analysis (WGCNA), these genes include FN1, ATP9A, ECM1 from skin studies most of these genes would not be classified as immune genes so their role in tick resistance is interesting. For spleen studies, there are GIMAP7, CR2, RHOT1, SRGN, LAPTM5. it is interesting that none are the same as for the skin dataset and that only one of the genes (GIMAP7) was also one of the genes identified as differing significantly between the cattle breed groups. My finding may suggest that these could be the potential target for develop to biomarker in genetic selection in tick resistance in further study

5.3.2 Directions for Immune-Related Genes in Pathogen Defense Mechanisms in *Bos taurus* cattle

The list of genes in Chapter 4 highlights their potential association with the immune system and extracellular matrix in *Bos taurus* cattle. These genes could play a significant role in various biological processes critical for maintaining health and combating infections. In future studies, a comprehensive investigation of these genes should be conducted to understand their specific pathways and functions related to host defence mechanisms. The research should not be limited to tick resistance in cattle but it should extend to other pathogens as well, providing a broader understanding of how these genes contribute to overall immune response. Additionally, it is important to assess whether genetic variations within these genes lead to changes in their expression levels or functional outcomes. Genetics variations might influence an individual's susceptibility or resistance to different pathogens. The example genetic variation study of genes, the previous study investigated the first knockout zebrafish model for fibronectin 1 (FN1) in relation to amyloid toxicity and identifies cellular changes associated with fibronectin activity (Bhattarai et al., 2024). (Periñán et al., 2021) studied that the mutation of RHOT1 genes identified as potentially linked to Parkinson's disease risk in human. Additional research, including alternative splicing analyses, isoform variation and more extensive validation studies, along with transcriptome-level information on immune-related genes, would enhance our understanding of the molecular mechanisms underlying the host immune response in cattle.

5.4 Future research approaches

In summary, the results of these studies provide comprehensive insights into integrated genes/QTLs from multiple investigations and identify potential genes affecting host responses in cattle to pathogens, particularly cattle ticks. These preliminary data will aid in predicting or accelerating the discovery of genes or biomarkers associated with host responses to pathogens in cattle. Further analyses, such as identifying tissue-specific novel isoforms and their respective biological functions in cattle, should be undertaken. CRISPR/Cas9 gene editing technologies, as discussed by Cui and Yu (2016), are provided to explore host resistance to ticks in cattle and to develop targeted strategies for enhancing tick resistance. The ability to edit host genes using CRISPR/Cas9 also presents opportunities for future research. In 2023, Mueller et al. (2023) contributed to understanding the function of NANOS3 in cattle, providing valuable insights that could potentially lead to the development of innovative breeding technologies using germline complementation. Confirmation of SNPs associated with tick resistance in cattle can be effectively achieved through studies using knockout mice. For instance, Porto Neto et al. (2012) studied how genetic variants on BTA14 and near RIPK2 are linked to tick infestation. Their objective was to elucidate the role of the RIPK2 gene in antibody production against tick salivary gland antigens, employing a mouse knockout model for their experiments (Porto Neto et al., 2012). Recent advantages in novel approaches and technologies (such as transcriptomics, proteomics, immune-molecular profiling, exploration of naturally acquired resistance, and the development of animal models) have the potential to significantly enhance studies on naturally resistant cattle breeds against ticks and tick-borne pathogens. These integrated approaches could pave the way for more effective tick control strategies and contribute to livestock health and productivity. Additional understanding of the gene mechanism interactions between host response to tick in susceptibility or resistance groups of cattle could reveal key genes or biomarkers essential for developing effective tick control strategies and broad-spectrum anti-tick vaccines.

Appendix

Appendix A: The 100 genes overlapping in three skin gene expression studies

Gene name	GeneID	Official Symbol	Official Full Name
Adrenomedullin	280713	ADM	bta-mir-6518
arachidonate lipoxygenase, epidermal	787450	ALOX12E	arachidonate 12-lipoxygenase, epidermal-type
Apolipoprotein D	613972	APOD	apolipoprotein D
arylsulfatase E (chondrodysplasia punctata 1)	505899	ARSL	arylsulfatase L
MHC class I antigen (BoLa)	790811	BOLA	MHC class I antigen
major histocompatibility complex, class II, DRB3	282530	BOLA-DRB3	major histocompatibility complex, class II, DRB3
Complement component 1, q subcomponent, A chain	534961	C1QA	omplement C1q A chain
Complement component 1, q subcomponent, B chain	617435	C1QB	complement C1q B chain
complement component 1, q subcomponent, C chain	509968	C1QC	complement C1q C chain
Complement component 3	280677	C3	complement C3p
coiled-coil domain containing 80	515235	CCDC80	coiled-coil domain containing 80
Chemokine (C–C motif) ligand 2	281043	CCL2	chemokine (C-C motif) ligand 2
chemokine (C-C motif) ligand 3	282170	CCL3	chemokine (C-C motif) ligand 3
chemokine (C-C motif) ligand 8	281044	CCL8	C-C motif chemokine 8
chemokine (C-C motif) receptor 1	407771	CCR1	C-C chemokine receptor type 1
CD14 molecule	281048	CD14	monocyte differentiation antigen CD14 precursor
CD53 molecule	505040	CD53	CD53 molecule

Gene name	GeneID	Official Symbol	Official Full Name
CD68 molecule	504960	CD68	CD68 molecule
CD86 molecule	414345	CD86	CD86 molecule
complement factor B	514076	CFB	complement factor B
complement factor D	505647	CFD	Complement factor D
C-type lectin domain family 4, member G	507332	CLEC4G	C-type lectin domain family 4 member G
collagen, type XI, alpha 1	287013	COL11A1	collagen alpha-1(XI) chain
collagen, type I, alpha 2	282188	COL1A2	Collagen alpha-2(I) chain
collagen, type VI, alpha 3	530657	COL6A3	Uncharacterized protein
carnitine O-octanoyltransferase	281092	CROT	carnitine O-octanoyltransferase
cathepsin K	513038	CTSK	cathepsin K
cathepsin S	327711	CTSS	cathepsin S
Chemokine (C-X-C motif) ligand 2	281214	CXCL2	chemokine (C-X-C motif) ligand 2
Chemokine (C-X-C motif) ligand 5	281735	CXCL5	chemokine (C-X-C motif) ligand 5
interleukin 8	280828	CXCL8	C-X-C motif chemokine ligand 8
Cytochrome P450, family 11, subfamily A, polypeptide 1	338048	CYP11A1	cytochrome P450, family 11, subfamily A, polypeptide 1
cytochrome P450, family 2, subfamily B, polypeptide 6	504769	CYP2B6	cytochrome P450 subfamily 2B
TIGR TC288696	101109448	DDIT4L	DNA damage inducible transcript 4 like
2,4-dienoyl-CoA reductase 1	509952	DECR1	2,4-dienoyl-CoA reductase 1
Dehydrogenase/reductase (SDR family) member 7	514044	DHRS7	dehydrogenase/reductase 7
DnaJ (Hsp40) homolog, subfamily C, member 12	281259	DNAJC12	DnaJ heat shock protein family (Hsp40) member C12
E74-like factor 5 (ets domain transcription factor)	539420	ELF5	E74 like ETS transcription factor 5

Gene name	GeneID	Official Symbol	Official Full Name
similar to ectonucleotide pyrophosphatase/phosphodiesterase 5 (putative function)	512304	ENPP5	ectonucleotide pyrophosphatase/phosphodiesterase family member 5
epoxide hydrolase 2, cytoplasmic, transcript variant 1 (LOC511716),	511716	EPHX2	epoxide hydrolase 2
v-ets erythroblastosis virus E26 oncogene homolog 1 (avian)	540313	ETS1	ETS proto-oncogene 1, transcription factor
Coagulation factor XIII, A1 polypeptide	617881	F13A1	coagulation factor XIII A chain
Fc fragment of IgG, high affinity Ia, receptor (CD64)	282227	FCGR1A	Fc fragment of IgG receptor Ia
fibronectin 1 (FN1), transcript variant 1	280794	FN1	fibronectin 1
v-fos FBJ murine osteosarcoma viral oncogene homolog	280795	FOS	Fos proto-oncogene, AP-1 transcription factor subunit
guanylate binding protein 5	516949	GBP5	guanylate binding protein 5
SIMILAR TO GTPASE, IMAP FAMILY MEMBER 7	530031	GIMAP7	GTPase, IMAP family member 7
G protein-coupled receptor 34	100139092	GPR34	G protein-coupled receptor 34
glutathione S-transferase A1	777644	GSTA1	glutathione S-transferase alpha 1
glutathione S-transferase A1	281805	GSTA2	glutathione S-transferase alpha 2
hemoglobin, beta [beta globin]	280813	HBB	hemoglobin, beta
4-Hydroxyphenylpyruvate dioxygenase	516058	HPD	4-hydroxyphenylpyruvate dioxygenase
Hydroxysteroid (17-beta) dehydrogenase 13	618192	HSD17B13	hydroxysteroid 17-beta dehydrogenase 13
inteferon-induced membrane protein Leu-13/9-28	353510	IFITM1	interferon induced transmembrane protein 1 (9-27)
IgG2a heavy chain constant region	404109	IGCGAMMA	immunoglobulin heavy constant gamma 2
Homo sapiens: insulin like growth factor binding protein 5	3488	IGFBP5	insulin like growth factor binding protein 5

Gene name	GeneID	Official Symbol	Official Full Name
v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog	280832	KIT	KIT proto-oncogene, receptor tyrosine kinase
keratin 18	506480	KRT18	keratin 18
cytokeratin 19	514812	KRT19	keratin 19
Regakine 1	504773	LOC504773	regakine 1
cytochrome P450, family 4, subfamily F, polypeptide 3	534967	LOC534967	leukotriene-B(4) omega-hydroxylase 2-like
Lipoprotein lipase	280843	LPL	lipoprotein lipase
hematopoietic cell-specific Lyn substrate 1	508445	HCLS1	hematopoietic cell-specific Lyn substrate 1
lymphatic vessel endothelial hyaluronan receptor 1	404179	LYVE1	lymphatic vessel endothelial hyaluronan receptor 1
Lysozyme (renal amyloidosis)	781349	LYZ1	lysozyme (renal amyloidosis)
Lysozyme (renal amyloidosis)	781349	LYZ1	lysozyme (renal amyloidosis)
mal, T-cell differentiation protein	510077	MAL	mal, T cell differentiation protein
myristoylated alanine-rich protein kinase C substrate	613548	MARCKS	myristoylated alanine rich protein kinase C substrate
hypothetical LOC507942	507942	MGC152281	uncharacterized LOC507942
similar to mannose receptor C1	4360	MRC1	mannose receptor C-type 1
myxovirus (influenza virus) resistance 1, interferon-inducible protein p78 (mouse)	280872	MX1	MX dynamin like GTPase 1
oncostatin M receptor	514720	OSMR	oncostatin M receptor
Procollagen C-endopeptidase enhancer	504471	PCOLCE	procollagen C-endopeptidase enhancer
phosphodiesterase 4B, cAMP-specific	100124505	PDE4B	phosphodiesterase 4B
platelet-derived growth factor receptor-like	515017	PDGFRL	platelet derived growth factor receptor like
podoplanin	509732	PDPN	podoplanin
Peptidoglycan recognition protein 1	282305	PGLYRP1	peptidoglycan recognition protein 1

Gene name	GeneID	Official Symbol	Official Full Name
Phospholipase A2, group VII (platelet-activating factor acetylhydrolase, plasma)	282311	PLA2G7	phospholipase A2 group VII
Bos taurus paraoxonase 3, mRNA.	510953	PON3	paraoxonase 3
Homo sapiens: ras homologue gene family, member Q (RHOQ), mRNA NM_012249	23433	RHOQ	ras homolog family member Q
ribonuclease, RNase A family, k6	282341	RNASE6	ribonuclease A family member k6
Bos taurus serum amyloid A 3, mRNA.	281474	SAA3	serum amyloid A 3
SAM domain, SH3 domain and nuclear localization signals 1	534171	SAMSN1	SAM domain, SH3 domain and nuclear localization signals 1
similar to Selenoprotein M precursor (Protein SelM)	787736	SELENOM	selenoprotein M
serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 1	280699	SERPINA1	serpin family A member 1
Serpin peptidase inhibitor, clade F	20317	Serpinf1	serine (or cysteine) peptidase inhibitor, clade F, member 1
shisa homolog 2 (Xenopus laevis)	617336	SHISA2	shisa family member 2
signal-regulatory protein alpha	327666	SIRPA	signal regulatory protein alpha
similar to solute carrier family 16, member 11	618274	SLC16A11	solute carrier family 16 member 11
solute carrier family 6, member 16	785362	SLC6A16	solute carrier family 6 member 16
proteoglycan 1, secretory granule; Proteoglycan 1,	509501	SRGN	serglycin
Tryptophan 2,3-dioxygenase	530397	TDO2	tryptophan 2,3-dioxygenas
thyroid hormone responsive (SPOT14 homolog, rat)	515940	THRSP	thyroid hormone responsive
TIMP metallopeptidase inhibitor 1	282092	TIMP1	TIMP metallopeptidase inhibitor 1
toll-like receptor 2	281534	TLR2	toll like receptor 2
tenascin C	540664	TNC	tenascin C

Gene name	GeneID	Official Symbol	Official Full Name
calcium channel protein transient receptor potential vanilloid 6	614878	TRPV6	transient receptor potential cation channel subfamily V member 6
TYRO protein tyrosine kinase binding protein	282390	TYROBP	transmembrane immune signaling adaptor TYROBP
similar to ubiquitin-specific protease 7	514813	USP7	ubiquitin specific peptidase 7
versican	282662	VCAN	versican

Appendix B: The 114 genes overlapping in two blood gene expression studies

Gene name	GeneID	Official Symbol	Official Full Name
4-aminobutyrate aminotransferase	280969	ABAT	4-aminobutyrate aminotransferase
acetyl-CoA acyltransferase 2	522006	ACAA2	acetyl-CoA acyltransferase 2
DELTA-AMINOLEVULINIC ACID DEHYDRATASE ISOFORM B	510679	ALAD	aminolevulinate dehydratase
Rho guanine nucleotide exchange factor (GEF) 12	503575	ARHGEF12	Rho guanine nucleotide exchange factor 12
ariadne RBR E3 ubiquitin protein ligase 2	505320	ARIH2	ariadne RBR E3 ubiquitin protein ligase 2
ATP SYNTHASE, H+ TRANSPORTING, MITOCHONDRIAL F1 COMPLEX, BETA POLYPEPTIDE	327675	ATP5F1B	ATP synthase F1 subunit beta
ATP SYNTHASE, H+ TRANSPORTING, MITOCHONDRIAL F1 COMPLEX, O SUBUNIT	281640	ATP5PO	ATP synthase peripheral stalk subunit OSCP
SIMILAR TO ATPASE, H+ TRANSPORTING, LYSOSOMAL V0 SUBUNIT A ISOFORM 4	518974	ATP6V0A4	ATPase H+ transporting V0 subunit a4
ATPASE, H+ TRANSPORTING, LYSOSOMAL 9KDA, V0 SUBUNIT E	338075	ATP6V0E1	ATPase H+ transporting V0 subunit e1
SIMILAR TO ATF-LIKE BASIC LEUCINE ZIPPER TRANSCRIPTION FACTOR B-ATF	617628	BATF	basic leucine zipper ATF-like transcription factor
BCL2 apoptosis regulator	281020	BCL2	BCL2 apoptosis regulator
HYPOTHETICAL PROTEIN LOC614339	614339	C18H16orf87	chromosome 18 C16orf87 homolog
SIMILAR TO CALMODULIN REGULATED SPECTRIN-ASSOCIATED PROTEIN 1	506553	CAMSAP1	calmodulin regulated spectrin associated protein 1
SIMILAR TO GENE MODEL 631, (NCBI) (PREDICTED)	506935	CCDC32	coiled-coil domain containing 32

Gene name	GeneID	Official Symbol	Official Full Name
CYCLIN B2	281668	CCNB2	cyclin B2
CHEMOKINE RECEPTOR 7	510668	CCR7	C-C motif chemokine receptor 7
SIMILAR TO CDC42 SMALL EFFECTOR 2	789618	CDC42SE2	CDC42 small effector 2
CORONIN, ACTIN BINDING PROTEIN, 1B	510128	CORO1B	coronin 1B
COMPLEMENT RECEPTOR TYPE 2	407126	CR2	complement C3d receptor 2
SIMILAR TO CAMP RESPONSIVE ELEMENT BINDING PROTEIN 3-LIKE 1	513105	CREB3L1	cAMP responsive element binding protein 3 like 1
SIMILAR TO CG2843-PA	512527	CWC25	CWC25 spliceosome associated protein homolog
CHEMOKINE (C-X-C MOTIF) RECEPTOR 4	281736	CXCR4	C-X-C motif chemokine receptor 4
Similar to cytoplasmic FMR1 interacting protein 1	100141021	CYFIP1	cytoplasmic FMR1 interacting protein 1
SIMILAR TO DEAD (ASP-GLU-ALA-ASP) BOX POLYPEPTIDE 1	510816	DDX1	DEAD-box helicase 1
SIMILAR TO 2,4-DIENOYL-COA REDUCTASE, MITOCHONDRIAL PRECURSOR	509952	DECR1	2,4-dienoyl-CoA reductase 1
SIMILAR TO AXONEMAL DYNEIN HEAVY CHAIN 7	516576	DNAH14	dynein axonemal heavy chain 14
SIMILAR TO DYSTROBREVIN BINDING PROTEIN 1 ISOFORM A	506612	DTNBP1	dystrobrevin binding protein 1
SIMILAR TO EXTRACELLULAR MATRIX PROTEIN 1 ISOFORM 1 PRECURSOR	524222	ECM1	extracellular matrix protein 1
SIMILAR TO INHIBITOR OF KAPPA LIGHT POLYPEPTIDE GENE ENHANCER IN B-CELLS, KINASE COMPLEX-ASSOCIATED PROTEIN	505465	ELP1	elongator acetyltransferase complex subunit 1
ENDOPLASMIC RETICULUM PROTEIN ERP29	613357	ERP29	endoplasmic reticulum protein 29
SIMILAR TO EYES ABSENT 3 ISOFORM B	514364	EYA3	EYA transcriptional coactivator and phosphatase 3

Gene name	GeneID	Official Symbol	Official Full Name
SIMILAR TO PROTEINASE ACTIVATED RECEPTOR 1 PRECURSOR (PAR-1) (THROMBIN RECEPTOR)	526585	F2R	coagulation factor II thrombin receptor
SIMILAR TO CG8312-PA, ISOFORM A	539374	FAM43A	family with sequence similarity 43 member A
SIMILAR TO FUCOSE-1-PHOSPHATE GUANYLTRANSFERASE	100138313	FPGT	fucose-1-phosphate guanylyltransferase
GATA binding protein 2	506809	GATA2	GATA binding protein 2
SIMILAR TO ELONGATION FACTOR G 1	513309	GFM1	G elongation factor mitochondrial 1
SIMILAR TO GTPASE, IMAP FAMILY MEMBER 7	530031	GIMAP7	GTPase, IMAP family member 7
GEMININ-LIKE	526377	GMNN	geminin DNA replication inhibitor
SIMILAR TO GMP REDUCTASE 2 (GUANOSINE 5-MONOPHOSPHATE OXIDOREDUCTASE 2)	515837	GMPR2	guanosine monophosphate reductase 2
Similar to Granzyme B	281731	GZMB	granzyme B (granzyme 2, cytotoxic T-lymphocyte-associated serine esterase 1)
SIMILAR TO HYALURONAN BINDING PROTEIN 4	514807	HABP4	hyaluronan binding protein 4
SIMILAR TO SHORT CHAIN 3-HYDROXYACYL-COA DEHYDROGENASE, MITOCHONDRIAL PRECURSOR	532785	HADH	hydroxyacyl-CoA dehydrogenase
SIMILAR TO HISTIDYL-TRNA SYNTHETASE HOMOLOG (HISTIDINE--TRNA LIGASE HOMOLOG) (HISRS)	615182	HARS2	histidyl-tRNA synthetase 2, mitochondrial
SIMILAR TO HMBA-INDUCIBLE	539696	HEXIM1	HEXIM P-TEFb complex subunit 1
SIMILAR TO HIGH-MOBILITY GROUP PROTEIN 2-LIKE 1 ISOFORM B	505539	HMGXB4	HMG-box containing 4

Gene name	GeneID	Official Symbol	Official Full Name
HYDROXYACYL-COENZYME A DEHYDROGENASE, TYPE II HYDROXYACYL-COENZYME A	281809	HSD17B10	hydroxysteroid 17-beta dehydrogenase 10
HEAT SHOCK TRANSCRIPTION FACTOR 2 BINDING PROTEIN	617142	HSF2BP	heat shock transcription factor 2 binding protein
SIMILAR TO INTERFERON-INDUCED 35 KDA PROTEIN (IFP 35)	510697	IFI35	interferon induced protein 35
INTERLEUKIN 2 RECEPTOR, ALPHA	281861	IL2RA	interleukin 2 receptor subunit alpha
SIMILAR TO TYPE II INOSITOL-3,4-BISPHOSPHATE 4-PHOSPHATASE	534793	INPP4B	inositol polyphosphate-4-phosphatase type II B
integrator complex subunit 8	100124522	INTS8	integrator complex subunit 8
SIMILAR TO RAS GTPASE-ACTIVATING-LIKE PROTEIN IQGAP2	782315	IQGAP2	IQ motif containing GTPase activating protein 2
SIMILAR TO INTERLEUKIN-1 RECEPTOR-ASSOCIATED KINASE 4 (IRAK-4) (NY-REN-64 ANTIGEN)	533692	IRAK4	interleukin 1 receptor associated kinase 4
SIMILAR TO INOSINE TRIPHOSPHATASE ISOFORM A	613653	ITPA	inosine triphosphatase
SIMILAR TO INTERMEDIATE CONDUCTANCE CALCIUM-ACTIVATED POTASSIUM CHANNEL PROTEIN 4	534591	KCNN4	potassium calcium-activated channel subfamily N member 4
SIMILAR TO KELCH DOMAIN CONTAINING 2	535436	KLHDC2	kelch domain containing 2
SIMILAR TO L(3)MBT-LIKE 2 ISOFORM A	513297	L3MBTL2	L3MBTL histone methyl-lysine binding protein 2
laminin subunit alpha 4	529670	LAMA4	laminin subunit alpha 4
LINKER FOR ACTIVATION OF T CELLS	514735	LAT	linker for activation of T cells

Gene name	GeneID	Official Symbol	Official Full Name
limb bud and heart development	616148	LBH	LBH regulator of WNT signaling pathway
SIMILAR TO RHOMBOTIN-2 (CYSTEINE-RICH PROTEIN TTG-2) (T-CELL TRANSLOCATION PROTEIN 2) (LIM-ONLY PROTEIN 2)	614876	LMO2	LIM domain only 2
SIMILAR TO LYMPHOCYTE ANTIGEN 86 PRECURSOR (MD-1 PROTEIN)	613856	LY86	lymphocyte antigen 86
SIMILAR TO LIMKAIN B1	506615	MARF1	meiosis regulator and mRNA stability factor 1
MALATE DEHYDROGENASE 2, MITOCHONDRIAL	281306	MDH2	malate dehydrogenase 2
SIMILAR TO CHROMOSOME 17 OPEN READING FRAME 37	505710	MIEN1	migration and invasion enhancer 1
MNAT CDK-activating kinase assembly factor 1	534176	MNAT1	MNAT1 component of CDK activating kinase
SIMILAR TO BRAIN PROTEIN 44	616718	MPC2	mitochondrial pyruvate carrier 2
SIMILAR TO N-METHYLPURINE-DNA GLYCOSYLASE ISOFORM B	618306	MPG	N-methylpurine DNA glycosylase
SIMILAR TO MITOCHONDRIAL RIBOSOMAL PROTEIN L51	513622	MRPL51	mitochondrial ribosomal protein L51
methylenetetrahydrofolate dehydrogenase (NADP+ dependent) 2, methenyltetrahydrofolate cyclohydrolase	517539	MTHFD2	methylenetetrahydrofolate dehydrogenase (NADP+ dependent) 2, methenyltetrahydrofolate cyclohydrolase
SIMILAR TO N-MYC DOWNSTREAM REGULATED GENE 3	514399	NDRG3	NDRG family member 3
CELL DEATH-REGULATORY PROTEIN GRIM19	338084	NDUFA13	NADH:ubiquinone oxidoreductase subunit A13

Gene name	GeneID	Official Symbol	Official Full Name
NADH DEHYDROGENASE (UBIQUINONE) 1 BETA SUBCOMPLEX, 3, 12KDA	338073	NDUFB3	NADH:ubiquinone oxidoreductase subunit B3
NADH DEHYDROGENASE (UBIQUINONE) FLAVOPROTEIN 1, 51KDA	287014	NDUFV1	NADH:ubiquinone oxidoreductase core subunit V1
SIMILAR TO NEBULIN	407121	NEB	nebulin
SIMILAR TO EXPRESSED SEQUENCE C78541	615252	NOL12	nucleolar protein 12
SIMILAR TO CG4497-PA	539712	PABIR1	PP2A Aalpha (PPP2R1A) and B55A (PPP2R2A) interacting phosphatase regulator 1
SIMILAR TO PROGRAMMED CELL DEATH 1 LIGAND 2	539392	PDCD1LG2	programmed cell death 1 ligand 2
SIMILAR TO PEROXISOMAL TRANS-2-ENOYL-COA REDUCTASE	617037	PECR	peroxisomal trans-2-enoyl-CoA reductase
SIMILAR TO PELLINO PROTEIN	535054	PELI1	pellino E3 ubiquitin protein ligase 1
SIMILAR TO PHD FINGER PROTEIN 12	507057	PHF12	PHD finger protein 12
SIMILAR TO PHOSPHATIDYLINOSITOL GLYCAN, CLASS N	525095	PIGN	phosphatidylinositol glycan anchor biosynthesis class N
SIMILAR TO SERINE-THREONINE PROTEIN KINASE PIM-2 ISOFORM 1	508424	PIM2	Pim-2 proto-oncogene, serine/threonine kinase
SIMILAR TO METALLOPROTEASE 1	513230	PITRM1	pitrilysin metallopeptidase 1
phospholipase D family, member 4	789315	PLD4	phospholipase D family member 4
pleckstrin homology domain containing, family A (phosphoinositide binding specific) member 1	513040	PLEKHA1	pleckstrin homology domain containing A1
SIMILAR TO RNA POLYMERASE I ASSOCIATED FACTOR 53	511587	POLR1E	RNA polymerase I subunit E
SIMILAR TO PEPTIDYLPROLYL ISOMERASE-LIKE PROTEIN 3 ISOFORM PPIL3B	615703	PPIL3	peptidylprolyl isomerase like 3

Gene name	GeneID	Official Symbol	Official Full Name
PROTEIN PHOSPHATASE 1, REGULATORY (INHIBITOR) SUBUNIT 1B	282459	PPP1R1B	protein phosphatase 1 regulatory inhibitor subunit 1B
protein phosphatase 3, catalytic subunit, alpha isozyme	286852	PPP3CA	protein phosphatase 3 catalytic subunit alpha
RAP1, GTP-GDP dissociation stimulator 1	282516	RAP1GDS1	Rap1 GTPase-GDP dissociation stimulator 1
SIMILAR TO RETINOBLASTOMA-ASSOCIATED PROTEIN (PP110) (P105-RB) (RB)	534712	RB1	RB transcriptional corepressor 1
SIMILAR TO RNA PSEUDOURIDYLATE SYNTHAS	534606	RPUSD3	RNA pseudouridine synthase D3
ribonucleotide reductase regulatory subunit M2	508167	RRM2	ribonucleotide reductase regulatory subunit M2
SIMILAR TO SH2 DOMAIN CONTAINING 3C ISOFORM 2	515820	SH2D3C	SH2 domain containing 3C
SOLUTE CARRIER FAMILY 25 MEMBER 5	282479	SLC25A5	solute carrier family 25 member 5
SIMILAR TO SIGNAL TRANSDUCER AND ACTIVATOR OF TRANSCRIPTION 4	515988	STAT4	signal transducer and activator of transcription 4
SIGNAL TRANSDUCER AND ACTIVATOR OF TRANSCRIPTION 5B	282376	STAT5B	signal transducer and activator of transcription 5B
SUCCINATE-COA LIGASE, GDP-FORMING, ALPHA SUBUNIT	509983	SUCLG1	succinate-CoA ligase GDP/ADP-forming subunit alpha
SIMILAR TO NESPRIN-2 (NUCLEAR ENVELOPE SPECTRIN REPEAT PROTEIN 2) (SYNE-2) (SYNAPTIC NUCLEAR ENVELOPE PROTEIN 2) (NUCLEUS AND ACTIN CONNECTING ELEMENT PROTEIN) (NUANCE PROTEIN)	540504	SYNE2	spectrin repeat containing nuclear envelope protein 2

Gene name	GeneID	Official Symbol	Official Full Name
SIMILAR TO TRANSMEMBRANE 9 SUPERFAMILY PROTEIN MEMBER 2 PRECURSOR	514455	TM9SF2	transmembrane 9 superfamily member 2
tripartite motif containing 14	522632	TRIM14	tripartite motif containing 14
tetratricopeptide repeat domain 39C	532895	TTC39C	tetratricopeptide repeat domain 39C
THIOREDOXIN 2	281557	TXN2	thioredoxin 2
SIMILAR TO UBIQUITIN PROTEIN LIGASE E3B ISOFORM A	512750	UBE3B	ubiquitin protein ligase E3B
UDP-GLUCOSE DEHYDROGENASE	281564	UGDH	UDP-glucose 6-dehydrogenase
UBIQUINOL-CYTOCHROME C REDUCTASE (6.4KD) SUBUNIT	281570	UQCR11	ubiquinol-cytochrome c reductase, complex III subunit XI
SIMILAR TO VESICLE AMINE TRANSPORT PROTEIN 1	510698	VAT1	vesicle amine transport 1
VERSICAN	282662	VCAN	versican
SIMILAR TO PROTEIN C14ORF133 HOMOLOG	508872	VIPAS39	VPS33B interacting protein, apical-basolateral polarity regulator, spe-39 homolog
SIMILAR TO WD REPEAT DOMAIN 76	515302	WDR76	WD repeat domain 76
SIMILAR TO ZINC FINGER, DHHC DOMAIN CONTAINING 13 ISOFORM 1	514176	ZDHHC13	zinc finger DHHC-type palmitoyltransferase 13
ZFP36 ring finger protein-like 2	767898	ZFP36L2	ZFP36 ring finger protein like 2
zinc finger protein 462	515561	ZNF462	zinc finger protein 462

Appendix C: The 88 SNPs identified from this list of candidate genes

Gene Name	SNP Name	British AA	British AB	British BB	European AA	European AB	European BB	Hill AA	Hill AB	Hill BB	chi-square	df	p_value
ALPL	Hapmap46765-BTA-114976	2	9	2	1	6	1	2	5	2	0.78846	4	0.94
ALPL	Hapmap47186-BTA-114970	1	7	6	2	3	4	2	4	1	3.3157	4	0.5064
ATP9A	ARS-BFGL-NGS-35502	1	5	5	0	2	8	0	1	9	6.1968	4	0.1849
ATP9A	ARS-BFGL-NGS-36793	4	5	3	2	4	3	3	4	2	0.50125	4	0.9734
ATP9A	ARS-BFGL-NGS-78930	1	3	10	1	2	4	0	5	3	4.6942	4	0.3201
ATP9A	BovineHD1300023198	0	2	12	1	1	8	1	4	5	5.1375	4	0.2735
ATP9A	BovineHD1300023202	3	3	0	1	5	0	3	3	0	1.8701	2	0.3926
ATP9A	BovineHD1300023209	0	1	10	0	1	6	0	0	8	1.1255	2	0.5696
ATP9A	BovineHD1300023225	1	7	4	3	4	3	1	6	3	2.4345	4	0.6564
ATP9A	BovineHD4100010464	3	5	6	0	4	5	0	3	7	5.1362	4	0.2736
CA2	BovineHD1400022284	3	5	6	2	7	1	0	8	2	6.9403	4	0.1391
CA2	BovineHD1400022287	0	3	11	1	5	4	0	3	7	5.2987	4	0.258
DKK1	BovineHD2600001559	9	4	0	8	2	0	8	0	1	5.5226	4	0.2378
GIMAP7	Hapmap22875-BTA-155031	0	7	7	1	4	5	2	7	1	6.5094	4	0.1642

Gene Name	SNP Name	British AA	British AB	British BB	European AA	European AB	European BB	Hill AA	Hill AB	Hill BB	chi-square	df	p_value
GIMAP7	Hapmap59459-rs29011115	1	10	3	0	1	9	0	7	3	13.222	4	0.0102
GSTM1	BovineHD0300010553	1	8	5	2	6	2	0	4	6	4.9319	4	0.2944
HOXD1	ARS-BFGL-NGS-118009	1	7	2	7	2	0	1	4	5	16.693	4	0.0022
ITGA11	10-15108992-A-G-rs384947169	12	1	0	10	0	0	10	0	0	1.5865	2	0.4524
ITGA11	ARS-BFGL-NGS-119197	5	8	1	9	1	0	3	6	0	9.5654	4	0.0484
ITGA11	ARS-BFGL-NGS-29149	1	3	1	1	0	8	0	2	2	8.48	4	0.0755
ITGA11	ARS-BFGL-NGS-76838	0	4	10	1	5	2	1	5	3	6.1227	4	0.1902
ITGA11	BovineHD1000005028	1	7	6	4	4	2	2	5	2	4.4951	4	0.3431
KCNK17	BovineHD2300003313	0	5	9	0	6	4	2	5	3	7.1643	4	0.1275
LAPTM5	BovineHD0200035769	7	6	1	3	6	0	3	6	1	2.1839	4	0.702
LAPTM5	Hapmap40752-BTA-49567	5	6	2	0	5	4	1	5	4	6.6773	4	0.154
MYOC	BovineHD1600011364	0	4	10	0	2	8	0	4	6	0.97143	2	0.6153
PLA2G7	ARS-BFGL-NGS-111955	1	12	1	0	0	9	1	5	4	19.802	4	0.0005
PLA2G7	BovineHD2300005156	1	10	3	5	5	0	6	3	1	9.7143	4	0.0455
PLA2G7	BovineHD2300005160	2	8	4	3	6	1	6	3	1	6.5446	4	0.162
PRKG1	ARS-BFGL-NGS-10114	6	7	1	7	3	0	6	3	1	2.5879	4	0.629

Gene Name	SNP Name	British AA	British AB	British BB	European AA	European AB	European BB	Hill AA	Hill AB	Hill BB	chi-square	df	p_value
PRKG1	ARS-BFGL-NGS-106308	0	4	10	0	0	9	1	1	8	6.0552	4	0.1951
PRKG1	ARS-BFGL-NGS-109705	1	6	6	3	2	4	3	3	4	3.0547	4	0.5487
PRKG1	ARS-BFGL-NGS-110280	0	6	8	1	3	6	1	1	7	3.7021	4	0.4478
PRKG1	ARS-BFGL-NGS-43507	7	6	1	4	4	1	9	1	0	5.4329	4	0.2457
PRKG1	ARS-BFGL-NGS-4669	0	1	13	0	0	9	0	0	9	1.3272	2	0.515
PRKG1	ARS-BFGL-NGS-47701	8	6	0	6	3	1	2	4	4	9.038	4	0.0602
PRKG1	ARS-BFGL-NGS-58471	4	7	3	5	3	2	8	2	0	6.7771	4	0.1481
PRKG1	ARS-BFGL-NGS-73895	3	6	4	2	3	4	8	2	0	10.715	4	0.03
PRKG1	ARS-BFGL-NGS-82901	2	9	3	5	3	2	6	3	1	6.3716	4	0.1731
PRKG1	ARS-USMARC-Parent-DQ990834-rs29013727	2	7	5	2	5	2	4	4	2	2.5037	4	0.644
PRKG1	BovineHD2600001571	1	3	10	0	5	5	0	5	5	3.8297	4	0.4295
PRKG1	BovineHD2600001576	1	4	9	0	4	5	0	0	10	6.9296	4	0.1397
PRKG1	BovineHD2600001618	6	8	0	3	5	1	1	2	7	16.835	4	0.0021
PRKG1	BovineHD2600001624	3	5	6	1	4	4	0	1	9	6.9321	4	0.1395
PRKG1	BovineHD2600001653	7	7	0	6	2	2	7	2	1	5.3576	4	0.2525

Gene Name	SNP Name	British AA	British AB	British BB	European AA	European AB	European BB	Hill AA	Hill AB	Hill BB	chi-square	df	p_value
PRKG1	BovineHD2600001677	7	7	0	4	6	0	4	4	2	5.3867	4	0.2499
PRKG1	BovineHD2600001699	0	6	8	0	3	7	2	5	3	6.8925	4	0.1417
PRKG1	BovineHD2600001755	4	8	2	0	6	4	3	3	4	5.9167	4	0.2055
PRKG1	BovineHD2600001760	2	7	1	1	3	4	1	3	4	4.3694	4	0.3583
PRKG1	BovineHD2600001782	1	5	8	0	1	9	8	1	1	24.216	4	0.00007
PRKG1	BovineHD2600001842	0	8	6	3	3	3	1	1	8	11.028	4	0.0263
PRKG1	BovineHD2600001851	0	7	7	1	5	3	0	0	7	9.3219	4	0.0535
PRKG1	BovineHD2600001852	7	7	0	4	5	1	8	1	1	5.9328	4	0.2042
PRKG1	BovineHD2600001901	8	5	0	4	3	2	5	2	2	3.6128	4	0.4609
PRKG1	BovineHD2600001918	1	8	5	2	4	1	2	4	3	2.4762	4	0.6489
PRKG1	BovineHD2600001921	9	5	0	4	3	3	8	1	1	7.3435	4	0.1188
PRKG1	BovineHD2600001940	9	5	0	6	1	3	5	4	1	6.5571	4	0.1612
PRKG1	BovineHD2600001964	6	7	1	5	4	0	2	2	6	13.315	4	0.0098
PRKG1	BovineHD2600001969	0	7	7	0	4	6	0	1	9	4.2242	4	0.121
PRKG1	BovineHD2600001993	4	8	2	4	2	3	4	5	1	3.5881	4	0.4646
PRKG1	BovineHD2600002013	5	7	2	6	3	0	8	1	1	6.2075	4	0.1842

Gene Name	SNP Name	British AA	British AB	British BB	European AA	European AB	European BB	Hill AA	Hill AB	Hill BB	chi-square	df	p_value
PRKG1	BovineHD2600002046	1	4	9	4	2	4	1	0	9	8.7134	4	0.0687
PRKG1	BovineHD4100017639	2	5	6	1	6	3	1	1	7	5.4023	4	0.2485
PRKG1	BovineHD4100017647	7	6	1	5	3	2	3	5	2	2.0678	4	0.7233
PRKG1	BTA-113588-no-rs	0	3	11	2	4	4	3	5	2	9.5638	4	0.0485
PRKG1	BTA-113590-no-rs	0	1	13	0	1	9	1	4	5	7.9873	4	0.092
PRKG1	BTA-61968-no-rs	5	6	3	6	1	1	8	1	0	7.6881	4	0.1037
PRKG1	BTB-00922910	0	10	4	1	6	3	3	5	2	5.1694	4	0.2704
PRKG1	BTB-01078331	6	7	1	9	0	1	9	0	1	12.669	4	0.013
PRKG1	BTB-01078829	4	6	3	2	4	2	6	3	0	4.7944	4	0.3091
PRKG1	DB-1374-seq-rs446899684	0	0	14	0	0	10	0	0	10	0.94118	2	0.6246
PRKG1	Hapmap33924-BES8_Contig245_678	10	4	0	7	3	0	4	5	1	4.3599	4	0.3595
PRKG1	Hapmap48832-BTA-61957	0	6	8	3	5	1	3	4	3	8.0902	4	0.0883
PRKG1	Hapmap53362-rs29013727	2	7	5	2	5	3	4	4	2	2.3557	4	0.6706
SATB2	BovineHD0200025066	5	5	4	4	3	2	0	3	7	7.591	4	0.1078
SATB2	BovineHD0200025071	13	1	0	9	1	0	10	0	0	0.97143	2	0.6153
SATB2	BovineHD0200025079	0	6	8	0	8	2	0	1	9	9.9529	2	0.0069
SATB2	BovineHD0200025086	9	5	0	5	5	0	9	1	0	3.7782	2	0.1512
SATB2	BovineHD0200025097	8	5	1	1	4	4	1	7	2	10.842	4	0.0284

Gene Name	SNP Name	British AA	British AB	British BB	European AA	European AB	European BB	Hill AA	Hill AB	Hill BB	chi-square	df	p_value
SATB2	BovineHD0200025113	1	4	9	1	5	3	2	6	2	5.2891	4	0.2589
SERINC5	ARS-BFGL-NGS-119397	1	6	4	0	7	2	1	4	5	3.0677	4	0.5466
SERINC5	ARS-BFGL-NGS-66090	7	5	2	4	4	1	8	2	0	3.538	4	0.4721
SERINC5	BovineHD4100007857	7	4	3	3	6	1	8	2	0	7.2857	4	0.1215
SERINC5	BTB-00408221	8	5	1	6	4	0	6	3	1	1.0686	4	0.8992
SIRPA	ARS-BFGL-NGS-116998	3	6	4	1	4	4	0	2	7	5.5906	4	0.2319
SIRPA	BovineHD1300015209	3	5	6	2	3	4	2	1	7	2.5469	4	0.6363
SIRPA	BovineHD1300015212	5	6	3	2	5	3	4	2	4	2.6969	4	0.6098
TYROBP	BovineHD1800013787	0	4	10	2	4	3	0	3	7	7.316	4	0.1201

Appendix D: The significant pathways of significant modules in skin datasets. This table shows the genes involved in each pathway and the value of Enrichment FDR and Fold Enrichment

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
blue	1.47E-07	22	363	5.150657	Amyotrophic lateral sclerosis	ATP5F1D ATP5F1A NUP133 FUS SDHB SEH1L MAP1LC3B TUBA1B NDUFB3 ATG2B SQSTM1 OPTN VAPB PSMD7 SOD1 NDUFB9 ATP5PD PSMD14 ACTB COX5B
blue	1.031E-08	29	547	4.505648	MRNA metabolic proc.	UPF2 CSDE1 SRRT SYF2 CELF1 PAPOLA EXOSC9 FUS CCNT1 PRKACA EIF3E CPSF6 RBM8A SLU7 PUM2 XRN2 SRSF9 WBP11 IK PCF11 PRPF18 APP AHCYL1 NSRP1 STAT3 THRAP3 SNRPC TENT5C
blue	1.495E-07	25	472	4.501368	Pathways of neurodegeneration	ATP5F1D ATP5F1A FUS SDHB MAP1LC3B TUBA1B CSNK2A1 NDUFB3 ATG2B SQSTM1 OPTN UBA1 VAPB APP PSMD7 SOD1 NDUFB9 ATP5PD PSMD14 FZD3 SLC25A5 NRAS COX5B
blue	3.908E-09	33	675	4.154863	Reg. of catabolic proc.	UBXN2A MSN EXOSC9 FUS PRKACA HMGCR FBXL20 CRTC3 PUM2 XPO1 UCHL5 EIF3H RAD23A EP300 OPTN APP WDR91 EGLN2 WAC PTPN3 MYCBP2 EIF4G2 STAT3 PSMD14 PTTG1IP THRAP3 COP1 RALB JMJD8 SLC25A5 TENT5C CIDEA OPHN1
blue	2.096E-10	38	784	4.119211	RNA processing	SRRT SYF2 ERCC2 CELF1 HNRNPUL1 CCNL1 PAPOLA TARBP1 TRUB2 EXOSC9 CLK1 FUS CCNT1 RTCA PRKACA CPSF6 RPUSD3 RBM8A

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
						SLU7 HNRNPH1 PUM2 XRN2 SRSF9 AHNAK WBP11 IK PCF11 PRPF18 APP WDR43 NOL8 AHCYL1 NSRP1 STAT3 THRAP3 SNRPC C1D
blue	9.018E-12	48	1070	3.812449	Nuclear protein-containing complex	ANAPC16 UPF2 GTF2B SNRNP48 SYF2 MED13L ERCC2 ATP5F1A UBA2 COPS5 ASH1L JUN EXOSC9 NUP133 CHTOP COPS8 CCNT1 RPA2 CPSF6 POLA1 RBM8A SLU7 HNRNPH1 SEH1L HCFC1 JARID2 UCHL5 PBX1 BPTF SMARCC2 WBP11 IK PCF11 WDR61 CBFB EP300 PRPF18 KMT2A ASXL2 YY1 STAT3 THRAP3 SNRPC ACTB YAP1 WDR5 SMAD5
blue	1.99E-10	46	1113	3.512442	Intracellular transport	CRYAB ATP5F1D KIF1C CDC42 ATP5F1A AP4M1 DNAJC13 MSN SCP2 ASPSCR1 NUP133 CHTOP NUTF2 PRKACA PABPN1 TLK1 CPSF6 HOOK3 FIS1 LMAN2 SLU7 AP3D1 GRTP1 KPNA1 HSPB1 GDI1 CHMP5 RANGAP1 XPO1 CHMP4B CD74 APP WDR91 SSR3 AHCYL1 RAB21 USP6NL C19H17orf75 ATP5PD STAT3 PTTG1IP NPC2 MKKS RAN
blue	1.295E-09	44	1097	3.40873	RNA binding	UPF2 CSDE1 CELF1 HNRNPUL1 MSN PAPOLA TARBP1 TRUB2 EIF3F MRPL20 EXOSC9 CHTOP FUS CCNT1 EIF2S2 RPS9 EIF3E MRPL13 PABPN1 CPSF6 NUDT4 RPUSD3 PAIP2 RBM8A SLU7 HNRNPH1 PUM2 XRN2 TUBA1B RANGAP1 IARS2 SRSF9 EIF3H PCF11 EP300

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
						NCBP3 ACTN1 NOL8 NSRP1 EIF4G2 YY1 STAT3 SNRPC C1D
blue	1.165E-07	36	917	3.336412	Supramolecular complex	ANAPC16 UPF2 CRYAB KIF1C CDC42 CELF1 TCP1 TPM4 NUP133 CHTOP HOOK3 RPUSD3 KIF23 COL5A3 SEH1L PUM2 HSPB1 TUBA1B RANGAP1 KRT79 AJUBA DEK XPO1 AHNAK SQSTM1 KRT75 CAPRIN1 ACTN1 CCT2 VCL CENPF ACTB KRT89 KRTAP13-1
blue	2.045E-07	37	991	3.173033	Protein transport	BET1L TCP1 AP4M1 ASPSCR1 NUP133 NUTF2 PRKACA TLK1 HMGCR FIS1 SLU7 AP3D1 GRTP1 SEH1L KPNA1 VMP1 HSPB1 GDI1 RANGAP1 XPO1 NDFIP2 CD74 RRBP1 OPTN ACSL3 KTN1 WWP2 SSR3 RAB4A AHCYL1 RAB21 USP6NL MPC2 C19H17orf75 STAT3 PTTG1IP CENPF
blue	6.804E-08	40	1080	3.147624	Establishment of protein localization	BET1L TCP1 AP4M1 ASPSCR1 NUP133 NUTF2 LAMP3 PRKACA TLK1 HMGCR FIS1 SLU7 AP3D1 GRTP1 SEH1L KPNA1 VMP1 HSPB1 GDI1 HK1 RANGAP1 XPO1 NDFIP2 CD74 RRBP1 OPTN ACSL3 KTN1 WWP2 SSR3 RAB4A AHCYL1 CCT2 RAB21 USP6NL MPC2 C19H17orf75 STAT3 PTTG1IP CENPF
blue	4.763E-08	45	1297	2.948622	Cellular protein localization	CDC42 TCP1 AP4M1 MSN ASPSCR1 NUP133 NUTF2 LAMP3 PRKACA TLK1 HOOK3 WWTR1 FIS1 SLU7 AP3D1 GRTP1 SEH1L KPNA1 MPP5 HSPB1 GDI1 HK1 RANGAP1 AJUBA XPO1

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
						CTCF IK CD74 SQSTM1 OPTN ACSL3 SUMO3 SSR3 NOL8 MTCH2 AHCYL1 CCT2 RAB21 USP6NL C19H17orf75 STAT3 PTTG1IP VCL ACTB YAP1
blue	1.414E-09	53	1531	2.942031	Establishment of localization in cell	CRYAB ATP5F1D KIF1C CDC42 ATP5F1A AP4M1 DNAJC13 MSN SCP2 ASPSCR1 NUP133 CHTOP NUTF2 PRKACA PABPN1 TLK1 CPSF6 HOOK3 FIS1 LMAN2 FBXL20 SLU7 AP3D1 GRTP1 SEH1L KPNA1 HSPB1 GDI1 CHMP5 RANGAP1 XPO1 CHMP4B CD74 DNAJC5 APP WDR91 SSR3 AHCYL1 RAB21 USP6NL MPC2 C19H17orf75 ATP5PD STAT3 PTTG1IP NPC2 CENPF ACTB MKKS RAN OPHN1
blue	5.065E-08	45	1303	2.935044	Cellular macromolecule localization	CDC42 TCP1 AP4M1 MSN ASPSCR1 NUP133 NUTF2 LAMP3 PRKACA TLK1 HOOK3 WWTR1 FIS1 SLU7 AP3D1 GRTP1 SEH1L KPNA1 MPP5 HSPB1 GDI1 HK1 RANGAP1 AJUBA XPO1 CTCF IK CD74 SQSTM1 OPTN ACSL3 SUMO3 SSR3 NOL8 MTCH2 AHCYL1 CCT2 RAB21 USP6NL C19H17orf75 STAT3 PTTG1IP VCL ACTB YAP1
blue	3.919E-11	62	1800	2.92729	Reg. of cellular protein metabolic proc.	LTN1 CRYAB SH3RF1 THBS1 UBXN2A PPP4R2 UBA2 MSN CCNL1 JUN ASPSCR1 EXOSC9 CHTOP CCNT1 PRKACA EIF3E MRPL13 ARRDC3 WWTR1 HMGCR FIS1 KIDINS220 RPUSD3 PAIP2 ARPP19 HSPB1

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
						CSNK2A1 AJUBA JARID2 MAP4K4 XPO1 UCHL5 CTCF EIF3H NDFIP2 RAD23A CD74 SQSTM1 WDR61 EP300 RPS6KA3 APP WDR91 KMT2A WAC SDR16C5 PTPN3 MYCBP2 VPS25 PPP2R5C EIF4G2 STAT3 PSMD14 PTTG1IP CACUL1 STK3 COP1 RALB CAB39 WDR5 PAK2 OPHN1
blue	9.018E-12	66	1925	2.9138	Reg. of protein metabolic proc.	LTN1 CRYAB SH3RF1 THBS1 UBXN2A PPP4R2 UBA2 ITM2B MSN CCNL1 JUN ASPSCR1 EXOSC9 CHTOP CCNT1 PRKACA EIF3E MRPL13 ARRDC3 WWTR1 HMGCR FIS1 KIDINS220 RPUSD3 PAIP2 FBXL20 ARPP19 HSPB1 CSNK2A1 AJUBA JARID2 MAP4K4 XPO1 UCHL5 CTCF EIF3H NDFIP2 RAD23A CD74 SQSTM1 WDR61 EP300 RPS6KA3 APP WDR91 EGLN2 KMT2A WAC SDR16C5 PTPN3 MYCBP2 VPS25 PPP2R5C EIF4G2 STAT3 PSMD14 PTTG1IP CACUL1 STK3 COP1 RALB CAB39 WDR5 PAK2 PFDN1 OPHN1
blue	1.292E-07	45	1354	2.824492	Nitrogen compound transport	BET1L TCP1 AP4M1 ASPSCR1 NUP133 CHTOP NUTF2 PRKACA PABPN1 TLK1 CPSF6 HMGCR FIS1 RBM8A SLU7 AP3D1 GRTP1 SEH1L KPNA1 VMP1 HSPB1 GDI1 RANGAP1 XPO1 SLC1A5 NDFIP2 CD74 RRBP1 OPTN ACSL3 KTN1 WWP2 SSR3 RAB4A AHCYL1 RAB21

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
						USP6NL MPC2 C19H17orf75 STAT3 PTTG1IP CENPF ACTB SLC25A5
blue	1.598E-09	57	1743	2.779227	Protein localization	CDC42 BET1L TCP1 AP4M1 MSN ASPSCR1 NUP133 NUTF2 LAMP3 PRKACA TLK1 HOOK3 WWTR1 HMGCR FIS1 SLU7 AP3D1 GRTP1 SEH1L KPNA1 VMP1 MPP5 HSPB1 GDI1 HK1 RANGAP1 AJUBA XPO1 CTCF NDFIP2 IK CD74 RRBP1 SQSTM1 OPTN ACSL3 SUMO3 KTN1 WWP2 SSR3 NOL8 MTCH2 MYCBP2 RAB4A AHCYL1 CCT2 RAB21 USP6NL MPC2 C19H17orf75 STAT3 PTTG1IP VCL CENPF ACTB YAP1 PAK2
blue	6.804E-08	57	1946	2.489308	Reg. of localization	CRYAB SLK CDC42 THBS1 ATP5F1A TCP1 DNAJC13 MSN SCP2 JUN ASPSCR1 SEMA5A PRKACA CPSF6 WWTR1 HMGCR FIS1 LMAN2 FBXL20 AP3D1 LAMA2 VMP1 HSPB1 GDI1 RANGAP1 AJUBA MAP4K4 XPO1 AHNAK SARAF NDFIP2 NISCH CD74 DOCK5 SQSTM1 DNAJC5 ACSL3 SUMO3 KLHL24 APP ACTN1 WWP2 PTPN3 MYCBP2 RAB4A AHCYL1 CCT2 MPC2 STAT3 VCL PKN2 ACTB MKKS YAP1 SLC25A5 CIDEA OPHN1
blue	3.213E-07	54	1883	2.437193	Catabolic proc.	LTN1 UPF2 CSDE1 UBXN2A MSN EXOSC9 FUS PRKACA EIF3E NUDT4 WWTR1 HMGCR FIS1 RBM8A FBXL20 CRTC3 PUM2 XRN2 VMP1 LYPLA2 HK1 ATG2B XPO1 UCHL5

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
						EIF3H PCNP RAD23A SQSTM1 EP300 OPTN ECHS1 APP WDR91 EGLN2 WAC WWP2 PTPN3 MYCBP2 PPP2R5C UBR3 EIF4G2 STAT3 PSMD14 UFC1 PTTG1IP CACUL1 THRAP3 COP1 RALB JMJD8 SLC25A5 TENT5C CIDEA OPHN1
brown	0.0001211	20	576	4.150066	Reg. of protein localization	SWAP70 SETD2 TXN CCT4 AFDN PTPN1 PCM1 LRIG2 P2RX7 SRI ATP2C1 USP8 WAPL CCDC88A DVL3 DLG1 TFAP2B PARK7 SELENOK NDFIP1
brown	0.0001561	27	1080	2.988048	Establishment of protein localization	RAB10 TXN EMC3 UFD1 CCT4 MFN2 GDI2 PTPN1 PCM1 P2RX7 SRI ATP2C1 CLSTN1 DNM2 ARFGEF1 TRAK1 SAR1A TSPO TFAP2B PARK7 ABCA1 SELENOK CD24 RABGAP1 SEC62 NDFIP1
brown	0.0001999	28	1171	2.857911	Reg. of protein modification proc.	ITCH RB1CC1 PROM2 SWAP70 TXN AFAP1L2 PTPN1 AKAP9 TADA3 ROCK1 PRKAR1A P2RX7 IVNS1ABP ARFGEF1 MOB1B CCDC88A TSPO DVL3 PPP2R5E DLG1 PARK7 TRIB1 MNAT1 MAP2K1 CD24 NDFIP1 PRNP PMEPA1
brown	1.548E-05	38	1600	2.838645	Nervous system development	NDRG1 PRDM1 SETD2 MACO1 MAFB ADGRL1 ATRN ARL3 TOP2B TRIO CRABP2 PCM1 ROCK1 LRIG2 DDX6 HIPK1 EIF2B5 CLSTN1 FA2H JAG1 SMARCE1 DHX30 NDE1 TRAK1 BSG THRB SUZ12 ACSL4 DVL3

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
						RBFOX2 DLG1 TFAP2B HES2 CHD7 ALDH5A1 MYCN MNAT1 MAP2K1
brown	0.0001548	30	1272	2.818913	Protein-containing complex assembly	RPL5 SRPK2 COX17 SMARCA5 OTUD6B ARL3 GRB2 UBTF RBM5 MITF AKAP9 GTF2H5 P2RX7 DHX29 G3BP2 ARFGEF1 DHX30 NDE1 FCHO2 NDUFS7 DLG1 PARK7 ABCA1 ADD1 TP53BP1 NAP1L4 FMN1 PRNP NA PMEPA1
brown	9.447E-06	41	1743	2.811474	Protein localization	RAB10 SWAP70 INSIG2 SETD2 TXN EMC3 UFD1 ARL3 CCT4 MFN2 GDI2 AFDN PTPN1 PCM1 ROCK1 LRIG2 P2RX7 SRI ATP2C1 CLSTN1 USP8 WAPL DNM2 ATP1B3 ARFGEF1 TRAK1 TM9SF3 SAR1A CCDC88A TSPO DVL3 DLG1 TFAP2B PARK7 ABCA1 SELENOK CD24 RABGAP1 SEC62 NDFIP1
brown	0.0001211	33	1424	2.76982	Protein-containing complex subunit organization	SWAP70 RPL5 SRPK2 COX17 SETD2 SMARCA5 OTUD6B ARL3 GRB2 UBTF RBM5 MITF AKAP9 GTF2H5 P2RX7 DHX29 G3BP2 ARFGEF1 SMARCE1 DHX30 NDE1 FCHO2 NDUFS7 DLG1 PARK7 ABCA1 ADD1 TP53BP1 NAP1L4 FMN1 PRNP NA PMEPA1
brown	0.0003777	29	1297	2.672425	Cellular protein localization	RAB10 INSIG2 SETD2 TXN EMC3 UFD1 ARL3 CCT4 MFN2 PTPN1 PCM1 ROCK1 LRIG2 ATP2C1 CLSTN1 WAPL ATP1B3 TRAK1 TM9SF3 SAR1A CCDC88A TSPO DVL3 DLG1 PARK7 CD24 RABGAP1 SEC62

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
brown	7.555E-05	39	1800	2.589641	Reg. of cellular protein metabolic proc.	ITCH RB1CC1 PROM2 SWAP70 TXN AFAP1L2 PRELID1 ARIH2 PTPN1 AKAP9 TADA3 ROCK1 LRIG2 PRKAR1A P2RX7 DDX6 DHX29 EIF3K IVNS1ABP NGRN ARFGEF1 TCOF1 MOB1B CCDC88A TSPO TNRC6C DPH5 DVL3 PPP2R5E DLG1 PARK7 TRIB1 MNAT1 MAP2K1 DPH7 CD24 NDFIP1 PRNP PMEPA1
brown	0.0001548	35	1631	2.564848	Reg. of cellular component organization	SUPV3L1 PROM2 SWAP70 ADGRL1 MPHOSPH8 ARL3 CCT4 GRB2 PRELID1 CRABP2 PTPN1 AKAP9 NOTCH2 ROCK1 LRIG2 P2RX7 RAE1 CLSTN1 WAPL DNM2 OSGIN2 SYNPO STMN1 G3BP2 ARFGEF1 CCDC88A TSPO DLG1 PARK7 ADD1 RNF5 FMN1 MNAT1 MAP2K1 PMEPA1
brown	0.0001548	35	1651	2.533778	Purine ribonucleoside triphosphate binding	ATP11B RAB10 SWAP70 SRPK2 SMARCA5 DDX24 GTPBP3 ARL3 CCT4 TOP2B MFN2 TRIO MYO9A ROCK1 P2RX7 DDX6 DHX29 SRXN1 HIPK1 ATP2C1 DNM2 SMC3 CCT8 EIF2S3 HIPK3 DHX30 HSPA14 SAR1A RARS2 CSNK1G3 ABCA1 CHD7 TRIB1 MAP2K1 IRGQ
brown	0.0001548	36	1702	2.528078	Purine ribonucleotide binding	ATP11B RAB10 SWAP70 SRPK2 SMARCA5 DDX24 GTPBP3 ARL3 CCT4 TOP2B MFN2 TRIO MYO9A ROCK1 PRKAR1A P2RX7 DDX6 DHX29 SRXN1 HIPK1 ATP2C1 DNM2 SMC3 CCT8 EIF2S3 HIPK3 DHX30 HSPA14 SAR1A

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
						RARS2 CSNK1G3 ABCA1 CHD7 TRIB1 MAP2K1 IRGQ
brown	0.0001548	36	1716	2.507453	Ribonucleotide binding	ATP11B RAB10 SWAP70 SRPK2 SMARCA5 DDX24 GTPBP3 ARL3 CCT4 TOP2B MFN2 TRIO MYO9A ROCK1 PRKAR1A P2RX7 DDX6 DHX29 SRXN1 HIPK1 ATP2C1 DNM2 SMC3 CCT8 EIF2S3 HIPK3 DHX30 HSPA14 SAR1A RARS2 CSNK1G3 ABCA1 CHD7 TRIB1 MAP2K1 IRGQ
brown	0.0001548	36	1722	2.498716	Purine nucleotide binding	ATP11B RAB10 SWAP70 SRPK2 SMARCA5 DDX24 GTPBP3 ARL3 CCT4 TOP2B MFN2 TRIO MYO9A ROCK1 PRKAR1A P2RX7 DDX6 DHX29 SRXN1 HIPK1 ATP2C1 DNM2 SMC3 CCT8 EIF2S3 HIPK3 DHX30 HSPA14 SAR1A RARS2 CSNK1G3 ABCA1 CHD7 TRIB1 MAP2K1 IRGQ
brown	0.000174	36	1746	2.464369	Cytoskeleton	NDRG1 SWAP70 ARL3 CCT4 ABLIM1 MFN2 HAUS2 PCM1 MYO9A AKAP9 TADA3 PROCR ROCK1 PRKAR1A RBBP6 RAE1 HIPK1 PLEC WAPL DNM2 SYNPO SMC3 YEATS2 NGRN HMMR MAP4 NDE1 KRT32 CCDC88A FCHO2 DLG1 ADD1 TBL1XR1 DYNLL2 FMN1 TMEM63B
brown	0.0001548	38	1866	2.433994	Reg. of multicellular organismal proc.	ITCH PRDM1 ANKRD42 SETD2 MAFB ATRN UFD1 AFAP1L2 PRELID1 MFN2 CRABP2 MITF PCM1 AKAP9 NOTCH2 PROCR ROCK1 P2RX7

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
						SRI CLSTN1 FA2H JAG1 THRB RBFOX2 DLG1 RIN2 PARK7 ADD1 HES2 TP53BP1 FH CHD7 TRIB1 MYCN SELENOK MAP2K1 CD24 NDFIP1
brown	0.0002515	38	1948	2.331536	Neg. reg. of cellular metabolic proc.	PRDM1 SWAP70 INSIG2 TXN MPHOSPH8 PTPN1 MITF NOTCH2 ROCK1 LRIG2 PRKAR1A DDX6 WAPL IVNS1ABP YEATS2 SMARCE1 RRP1B TRPS1 THRB TSPO TNRC6C SUZ12 RBFOX2 GON4L DLG1 MARF1 TFAP2B PARK7 TP53BP1 TBL1XR1 TRIB1 MYCN RNF5 CREBBP NA LCOR NDFIP1 PMEPA1
brown	0.0002515	38	1949	2.33034	Nucleotide binding	ATP11B RAB10 SWAP70 SRPK2 SMARCA5 DDX24 GTPBP3 ARL3 CCT4 TOP2B MFN2 TRIO MYO9A ROCK1 PRKAR1A P2RX7 DDX6 DHX29 SRXN1 HIPK1 ATP2C1 DNM2 SMC3 CCT8 EIF2S3 HIPK3 DHX30 HSPA14 SAR1A RARS2 CSNK1G3 ACSL4 ABCA1 CHD7 TRIB1 IMPDH2 MAP2K1 IRGQ
brown	0.0002515	38	1949	2.33034	Nucleoside phosphate binding	ATP11B RAB10 SWAP70 SRPK2 SMARCA5 DDX24 GTPBP3 ARL3 CCT4 TOP2B MFN2 TRIO MYO9A ROCK1 PRKAR1A P2RX7 DDX6 DHX29 SRXN1 HIPK1 ATP2C1 DNM2 SMC3 CCT8 EIF2S3 HIPK3 DHX30 HSPA14 SAR1A RARS2 CSNK1G3 ACSL4 ABCA1 CHD7 TRIB1 IMPDH2 MAP2K1 IRGQ
greenyellow	1.656E-06	5	17	68.93382	Complex of collagen trimers	COL1A1 COL1A2 COL3A1 COL5A2 COL5A1

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
greenyellow	7.278E-08	7	34	48.25368	ECM proteoglycans	COL6A1 COL1A1 COL1A2 SPARC COL5A2 COL6A3 COL5A1
greenyellow	9.497E-07	6	31	45.3629	Collagen chain trimerization	COL6A1 COL1A1 COL1A2 COL5A2 COL6A3 COL5A1
greenyellow	1.229E-06	6	33	42.61364	Collagen degradation	COL6A1 COL1A1 COL1A2 COL5A2 COL6A3 COL5A1
greenyellow	2.058E-07	7	40	41.01563	Assembly of collagen fibrils and other multimeric structures	COL6A1 COL1A1 COL1A2 CTSS COL5A2 COL6A3 COL5A1
greenyellow	3.283E-06	6	40	35.15625	Integrin cell surface interactions	COL6A1 COL1A1 COL1A2 COL5A2 COL6A3 COL5A1
greenyellow	4.668E-06	6	43	32.70349	Collagen trimer	COL1A1 COL1A2 COL3A1 COL5A2 COL6A3 COL5A1
greenyellow	1.037E-06	7	56	29.29688	Collagen biosynthesis and modifying enzymes	P4HA2 COL6A1 COL1A1 COL1A2 COL5A2 COL6A3 COL5A1
greenyellow	2.552E-07	8	71	26.40845	Collagen formation	P4HA2 COL6A1 COL1A1 COL1A2 CTSS COL5A2 COL6A3 COL5A1
greenyellow	6.258E-14	16	165	22.72727	Collagen-containing extracellular matrix	LTBP3 FN1 CCDC80 LAMB1 DAG1 COL6A1 COL1A1 COL1A2 SPARC FMOD CASK COL3A1 COL5A2 COL6A3 COL5A1 COL6A5
greenyellow	3.283E-06	8	105	17.85714	Amoebiasis	FN1 LAMB1 PRKACB COL1A1 COL1A2 RELA RAB5B COL3A1

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
greenyellow	5.377E-08	12	199	14.13317	Extracellular matrix organization	P4HA2 LTBP3 LAMB1 COL6A1 COL1A1 COL1A2 SPARC CAPNS1 CTSS COL5A2 COL6A3 COL5A1
greenyellow	3.898E-10	16	307	12.21498	Extracellular matrix	LTBP3 FN1 CCDC80 LAMB1 DAG1 COL6A1 COL1A1 COL1A2 SPARC FMOD CASK COL3A1 COL5A2 COL6A3 COL5A1 COL6A5
greenyellow	3.898E-10	16	308	12.17532	External encapsulating structure	LTBP3 FN1 CCDC80 LAMB1 DAG1 COL6A1 COL1A1 COL1A2 SPARC FMOD CASK COL3A1 COL5A2 COL6A3 COL5A1 COL6A5
greenyellow	2.173E-07	12	239	11.76778	Extracellular matrix organization	LTBP3 CCDC80 LAMB1 DAG1 COL1A1 COL1A2 FMOD PHLDB1 CTSS COL3A1 COL5A2 COL5A1
greenyellow	2.173E-07	12	239	11.76778	External encapsulating structure organization	LTBP3 CCDC80 LAMB1 DAG1 COL1A1 COL1A2 FMOD PHLDB1 CTSS COL3A1 COL5A2 COL5A1
greenyellow	2.173E-07	12	241	11.67012	Extracellular structure organization	LTBP3 CCDC80 LAMB1 DAG1 COL1A1 COL1A2 FMOD PHLDB1 CTSS COL3A1 COL5A2 COL5A1
greenyellow	8.842E-07	24	1362	4.129956	Tissue development	PI16 ROCK2 GPI LTBP3 FN1 GRSF1 MEF2A LAMB1 DAG1 PRKACB XDH COL1A1 COL1A2 RELA BMI1 PHLDB1 IL6ST PTPRS VIM TXNIP COL3A1 COL5A2 COL5A1 GPX1
greenyellow	6.043E-07	29	1925	3.530844	Reg. of protein metabolic proc.	CST3 GPNMB CNOT2 ROCK2 DAP ERP29 FN1 CTBP1 PURA PRKACB XDH ENO1 RELA WWP1 FXR1 CTDSP2 EEF2K IL6ST VIM

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
						PSENEN SETD5 HBS1L UFL1 NRDC ASPH COL6A3 STYX CNOT6L GPX1
greenyellow	1.734E-06	27	1800	3.515625	Reg. of cellular protein metabolic proc.	CST3 GPNMB CNOT2 ROCK2 DAP ERP29 FN1 CTBP1 PURA XDH ENO1 RELA FXR1 CTDSP2 EEF2K IL6ST VIM PSENEN SETD5 HBS1L UFL1 NRDC ASPH COL6A3 STYX CNOT6L GPX1
magenta	1.046E-06	24	1131	4.301384	Neg. reg. of macromolecule biosynthetic proc.	MTDH RREB1 KDM5B CBX5 SMAD4 NFIC PFDN5 ZBTB16 NCOR1 CHD3 CALR CC2D1B CLTC KLF3 TNRC6A CNOT7 VGLL4 SPEN PUM1 MEF2C RC3H2 MAF KCTD1 ZNF148
magenta	1.633E-05	20	965	4.201092	Neg. reg. of transcription, DNA-templated	MTDH RREB1 KDM5B CBX5 SMAD4 NFIC PFDN5 ZBTB16 NCOR1 CHD3 CALR CC2D1B KLF3 CNOT7 VGLL4 SPEN MEF2C MAF KCTD1 ZNF148
magenta	1.633E-05	20	966	4.196743	Neg. reg. of RNA biosynthetic proc.	MTDH RREB1 KDM5B CBX5 SMAD4 NFIC PFDN5 ZBTB16 NCOR1 CHD3 CALR CC2D1B KLF3 CNOT7 VGLL4 SPEN MEF2C MAF KCTD1 ZNF148
magenta	1.633E-05	20	966	4.196743	Neg. reg. of nucleic acid-templated transcription	MTDH RREB1 KDM5B CBX5 SMAD4 NFIC PFDN5 ZBTB16 NCOR1 CHD3 CALR CC2D1B KLF3 CNOT7 VGLL4 SPEN MEF2C MAF KCTD1 ZNF148
magenta	1.654E-06	24	1167	4.168693	Neg. reg. of cellular biosynthetic proc.	ACADVL MTDH RREB1 KDM5B CBX5 SMAD4 NFIC PFDN5 ZBTB16 NCOR1 CHD3 CALR

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
						CC2D1B KLF3 TNRC6A CNOT7 VGLL4 SPEN PUM1 MEF2C RC3H2 MAF KCTD1 ZNF148
magenta	0.0006611	15	730	4.165124	Nucleolus	SLBP ACADVL MTDH SBDS RREB1 CBX5 TTC3 NFIC RNF20 ARL14EP MYSM1 CHD3 ZZZ3 RPP14 WRN
magenta	8.444E-07	27	1386	3.948754	Pos. reg. of RNA metabolic proc.	PRKAA1 NSD3 PICALM LUM MTDH PCGF5 RREB1 SMAD4 NFIC RNF20 PIK3R1 ZBTB16 MYSM1 TMF1 TNRC6A CNOT7 PUM1 LYL1 BARX2 MEF2C RTRAF RC3H2 FHOD1 ARMCX3 MAF ZNF148
magenta	3.882E-05	21	1128	3.773721	Neg. reg. of nucleobase-containing compound metabolic proc.	MTDH RREB1 KDM5B CBX5 SMAD4 NFIC PFDN5 RNF20 ZBTB16 NCOR1 CHD3 CALR CC2D1B KLF3 CNOT7 VGLL4 SPEN MEF2C MAF KCTD1 ZNF148
magenta	7.42E-06	24	1291	3.768292	Pos. reg. of transcription, DNA-templated	PRKAA1 NSD3 PICALM LUM MTDH PCGF5 RREB1 SMAD4 NFIC RNF20 PIK3R1 ZBTB16 MYSM1 TMF1 CNOT7 LYL1 BARX2 MEF2C RTRAF FHOD1 ARMCX3 MAF ZNF148
magenta	7.42E-06	24	1291	3.768292	Pos. reg. of nucleic acid-templated transcription	PRKAA1 NSD3 PICALM LUM MTDH PCGF5 RREB1 SMAD4 NFIC RNF20 PIK3R1 ZBTB16 MYSM1 TMF1 CNOT7 LYL1 BARX2 MEF2C RTRAF FHOD1 ARMCX3 MAF ZNF148
magenta	7.42E-06	24	1292	3.765375	Pos. reg. of RNA biosynthetic proc.	PRKAA1 NSD3 PICALM LUM MTDH PCGF5 RREB1 SMAD4 NFIC RNF20 PIK3R1 ZBTB16

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
						MYSM1 TMF1 CNOT7 LYL1 BARX2 MEF2C RTRAF FHOD1 ARMCX3 MAF ZNF148
magenta	3.467E-07	32	1798	3.607612	Neg. reg. of nitrogen compound metabolic proc.	PRKAA1 PICALM MTDH RREB1 KDM5B CBX5 SMAD4 NFIC PFDN5 RNF20 ZBTB16 CD44 CD109 NCOR1 CHD3 CALR CC2D1B CLTC KLF3 TNRC6A CNOT7 VGLL4 SPEN SNX3 PUM1 MEF2C RTRAF RC3H2 MAF KCTD1 ZNF148 CTDSPL
magenta	2.775E-07	34	1948	3.537932	Neg. reg. of cellular metabolic proc.	PRKAA1 PICALM ACADVL MTDH RREB1 KDM5B CBX5 SMAD4 NFIC PFDN5 RNF20 ZFYVE1 ZBTB16 CD44 CD109 NCOR1 ANKRD13A CHD3 CALR CC2D1B KLF3 TNRC6A CNOT7 VGLL4 SPEN PUM1 MEF2C RTRAF RC3H2 CHMP4A MAF KCTD1 ZNF148 CTDSPL
pink	0.0284563	3	14	47.26891	Proteasome assembly	PSMD10 PSMD11 PSMG4
pink	0.0284563	3	15	44.11765	Pos. reg. by host of viral transcription	SP1 SNW1 ZNF639
pink	0.0420096	4	54	16.33987	WP1079 Proteasome Degradation	PSMD10 UBB PSMD11 PSMB4
pink	0.0420096	4	55	16.04278	Viral gene expression	EIF3G SP1 SNW1 ZNF639
pink	0.0420096	6	161	8.22068	MRNA Splicing-Major Pathway	SF3B2 PPIL3 HNRNPA2B1 SNW1 SF3B1 CWC25

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
pink	0.0420096	6	167	7.925326	MRNA Splicing	SF3B2 PPIL3 HNRNPA2B1 SNW1 SF3B1 CWC25
pink	0.0438379	6	179	7.394019	Cilium Assembly	CEP89 NPHP3 HAUS5 ARF4 HAUS6 PCNT
pink	0.0163748	9	306	6.487889	Nuclear speck	SF3B2 SRSF10 SNW1 SF3B1 PACSIN2 HAUS6 CWC25 RAF1 NRIP1
pink	0.0438379	7	265	5.826859	RNA splicing, via transesterification reactions	SFSWAP SF3B2 HNRNPA2B1 SRSF10 SNW1 SF3B1 CWC25
pink	0.0438379	7	265	5.826859	RNA splicing, via transesterification reactions with bulged adenosine as nucleop	SFSWAP SF3B2 HNRNPA2B1 SRSF10 SNW1 SF3B1 CWC25
pink	0.0438379	7	265	5.826859	MRNA splicing, via spliceosome	SFSWAP SF3B2 HNRNPA2B1 SRSF10 SNW1 SF3B1 CWC25
pink	0.0438379	11	647	3.750341	Cellular protein catabolic proc.	LAMP2 CUL5 PSMD10 TLK2 UBB PSMD11 NUDT15 PSMB4 GET4 SMARCC1 FBXL14
pink	0.0006888	21	1256	3.688179	Catalytic complex	MORF4L2 KANSL2 SF3B2 DCAF8 DCAF1 SNW1 SF3B1 WDR74 REV3L CFLAR CUL5 POLR3A PSMD10 ING4 PSMD11 PDHA1 PSMB4 NDUFB2 CWC25 SMARCC1 NRIP1
pink	0.0467287	11	669	3.627011	Intracellular protein-containing complex	MORF4L2 KANSL2 DCAF8 DCAF1 REV3L CUL5 POLR3A PSMD10 ING4 PSMD11 PSMB4

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
pink	0.0438379	12	768	3.446691	Protein catabolic proc.	LAMP2 EGFR CUL5 PSMD10 TLK2 UBB PSMD11 NUDT15 PSMB4 GET4 SMARCC1 FBXL14
turqiouse	3.178E-21	24	62	17.75673	WP1071 Cytoplasmic Ribosomal Proteins	RPS17 RPS12 RPL14 RPS18 RPL35 RPL4 RPL27A RPS14 RPS2 RPSA RPS19 RPL35A RPL34 RPL22 RPS8 RPL18 RPS27A RPS5 RPS7 RPL30 RPLP0 RPL11 RPS29
turqiouse	1.231E-19	28	109	11.78352	Nonsense Mediated Decay NMD independent of the Exon Junction Complex EJC	RPS12 RPL21 RPL14 RPS18 RPL35 RPL4 RPL13A RPL27A RPL10 RPS2 RPSA RPS19 EIF4G1 RPL35A RPL34 RPL22 RPS8 RPL18 RPS27A RPS5 RPS7 RPL30 RPL36A RPL11 RPS16 RPS6 RPS4Y1 PABPC1
turqiouse	4.648E-18	26	103	11.57923	GTP hydrolysis and joining of the 60S ribosomal subunit	RPS12 RPL21 RPL14 RPS18 RPL35 RPL4 RPL13A RPL27A RPL10 RPS2 RPSA RPS19 RPL35A RPL34 RPL22 RPS8 RPL18 RPS27A RPS5 RPS7 RPL30 RPL36A RPL11 RPS16 RPS6 RPS4Y1
turqiouse	2.593E-20	30	123	11.18819	L13a-mediated translational silencing of Ceruloplasmin expression	RPS12 RPL21 RPL14 EIF3A RPS18 RPL35 RPL4 RPL13A RPL27A RPL10 EIF4E RPS2 RPSA RPS19 EIF4G1 RPL35A RPL34 RPL22 RPS8 RPL18 RPS27A RPS5 RPS7 RPL30 RPL36A RPL11 RPS16 RPS6 RPS4Y1 PABPC1
turqiouse	3.657E-20	30	125	11.00917	Eukaryotic Translation Initiation	RPS12 RPL21 RPL14 EIF3A RPS18 RPL35 RPL4 RPL13A RPL27A RPL10 EIF4E RPS2 RPSA RPS19 EIF4G1 RPL35A RPL34 RPL22 RPS8

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
						RPL18 RPS27A RPS5 RPS7 RPL30 RPL36A RPL11 RPS16 RPS6 RPS4Y1 PABPC1
turqiouse	3.657E-20	30	125	11.00917	Cap-dependent Translation Initiation	RPS12 RPL21 RPL14 EIF3A RPS18 RPL35 RPL4 RPL13A RPL27A RPL10 EIF4E RPS2 RPSA RPS19 EIF4G1 RPL35A RPL34 RPL22 RPS8 RPL18 RPS27A RPS5 RPS7 RPL30 RPL36A RPL11 RPS16 RPS6 RPS4Y1 PABPC1
turqiouse	2.2E-22	34	143	10.90652	Ribosome	RPS12 RPL21 RPL14 RPS18 RPL35 RPL4 RPL13A RPL27A RPL10 RPS14 RPS2 RPSA RPS19 RPL35A RPL34 RPL22 RPS8 MRPS10 RPL18 RPS27A RPS5 RPS7 RPL30 RPLP0 RPL36A RPL11 RPS16 RPS6 RPS4Y1 RPS26 RPL23A RPS29
turqiouse	4.944E-18	27	114	10.86432	Formation of a pool of free 40S subunits	RPS12 RPL21 RPL14 EIF3A RPS18 RPL35 RPL4 RPL13A RPL27A RPL10 RPS2 RPSA RPS19 RPL35A RPL34 RPL22 RPS8 RPL18 RPS27A RPS5 RPS7 RPL30 RPL36A RPL11 RPS16 RPS6 RPS4Y1
turqiouse	3.875E-19	29	125	10.6422	Nonsense-Mediated Decay NMD	RPS12 RPL21 RPL14 RPS18 RPL35 RPL4 RPL13A RPL27A RPL10 RPS2 RPSA RPS19 EIF4G1 RPL35A RPL34 RPL22 RPS8 RPL18 RPS27A RPS5 RPS7 RPL30 RPL36A UPF3A RPL11 RPS16 RPS6 RPS4Y1 PABPC1
turqiouse	4.001E-22	44	274	7.366236	Prion disease	UQCR10 PSMD13 KIF5B PSMA3 COX7A2 PSMC1 ATP5MC2 PSMA1 HSPA5 NDUFA13 UQCR11 ATF6B NDUFA5 UQCRH PSMA6

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
						NDUFS5 PSMB5 COX6A1 VDAC2 VDAC1 HSPA8 ATP5F1B PSMB6 PSMA4 NDUFA2 PSMD6 NDUFB1 COX5A ATP5PO MAPK14 UQCRC2 NDUFV1 NOTCH1 HSPA1A TUBA3E PSMC2 SLC25A6 TUBB2A COX8A RAC1 COX6B1
turqiouse	4.001E-22	44	275	7.33945	Parkinson disease	UQCR10 SLC39A6 TP53 PSMD13 KIF5B PSMA3 COX7A2 ATF6 PSMC1 ATP5MC2 PSMA1 HSPA5 NDUFA13 UQCR11 NDUFA5 UQCRH PSMA6 NDUFS5 PSMB5 COX6A1 VDAC2 UBE2L3 VDAC1 ATP5F1B PSMB6 PSMA4 NDUFA2 PSMD6 NDUFB1 RPS27A COX5A ATP5PO UQCRC2 NDUFV1 CALM1 TUBA3E PSMC2 SLC25A6 TUBB2A COX8A COX6B1
turqiouse	9.435E-25	53	363	6.6975	Amyotrophic lateral sclerosis	UQCR10 TP53 HNRNPA1 NOS1 PSMD13 KIF5B PSMA3 DCTN1 PFN1 COX7A2 ATF6 PSMC1 ATP5MC2 ACTG1 DCTN5 PSMA1 MATR3 HSPA5 NDUFA13 UQCR11 NDUFA5 UQCRH PSMA6 NDUFS5 DCTN2 WIPI2 PSMB5 COX6A1 VDAC1 ATP5F1B PSMB6 PSMA4 VCP NDUFA2 PSMD6 NDUFB1 COX5A ATP5PO ANXA7 MAPK14 HNRNPA3 UQCRC2 NDUFV1 ULK2 TUBA3E PSMC2 TUBB2A COX8A RAC1 COX6B1
turqiouse	2.593E-20	44	308	6.55308	Huntington disease	UQCR10 TP53 PSMD13 KIF5B PSMA3 DCTN1 COX7A2 PSMC1 ATP5MC2 DCTN5 PSMA1

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
						NDUFA13 UQCR11 NDUFA5 UQCRH PSMA6 NDUFS5 DCTN2 WIPI2 HDAC1 PSMB5 COX6A1 VDAC2 VDAC1 ATP5F1B PSMB6 PSMA4 NDUFA2 PSMD6 NDUFB1 COX5A ATP5PO UQCRC2 NDUFV1 ULK2 TUBA3E PSMC2 SLC25A6 TUBB2A COX8A COX6B1
turqiouse	4.884E-22	51	386	6.06075	Alzheimer disease	UQCR10 SLC39A6 ADAM17 NOS1 PSMD13 KIF5B PSMA3 COX7A2 ATF6 PSMC1 ATP5MC2 PSMA1 NDUFA13 UQCR11 NDUFA5 UQCRH PSMA6 NAE1 NDUFS5 RTN4 WIPI2 INSR PSMB5 COX6A1 VDAC2 VDAC1 ATP5F1B PSMB6 PSMA4 GAPDH NDUFA2 PSMD6 NDUFB1 COX5A ATP5PO RTN3 APH1A UQCRC2 NDUFV1 CALM1 ULK2 TUBA3E CSNK1B PSMC2 SLC25A6 TUBB2A COX8A COX6B1
turqiouse	2.593E-20	55	495	5.09684	Metabolism of RNA	CDC40 RPS12 NOP14 RPL21 TRA2B RPL14 GPKOW RPS18 RPL35 RPL4 RPL13A RPL27A RPL10 HNRNPU U2SURP EIF4E RPS2 RPSA SF3B3 SRSF5 RPS19 EIF4G1 HSPA8 SRRM1 HNRNPD RPL35A RPL34 RPL22 DDX21 RPS8 RPL18 RPS27A SRSF11 UTP6 RPS5 NCL RPS7 RPL30 YWHAB DDX5 RPL36A PCBP2 UPF3A RPL11 SET HNRNPM RPS16 HNRNPK HNRNPA3 RPS6 HSPA1A RPS4Y1 PTBP1 PABPC1 PRPF40A

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
turqiouse	3.488E-19	52	472	5.053646	Pathways of neurodegeneration	UQCR10 NOS1 PSMD13 KIF5B PSMA3 DCTN1 COX7A2 ATF6 PSMC1 ATP5MC2 DCTN5 PSMA1 HSPA5 NDUFA13 UQCR11 NDUFA5 UQCRH PSMA6 NDUFS5 DCTN2 WIPI2 PSMB5 COX6A1 VDAC2 UBE2L3 VDAC1 ATP5F1B PSMB6 PSMA4 VCP NDUFA2 PSMD6 NDUFB1 RPS27A COX5A ATP5PO MAPK14 UQCRC2 NDUFV1 CALM1 ULK2 TUBA3E CSNK1B PSMC2 SLC25A6 TUBB2A COX8A RAC1 COX6B1
turqiouse	9.435E-25	76	773	4.510011	Ribonucleoprotein complex	CDC40 RPS12 NOP14 HNRNPA1 RPL21 TRA2B RPL14 EIF3A GPKOW RPS18 RPL35 RPL4 RPL13A RPL27A SYNCRIP RPL10 HNRNPU GCFC2 RPS14 EIF4E RPS2 LARP7 PSMA6 RPSA DDX17 SF3B3 RPS19 EIF4G1 IQGAP1 HSPA8 HNRNPAB SRRM1 PRPF3 HNRNPD RPL35A GAPDH MRPS10 NPM1 RPL18 RPS27A UTP6 RPS5 NCL RPL30 RALY LUC7L PABPC4 YBX1 RPLP0 RBMS2 DDX5 RPL7L1 NOP56 RNPC3 RPL36A DAZAP1 DHX9 RPL11 HNRNPM RPS16 HNRNPK ESRP1 RPS6 RPS4Y1 PNN RPS26 UTP23 PKP1 PABPC1 RPL23A PKP3 PRPF40A RPS29
turqiouse	9.623E-33	103	1097	4.306992	RNA binding	KHDC4 STAU1 RBM4B LSM14A TP53 HNRNPA1 PKM TRA2B RPL14 EIF1 EIF3A GPKOW ZRANB2 RPS18 RPL35 HSP90B1

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
						RBM39 ZNF106 SETD1B AIMP1 RPL4 RPL13A DHX36 RBM6 SYNCRIP MATR3 HNRNPU SON SRSF6 RPS14 U2SURP EIF4E RPS2 ZC3H10 LARP7 PSMA6 DDX17 EIF4E2 EIF6 SRSF5 EEF1G RPS19 LARP1B ENDOU EIF4G1 SERBP1 HNRNPAB HNRNPD CCNT2 RPL35A RPL22 EEF1A1 DDX21 DUS2 HNRNPDL PNLDC1 NPM1 RPL18 SRSF11 YTHDC1 UTP6 RPS5 NCL RBM43 RPL30 RALY SEC63 LUC7L PABPC4 YBX1 RPLP0 RBM26 RBMS2 DHX40 DDX5 RPL7L1 NOP56 RNPC3 TUFM DAZAP1 DHX9 PCBP2 UPF3A RPL11 HNRNPM RPS16 HNRNPK NOP53 HNRNPA3 ESRP1 RPS4Y1 RBM3 RPS26 UTP23 EARS2 PTBP1 PABPC1 RPL23A SECISBP2L PRPF40A RPS29
turqiouse	3.168E-19	91	1348	3.096671	Metabolism of proteins	PARP1 TP53 RPS12 RPL21 RTF1 PDIA6 RPL14 PSMD13 EIF3A DYNC1I2 RPS18 PSMA3 RPL35 HSP90B1 DCTN1 RPL4 ATF6 DOHH RPL13A RPL27A PSMC1 PIGG P4HB DCTN5 PSMA1 TRAPPC10 ASB12 RPL10 UCHL3 GOLGB1 GNG10 STAG2 GAN EIF4E RPS2 PSMA6 USP2 RPSA DDX17 NAE1 DCTN2 RPS19 HDAC1 PSMB5 EIF4G1 VDAC2 UBE2L3 VDAC1 HSPA8 PSMB6 COPS2 HERC2 EXOC4 RPL35A RPL34 RPL22 PSMA4 EEF1A1 PRSS23 RPS8 MRPS10 NPM1 SPTAN1 RPL18 RPS27A RPS5 RPS7

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
						RPL30 COPG2 RXRA DDX5 ANK3 RPL36A FOXK1 NR3C1 UBE2K RPL11 RPS16 RPS6 SKP1 RPS4Y1 CALM1 TUBA3E MUC20 NEDD8 PABPC1 PSMC2 TUBB2A
yellow	2.91E-07	6	14	57.14286	Cornified envelope	CSTA DSP KRT10 KRT2 KRT1
yellow	2.44E-06	9	82	14.63415	Formation of the cornified envelope	KRT80 DSG1 DSP JUP SPINK5 PERP CDSN KRT2 KRT1
yellow	1.95E-06	11	137	10.7056	Epidermal cell differentiation	HES1 NCOA3 TMEM79 CSTA DSP TP63 SPINK5 KRT10 ZFP36L1 KRT2
yellow	2.80E-07	14	198	9.427609	Skin development	ASPRV1 NCOA3 TMEM79 CSTA DSP TP63 ITGA6 JUP KRT10 CDSN ZFP36L1 KRT2 KRT1
yellow	2.37E-06	16	347	6.147935	Developmental Biology	CXCL12 EPHA1 KRT80 EZR DSG1 DSP ENAH TCF4 JUP SPINK5 PERP CDSN TLN1 UBC KRT2 KRT1
yellow	4.72E-07	22	600	4.888889	Cell-cell adhesion	HES1 BCL6 CXCL12 PPARA NEXN EZR CSTA S100A8 DSG1 DSP SDC4 ITGB1 GATA3 JUP NECTIN4 DSC1 PERP ZBTB7B CDSN ZFP36L1 TLN1
yellow	8.49E-09	32	999	4.270938	Cell adhesion	HES1 BCL6 CXCL12 EPHA1 EMP2 PPARA NEXN RASA1 EZR CSTA FLNA POSTN S100A8 DSG1 PRKX DSP SDC4 ITGB1 RAB1A GATA3 ITGA6 JUP NECTIN4 LYPD3 SPINK5 DSC1 PERP ZBTB7B CDSN ZFP36L1 TLN1
yellow	8.49E-09	32	1004	4.249668	Biological adhesion	HES1 BCL6 CXCL12 EPHA1 EMP2 PPARA NEXN RASA1 EZR CSTA FLNA POSTN S100A8 DSG1 PRKX DSP SDC4 ITGB1 RAB1A GATA3

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
						ITGA6 JUP NECTIN4 LYPD3 SPINK5 DSC1 PERP ZBTB7B CDSN ZFP36L1 TLN1
yellow	5.02E-07	27	908	3.964758	Neg. reg. of signal transduction	ERBIN PDE4D LY6G6E NDRG2 BCL6 AMFR ECM1 CXCL12 CTNNBIP1 URI1 PPARA CRIM1 RASA1 YBX3 EZR RASIP1 TP63 RORA ITGB1 PHB GATA3 ITGA6 ARRB1 TSC2 TCF7L2 PHIP CSNK1E
yellow	4.72E-07	28	961	3.884842	Neg. reg. of cell communication	ERBIN PDE4D LY6G6E NDRG2 BCL6 AMFR ECM1 CXCL12 CTNNBIP1 URI1 PPARA CRIM1 RASA1 YBX3 EZR RASIP1 RAB11FIP1 TP63 RORA ITGB1 PHB GATA3 ITGA6 ARRB1 TSC2 TCF7L2 PHIP CSNK1E
yellow	1.57E-08	34	1169	3.877958	Cytoskeleton organization	BCL6 MAP1B CCDC120 CAPG EPHA1 EMP2 KRT14 NEXN RASA1 CDC42BPG EZR LIMK2 FLNA SS18 CHMP3 CAPZA1 POF1B CCL27 DSP ITGB1 ENAH CLIP1 CCSER2 MAP2 AMOTL1 TPPP3 ARRB1 PHIP TLN1 TUBA4A KRT2 ARF6 BICD2
yellow	4.72E-07	28	964	3.872752	Neg. reg. of signaling	ERBIN PDE4D LY6G6E NDRG2 BCL6 AMFR ECM1 CXCL12 CTNNBIP1 URI1 PPARA CRIM1 RASA1 YBX3 EZR RASIP1 RAB11FIP1 TP63 RORA ITGB1 PHB GATA3 ITGA6 ARRB1 TSC2 TCF7L2 PHIP CSNK1E
yellow	3.61E-09	41	1514	3.610744	Enzyme binding	PDE4D HES1 KDM4A AMFR ECM1 EPHA1 GABARAPL2 EMP2 URI1 SHOC2 PPARA SORT1 RASA1 PPP6R3 EZR CSTA FLNA

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
						RASIP1 RAB11FIP1 EXOC5 CHMP3 SRSF1 MST1R DSP HACD3 ITGB1 LRPPRC PHB JUP MAP2 UQCRC1 TOP2A ARRB1 TSC2 ZBTB7B TCF7L2 TUBA4A UBC ARF6 BICD2 YIPF1
yellow	1.07E-06	30	1159	3.451251	Neg. reg. of response to stimulus	ERBIN PDE4D LY6G6E NDRG2 BCL6 AMFR ECM1 CXCL12 CTNNBIP1 URI1 PPARA CRIM1 RASA1 YBX3 EZR RASIP1 TP63 RORA ITGB1 PHB GATA3 ITGA6 SPINK5 ARRB1 TSC2 ZBTB7B TCF7L2 PHIP KRT1 CSNK1E
yellow	9.50E-07	33	1362	3.230543	Tissue development	HES1 ECM1 VEZF1 KRTDAP CTNNBIP1 NCOA3 KRT14 PPARA TMEM79 YBX3 EZR CSTA BRD2 FLNA RASIP1 POSTN EXOC5 PRKX POF1B DSP SDC4 TP63 ITGB1 GATA3 ITGA6 SPINK5 TPPP3 KRT10 TCF7L2 ZFP36L1 HECTD1 KRT2
yellow	2.80E-07	42	1925	2.909091	Reg. of protein metabolic proc.	HES1 BCL6 KDM4A SYAP1 GSPT1 SETD7 TARDBP ECM1 ACSL1 EPHA1 GABARAPL2 EMP2 CLPX CRIM1 DBI PPP6R3 EZR LIMK2 CSTA FLNA RASIP1 S100A8 CSNK2A2 MST1R TP63 RAB1A PHB GATA3 CSTB SPINK5 CPEB2 ARRB1 ZBTB7B TCF7L2 BRD7 PITHD1 PRLR ZFP36L1 VPS28 GNL3L WFDC18 CSNK1E
yellow	9.50E-07	38	1746	2.901871	Cytoskeleton	MAP1B CCDC120 CAPG KRT80 SHTN1 KRT14 NEXN PHAX CDC42BPG EZR LIMK2 FLNA SS18 S100A8 CALD1 CAPZA1 POF1B MST1R DSP TP63 ENAH LRPPRC CDK16 CLIP1

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
						CCSER2 JUP MAP2 TOP2A TPPP3 KRT10 TLN1 IFFO2 TUBA4A KRT5 KRT2 KRT1 BICD2
yellow	7.22E-07	39	1798	2.892102	Neg. reg. of nitrogen compound metabolic proc.	HES1 BCL6 TFAP2C KDM4A TARDBP ECM1 ZNF205 CTNNBIP1 EPHA1 GABARAPL2 ZMYM5 URI1 PPARA CRIM1 TRIM29 DBI PHAX YBX3 EZR CSTA FLNA RASIP1 CSNK2A2 HMGB2 TP63 PHB GATA3 RUNX1T1 CSTB SPINK5 CPEB2 ARRB1 DNMT3A ZBTB7B TCF7L2 BRD7 ZFP36L1 VPS28 GNL3L
yellow	2.91E-07	42	1946	2.877698	Reg. of localization	PDE4D BCL6 MAP1B ITPR2 TARDBP ECM1 CXCL12 ZNF205 PLIN2 EPHA1 EMP2 SHTN1 PPARA ARSB NEXN TRIM29 DBI EZR LIMK2 FLNA RAB11FIP1 S100A8 CHMP3 PRKX CCL27 DSP SDC4 ITGB1 CDK16 GATA3 ITGA6 JUP AMOTL1 ARRB1 TCF7L2 VAMP8 ZFP36L1 VPS28 GNL3L HECTD1 ARF6 CSNK1E
yellow	1.54E-06	39	1866	2.78671	Reg. of multicellular organismal proc.	ERBIN PDE4D HES1 NDRG2 BCL6 ECM1 VEZF1 ACSL1 CTNNBIP1 EPHA1 EMP2 NCOA3 NPR3 PPARA ARSB YBX3 SUCO EZR RAB11FIP1 IL37 PRKX HMGB2 DSP TP63 RORA ITGB1 RAB1A PHB GATA3 JUP IL33 ARRB1 ZBTB7B TCF7L2 PITHD1 VAMP8 ZFP36L1 ARF6

Appendix E: The significant pathways of significant modules in spleen datasets. This table shows the genes involved in each pathway and the value of Enrichment FDR and Fold Enrichment

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
blue	1.51E-22	21	62	29.886148	WP1071 Cytoplasmic Ribosomal Proteins	RPS17 RPL14 RPS18 RPL35 RPL27A RPS9 RPS14 RPS2 RPSA RPL35A RPL34 RPS8 RPL18 RPS5 RPS7 RPL30 RPLP0 RPS15 RPL11 RPL28 RPS29
blue	8.94E-22	24	103	20.55968	GTP hydrolysis and joining of the 60S ribosomal subunit	RPL21 RPL14 RPS18 RPL35 RPL3 RPL13A RPL27A RPS9 RPL10 RPS2 RPSA RPL35A RPL34 RPS8 RPL18 RPS5 RPS7 RPL30 RPL36A RPS15 RPL11 RPS16 RPL28 RPS6
blue	5.97E-30	33	143	20.361991	Ribosome	RPL21 RPL14 RPS18 RPL35 RPL3 RPL13A RPL27A RPS9 RPL22L1 RPL10 RPS14 RPS2 RPSA RPL35A RPL34 RPS8 MRPS10 RPL18 RPS5 RPS7 RPL30 RPLP0 RPL36A RPS15 RPL11 RPS16 RPL17 RPL28 RPS6 MRPS18A RPS26 RPL23A RPS29
blue	2.45E-21	24	108	19.607843	SRP-dependent cotranslational protein targeting to membrane	RPL21 RPL14 RPS18 RPL35 RPL3 RPL13A RPL27A RPS9 RPL10 RPS2 RPSA RPL35A RPL34 RPS8 RPL18 RPS5 RPS7 RPL30 RPL36A RPS15 RPL11 RPS16 RPL28 RPS6

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
blue	2.60E-21	24	109	19.427955	Nonsense Mediated Decay NMD independent of the Exon Junction Complex EJC	RPL21 RPL14 RPS18 RPL35 RPL3 RPL13A RPL27A RPS9 RPL10 RPS2 RPSA RPL35A RPL34 RPS8 RPL18 RPS5 RPS7 RPL30 RPL36A RPS15 RPL11 RPS16 RPL28 RPS6
blue	6.18E-21	24	114	18.575851	Formation of a pool of free 40S subunits	RPL21 RPL14 RPS18 RPL35 RPL3 RPL13A RPL27A RPS9 RPL10 RPS2 RPSA RPL35A RPL34 RPS8 RPL18 RPS5 RPS7 RPL30 RPL36A RPS15 RPL11 RPS16 RPL28 RPS6
blue	3.81E-20	24	123	17.216643	L13a-mediated translational silencing of Ceruloplasmin expression	RPL21 RPL14 RPS18 RPL35 RPL3 RPL13A RPL27A RPS9 RPL10 RPS2 RPSA RPL35A RPL34 RPS8 RPL18 RPS5 RPS7 RPL30 RPL36A RPS15 RPL11 RPS16 RPL28 RPS6
blue	3.97E-20	24	125	16.941176	Eukaryotic Translation Initiation	RPL21 RPL14 RPS18 RPL35 RPL3 RPL13A RPL27A RPS9 RPL10 RPS2 RPSA RPL35A RPL34 RPS8 RPL18 RPS5 RPS7 RPL30 RPL36A RPS15 RPL11 RPS16 RPL28 RPS6
blue	3.97E-20	24	125	16.941176	Cap-dependent Translation Initiation	RPL21 RPL14 RPS18 RPL35 RPL3 RPL13A RPL27A RPS9 RPL10 RPS2 RPSA RPL35A RPL34 RPS8 RPL18 RPS5 RPS7 RPL30 RPL36A RPS15 RPL11 RPS16 RPL28 RPS6

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
blue	3.97E-20	24	125	16.941176	Nonsense-Mediated Decay NMD	RPL21 RPL14 RPS18 RPL35 RPL3 RPL13A RPL27A RPS9 RPL10 RPS2 RPSA RPL35A RPL34 RPS8 RPL18 RPS5 RPS7 RPL30 RPL36A RPS15 RPL11 RPS16 RPL28 RPS6
blue	6.78E-26	36	244	13.018322	Coronavirus disease	RPL21 RPL14 RPS18 RPL35 RPL3 RPL13A RPL27A RPS9 RPL22L1 RPL10 RPS14 RPS2 RPSA C1QB RELA RPL35A RPL34 RPS8 RPL18 RPS5 RPS7 RPL30 RPLP0 IL6R RPL36A RPS15 NFKB1 RPL11 RPS16 RPL17 RPL28 RPS6 MX1 RPS26 RPL23A RPS29
blue	2.19E-18	26	186	12.333966	Major pathway of rRNA processing in the nucleolus and cytosol	RPL21 RPL14 RPS18 RPL35 RPL3 RPL13A RPL27A WDR36 NOP58 RPS9 RPL10 RPS2 RPSA RPL35A RPL34 RPS8 RPL18 RPS5 RPS7 RPL30 RPL36A RPS15 RPL11 RPS16 RPL28 RPS6
blue	2.19E-18	26	186	12.333966	RRNA processing	RPL21 RPL14 RPS18 RPL35 RPL3 RPL13A RPL27A WDR36 NOP58 RPS9 RPL10 RPS2 RPSA RPL35A RPL34 RPS8 RPL18 RPS5 RPS7 RPL30 RPL36A RPS15 RPL11 RPS16 RPL28 RPS6
blue	2.19E-18	26	186	12.333966	RRNA processing in the nucleus and cytosol	RPL21 RPL14 RPS18 RPL35 RPL3 RPL13A RPL27A WDR36 NOP58 RPS9 RPL10 RPS2 RPSA RPL35A RPL34 RPS8

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
						RPL18 RPS5 RPS7 RPL30 RPL36A RPS15 RPL11 RPS16 RPL28 RPS6
blue	3.53E-17	26	208	11.029412	Translation	RPL21 RPL14 RPS18 RPL35 RPL3 RPL13A RPL27A RPS9 RPL10 RPS2 RPSA RPL35A RPL34 RPS8 MRPS10 RPL18 RPS5 RPS7 RPL30 RPL36A RPS15 RPL11 RPS16 RPL28 RPS6 MRPS18A
blue	3.73E-18	28	232	10.649087	Cytosolic ribosome	RPL21 RPL14 RPS18 RPL35 RPL3 RPL13A RPL27A RPS9 RPL10 RPS14 RPSA RPL35A RPL34 RPL18 RPS5 RPL30 RPLP0 RPL36A RPS15 RPL11 RPS16 RPL17 RPL28 RPS6 RPS26 RPL23A RPS29
blue	6.66E-20	34	331	9.0634441	Structural constituent of ribosome	RPS17 RPL21 RPL14 RPS18 RPL35 RPL3 RPL13A RPL27A RPS9 RPL22L1 RPL10 RPS14 RPS2 RPSA RPL35A RPL34 RPS8 RPL18 RPS5 RPS7 RPL30 RPLP0 RPL36A RPS15 RPL11 RPS16 RPL17 RPL28 RPS6 MRPS18A RPS26 RPL23A RPS29
blue	6.18E-21	37	375	8.7058824	Ribosome	RPS17 RPL21 RPL14 RPS18 RPL35 RPL3 RPL13A RPL27A RPS9 RPL22L1 RPL10 RPS14 RPS2 RPSA BTF3 RPL35A RPL34 RPS8 MRPS10 RPL18 RPS5 RPS7 RPL30 RPLP0 RPL36A RPS15 RPL11

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
						RPS16 RPL17 RPL28 RPS6 MRPS18A NUFIP2 RPS26 RPL23A RPS29
blue	1.97E-18	47	773	5.3648885	Ribonucleoprotein complex	RPL21 RPL14 GPKOW RPS18 EIF3G RPL35 RPL3 RPL13A RPL27A WDR36 NOP58 RPS9 NVL RPL10 RPS14 RPS2 RPSA BTF3 AKAP8L SNRPA1 G3BP2 RPL35A GAPDH MRPS10 RPL18 RPS5 RPL30 RALY RPLP0 CWF19L2 RPL36A RPS15 RPL11 RPS16 HNRNPK RPL17 RPL28 RPS6 MRPS18A NUFIP2 RPS26 ILF3 RBMS3 RPL23A RPS29
brown	2.98E-09	41	1009	3.443584	Pos. reg. of transcription by RNA polymerase II	CREBRF COPS5 KAT6B EPAS1 UBP1 MLXIP IRF5 MED6 SMARCA2 NUCKS1 HEYL IRF2 MEF2A FOXO3 RBM15 TNIP1 PBX1 TMF1 FOXJ3 CBFB EP300 EOMES GATA3 THRB RXRA AHI1 MAX BCL11B CNOT7 NFIX NCK1 ARID4A DHX9 ASXL2 SPI1 SUB1 PHIP SETX DDX3X SMAD2
brown	4.13E-09	41	1034	3.3603252	Post-translational protein modification	CST3 FBXW2 SENP2 DNMT1 COPS5 OTULIN EPAS1 SPARCL1 NUP133 COPS8 RAD21 XPC NUP98 PSMA6 SEH1L USP47 ALG5 TNIP1 PSMB6 VDAC3 SDC2 CALM3 RAD23A SPTAN1 CFP EP300 DYNC1H1 RAB1A UBA1 THRB RXRA SMC5 MIA3 ASXL2

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
						PSMC5 MAN1A1 DYNLL2 BIRC6 PSMD8 DCUN1D3 IGFBP5
brown	2.44E-09	44	1125	3.3145009	Neg. reg. of cellular macromolecule biosynthetic proc.	PRDM1 DNMT1 KAT6B TARDBP CIRBP SMARCA2 PAIP2 PRDX5 HEYL PURA MEF2A USP47 FOXO3 WAPL AEBP1 RBM15 SMARCC2 BANK1 CBFB EP300 GATA3 BRMS1 THRB RXRA MAX TNRC6A CNOT7 NFIX NCK1 ARID4A PUM1 PDCD4 SUDS3 TXNIP SET PSMC5 SPI1 ID3 ZNF217 ZNF557 DDX3X CIR1 SMAD2 IGFBP5
brown	2.85E-10	52	1386	3.1794947	Pos. reg. of RNA metabolic proc.	TRA2B CREBRF COPS5 KAT6B TARDBP EPAS1 UBP1 MLXIP IRF5 MED6 XPC CIRBP SMARCA2 NUCKS1 NUP98 HEYL IRF2 MEF2A FOXO3 ARHGEF11 RBM15 TNIP1 PBX1 SMARCC2 TMF1 FOXJ3 CBFB EP300 EOMES GATA3 THRB RXRA AHI1 MAX TNRC6A BCL11B CNOT7 NFIX ACTN1 NCK1 ARID4A PUM1 DHX9 ASXL2 SPI1 SUB1 PHIP SETX SUPT3H DDX3X SMAD2
brown	2.98E-09	47	1291	3.0852447	Pos. reg. of transcription, DNA-templated	CREBRF COPS5 KAT6B EPAS1 UBP1 MLXIP IRF5 MED6 XPC SMARCA2 NUCKS1 NUP98 HEYL IRF2 MEF2A FOXO3 ARHGEF11 RBM15 TNIP1 PBX1

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
						SMARCC2 TMF1 FOXJ3 CBFB EP300 EOMES GATA3 THRB RXRA AHI1 MAX BCL11B CNOT7 NFIX ACTN1 NCK1 ARID4A DHX9 ASXL2 SPI1 SUB1 PHIP SETX SUPT3H DDX3X SMAD2
brown	2.98E-09	47	1291	3.0852447	Pos. reg. of nucleic acid-templated transcription	CREBRF COPS5 KAT6B EPAS1 UBP1 MLXIP IRF5 MED6 XPC SMARCA2 NUCKS1 NUP98 HEYL IRF2 MEF2A FOXO3 ARHGEF11 RBM15 TNIP1 PBX1 SMARCC2 TMF1 FOXJ3 CBFB EP300 EOMES GATA3 THRB RXRA AHI1 MAX BCL11B CNOT7 NFIX ACTN1 NCK1 ARID4A DHX9 ASXL2 SPI1 SUB1 PHIP SETX SUPT3H DDX3X SMAD2
brown	2.98E-09	47	1292	3.0828567	Pos. reg. of RNA biosynthetic proc.	CREBRF COPS5 KAT6B EPAS1 UBP1 MLXIP IRF5 MED6 XPC SMARCA2 NUCKS1 NUP98 HEYL IRF2 MEF2A FOXO3 ARHGEF11 RBM15 TNIP1 PBX1 SMARCC2 TMF1 FOXJ3 CBFB EP300 EOMES GATA3 THRB RXRA AHI1 MAX BCL11B CNOT7 NFIX ACTN1 NCK1 ARID4A DHX9 ASXL2 SPI1 SUB1 PHIP SETX SUPT3H DDX3X SMAD2

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
brown	3.43E-09	48	1348	3.0176533	Metabolism of proteins	CST3 FBXW2 SENP2 CREBRF DNMT1 COPS5 OTULIN EPAS1 SPARCL1 NUP133 COPS8 EIF3E EIF4B RAD21 XPC NUP98 SRP68 PSMA6 SEH1L USP47 EIF3K ALG5 TNIP1 PSMB6 VDAC3 EIF2S3 SDC2 CALM3 OXA1L RAD23A SPTAN1 CFP EP300 DYNC1H1 RAB1A UBA1 THRB RXRA SMC5 MIA3 ASXL2 PSMC5 MAN1A1 DYNLL2 BIRC6 PSMD8 DCUN1D3 IGFBP5
brown	1.86E-09	51	1440	3.0014124	Pos. reg. of macromolecule biosynthetic proc.	CREBRF COPS5 KAT6B BCAR3 EPAS1 UBP1 MLXIP IRF5 EIF3E MED6 XPC SMARCA2 NUCKS1 NUP98 HEYL IRF2 MEF2A FOXO3 ARHGEF11 RBM15 TNIP1 PBX1 SMARCC2 TMF1 FOXJ3 CBFB EP300 EOMES GATA3 THRB RXRA AHI1 MAX BCL11B CNOT7 NFIX ACTN1 NCK1 VIM ARID4A DHX9 COA3 ASXL2 SPI1 SUB1 PHIP SETX SUPT3H DDX3X SMAD2
brown	3.61E-11	63	1798	2.969401	Neg. reg. of nitrogen compound metabolic proc.	MYD88 CST3 PRDM1 SENP2 PTN DNMT1 KAT6B TARDBP YWHAE CIRBP SMARCA2 CORO1C PAIP2 PRDX5 HEYL PURA NTRK2 MEF2A TIMP2 USP47 RTN4 FOXO3 AEBP1 XDH RBM15 SMARCC2 CD27 BANK1

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
						TLK2 CBFB EP300 GATA3 BRMS1 THRB RXRA MAX TNRC6A CNOT7 A2M NFIX NCK1 ARID4A PUM1 PDCD4 SUDS3 DHX9 TXNIP PFDN2 TIMP3 SET PSMC5 SPI1 SUB1 IBTK BIRC6 ID3 ZNF217 ZNF557 DCUN1D3 DDX3X CIR1 SMAD2 IGFBP5
brown	1.19E-09	53	1514	2.9666614	Enzyme binding	PRDM1 ATP2A2 PTN COPS5 BCAR3 MSN EPAS1 IRF5 TNPO3 YWHAE RBBP4 CCDC50 CORO1C PITPNM1 NTRK2 MEF2A TIMP2 RTN4 PPP1CC CYB5A ATP6AP1 SPAG9 TNIP1 LCK CSF1R CAP1 WDFY3 CHMP3 CALM3 ACTA2 AKAP12 RAD23A BANK1 LYST HACD4 THRB RXRA A2M NCK1 UQCRC1 ZNF326 SUDS3 DHX9 TXNIP ABCA1 SPI1 CACUL1 IBTK PPP1R3E CIR1 RABGAP1 SMAD2 SH3PXD2A
brown	6.14E-11	65	1948	2.827759	Neg. reg. of cellular metabolic proc.	MYD88 CST3 PRDM1 SENP2 PTN DNMT1 KAT6B TARDBP YWHAE CIRBP SMARCA2 CORO1C PAIP2 PRDX5 HEYL PURA NTRK2 MEF2A TIMP2 USP47 ZFYVE1 RTN4 FOXO3 WAPL AEBP1 XDH RBM15 SMARCC2 CD27 BANK1 TLK2 CBFB EP300 GATA3 BRMS1 THRB RXRA MAX

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
						TNRC6A CNOT7 A2M NFIX NCK1 ARID4A PUM1 PDCD4 SUDS3 DHX9 TXNIP TIMP3 SMG1 SET PSMC5 SPI1 SUB1 IBTK BIRC6 ID3 ZNF217 ZNF557 DCUN1D3 DDX3X CIR1 SMAD2 IGFBP5
brown	1.00E-09	59	1800	2.7777778	Reg. of cellular protein metabolic proc.	CST3 HIP1 SENP2 PTN DNMT1 BCAR3 TPD52L1 MSN TARDBP EPAS1 ASPSCR1 NEMF ABI1 YWHAE EIF3E CIRBP CORO1C PAIP2 PURA APH1B NTRK2 TIMP2 USP47 EIF3K XDH TNIP1 LCK CSF1R CALM3 CD27 RAD23A BANK1 TMF1 TLK2 EP300 FXR1 GATA3 BRMS1 TNRC6A CNOT7 A2M NCK1 VIM PUM1 DHX9 COA3 TIMP3 PELI1 PSMC5 NRDC SPI1 CACUL1 IBTK BIRC6 DCUN1D3 MTCH1 DDX3X IGFBP5 RAP2B
cyan	1.01E-10	8	53	83.857442	Endoplasmic reticulum lumen	PDIA6 HSP90B1 P4HB ERP29 HSPA5 TXNDC5 MANF MZB1
cyan	0.0014338	3	22	75.757576	Immunoglobulin complex, circulating	JCHAIN
cyan	0.0018533	3	26	64.102564	Immunoglobulin complex	JCHAIN
cyan	0.0018533	3	26	64.102564	Immunoglobulin receptor binding	JCHAIN

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
cyan	0.0029933	3	34	49.019608	Asthma	FCER1G
cyan	2.74E-05	5	59	47.080979	Peptidyl-proline modification	P4HB PRDX4 FKBP2 FKBP11 PPIB
cyan	4.38E-07	7	107	36.344756	Endoplasmic reticulum protein-containing complex	P4HB HSPA5 SPCS3 PPIB DERL3 TMEM258 MZB1
cyan	1.12E-08	9	167	29.94012	Protein processing in endoplasmic reticulum	PDIA6 HSP90B1 P4HB ERP29 HSPA5 RRBP1 TXNDC5 DERL3 SSR1
cyan	0.002415	4	93	23.894863	Fc gamma R-mediated phagocytosis	FCGR3A ACTR3
cyan	0.0001019	6	152	21.929825	Protein folding	HSP90B1 P4HB PRDX4 FKBP2 FKBP11 PPIB
cyan	0.0014338	5	153	18.15541	Isomerase activity	PDIA6 P4HB FKBP2 FKBP11 PPIB
cyan	0.0018533	5	172	16.149871	Immune response-regulating cell surface receptor signaling pathway	FCGR3A ADA FCER1G
cyan	0.0005045	6	210	15.873016	Response to endoplasmic reticulum stress	HSP90B1 P4HB ERP29 HSPA5 DERL3 MANF
cyan	0.0021469	5	182	15.262515	Tuberculosis	IL18 FCGR3A FCER1G
cyan	5.09E-10	18	1129	8.8573959	Endoplasmic reticulum	TMEM50A TMED9 PDIA6 HSP90B1 P4HB PRDX4 ERP29 TPD52 HSPA5 SPCS3 RRBP1 PPIB TXNDC5 DERL3 SSR1 MANF TMEM258 MZB1

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
cyan	0.0021469	8	646	6.879945	Leukocyte activation	IL18 NDRG1 FCGR3A ADA TPD52 MZB1
cyan	0.0023777	8	661	6.7238191	Cell surface	FCGR3A ADA P4HB ERP29 HSPA5 FCER1G
cyan	0.0014338	9	751	6.6577896	Organelle subcompartment	ARAP1 HSP90B1 ATP8B2 HSPA5 SPCS3 RRBP1 DERL3 SSR1 TMEM258
cyan	0.0003934	11	1006	6.0746631	Reg. of immune system proc.	IL18 HCLS1 FCGR3A LRCH1 ADA PARP14 TRIB1 FCER1G MZB1
cyan	0.0018533	12	1509	4.4179368	Immune System	SLAMF7 FCGR3A HSP90B1 ACTR3 P4HB PRDX4 HSPA5 PNP PDXK TXNDC5 FCER1G LAP
green	4.54E-05	15	338	5.3901253	Autophagy	KDR HTT MTDH EI24 TRIM13 RBM8A RAB7A EIF4G1 QSOX1 MCL1 PLK2 WAC UFL1 VAMP8 TP53INP1
green	4.54E-05	15	338	5.3901253	Process utilizing autophagic mechanism	KDR HTT MTDH EI24 TRIM13 RBM8A RAB7A EIF4G1 QSOX1 MCL1 PLK2 WAC UFL1 VAMP8 TP53INP1
green	3.43E-06	24	672	4.3377675	Neg. reg. of gene expression	KDR DYRK1A APEX1 ATF7IP PPARA RBM8A TTC37 NAV3 RNPS1 PUM2 CITED2 SERPINB1 EIF4G1 GIGYF2 EXOSC4 PTBP3 APP INPPL1 PTPN6 PDGFB DDIT3 MKKS FAM172A TP53INP1
green	7.76E-06	23	675	4.1385515	Reg. of catabolic proc.	KDR LPCAT1 HTT APEX1 MTDH PPARA TRIM13 RAB7A UBXN1 SF3B3 PUM2 EIF4G1 GIGYF2 QSOX1 SGTA

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
						MCL1 PLK2 HERPUD1 APP WAC WNT5A UFL1 TP53INP1
green	4.54E-05	27	1037	3.1623454	Macromolecule catabolic proc.	LPCAT1 APEX1 PSMD4 PSMB1 TRIM13 RBM8A PSMB9 TTC37 RNPS1 RAB7A UBXN1 SF3B3 PUM2 USP8 UBE2L3 GIGYF2 EXOSC4 SGTA PLK2 HERPUD1 WAC USP16 WNT5A UFL1 RMND5A DDIT3 ELOC
green	1.10E-06	39	1509	3.1390604	Immune System	YWHAZ RHOG PPP2CA LPCAT1 PKM FOS GDI2 ARPC3 MAP3K14 VAT1 ATP6V1F RAB37 TAP1 ATP7A LIFR EIF4E2 SEC24B SERPINB1 EIF4G1 IFNGR2 UBE2L3 CD86 HERC4 QSOX1 CPNE3 BAIAP2 INPPL1 SMARCA4 LNPEP CYBB PTPN6 UFL1 ARHGAP45 PDGFB VAMP8 FBXO21 C4A ELOC
green	4.54E-05	28	1099	3.0944583	Reg. of cell death	YWHAZ KDR PRKCA HTT APEX1 MTDH NAIP MYBBP1A NET1 GRK5 PPARA TRIM13 PDCD5 RNPS1 FRZB CITED2 EIF4G1 CTNNA1 MCL1 ANP32E PLK2 HERPUD1 RBM25 ITM2C RAMP2 WNT5A DDIT3 TP53INP1
green	6.30E-05	27	1061	3.0908127	Pos. reg. of signal transduction	KDR PRKCA HTT NDST1 MAPK8IP3 MTDH AFAP1L2 MAP3K14 NET1 TRIM13 LIMS1 PUM2 CITED2 USP8 CTNNA1 MCL1 ERN1 TAOK2 APP

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
						ITM2C SMARCA4 WNT5A PTPN6 COL3A1 TCF7L2 PDGFB DDIT3
green	4.46E-05	29	1161	3.0338219	Pos. reg. of molecular function	RHOG PPP2CA KDR HTT MTDH ARHGAP42 AFAP1L2 CHTOP GDI2 PABPN1 MAP3K14 NET1 TRIM13 PDCD5 DBI ATP7A LIMS1 EVI5 EIF4G1 TAOK2 APP SMARCA4 DOCK7 WNT5A ARHGAP45 TCF7L2 PDGFB DDIT3 DNAJC24
green	1.51E-05	33	1351	2.9667633	Cell death	YWHAZ KDR PRKCA HTT PKM APEX1 MTDH NAIP TCHP MYBBP1A NET1 GRK5 PPARA TRIM13 PDCD5 RNPS1 FRZB CITED2 EIF4G1 CTNNA1 MCL1 ANP32E PLK2 ERN1 HERPUD1 RBM25 APP ITM2C EMP3 RAMP2 WNT5A DDIT3 TP53INP1
green	5.33E-06	38	1604	2.8774218	Cellular catabolic proc.	KDR HTT APEX1 MTDH ESD PSMD4 EI24 PSMB1 TRIM13 RBM8A PSMB9 TTC37 RNPS1 RAB7A UBXN1 PUM2 USP8 PDE4A EIF4G1 UBE2L3 GIGYF2 QSOX1 EXOSC4 SGTA MCL1 PLK2 HERPUD1 APP WAC LNPEP USP16 UFL1 RMND5A SAMHD1 VAMP8 DDIT3 ELOC TP53INP1
green	2.67E-05	34	1480	2.7902396	Pos. reg. of response to stimulus	KDR PRKCA HTT NDST1 MAPK8IP3 MTDH AFAP1L2 MAP3K14 NET1

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
						TRIM13 C1R LIMS1 PUM2 CITED2 USP8 ECHDC3 CTNNA1 SGTA MCL1 ERN1 HERPUD1 TAOK2 APP ITM2C SMARCA4 WNT5A PTPN6 COL3A1 TCF7L2 PDGFB VAMP8 DDIT3 C4A
green	2.66E-06	43	1883	2.7735911	Catabolic proc.	KDR LPCAT1 HTT PKM APEX1 MTDH ESD PSMD4 EI24 PSMB1 PPARA TRIM13 RBM8A PSMB9 TTC37 RNPS1 RAB7A UBXN1 SF3B3 PUM2 USP8 PDE4A EIF4G1 UBE2L3 GIGYF2 QSOX1 EXOSC4 SGTA MCL1 PLK2 HERPUD1 APP WAC LNPEP USP16 WNT5A UFL1 RMND5A SAMHD1 VAMP8 DDIT3 ELOC TP53INP1
green	7.67E-06	40	1798	2.7020576	Neg. reg. of nitrogen compound metabolic proc.	SP3 LPCAT1 DYRK1A GATAD2A MTDH ATF7IP NAIP BOD1L1 MYBBP1A PPARA DBI RNPS1 UBXN1 SF3B3 LIMS1 CITED2 SERPINB1 EIF4G1 GIGYF2 YEATS2 SGTA PTBP3 KLF3 APP NELFA WAC SMARCA4 NR3C1 WNT5A PTPN6 UFL1 TCF7L2 PDGFB DDIT3 C4A ZBTB34 KLF7 CGGBP1 PFDN1
green	1.64E-05	38	1732	2.6647717	Reg. of catalytic activity	RHOG PPP2CA KDR HTT NAIP BOD1L1 ARHGAP42 AFAP1L2 CHTOP GDI2 ARHGDIB DENND5A PABPN1

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
						MAP3K14 NET1 PSMB9 PDCD5 DBI ATP7A UBXN1 LIMS1 EVI5 SERPINB1 RHOH ANP32E TAOK2 APP ARHGEF6 DOCK7 WNT5A PTPN6 ARHGAP45 PDGFB ARHGDIA MKKS C4A DNAJC24
green	7.67E-06	42	1948	2.6186933	Neg. reg. of cellular metabolic proc.	SP3 TIPIN LPCAT1 DYRK1A GATAD2A MTDH ATF7IP NAIP BOD1L1 MYBBP1A PPARA DBI RNPS1 UBXN1 LIMS1 CITED2 SERPINB1 EIF4G1 RHOH GIGYF2 QSOX1 YEATS2 SGTA MCL1 PTBP3 KLF3 APP NELFA WAC SMARCA4 NR3C1 WNT5A PTPN6 UFL1 TCF7L2 PDGFB DDIT3 C4A ZBTB34 KLF7 CGGBP1
green	7.71E-05	34	1583	2.608689	Organic substance catabolic proc.	LPCAT1 PKM APEX1 ESD PSMD4 PSMB1 PPARA TRIM13 RBM8A PSMB9 TTC37 RNPS1 RAB7A UBXN1 SF3B3 PUM2 USP8 PDE4A UBE2L3 GIGYF2 EXOSC4 SGTA PLK2 HERPUD1 APP WAC LNPEP USP16 WNT5A UFL1 RMND5A SAMHD1 DDIT3 ELOC
green	1.51E-05	41	1947	2.5576564	Intracellular signal transduction	YWHAZ RHOG TIPIN KDR PRKCA HTT NDST1 DYRK1A SRPK2 MAPK8IP3 MTDH ARHGAP42 GDI2

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
						ARHGDIB MOB1A MAP3K14 MYBBP1A NET1 PPARA TRIM13 LIMS1 PUM2 USP8 PDE4A RHOH MCL1 CXCR6 ERN1 HERPUD1 TAOK2 APP ARHGEF6 DOCK7 WNT5A PTPN6 ARHGAP45 COL3A1 TCF7L2 PDGFB ARHGDIA DDIT3
green	2.67E-05	40	1925	2.523792	Reg. of protein metabolic proc.	RBM4B PPP2CA KDR LPCAT1 HTT PKM DYRK1A NAIP BOD1L1 AFAP1L2 CHTOP MAP3K14 TADA3 PSMB9 PDCD5 DBI RAB7A UBXN1 SF3B3 SERPINB1 EIF4G1 GIGYF2 SGTA PLK2 HERPUD1 TAOK2 APP WAC SETD5 DOCK7 GSN USP16 WNT5A PTPN6 UFL1 TCF7L2 PDGFB C4A PFDN1
green	7.71E-05	36	1735	2.5201554	Reg. of developmental proc.	RHOG PPP2CA KDR PRKCA PKM MTDH FOS SEMA5A CBLN2 VAT1 POFUT2 PPARA SEC24B LIMS1 FRZB CITED2 EIF4G1 RHOH CTNNA1 PTBP3 PLK2 TAOK2 APP BAIAP2 RAMP2 WNT5A PTPN6 UFL1 USF3 TCF7L2 PDGFB CHD7 DDIT3 MKKS KLF7 TP53INP1
greentellow	0.0410689	3	39	25.083612	Body morphogenesis	PDGFRA CLASP1 BRAF
greentellow	0.0410689	4	96	13.586957	Rab reg. of trafficking	DENND1A TRAPPC10 RAB4A TSC2

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
greentellow	0.0311302	5	136	11.988491	Neg. reg. of Wnt signaling pathway	AMFR UBAC2 CBY1 TLE3 TSC2
greentellow	0.0410689	5	174	9.3703148	Bta-mir-2382-5p target gene	UBAC2 ACADS EPS15L1 ACVR1 APH1A
greentellow	0.0410689	5	179	9.1085742	Endosomal transport	DENND1A USP7 TRAPPC10 EPS15L1 SNX2
greentellow	0.0076069	9	388	7.5638727	Guanyl nucleotide binding	RIT1 GSPT1 MFN2 EIF5 ARL4D RAB4A TUBA1C GNA13 NME2
greentellow	0.0410689	6	264	7.4110672	Ras protein signal transduction	RIT1 DENND1A MFN2 CYTH1 RAB4A GNA13
greentellow	0.0180794	8	358	7.2868594	Small GTPase mediated signal transduction	RIT1 DENND1A MFN2 CYTH1 RAB4A TSC2 GNA13
greentellow	0.0410689	6	275	7.1146245	Parkinson disease	NDUFS4 MFN2 COX6A1 CAMK2D NDUFA2 TUBA1C
greentellow	0.0184719	8	380	6.8649886	Guanyl ribonucleotide binding	RIT1 GSPT1 MFN2 EIF5 ARL4D RAB4A TUBA1C NME2
greentellow	0.0416265	6	286	6.8409851	GTPase activity	RIT1 GSPT1 MFN2 RAB4A TUBA1C GNA13
greentellow	0.0410689	7	365	6.2537225	GTP binding	RIT1 GSPT1 MFN2 EIF5 ARL4D RAB4A TUBA1C
greentellow	0.0410689	8	508	5.1352277	Protein homodimerization activity	CACYBP GOLGA5 PDGFRA SLC11A1 TSC2 TRIM8 SNX2 ALDH3A2

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
greentellow	0.0311302	13	1165	3.6387386	Vesicle-mediated transport	BLZF1 DENND1A GOLGA5 USP7 TRAPPC10 EPS15L1 SLC11A1 ITGB2 RAB4A CLASP1 CNN2 BRAF SNX2
greentellow	0.0410689	14	1514	3.0153351	Enzyme binding	CACYBP BLZF1 DENND1A AMFR GOLGA5 MFN2 USP7 NEDD9 URI1 TSPAN17 NOP56 TSC2 APH1A BFAR
grey	2.06E-14	72	1070	3.1104647	Nuclear protein-containing complex	RANBP2 ANAPC16 SSRP1 INTS1 ANAPC4 OGT NOP14 CDK13 RTF1 RFXANK EXOSC10 RBBP7 TAF3 MCM3AP ASH1L JUN SETD1B KANSL3 XRCC6 DRAP1 TFDP2 RBM5 RCOR3 CDK19 SF3A3 CPSF6 ENY2 NCOA6 POLA1 LSM4 BRD8 ATF6B LARP7 SF3B5 HMGXB4 FLNA MED13 JARID2 RPA3 NBN MED1 GTF2A1 PRPF3 MCM2 KMT2D WBP11 SRSF1 PCF11 SMARCE1 CTR9 UTP6 CTNNB1 MLH1 ING4 MXI1 KMT2A AFF1 RAD51D BAZ1A ZNFX1 SUPT16H MED12 DDX23 RBM17 POU2F1 TAF12 ANAPC1 RBM22 SHMT2 CDK6 SMAD5
grey	1.11E-13	74	1171	2.9211333	Reg. of protein modification proc.	TAOK1 OGT STAP1 AKT2 NSD3 RTF1 SPRY2 SWAP70 PPP4R2 APOA1 C23H6orf89 SYAP1 CCNL1 JUN OSBPL8 DUSP6 CCNL2 DIP2B PPP2R5D PIAS1 PTK2B XRCC6 ARRDC3 EMP2 WWTR1

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
						SH2B3 JDP2 RNF20 KRAS ITGB3 CCND3 SNX25 IGF1 PFN2 LRP6 FBLN1 CSNK2A1 INSR JARID2 MAP4K4 PPP1R15B NBN PUM3 CHORDC1 CTCF DNAJB2 GNL3 MTOR CTR9 MOB1B CDYL CTNNB1 PHACTR2 PKN1 RASSF2 EEF2K NUPR1 KMT2A APC ZNF451 PPP2R5E VPS25 PPP2R1B BIRC3 FNIP1 MALT1 PDCD10 MAP2K1 ATG7 XIAP CD81 PAK2 PRNP SOX4
grey	2.06E-14	79	1256	2.9074617	Catalytic complex	ANAPC16 ANAPC4 OGT CDK13 RTF1 PPP4R2 EXOSC10 RBBP7 TAF3 DPM3 ASH1L UBE2D2 SETD1B KANSL3 NAA15 PPP2R5D PSME2 UBR2 SUCLG1 DCAF1 PRMT1 ARIH1 RCOR3 CUL2 DYRK2 SF3A3 ENY2 NCF4 NDUFA13 NCOA6 POLA1 SDHB BRD8 RNF20 CCND3 CDK1 HMGXB4 KBTBD8 SNX4 SEC11C INSR NDUFB3 DICER1 JARID2 PPP1R15B GTF2A1 PRKAB2 KMT2D NDUFA12 SRSF1 PSMD6 AGO4 SMARCE1 CTR9 DNAH7 ING4 RNF19A KMT2A SMC6 PSMD7 PPP2R5E BAZ1A ZNFX1 PPP2R1B GNAI2 DCAF6 NDUFV1 DDX23 NDUFS1 CBR4

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
						TRMT10C ENC1 TAF12 ANAPC1 RBM22 DYNLT3 SLA2 CDK6
grey	7.88E-12	66	1063	2.8700352	Peptidyl-amino acid modification	RANBP2 PARP1 OGT STAP1 AKT2 NSD3 RTF1 SPRY2 APOA1 MCM3AP DPH3 ASH1L SETD1B KANSL3 NAA15 WNK1 DIP2B DOHH UGGT1 PIAS1 PTK2B DCAF1 PRMT1 DYRK2 NCOA3 PRDX3 BRD8 SH2B3 ITGB3 CDK1 IGF1 PFN2 SENP1 INSR JARID2 CTCF KMT2D GNL3 MTOR CTR9 CDC7 CTNNB1 PKN1 ING4 SUMO3 RASSF2 EFEMP1 MKNK2 KMT2A ZNF451 DPH5 VPS25 METAP2 CSK PKN2 FNIP1 ASPH TAF12 PDCD10 MAP2K1 DPH7 CD81 FKBP5 PAK2 PRNP SOX4
grey	9.61E-12	68	1123	2.7990181	Reg. of cell population proliferation	NDRG2 AKT2 SPRY2 EAPP JUN PIAS1 PTK2B XRCC6 PRMT1 TNS2 EMP2 WWTR1 MEIS1 SRSF6 EIF2AK2 PRDX3 MAGED1 VASH1 SH2B3 ITGA4 KRAS ITGB3 CCND3 CTSH IKZF3 IGF1 CNN1 FBLN1 INSR JAG1 JARID2 RPA3 MED1 CTCF DNAJB2 PTGDS TOB2 HMOX1 ITGB1 CDC7 CTNNB1 PKN1 ING4 NUPR1 KMT2A NFATC2 APC FOXJ2 ALOX5 MTSS1 AIF1 GNAI2 DPT CSK TNFRSF13C VAV3 PDCD10 PDCD1LG2

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
						CCAR1 MAP2K1 CCL2 CDK6 NRAS LGMN CD81 PLA2G2D1 SOX4 MAZ
grey	1.21E-12	75	1259	2.7536713	Programmed cell death	SELENOS BCL2A1 TAOK1 PARP1 OGT AKT2 SPRY2 DOCK8 VPS35 ANO6 ELMO1 JUN SKIL DUSP6 NAA15 PTK2B RBM5 IRF3 DYRK2 IFI6 NDUFA13 EIF2AK2 PRDX3 MAGED1 ITGA4 KRAS ITPK1 CDK1 LAMP1 CTSH IKZF3 IGF1 FLNA DNM1L PYGL LRP6 ZFAND6 SENP1 MAP4K4 STK17B NBN MED1 PPT1 BCL2L13 MCM2 AGO4 HMOX1 ITGB1 CTNNB1 RGL2 PKN1 ING4 RASSF2 NUPR1 APC DFFB PPP2R1B IDO1 CASP4 NDUFS1 BIRC3 FNIP1 MALT1 PDCD10 CCAR1 ATG7 CCL2 XIAP LGMN PAK2 PRNP SOX4 MAZ GPX4
grey	3.76E-18	107	1800	2.7478172	Reg. of cellular protein metabolic proc.	TAOK1 OGT STAP1 ZNF598 AKT2 NSD3 RTF1 SPRY2 SWAP70 EIF3D PPP4R2 APOA1 C23H6orf89 VPS35 SYAP1 CCNL1 DPH3 JUN OSBPL8 DUSP6 CCNL2 DIP2B PPP2R5D PIAS1 PTK2B XRCC6 MRPL13 ARRDC3 EMP2 IFI6 NDUFA13 WWTR1 RPUSD3 LRIG2 EIF2AK2 NCBP2 SH2B3 JDP2 RNF20 EIF4E KRAS ITGB3 CCND3 SNX25

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
						CTSH KBTBD8 IGF1 PFN2 RGP1 LRP6 FBLN1 CSNK2A1 SENP1 INSR JARID2 MAP4K4 PPP1R15B NBN PUM3 CHORDC1 CTCF DNAJB2 GNL3 AGO4 TCOF1 MTOR CTR9 TRIB2 MOB1B CDYL CTNNB1 PHACTR2 PKN1 RASSF2 EEF2K RNF19A NUPR1 KMT2A TNRC6C CTSZ APC ZNF451 DPH5 PPP2R5E VPS25 PPP2R1B EIF4G2 TNRC6B PITHD1 TRMT10C BIRC3 FNIP1 ASPH ENC1 MALT1 SHMT2 PDCD10 MAP2K1 ATG7 DPH7 XIAP LGMN CD81 PAK2 PRNP SERPINB6 SOX4
grey	1.20E-18	113	1925	2.7134653	Reg. of protein metabolic proc.	TAOK1 OGT STAP1 ZNF598 AKT2 NSD3 RTF1 SPRY2 SWAP70 EIF3D PPP4R2 APOA1 C23H6orf89 VPS35 SYAP1 CCNL1 DPH3 JUN OSBPL8 DUSP6 CCNL2 DIP2B PPP2R5D PIAS1 PTK2B XRCC6 MRPL13 ARRDC3 EMP2 IFI6 NDUFA13 WWTR1 RPUSD3 LRIG2 EIF2AK2 NCBP2 SH2B3 JDP2 RNF20 EIF4E KRAS ITGB3 CCND3 SNX25 PFDN6 CTSH KBTBD8 IGF1 PFN2 FLNA RGP1 LRP6 PRKACB FBLN1 CSNK2A1 SENP1 INSR JARID2

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
						MAP4K4 PPP1R15B NBN PUM3 CHORDC1 CTCF DNAJB2 GNL3 AGO4 TCOF1 MTOR CTR9 TRIB2 MOB1B CDYL CTNNB1 PHACTR2 PKN1 RASSF2 EEF2K RNF19A NUPR1 KMT2A TNRC6C CTSZ SNF8 APC ZNF451 DPH5 PPP2R5E VPS25 PPP2R1B EIF4G2 HECW2 TNRC6B ABCG1 PITHD1 TRMT10C BIRC3 FNIP1 ASPH ENC1 MALT1 SHMT2 PDCD10 MAP2K1 ATG7 DPH7 XIAP LGMN CD81 PAK2 PRNP SERPINB6 SOX4
grey	1.09E-11	71	1212	2.7078979	Apoptotic proc.	SELENOS BCL2A1 TAOK1 PARP1 OGT AKT2 SPRY2 DOCK8 VPS35 ANO6 ELMO1 JUN SKIL DUSP6 NAA15 PTK2B RBM5 IRF3 DYRK2 IFI6 NDUFA13 EIF2AK2 PRDX3 MAGED1 ITGA4 KRAS CDK1 CTSH IKZF3 IGF1 FLNA DNM1L LRP6 ZFAND6 SENP1 MAP4K4 STK17B NBN MED1 PPT1 BCL2L13 MCM2 AGO4 HMOX1 ITGB1 CTNNB1 RGL2 PKN1 ING4 RASSF2 NUPR1 APC DFFB PPP2R1B IDO1 CASP4 NDUFS1 BIRC3 FNIP1 MALT1 PDCD10 CCAR1 ATG7 CCL2 XIAP LGMN PAK2 PRNP SOX4 MAZ

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
grey	1.52E-14	90	1541	2.6997057	Phosphorylation	RANBP2 N4BP2 TAOK1 OGT STAP1 RPS6KC1 AKT2 PPIP5K2 CDK13 SPRY2 APOA1 NME7 SYAP1 CCNL1 JUN MAP3K2 OSBPL8 DGKE VAV2 DUSP6 CCNL2 WNK1 PPP2R5D PRKAB1 PTK2B XRCC6 DCAF1 RPS6KL1 EMP2 DYRK2 CDK19 WWTR1 EIF2AK2 PRDX3 SH2B3 KRAS ITPK1 ITGB3 CCND3 CDK1 SNX25 USP15 MAP2K7 IGF1 PFN2 MAP3K8 LRP6 PRKACB FBLN1 CSNK2A1 INSR MAP4K4 PPP1R15B STK17B NBN CHORDC1 DGKZ MAP3K1 PRKAB2 MTOR TRIB2 RBL2 MOB1B CDC7 SRPK1 PKN1 RASSF2 EFEMP1 EEF2K NUPR1 MKNK2 APC PACSIN2 VPS25 MKNK1 CSK PKN2 FNIP1 MALT1 RSRC1 PDCD10 JMJD8 MAP2K1 CSNK1B SLA2 CDK6 CD81 PAK2 PRNP SMAD5
grey	4.19E-12	75	1293	2.6812623	Protein phosphorylation	TAOK1 STAP1 RPS6KC1 AKT2 CDK13 SPRY2 APOA1 SYAP1 CCNL1 JUN MAP3K2 OSBPL8 DUSP6 CCNL2 WNK1 PPP2R5D PRKAB1 PTK2B XRCC6 DCAF1 RPS6KL1 EMP2 DYRK2 CDK19 WWTR1 EIF2AK2 SH2B3 KRAS ITGB3 CCND3 CDK1 SNX25 USP15 MAP2K7

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
						IGF1 PFN2 MAP3K8 LRP6 PRKACB FBLN1 CSNK2A1 INSR MAP4K4 PPP1R15B STK17B NBN CHORDC1 MAP3K1 PRKAB2 MTOR TRIB2 MOB1B CDC7 SRPK1 PKN1 RASSF2 EFEMP1 EEF2K MKNK2 APC VPS25 MKNK1 CSK PKN2 FNIP1 MALT1 RSRC1 PDCD10 MAP2K1 CSNK1B CDK6 CD81 PAK2 PRNP SMAD5
grey	1.19E-12	79	1368	2.6694239	Cell population proliferation	NDRG2 AKT2 SPRY2 DOCK8 APOA1 C23H6orf89 EAPP JUN PIAS1 PTK2B XRCC6 PRMT1 TNS2 EMP2 HOOK3 WWTR1 MEIS1 SRSF6 EIF2AK2 PRDX3 MAGED1 VASH1 RAP1B SH2B3 ITGA4 KRAS ITGB3 CCND3 KANK1 CTSH IKZF3 IGF1 CNN1 FBLN1 INSR JAG1 JARID2 RPA3 NBN MED1 CTCF DNAJB2 PTGDS TOB2 HMOX1 ITGB1 NDE1 CDC7 CTNNB1 PKN1 ING4 NUPR1 KMT2A NFATC2 APC DAZAP1 FOXJ2 ALOX5 MTSS1 AIF1 GNAI2 DPT CSK SOX5 TNFRSF13C MALT1 VAV3 PDCD10 PDCD1LG2 CCAR1 MAP2K1 CCL2 CDK6 NRAS LGMN CD81 PLA2G2D1 SOX4 MAZ

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
grey	1.59E-12	78	1351	2.6687987	Cell death	SELENOS BCL2A1 TAOK1 PARP1 OGT AKT2 SPRY2 DOCK8 VPS35 ANO6 ELMO1 JUN SKIL DUSP6 NAA15 PTK2B RBM5 IRF3 EMP2 DYRK2 IFI6 NDUFA13 SRSF6 EIF2AK2 PRDX3 MAGED1 ITGA4 KRAS ITPK1 CDK1 LAMP1 CTSH IKZF3 IGF1 FLNA DNM1L PYGL LRP6 ZFAND6 SENP1 MAP4K4 STK17B NBN MED1 PPT1 BCL2L13 MCM2 AGO4 HMOX1 ITGB1 CTNNB1 RGL2 PKN1 ING4 RASSF2 NUPR1 CTSZ APC DFFB PPP2R1B IDO1 CASP4 NDUFS1 BIRC3 FNIP1 MALT1 PDCD10 CCAR1 ATG7 CCL2 XIAP LGMN PAK2 PRNP SOX4 MAZ GPX4
grey	1.59E-15	99	1732	2.6421889	Reg. of catalytic activity	RANBP2 TAOK1 PARP1 STAP1 AKT2 SPRY2 PPP4R2 DOCK8 APOA1 C23H6orf89 IQSEC1 SYAP1 ELMO1 CCNL1 ARHGAP24 JUN ARHGAP29 OSBPL8 VAV2 CCNL2 WNK1 PPP2R5D PRKAB1 PTK2B XRCC6 GIT2 FGD5 ARRDC3 EMP2 RAP1GDS1 IFI6 NDUFA13 WWTR1 PRDX3 SH2B3 ITGB3 CCND3 GRPEL1 GRTP1 RALGPS2 CTSH IGF1 DNM1L RGP1

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
						LRP6 FBLN1 CSNK2A1 SENP1 RANGAP1 INSR MAP4K4 PPP1R15B NBN PPT1 CHORDC1 DGKZ PKP4 DNAJB2 MCM2 PRKAB2 PSMD6 MTOR RASAL2 ITGB1 RBL2 MOB1B CTNNB1 PHACTR2 RGL2 PKN1 DENND4A RASSF2 APC PSD4 RGL1 PPP2R5E VPS25 PPP2R1B RALBP1 RGS1 BIRC3 FNIP1 ASPH MALT1 VAV3 PDCD10 JMJD8 MAP2K1 TOR1AIP2 VAV1 XIAP ARHGEF18 SLA2 LGMN CD81 PAK2 PRNP SERPINB6
grey	1.39E-13	86	1509	2.6344246	Immune System	CD68 ASB8 APRT ANAPC4 PILRA AP1B1 AKT2 FYB1 CDK13 WIPF1 ANO6 DNAJC13 ELMO1 UBE2Q2 JUN UBE2D2 VAV2 DUSP6 PPP2R5D PIAS1 CEACAM1 SNAP23 PTK2B UBR2 XRCC6 DCAF1 COPB1 IRF3 ARIH1 RAB5C CD55 CUL2 NCF4 LAMTOR3 FBXW7 MKRN1 EIF2AK2 RAP1B HVCN1 ITGA4 EIF4E AP2A2 LAMP1 MAP2K7 CTSH CTSF KBTBD8 PYGL MAP3K8 PRKACB S100A12 DNM2 MAP3K1 SRP14 FBXO40 ATP6V0E1 RASAL2 BOLA-DMA ITGB1 CTNNB1 RNF19A CD36 NFATC2 PSMD7 EIF4E3

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
						TAX1BP1 PPP2R5E PPP2R1B ALOX5 HECW2 CASP4 CSK YPEL5 BIRC3 TNFRSF13C MALT1 ANAPC1 VAV3 PDCD1LG2 MAP2K1 ATG7 EIF4G3 NRAS PAK2 UBE2W
grey	5.75E-14	97	1798	2.4937827	Neg. reg. of nitrogen compound metabolic proc.	CREBZF RFX3 PARP1 OGT HIVEP1 RTF1 SPRY2 SWAP70 EXOSC10 ZNF436 RBBP7 TAF3 EAPP PSPC1 JUN SKIL DUSP6 PIAS1 XRCC6 DRAP1 TFDP2 NAB2 IRF3 MRPL13 RCOR3 IFI6 NDUFA13 WWTR1 LRIG2 SRSF6 EIF2AK2 SH2B3 GLIS2 JDP2 RNF20 EIF4E LARP7 ZBTB25 CCND3 SNX25 PFDN6 NKAP IGF1 FLNA RGP1 LRP6 FBLN1 JARID2 PPP1R15B NBN MED1 CHORDC1 CTCF ZFHX3 TOB2 AGO4 SMARCE1 SCAI CTR9 CC2D1B CDYL CTNNB1 PHACTR2 PKN1 ING4 MXI1 RUNX1T1 RASSF2 NUPR1 NFATC2 WWP2 VGLL4 TNRC6C SPEN CTSZ APC ZNF451 VPS25 TNRC6B SATB1 ID2 TP53BP1 THAP5 CENPF POU2F1 BIRC3 FNIP1 ENC1 XIAP SLA2 MLX PAK2 C1D SERPINB6 SOX4 MAZ SMAD5

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
grey	2.71E-14	103	1948	2.444133	Neg. reg. of cellular metabolic proc.	CREBZF RFX3 PARP1 OGT STAP1 HIVEP1 RTF1 SPRY2 SWAP70 EXOSC10 ZNF436 RBBP7 TAF3 EAPP PSPC1 JUN SKIL DUSP6 WNK1 PIAS1 XRCC6 DRAP1 TFDP2 NAB2 IRF3 MRPL13 RCOR3 IFI6 NDUFA13 WWTR1 LRIG2 SRSF6 EIF2AK2 PRDX3 SH2B3 GLIS2 JDP2 RNF20 EIF4E LARP7 ZBTB25 CCND3 SNX25 NKAP IGF1 FLNA RGP1 LRP6 FBLN1 JARID2 PPP1R15B NBN MED1 CHORDC1 CTCF ZFHX3 TOB2 AGO4 SMARCE1 MTOR SCAI CTR9 HMOX1 CC2D1B CDYL CTNNB1 PHACTR2 PKN1 ING4 MXI1 RUNX1T1 RASSF2 NUPR1 NFATC2 WWP2 VGLL4 TNRC6C SPEN CTSZ APC ZNF451 VPS25 EIF4G2 TNRC6B SATB1 ID2 TP53BP1 THAP5 CENPF POU2F1 BIRC3 FNIP1 ENC1 ATG7 XIAP SLA2 MLX PAK2 C1D SERPINB6 SOX4 MAZ SMAD5
grey	5.75E-14	102	1947	2.4216467	Intracellular signal transduction	ASB8 ERBIN SELENOS TAOK1 NDRG2 OGT AKT2 RFXANK SPRY2 DOCK8 APOA1 IQSEC1 SYAP1 ELMO1 ADGRL1 ASH1L ARHGAP24 JUN SKIL ARHGAP29 OSBPL8 DGKE VAV2

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
						DUSP6 WNK1 LITAF PTK2B UBR2 CDK5RAP2 TNS2 DYRK2 IFI6 NDUFA13 WWTR1 LAMTOR3 EIF2AK2 RAP1B SH2B3 TMEM100 PPM1D KRAS USP15 RALGPS2 KANK1 CTSH IGF1 DNM1L MAP3K8 FBLN1 ZFAND6 S100A12 INSR DICER1 MAP4K4 STK17B NBN NFAT5 DGKZ MTOR SCAI ITGB1 MOB1B CTNNB1 RGL2 SRPK1 USO1 ING4 DENND4A WDR76 RNF135 RASSF2 NUPR1 MKNK2 CASTOR1 NFATC2 APC PSD4 RGL1 AIF1 ADGRG6 RALBP1 TP53BP1 PKN2 FNIP1 MALT1 VAV3 PDCD10 RHEB JMJD8 MAP2K1 CCL2 VAV1 ARHGEF18 SLA2 NRAS PAK2 PRNP SOX4 MAZ CDC42EP4
grey	1.09E-11	96	1949	2.276858	Nucleotide binding	N4BP2 ATP10D APRT TAOK1 PARP1 ACAD10 RPS6KC1 AKT2 PPIP5K2 TUBA1A CDK13 MDN1 SWAP70 ARL5A EXOSC10 NME7 RAB28 ATL3 MBD1 PAPOLA MAP3K2 DGKE RUNX1 HSPH1 WNK1 SMC4 PTK2B SUCLG1 XRCC6 PRMT1 RPS6KL1 ENTPD4 RAB5C DYRK2 CDK19 POLA1 EIF2AK2 RAP1B KRAS ITPK1 GIMAP7 CDK1

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
						GRPEL1 MFN1 MAP2K7 RAB8B DNM1L HSPA9 PYGL MAP3K8 PRKACB CSNK2A1 INSR DICER1 GMPS MAP4K4 STK17B DNM2 CHORDC1 DGKZ MAP3K1 YTHDC2 MCM2 GNL3 DHX57 MTOR DHX30 TRIB2 CDC7 MLH1 DNAH7 SRPK1 PKN1 EEF2K MKNK2 ACSL4 RAD51D TUFM MKNK1 GNAI2 ABCG1 CSK NDUFV1 DDX23 PKN2 CBR4 RHEB MAP2K1 CSNK1B CDK6 NRAS PAK2 ARL16 RAB43 TWF2
grey	1.09E-11	96	1949	2.276858	Nucleoside phosphate binding	N4BP2 ATP10D APRT TAOK1 PARP1 ACAD10 RPS6KC1 AKT2 PPIP5K2 TUBA1A CDK13 MDN1 SWAP70 ARL5A EXOSC10 NME7 RAB28 ATL3 MBD1 PAPOLA MAP3K2 DGKE RUNX1 HSPH1 WNK1 SMC4 PTK2B SUCLG1 XRCC6 PRMT1 RPS6KL1 ENTPD4 RAB5C DYRK2 CDK19 POLA1 EIF2AK2 RAP1B KRAS ITPK1 GIMAP7 CDK1 GRPEL1 MFN1 MAP2K7 RAB8B DNM1L HSPA9 PYGL MAP3K8 PRKACB CSNK2A1 INSR DICER1 GMPS MAP4K4 STK17B DNM2 CHORDC1 DGKZ MAP3K1 YTHDC2

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
						MCM2 GNL3 DHX57 MTOR DHX30 TRIB2 CDC7 MLH1 DNAH7 SRPK1 PKN1 EEF2K MKNK2 ACSL4 RAD51D TUFM MKNK1 GNAI2 ABCG1 CSK NDUFV1 DDX23 PKN2 CBR4 RHEB MAP2K1 CSNK1B CDK6 NRAS PAK2 ARL16 RAB43 TWF2
magenta	8.11E-07	14	392	8.2417582	Neg. reg. of protein modification proc.	TNFAIP3 PHF1 PABIR2 CBL GGNBP2 ARPP19 USP4 CBLB SORL1 TSPO MYADM VPS28 PID1 NCOA7
magenta	5.02E-10	28	1171	5.5179662	Reg. of protein modification proc.	MMD TNFAIP3 TP53 PHF1 SASH1 PABIR2 FBH1 SYK TSC1 CBL SMAD4 GGNBP2 ARPP19 USP4 CBLB SORL1 ARFGEF1 ITGA1 PHB TSPO MYADM KAT7 NOP53 PTTG1IP VPS28 TFRC PID1 NCOA7
magenta	1.52E-06	18	756	5.4945055	Neg. reg. of cellular protein metabolic proc.	TNFAIP3 TP53 PHF1 PABIR2 TSC1 CBL GGNBP2 ARPP19 USP4 CBLB SORL1 RPS6KA3 TSPO MYADM NOP53 VPS28 PID1 NCOA7
magenta	8.11E-07	19	799	5.4876288	Neg. reg. of protein metabolic proc.	TNFAIP3 TP53 PHF1 PABIR2 TSC1 CBL GGNBP2 ARPP19 USP4 CBLB SORL1 PHB RPS6KA3 TSPO MYADM NOP53 VPS28 PID1 NCOA7
magenta	6.38E-07	21	964	5.0271305	Neg. reg. of signaling	TNFAIP3 TP53 SESN1 AP1AR TSC1 CYLD CBL SMAD4 GGNBP2 CD22

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
						CBLB SORL1 ITGA1 PHB MYADM UBR1 NOP53 PTTG1IP NOTCH1 PID1 TRAT1
magenta	2.60E-08	25	1154	4.9993334	Reg. of intracellular signal transduction	TNFAIP3 TP53 SASH1 SESN1 FBH1 SYK TSC1 C1QTNF1 CYLD CBL SMAD4 CYTH3 S100A8 SORL1 ARFGEF1 LAMTOR4 PHB MYADM UBR1 NOP53 PTTG1IP NOTCH1 TFRC TRAT1
magenta	3.57E-06	19	908	4.8288716	Neg. reg. of signal transduction	TNFAIP3 TP53 SESN1 TSC1 CYLD CBL SMAD4 GGNBP2 CD22 CBLB SORL1 ITGA1 PHB MYADM UBR1 NOP53 PTTG1IP NOTCH1 PID1
magenta	6.38E-07	23	1159	4.5795447	Neg. reg. of response to stimulus	TNFAIP3 TP53 SESN1 FBH1 TSC1 C1QTNF1 CYLD CBL SMAD4 GGNBP2 CD22 CBLB SORL1 ANXA1 ITGA1 PHB MYADM UBR1 NOP53 PTTG1IP NOTCH1 PID1 NCOA7
magenta	7.80E-06	19	961	4.562555	Neg. reg. of cell communication	TNFAIP3 TP53 SESN1 TSC1 CYLD CBL SMAD4 GGNBP2 CD22 CBLB SORL1 ITGA1 PHB MYADM UBR1 NOP53 PTTG1IP NOTCH1 PID1
magenta	3.31E-06	21	1106	4.3816943	Reg. of phosphate metabolic proc.	MMD TNFAIP3 TP53 SASH1 PABIR2 FBH1 SYK TSC1 PIK3R5 CBL SMAD4 GGNBP2 ARPP19 CBLB SORL1 ITGA1 PHB MYADM NOP53 TFRC PID1

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
magenta	1.28E-06	23	1242	4.2735043	Cytoplasmic vesicle	ATP11B AP1AR SYK TRAPPC8 AP3D1 MON2 CD22 ZFYVE9 ANKRD13A SORL1 LAMTOR4 ANXA1 ARHGAP1 CDK16 PHB YIPF3 SLC30A7 SYPL1 NOTCH1 PIK3C2B VPS28 TFRC
magenta	1.28E-06	23	1245	4.2632067	Intracellular vesicle	ATP11B AP1AR SYK TRAPPC8 AP3D1 MON2 CD22 ZFYVE9 ANKRD13A SORL1 LAMTOR4 ANXA1 ARHGAP1 CDK16 PHB YIPF3 SLC30A7 SYPL1 NOTCH1 PIK3C2B VPS28 TFRC
magenta	5.48E-10	35	1947	4.148394	Intracellular signal transduction	TNFAIP3 TP53 SASH1 SESN1 FAT4 FBH1 SYK WSB1 TSC1 ADCY7 C1QTNF1 CYLD CBL SMAD4 CYTH3 HIPK1 S100A8 CBLB SORL1 ARFGEF1 LAMTOR4 BRK1 ARHGAP1 PHB RPS6KA3 MYADM KAT7 UBR1 NOP53 PTTG1IP NOTCH1 PIK3C2B TFRC TRAT1
magenta	1.28E-06	24	1349	4.1056053	Vesicle	ATP11B AP1AR SYK TRAPPC8 AP3D1 MON2 CD22 ZFYVE9 ANKRD13A SORL1 LAMTOR4 ANXA1 ARHGAP1 CDK16 PHB YIPF3 SLC30A7 SYPL1 NOTCH1 PIK3C2B VPS28 TFRC SPRN
magenta	2.60E-08	31	1800	3.974359	Reg. of cellular protein metabolic proc.	MMD TNFAIP3 TP53 PHF1 SASH1 PABIR2 FBH1 SYK TSC1 CBL SMAD4 GGNBP2 ARPP19 USP4 S100A8 CBLB

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
						SORL1 ARFGEF1 ITGA1 PHB RPS6KA3 TSPO MYADM KAT7 HBS1L NOP53 PTTG1IP VPS28 TFRC PID1 NCOA7
magenta	4.87E-06	24	1480	3.7422037	Pos. reg. of response to stimulus	TNFAIP3 TP53 SASH1 FBH1 SYK CRTAM C1QTNF1 CYLD CBL SMAD4 CTSC CD22 S100A8 CD247 CBLB SORL1 ARFGEF1 LAMTOR4 ANXA1 PHB NOTCH1 TFRC TRAT1
magenta	1.44E-05	25	1706	3.3817296	Organic substance transport	ATP11B TP53 SCP2 SYK TSC1 C1QTNF1 AP3D1 ESYT1 RAE1 MON2 S100A8 SORL1 ARFGEF1 SLCO2B1 ANXA1 ARHGAP1 CDK16 TSPO SLC4A7 PTTG1IP VPS28 TFRC THOC7 PID1 SPRN
magenta	1.29E-05	27	1948	3.1985468	Neg. reg. of cellular metabolic proc.	TNFAIP3 TP53 PHF1 PABIR2 AP1AR FBH1 TSC1 CBL SMAD4 GGNBP2 ARPP19 USP4 CBLB ANKRD13A SORL1 PHB RPS6KA3 PDS5A TSPO MYADM KAT7 NOP53 NOTCH1 VPS28 PID1 NCOA7 TRAT1
purple	0.0006233	13	547	5.2813325	MRNA metabolic proc.	SUPV3L1 SCAF11 SECISBP2 SYNCRIP WTAP SF3B1 DDX1 SLTM FMR1 TRA2A HIPK3 IK ZBTB7A
purple	0.0007209	13	561	5.1495346	Reg. of cellular catabolic proc.	SUPV3L1 SECISBP2 SPTLC1 ATM MYLIP ATF6 KAT5 SYNCRIP FMR1 XPO1 RAB3GAP1 GABARAP COMMD1

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
purple	0.0013541	12	533	5.003127	Nucleoside-triphosphatase activity	GUF1 PIKFYVE CHD6 ATP2B1 RHOT1 ABCB7 BMS1 RAB14 RAB21 ABCE1 ENTPD7
purple	0.000296	15	675	4.9382716	Reg. of catabolic proc.	SUPV3L1 SECISBP2 SPTLC1 ATM MYLIP ATF6 KAT5 SYNCRIP FMR1 XPO1 RAB3GAP1 GABARAP SNX3 COMMD1 ZBTB7A
purple	0.0014085	15	847	3.9354585	Reg. of organelle organization	CD2AP PIKFYVE ATM CYFIP2 PIK3CA RASA1 PAK1 SSH2 CHMP5 RAB3GAP1 IK NCKAP1L SLAIN2 RNF4 YAP1
purple	0.0003701	19	1113	3.793551	Intracellular transport	CLTA PIKFYVE DCTN1 GOSR1 RHOT1 CHMP5 SNX6 NAPA XPO1 HEATR5B NFKBIA RAB14 SNX3 RAB21 ABCE1 CCDC186 DUS4L ITSN1 TIMM10
purple	0.000296	21	1297	3.5980468	Cellular protein localization	CD2AP CLTA PIKFYVE EPB41L2 DDX1 SNX6 NAPA XPO1 RAB3GAP1 IK NFKBIA RAB14 RAB21 CCDC66 COMMD1 ZBTB7A DUS4L ITSN1 TIMM10 IFT80 YAP1
purple	0.000296	21	1303	3.5814786	Cellular macromolecule localization	CD2AP CLTA PIKFYVE EPB41L2 DDX1 SNX6 NAPA XPO1 RAB3GAP1 IK NFKBIA RAB14 RAB21 CCDC66 COMMD1 ZBTB7A DUS4L ITSN1 TIMM10 IFT80 YAP1
purple	0.0013541	17	1062	3.5572295	Pos. reg. of protein metabolic proc.	GUF1 HES1 MYLIP KAT5 CYFIP2 PIK3CA PAK1 FMR1 RAB3GAP1

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
						CCNT2 NGRN GABARAP NFKBIA NCKAP1L COMMD1 PIK3CG CAB39
purple	0.0013541	18	1169	3.421728	Cytoskeleton organization	CD2AP DCTN1 EPB41L2 CYFIP2 STAG2 PIK3CA RASA1 PAK1 SSH2 CDC14B CHMP5 XPO1 GABARAP NCKAP1L CCDC66 SLAIN2 CDC42BPA RNF4
purple	0.000296	25	1743	3.1873526	Protein localization	CD2AP CLTA PIKFYVE EPB41L2 DDX1 SNX6 NAPA XPO1 RAB3GAP1 HEATR5B IK NFKBIA KTN1 RAB14 SNX3 RAB21 CCDC66 COMMD1 ZBTB7A CCDC186 DUS4L ITSN1 TIMM10 IFT80 YAP1
purple	0.000296	26	1883	3.0683897	Catabolic proc.	SUPV3L1 MGAT1 SECISBP2 PIKFYVE SPTLC1 ATM MYLIP IVD RSPRY1 ATF6 KAT5 SYNCRIP FMR1 XPO1 RAB3GAP1 GABARAP CTSS SNX3 COMMD1 ZBTB7A ANXA7 ITSN1 RNF4 PLA2G2D4 ENTPD7 UBE2D3
purple	0.00102	22	1604	3.0479357	Cellular catabolic proc.	SUPV3L1 MGAT1 SECISBP2 PIKFYVE SPTLC1 ATM MYLIP IVD RSPRY1 ATF6 KAT5 SYNCRIP FMR1 XPO1 RAB3GAP1 GABARAP CTSS COMMD1 ANXA7 RNF4 ENTPD7 UBE2D3
purple	0.0013541	22	1651	2.9611683	Purine ribonucleoside triphosphate binding	EIF4A1 GUF1 PIKFYVE ATM CHD6 ATP2B1 DDX1 RHOT1 PAK1 ABCB7

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
						BMS1 HIPK3 SMARCAD1 RAB14 RAB21 ABCE1 SIK2 PIK3CG CDC42BPA ENTPD7 UBE2D3
purple	0.0014933	21	1583	2.9479891	Organic substance catabolic proc.	SUPV3L1 MGAT1 SECISBP2 ATM MYLIP IVD RSPRY1 KAT5 SYNCRIP FMR1 XPO1 GABARAP CTSS SNX3 COMMD1 ZBTB7A ITSN1 RNF4 PLA2G2D4 ENTPD7 UBE2D3
purple	0.0014587	22	1702	2.8724377	Purine ribonucleotide binding	EIF4A1 GUF1 PIKFYVE ATM CHD6 ATP2B1 DDX1 RHOT1 PAK1 ABCB7 BMS1 HIPK3 SMARCAD1 RAB14 RAB21 ABCE1 SIK2 PIK3CG CDC42BPA ENTPD7 UBE2D3
purple	0.0014933	22	1716	2.8490028	Ribonucleotide binding	EIF4A1 GUF1 PIKFYVE ATM CHD6 ATP2B1 DDX1 RHOT1 PAK1 ABCB7 BMS1 HIPK3 SMARCAD1 RAB14 RAB21 ABCE1 SIK2 PIK3CG CDC42BPA ENTPD7 UBE2D3
purple	0.0008134	25	1962	2.8315778	Carbohydrate derivative binding	EIF4A1 GUF1 PIKFYVE ATM AAMP CHD6 SULF2 ATP2B1 DDX1 RHOT1 PAK1 ABCB7 BMS1 HIPK3 SMARCAD1 CTSS RAB14 RAB21 ABCE1 SIK2 PIK3CG CDC42BPA ENTPD7 UBE2D3
purple	0.0013541	24	1949	2.736446	Nucleotide binding	EIF4A1 GUF1 PIKFYVE ATM IVD CHD6 ATP2B1 DDX1 RHOT1 PAK1 ABCB7 BMS1 HIPK3 SMARCAD1

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
						RAB14 RAB21 ABCE1 SIK2 PIK3CG DUS4L CDC42BPA ENTPD7 UBE2D3
purple	0.0013541	24	1949	2.736446	Nucleoside phosphate binding	EIF4A1 GUF1 PIKFYVE ATM IVD CHD6 ATP2B1 DDX1 RHOT1 PAK1 ABCB7 BMS1 HIPK3 SMARCAD1 RAB14 RAB21 ABCE1 SIK2 PIK3CG DUS4L CDC42BPA ENTPD7 UBE2D3
red	0.0054423	4	27	27.952481	Neg. reg. of stress-activated MAPK cascade	FKTN GPS2 NCOR1 PTPN22
red	0.0076563	4	33	22.870212	Cytochrome complex assembly	SMIM20 FASTKD3 UQCC3 COA6
red	0.0076563	6	109	10.386014	Neg. reg. of MAPK cascade	P2RX7 FKTN GPS2 NCOR1 PSMD10 PTPN22
red	0.0006116	9	168	10.107817	Bta-mir-2366 target gene	C1S CLIC2 CPM PDK4 C3H1orf109 COTL1 COMMD7 RCAN3 PDHA1
red	0.0006116	10	217	8.6948961	Cellular carbohydrate metabolic proc.	SLC35B4 PFKFB3 NCOR1 PRKAG3 PDK4 MTMR2 PTGES3 DDB1 B3GALT5 MECP2
red	0.008321	7	179	7.378518	Cilium Assembly	CEP162 CEP89 NPHP3 PRKACA CNTRL HAUS6 TUBA3E
red	0.0023057	12	413	5.4822057	Bta-mir-2443 target gene	BET1L EIF5A LTBP3 ABR SLC22A1 SLC43A2 PRKAG3 COTL1 FXYD1 MAN2A2 DDB1 EMC10

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
red	0.0015619	13	460	5.3322395	Carbohydrate metabolic proc.	SLC35B4 PFKFB3 GANC P2RX7 NCOR1 PRKAG3 PDK4 MTMR2 PTGES3 MAN2A2 DDB1 B3GALT5 MECP2
red	0.0054423	11	392	5.2945707	Mitochondrion organization	ATPAF1 SMIM20 NDUFA8 CEP89 PNPT1 P2RX7 FASTKD3 UQCC3 PSMD10 LRRK2 NUBPL
red	0.0076563	15	791	3.5779882	RNA Polymerase II Transcription	ZNF570 SP1 MDM4 RICTOR SNW1 ALYREF NR1H3 SRSF9 NCOR1 PRKAG3 NPM1 ELF2 KMT2C ZNF770 MECP2
red	0.0076563	15	818	3.4598884	Carbohydrate derivative metabolic proc.	PFKFB3 GANC P2RX7 FKTN MGAT4A NCOR1 PRKAG3 DPY19L4 PDK4 LRRK2 MAN2A2 PDHA1 B3GALT5 PFAS ACOT11
red	0.0076563	16	892	3.3843811	Pos. reg. of catalytic activity	LMO4 RICTOR ABR P2RX7 NR1H3 CFLAR IQGAP1 PSMD10 LRRK2 GPR55 CCDC88A ARHGAP30 PTGES3 AHSA1 WRN AGTR1
red	0.008321	15	844	3.3533041	Gene expression Transcription	ZNF570 SP1 MDM4 RICTOR SNW1 ALYREF NR1H3 SRSF9 NCOR1 PRKAG3 NPM1 ELF2 KMT2C ZNF770 MECP2
red	0.0029865	20	1161	3.2502885	Pos. reg. of molecular function	LMO4 MAPRE1 NPHP3 RICTOR ABR P2RX7 NR1H3 CLIC2 CFLAR IQGAP1 MAVS PSMD10 LRRK2 GPR55

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
						CCDC88A ARHGAP30 PTGES3 AHSA1 WRN AGTR1
red	0.0076563	18	1117	3.0404892	Reg. of transport	BEST3 TTC39B PRKACA GRIPAP1 P2RX7 GPS2 NR1H3 CLIC2 SLC43A2 KCNE3 MAVS CALR LRRK2 FXYD1 PTPN22 STON2 DLL1 IPO5
red	0.0076563	20	1348	2.7993953	Small molecule metabolic proc.	HAO2 SLC35B4 TTC39B FAAH PFKFB3 P2RX7 NR1H3 NCOR1 PRKAG3 PDK4 LRRK2 MTMR2 PTGES3 MAN2A2 PDHA1 DDB1 PFAS SLC2A9 ACOT11 MECP2
red	0.0079302	23	1706	2.5437413	Organic substance transport	BET1L SEC24C SLC35B4 TTC39B PRKACA GRIPAP1 SLC22A1 PNPT1 P2RX7 GPS2 NR1H3 SLC43A2 MAVS AP4E1 CALR NPM1 LRRK2 PTPN22 ZDHHC17 SLC2A9 IPO5 TMED2
red	0.0054423	26	1946	2.5208943	Reg. of localization	LMO4 DNAJB6 BEST3 SP1 TTC39B ADGRG3 PRKACA GRIPAP1 P2RX7 GPS2 NR1H3 CLIC2 SLC43A2 KCNE3 MAVS CALR LRRK2 CCDC88A FXYD1 PTPN22 STON2 DLL1 IPO5 CTNNA3 XG MECP2
red	0.0076563	25	1925	2.4503798	Reg. of protein metabolic proc.	DNAJB6 EIF5A SP1 PRKACA RICTOR SNW1 PNPT1 P2RX7 FASTKD3 GPS2 NR1H3 CFLAR IQGAP1 PRKAG3 MAVS

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
						PSMD10 DUS2 CALR LRRK2 CCDC88A TIA1 PTPN22 DDB1 AGTR1 MECP2
turqiouse	5.24E-10	24	221	6.265228	Reg. of mRNA metabolic proc.	CELF1 CNOT2 RBM39 MBNL1 EXOSC9 FUS HNRNPU SON DDX17 PAN3 HSPA8 HNRNPD PNLDC1 YTHDC1 NCL YBX1 DDX5 PCBP4 RC3H2 THRAP3 ZFP36L1 MAPKAPK2 PABPC1 CARHSP1
turqiouse	4.52E-10	25	238	6.0601164	MRNA binding	LUC7L3 IREB2 CELF1 EXOSC9 HNRNPA2B1 FUS HNRNPU FUBP1 PCBP1 DDX6 YTHDC1 LUC7L ZC3H11A YBX1 DDX5 PCBP4 PARK7 RC3H2 SNRPC ZFP36L1 UTP23 PABPC1 CARHSP1
turqiouse	5.20E-10	28	306	5.2790347	Nuclear speck	PPP4R3B PRPF4B LUC7L3 RBM39 KAT6A SGK1 MOCS2 RREB1 SREK1 HNRNPU BCLAF1 CWC15 SLU7 SON PCBP1 RBBP6 DDX17 SRRM1 YTHDC1 SFPQ WBP4 MEF2C FNBP4 DDX46 THRAP3 MAML2 NRIP1 PRPF40A
turqiouse	2.06E-10	30	336	5.1510989	RNA splicing	PRPF4B LUC7L3 LGALS3 CELF1 RBM39 MBNL1 HNRNPA2B1 FUS SREK1 HNRNPU CWC15 GCFC2 SLU7 SON DDX17 PIK3R1 HSPA8 RBMX2 YTHDC1 NCL SFPQ LUC7L WBP4

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
						DDX5 TMBIM6 CLK4 METTL4 THRAP3 SNRPC PRPF40A
turqiouse	5.20E-15	46	547	4.8516383	MRNA metabolic proc.	UPF2 CSDE1 PRPF4B LUC7L3 CELF1 CNOT2 RBM39 MBNL1 EXOSC9 HNRNPA2B1 FUS HNRNPU CWC15 GCFC2 SLU7 SON RNMT RBBP6 DDX17 PAN3 AGO1 DKC1 HSPA8 SRRM1 HNRNPD RBMX2 PNLDC1 YTHDC1 NCL SFPQ LUC7L WBP4 YBX1 ATF4 AGO3 DDX5 PCBP4 STAT3 RC3H2 THRAP3 SNRPC ZFP36L1 MAPKAPK2 PABPC1 PRPF40A CARHSP1
turqiouse	9.30E-11	37	495	4.3123543	Metabolism of RNA	UPF2 RPS12 RPL5 LAS1L UTP11 CNOT2 RPL4 HNRNPA2B1 FUS HNRNPU CWC15 SLU7 PRKCD PCBP1 DDX6 RNMT U2SURP PPWD1 PAN3 RPS19 HSPA8 RPS24 SRRM1 HNRNPD RPL22 DDX21 RPS27A NCL WBP4 ZC3H11A DDX5 DDX46 SNRPC ZFP36L1 UBC PABPC1 PRPF40A
turqiouse	3.85E-12	46	675	3.9316239	Reg. of catabolic proc.	ITCH CNOT2 RARRES2 EXOSC9 PSMC1 ARIH2 FUS HUWE1 DAP HNRNPU AZIN1 PRKCD PAN3 PDCL3 EZR EIF6 TMEM259 RASIP1 EGFR DKC1 VDAC1 HNRNPD VCP PNLDC1

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
						OAZ2 YBX1 WDR91 EGLN2 CDKN1B OAZ1 PCBP4 FOXK1 NUDT15 PARK7 BCAP31 PSME1 STAT3 RC3H2 THRAP3 ZFP36L1 COP1 ADRA1A MAPKAPK2 VPS13C PABPC1 CARHSP1
turqiouse	1.96E-09	37	561	3.8050185	Reg. of cellular catabolic proc.	ITCH CNOT2 EXOSC9 PSMC1 ARIH2 FUS DAP HNRNPU PRKCD PAN3 PDCL3 EZR TMEM259 RASIP1 DKC1 VDAC1 HNRNPD VCP PNLDC1 YBX1 WDR91 PCBP4 FOXK1 NUDT15 PARK7 BCAP31 PSME1 STAT3 RC3H2 THRAP3 ZFP36L1 COP1 ADRA1A MAPKAPK2 VPS13C PABPC1 CARHSP1
turqiouse	1.15E-11	49	784	3.6057692	RNA processing	PRPF4B LAS1L LUC7L3 UTP11 LGALS3 CELF1 BMP4 RBM39 MBNL1 EXOSC9 HNRNPA2B1 FUS SREK1 HNRNPU CWC15 GCFC2 SLU7 SON RNMT RBBP6 DDX17 PAN3 PIK3R1 EGFR AGO1 DKC1 HSPA8 RPS24 SRRM1 RBMX2 YTHDC1 NCL SFPQ LUC7L WBP4 RRP1B AGO3 DDX5 TMBIM6 CLK4 ZNHIT6 STAT3 METTL4 THRAP3 SNRPC ZFP36L1 FTSJ1 UTP23 PRPF40A
turqiouse	2.47E-11	48	773	3.582446	Ribonucleoprotein complex	UPF2 EIF3J RPS12 RPL5 PRPF4B LAS1L LUC7L3 UTP11 LGALS3 CELF1

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
						RPL4 HNRNPA2B1 HNRNPU CWC15 GCFC2 SLU7 DDX6 PPWD1 DDX17 RPS19 AGO1 DKC1 HSPA8 RPS24 SRRM1 HNRNPD RBMX2 RPS27A NCL LUC7L WBP4 YBX1 RRP1B RBMS2 AGO3 DDX5 RNPC3 FAU ZNHIT6 SNRPC ZFP36L1 ACTB UTP23 PABPC1 PRPF40A
turqiouse	5.20E-15	64	1037	3.5605667	Macromolecule catabolic proc.	CHIA UPF2 ITCH HYAL1 CSDE1 UBE2H CNOT2 UFD1 EXOSC9 PSMA7 PSMC1 ARIH2 FUS HUWE1 OS9 STT3B HNRNPU AZIN1 TTC3 RBBP6 PAN3 PDCL3 EZR VPS4B TMEM259 LYPLA2 EGFR AGO1 CTSB DKC1 HSPA8 SPOPL HERC2 HNRNPD PSMA4 VCP PNLDC1 UBR4 RPS27A CASP8 OAZ2 YBX1 WDR91 EGLN2 PSMB10 CDKN1B AGO3 DDX5 OAZ1 PCBP4 EDEM3 NUDT15 ARRB1 PARK7 BCAP31 PSME1 RC3H2 THRAP3 ZFP36L1 GET4 COP1 MAPKAPK2 PABPC1 CARHSP1
turqiouse	1.47E-12	54	882	3.5321821	Cellular macromolecule catabolic proc.	UPF2 ITCH CSDE1 UBE2H CNOT2 UFD1 EXOSC9 PSMA7 PSMC1 ARIH2 FUS HUWE1 OS9 STT3B HNRNPU TTC3 RBBP6 PAN3 PDCL3 EZR VPS4B TMEM259 AGO1 CTSB DKC1 SPOPL

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
						HERC2 HNRNPD PSMA4 VCP PNLDC1 UBR4 RPS27A CASP8 YBX1 WDR91 PSMB10 AGO3 DDX5 PCBP4 EDEM3 NUDT15 ARRB1 PARK7 BCAP31 PSME1 RC3H2 THRAP3 ZFP36L1 GET4 COP1 MAPKAPK2 PABPC1 CARHSP1
turqiouse	1.41E-14	65	1097	3.4184139	RNA binding	KHDC4 UPF2 EIF3J CSDE1 RPL5 LUC7L3 IREB2 CELF1 RBM39 MBNL1 ANKRD17 EXOSC9 RPL4 HNRNPA2B1 FUS MATR3 SREK1 HNRNPU SLU7 SON FUBP1 PCBP1 DDX6 RNMT U2SURP DDX17 PAN3 EIF6 EEF1G RPS19 AGO1 DKC1 HNRNPD RPL22 DDX21 RBMX2 HNRNPDL PNLDC1 YTHDC1 NCL SFPQ LUC7L WBP4 CLTC ZC3H11A YBX1 EIF1AX EWSR1 RBMS2 AGO3 DDX5 PCBP4 RNPC3 PARK7 YY1 STAT3 RC3H2 SNRPC ZFP36L1 UTP23 PABPC1 PRPF40A CARHSP1
turqiouse	5.24E-10	56	1113	2.9027576	Intracellular transport	KIF1C SNX1 PTPN11 KIF5B TXN BMP4 UFD1 TMED10 HNRNPA2B1 PMPCB OS9 PCM1 SEC23IP HNRNPU TIMM22 SLU7 CORO1A PRKCD AFTPH EZR VPS4B PIK3R1 EIF6 RAMP1 TNPO2 GDI1 AP5M1 HSPA8 EXOC4

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
						RAB5B SEC61G VCP YTHDC1 OAZ2 GOSR2 SFPQ CLTC ZC3H11A OSBPL2 RAB13 WDR91 CNIH4 USP6NL CMTM6 AP1S2 PARK7 BCAP31 ATP5PD ANP32B STAT3 TBC1D5 REPS2 VPS13C SEC62 ARF6
turqiouse	1.24E-12	76	1583	2.7698139	Organic substance catabolic proc.	CHIA UPF2 ITCH HYAL1 CSDE1 PGK1 UBE2H NOS1 CNOT2 RARRES2 UFD1 EXOSC9 PSMA7 PSMC1 ARIH2 FUS HUWE1 HSD17B11 OS9 STT3B HNRNPU AZIN1 TTC3 PRKCD RBBP6 PAN3 PDCL3 EZR VPS4B EIF6 TMEM259 LYPLA2 EGFR IMPA1 AGO1 CTSB DKC1 HSPA8 SPOPL HERC2 HNRNPD PSMA4 CYP19A1 VCP PNLDC1 UBR4 RPS27A CASP8 OAZ2 BCKDHA YBX1 WDR91 EGLN2 PSMB10 CDKN1B PAOX AGO3 DDX5 OAZ1 PCBP4 EDEM3 FOXK1 NUDT15 ARRB1 PARK7 BCAP31 PSME1 STAT3 RC3H2 THRAP3 ZFP36L1 GET4 COP1 MAPKAPK2 PABPC1 CARHSP1
turqiouse	8.61E-13	85	1883	2.6042731	Catabolic proc.	CHIA UPF2 ITCH HYAL1 CSDE1 PGK1 UBE2H NOS1 CNOT2 RARRES2 UFD1 EXOSC9 PSMA7 PSMC1 ARIH2 FUS HUWE1 HSD17B11 DAP OS9 STT3B

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
						HNRNPU AZIN1 TTC3 PRKCD ATG3 RBBP6 LIX1 PAN3 PDCL3 EZR VPS4B EIF6 TMEM259 RASIP1 LYPLA2 EGFR IMPA1 AGO1 CTSB DKC1 VDAC1 HSPA8 RUBCNL SPOPL HERC2 HNRNPD PSMA4 CYP19A1 VCP PNLDC1 UBR4 RPS27A CASP8 OAZ2 BCKDHA YBX1 WDR91 EGLN2 PSMB10 CDKN1B PAOX AGO3 DDX5 OAZ1 PCBP4 EDEM3 FOXK1 NUDT15 ARRB1 PARK7 BCAP31 PSME1 STAT3 RC3H2 THRAP3 ZFP36L1 GET4 TBC1D5 COP1 ADRA1A MAPKAPK2 VPS13C PABPC1 CARHSP1
turqiouse	4.00E-10	68	1509	2.5997859	Immune System	CD274 ITCH UBE2H CALM2 IL4R NOS1 PTPN11 LGALS3 TXN JAK1 SH3KBP1 GRB2 PRDX6 KLRK1 ADAM10 ARIH2 CLU PGM2 HUWE1 HSP90AA1 SPTBN1 ASB12 UBE2Q1 STAT1 MAP3K3 PRKCD RBBP6 CAND1 PA2G4 PIK3R1 RASAL3 EGFR PANX1 ARG1 CTSB HSPA8 TAB2 HERC2 RAB5B CAPZA1 DYNC1LI1 UBR4 RPS27A HEXB CASP8 CYSTM1 TAPBP HMGB1 GRN BOLA-NC1 CMTM6 LYN AP1S2 ARRB1 BCAP31 LTB CALM1

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
						CREBBP UBC HECTD1 MUC20 ZNRF1 MRC1 MAPKAPK2 PDAP1 CFD RAP1A
turqiouse	5.26E-10	70	1604	2.5177441	Cellular catabolic proc.	UPF2 ITCH CSDE1 UBE2H NOS1 CNOT2 UFD1 EXOSC9 PSMA7 PSMC1 ARIH2 FUS HUWE1 DAP OS9 STT3B HNRNPU TTC3 PRKCD ATG3 RBBP6 LIX1 PAN3 PDCL3 EZR VPS4B TMEM259 RASIP1 LYPLA2 AGO1 CTSB DKC1 VDAC1 HSPA8 RUBCNL SPOPL HERC2 HNRNPD PSMA4 VCP PNLDC1 UBR4 RPS27A CASP8 BCKDHA YBX1 WDR91 PSMB10 AGO3 DDX5 PCBP4 EDEM3 FOXK1 NUDT15 ARRB1 PARK7 BCAP31 PSME1 STAT3 RC3H2 THRAP3 ZFP36L1 GET4 TBC1D5 COP1 ADRA1A MAPKAPK2 VPS13C PABPC1 CARHSP1
turqiouse	4.52E-10	80	1962	2.3523877	Carbohydrate derivative binding	GRK3 CHIA PGK1 RAB2A KIF1C PLXND1 NOS1 PRPF4B KIF5B CHD9 ITM2B JAK1 SNRK BMP4 SGK1 PSMC1 DMC1 OARD1 HSP90AA1 RALA DDX27 ABCA9 HNRNPU TOP1 MAP3K3 PRKCD SPOCK2 ZNF207 DDX6 NEK11 SUCLG2 DDX17 PAN3 MYH9 VPS4B CDK11B STK38 RAD50 XYLB EGFR CTSB ACBD3 HSPA8 DGKH SMC3 RAB5B GIMAP8 DDX21 GUCY1A1

Module	Enrichment FDR	No. of Genes	No. of genes in Pathway	Fold Enrichment	Pathway	Genes
						GPAA1 DYNC1LI1 ANXA6 OXSR1 VCP AARSD1 IFI47 STK40 APLP2 DNAJA1 CSNK1G2 RARS2 RAB13 AKT3 DDX5 CLK4 TGFBR2 NRP1 MTHFS LYN DDX46 STK3 ARHGAP5 KATNA1 MAPKAPK2 ATRX ARF6 RAP1A
turqiouse	5.92E-10	79	1946	2.3420824	Reg. of localization	HYAL1 BCL6 CALM2 PLXND1 PTPN11 COX17 KIF5B LGALS3 TXN CD47 BMP4 RARRES2 KLRK1 PRELID1 ADAM10 TMED10 RREB1 HSP90AA1 RALA SPTBN1 ENSA OS9 PCM1 AKAP9 HNRNPU AZIN1 MAP3K3 CORO1A GPC5 PRKCD ATG3 EZR VPS4B PIK3R1 LAMB1 EGFR GDI1 TRADD DKC1 VDAC1 DYSF RAB5B NISCH SPARC CYP19A1 OXSR1 VCP DOCK5 OAZ2 GOPC CLTC SCN3B ITGA6 KLHL24 ATF4 RAB13 AKT3 RIPOR2 CDKN1B TMBIM6 GRN TGFBR2 NRP1 LYN ARRB1 PARK7 BCAP31 CNTN1 MEF2C ANP32B STAT3 ZFP36L1 ACTB TBC1D5 HECTD1 ARF6 ARHGAP4 RAP1A CCL5

List of References

ACOSTA-RODRIGUEZ, R., ALONSO-MORALES, R., BALLADARES, S., FLORES-AGUILAR, H., GARCIA-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. *Veterinary Parasitology,* 127**,** 313-321.

ADRA, C. N., ZHU, S., KO, J.-L., GUILLEMOT, J.-C., CUERVO, A. M., KOBAYASHI, H., HORIUCHI, T., LELIAS, J.-M., ROWLEY, J. D. & LIM, B. 1996. LAPTM5: a novel lysosomal-associated multispanning membrane protein preferentially expressed in hematopoietic cells. *Genomics,* 35**,** 328-337.

AMARASINGHE, S. L., SU, S., DONG, X., ZAPPIA, L., RITCHIE, M. E. & GOUIL, Q. 2020. Opportunities and challenges in long-read sequencing data analysis. *Genome biology,* 21**,** 30.

AMOS, W. 2007. Mix and match — hybridization reveals hidden complexity in seal breeding behaviour. *Molecular Ecology,* 16**,** 3066-3068.

ANTHONY, R. & KIN FAI, A. 2015. PacBio Sequencing and Its Applications. *Genomics, Proteomics & Bioinformatics,* 13**,** 278-289.

ARENDT, A. D. C. 2022. *xlsx: Read, Write, Format Excel 2007 and Excel 97/2000/XP/2003 Files* [Online]. Available: https://github.com/colearendt/xlsx [Accessed].

ASCHARD, H., LUTZ, S., MAUS, B., DUELL, E. J., FINGERLIN, T. E., CHATTERJEE, N., KRAFT, P. & VAN STEEN, K. 2012. Challenges and opportunities in genome-wide environmental interaction (GWEI) studies. *Human genetics,* 131**,** 1591-1613.

BAGNALL, N., GOUGH, J., CADOGAN, L., BURNS, B. & KONGSUWAN, K. 2009. Expression of intracellular calcium signalling genes in cattle skin during tick infestation. *Parasite Immunol,* 31**,** 177-87.

BARENDSE, W. 2007. *Assessing Tick resistance in a bovine animal for selecting cattle for tick resistance by providing a nucleic acid from the bovine animal and assaying for the occurrence of a single nucleotide polymorphism (SNP)*. WO2007051248-A1, 1–146.

BARKER, S. & MURRELL, A. 2004. Systematics and evolution of ticks with list of valid genus and species names. *Parasitology,* 129 Suppl**,** S15-36.

BASIEL, B. L., HARDIE, L. C., HEINS, B. J. & DECHOW, C. D. 2021. Genetic parameters and genomic regions associated with horn fly resistance in organic Holstein cattle. *Journal of Dairy Science,* 104**,** 12724-12740.

BEHL, J. D., VERMA, N. K., TYAGI, N., MISHRA, P., BEHL, R. & JOSHI, B. K. 2012. The major histocompatibility complex in bovines: a review. *ISRN Vet Sci,* 2012**,** 872710.

BENNETT, G. F. 1969. Boophilus microplus (acarina: ixodidae): experimental infestations on cattle restrained from grooming. *Exp Parasitol,* 26**,** 323-8.

BERMINGHAM, M., BISHOP, S., WOOLLIAMS, J., PONG-WONG, R., ALLEN, A., MCBRIDE, S., RYDER, J., WRIGHT, D., SKUCE, R. & MCDOWELL, S. W. 2014. Genome-wide association study identifies novel loci associated with resistance to bovine tuberculosis. *Heredity,* 112**,** 543-551.

BEVERLEY, M. A. 1996. The history of the cattle tick Boophilus microptus in Australia and achievements in its control. *International Journal for Parasitology,* 26**,** 1341-1355.

BHALERAO, J., TYLZANOWSKI, P., FILIE, J. D., KOZAK, C. A. & MERREGAERT, J. 1995. Molecular cloning, characterization, and genetic mapping of the cDNA coding for a novel secretory protein of mouse. Demonstration of alternative splicing in skin and cartilage. *Journal of Biological Chemistry,* 270**,** 16385-16394.

BHATTARAI, P., GUNASEKARAN, T. I., BELLOY, M. E., REYES-DUMEYER, D., JÜLICH, D., TAYRAN, H., YILMAZ, E., FLAHERTY, D., TURGUTALP, B.,

SUKUMAR, G., ALBA, C., MCGRATH, E. M., HUPALO, D. N., BACIKOVA, D., LE GUEN, Y., LANTIGUA, R., MEDRANO, M., RIVERA, D., RECIO, P., NURIEL, T., ERTEKIN-TANER, N., TEICH, A. F., DICKSON, D. W., HOLLEY, S., GREICIUS, M., DALGARD, C. L., ZODY, M., MAYEUX, R., KIZIL, C. & VARDARAJAN, B. N. 2024. Rare genetic variation in fibronectin 1 (FN1) protects against APOEε4 in Alzheimer's disease. *Acta Neuropathologica,* 147**,** 70.

BOLORMAA, S., PRYCE, J. E., HAYES, B. J. & GODDARD, M. E. 2010. Multivariate analysis of a genome-wide association study in dairy cattle. *J Dairy Sci,* 93**,** 3818-33.

BRADLEY, D. G., LOFTUS, R. T., CUNNINGHAM, P. & MACHUGH, D. E. 1998. Genetics and domestic cattle origins. *Evolutionary Anthropology: Issues, News, and Reviews: Issues, News, and Reviews,* 6**,** 79-86.

BRAZ, C. U., ROWAN, T. N., SCHNABEL, R. D. & DECKER, J. E. 2021. Genome-wide association analyses identify genotype-by-environment interactions of growth traits in Simmental cattle. *Scientific Reports,* 11**,** 13335.

BRELOER, M. & FLEISCHER, B. 2008. CD83 regulates lymphocyte maturation, activation and homeostasis. *Trends in immunology,* 29**,** 186-194.

BUDELI, M., NEPHAWE, K., NORRIS, D., SELAPA, N., BERGH, L. & MAIWASHE, A. 2009. Genetic parameter estimates for tick resistance in Bonsmara cattle. *South African Journal of Animal Science,* 39**,** 321-327.

BUDELI, M. A., NEPHAWE, K., NORRIS, D., SELAPA, N., BERGH, L. & MAIWASHE, N. 2010. Genetic parameter estimates for tick resistance in Bonsmara cattle. *South African Journal of Animal Science,* 39.

BURGESS, D. J. 2018. Next regeneration sequencing for reference genomes. *Nature Reviews Genetics,* 19**,** 125-125.

BURROW, H., MANS, B., CARDOSO, F., BIRKETT, M., KOTZE, A., HAYES, B., MAPHOLI, N., DZAMA, K., MARUFU, M., GITHAKA, N. & DJIKENG, A. 2019. Towards a new phenotype for tick resistance in beef and dairy cattle: a review. *Animal Production Science,* 59.

BURROW, H. M. 2014. Genetic Aspects of Cattle Adaptation in the Tropics. 2 ed.: CABI.

BYRNE, A., BEAUDIN, A. E., OLSEN, H. E., JAIN, M., COLE, C., PALMER, T., DUBOIS, R. M., FORSBERG, E. C., AKESON, M. & VOLLMERS, C. 2017. Nanopore long-read RNAseq reveals widespread transcriptional variation among the surface receptors of individual B cells. *Nature communications,* 8**,** 16027.

CANALES, M., ENRÍQUEZ, A., RAMOS, E., CABRERA, D., DANDIE, H., SOTO, A., FALCÓN, V., RODRÍGUEZ, M. & DE LA FUENTE, J. 1997. Large-scale production in Pichia pastoris of the recombinant vaccine Gavac against cattle tick. *Vaccine,* 15**,** 414-22.

CANDELS, L. S., BECKER, S. & TRAUTWEIN, C. 2022. PLA2G7: a new player in shaping energy metabolism and lifespan. *Signal Transduction and Targeted Therapy,* 7**,** 195.

CARDOSO, F. F., MATIKA, O., DJIKENG, A., MAPHOLI, N., BURROW, H. M., YOKOO, M. J. I., CAMPOS, G. S., GULIAS-GOMES, C. C., RIGGIO, V., PONG-WONG, R., ENGLE, B., PORTO-NETO, L., MAIWASHE, A. & HAYES, B. J. 2021. Multiple Country and Breed Genomic Prediction of Tick Resistance in Beef Cattle. *Front Immunol,* 12**,** 620847.

CARVALHO, W. A., DOMINGUES, R., DE AZEVEDO PRATA, M. C., DA SILVA, M. V. G. B., DE OLIVEIRA, G. C., GUIMARÃES, S. E. F. & MACHADO, M. A. 2014. Microarray analysis of tick-infested skin in resistant and susceptible cattle confirms the role of inflammatory pathways in immune activation and larval rejection. *Veterinary Parasitology,* 205**,** 307-317.

CARVALHO, W. A., IANELLA, P., ARNOLDI, F. G., CAETANO, A. R., MARUYAMA, S. R., FERREIRA, B. R., CONTI, L. H., DA SILVA, M. R., PAULA, J. O., MAIA, A. A. & SANTOS, I. K. 2011. Haplotypes of the bovine IgG2 heavy gamma chain in tick-resistant and tick-susceptible breeds of cattle. *Immunogenetics,* 63**,** 319-24.

CARVALHO, W. A., MARUYAMA, S. R., FRANZIN, A. M., ABATEPAULO, A. R., ANDERSON, J. M., FERREIRA, B. R., RIBEIRO, J. M., MORÉ, D. D., AUGUSTO MENDES MAIA, A., VALENZUELA, J. G., GARCIA, G. R. & DE MIRANDA SANTOS, I. K. 2010. Rhipicephalus (Boophilus) microplus: clotting time in tick-infested skin varies according to local inflammation and gene expression patterns in tick salivary glands. *Exp Parasitol,* 124**,** 428-35.

CASTRO, C. J. & NG, T. F. F. 2017. U(50): A New Metric for Measuring Assembly Output Based on Non-Overlapping, Target-Specific Contigs. *J Comput Biol,* 24**,** 1071-1080.

CHAI, K., ZHANG, X., TANG, H., GU, H., YE, W., WANG, G., CHEN, S., WAN, F., LIANG, J. & SHEN, D. 2022. The Application of Consensus Weighted Gene Co-expression Network Analysis to Comparative Transcriptome Meta-Datasets of Multiple Sclerosis in Gray and White Matter. *Frontiers in neurology* [Online], 13. [Accessed 2022].

CHEN, N., FU, W., ZHAO, J., SHEN, J., CHEN, Q., ZHENG, Z., CHEN, H., SONSTEGARD, T. S., LEI, C. & JIANG, Y. 2020. BGVD: An Integrated Database for Bovine Sequencing Variations and Selective Signatures. *Genomics Proteomics Bioinformatics,* 18**,** 186-193.

CHEN, Y., YU, M., DAI, X., ZOGG, M., WEN, R., WEILER, H. & WANG, D. 2011. Critical role for Gimap5 in the survival of mouse hematopoietic stem and progenitor cells. *Journal of Experimental Medicine,* 208**,** 923-935.

CHEN, Z., CORLETT, R. T., JIAO, X., LIU, S.-J., CHARLES-DOMINIQUE, T., ZHANG, S., LI, H., LAI, R., LONG, C. & QUAN, R.-C. 2018. Prolonged milk provisioning in a jumping spider. *Science,* 362**,** 1052-1055.

COLGAN, A. M., CAMERON, A. D. S. & KRÖGER, C. 2017. If it transcribes, we can sequence it: mining the complexities of host–pathogen–environment interactions using RNA-seq. *Current Opinion in Microbiology,* 36**,** 37-46.

CONSTANTINOIU, C., CONSTANTINOIU, C., LEW-TABOR, A. E., LEW-TABOR, A. E., JACKSON, L. A., JORGENSEN, W. K., JORGENSEN, W. K., PIPER, E. K., PIPER, E. K., MAYER, D., JOHNSON, L. K., VENUS, B., JONSSON, N. N. & JONSSON, N. N. 2018. Local immune response to larvae of Rhipicephalus microplus in Santa Gertrudis cattle. *Parasite Immunology,* 40.

CONSTANTINOIU, C. C., JACKSON, L. A., JORGENSEN, W. K., LEW-TABOR, A. E., PIPER, E. K., MAYER, D. G., VENUS, B. & JONSSON, N. N. 2010. Local immune response against larvae of Rhipicephalus (Boophilus) microplus in Bos taurus indicus and Bos taurus taurus cattle. *Int J Parasitol,* 40**,** 865-75.

COOLS, J., MENTENS, N. & MARYNEN, P. 2001. A new family of small, palmitoylated, membrane-associated proteins, characterized by the presence of a cysteine-rich hydrophobic motif. *Febs Letters,* 492**,** 204-209.

CORRIER, D., VIZCAINO, O., TERRY, M., BETANCOURT, A., KUTTLER, K., CARSON, C., TREVINO, G. & RISTIC, M. 1979. Mortality, weight loss and anaemia in Bos taurus calves exposed to Boophilus microplus ticks in the tropics of Colombia. *Tropical Animal Health and Production,* 11**,** 215-221.

CUI, J., SHEN, N., LU, Z., XU, G., WANG, Y. & JIN, B. 2020. Analysis and comprehensive comparison of PacBio and nanopore-based RNA sequencing of the Arabidopsis transcriptome. *Plant Methods,* 16**,** 1-13.

CUI, Y. & YU, L. 2016. Application of the CRISPR/Cas9 gene editing technique to research on functional genomes of parasites. *Parasitol Int,* 65**,** 641-644.

CUNDIFF, L., GREGORY, K. & KOCH, R. 1998. Germplasm Evaluation in Beef Cattle-Cycle IV: Birth and Weaning Traits. *Journal of animal science,* 76**,** 2528-35.

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**,** 23-8.

DE ROOS, A. P., HAYES, B. J. & GODDARD, M. E. 2009. Reliability of genomic predictions across multiple populations. *Genetics,* 183**,** 1545-53.

DE SOUZA, M. M., ZERLOTINI, A., GEISTLINGER, L., TIZIOTO, P. C., TAYLOR, J. F., ROCHA, M. I. P., DINIZ, W. J. S., COUTINHO, L. L. & REGITANO, L. C. A. 2018. A comprehensive manually-curated compendium of bovine transcription factors. *Sci Rep,* 8**,** 13747.

DEPLEDGE, D. P., SRINIVAS, K. P., SADAOKA, T., BREADY, D., MORI, Y., PLACANTONAKIS, D. G., MOHR, I. & WILSON, A. C. 2019. Direct RNA sequencing on nanopore arrays redefines the transcriptional complexity of a viral pathogen. *Nature communications,* 10**,** 754.

DOBREVA, G., DAMBACHER, J. & GROSSCHEDL, R. 2003. SUMO modification of a novel MAR-binding protein, SATB2, modulates immunoglobulin μ gene expression. *Genes & development,* 17**,** 3048-3061.

DOMINGUES, R., VIANA, S., FAZA, D., TEIXEIRA, H. C., FERREIRA, A. P., GUIMARÃES, S., PRATA, M. C. A., FURLONG, J., VERNEQUE, R. & MACHADO, M. 2014. Expression of immune response genes in peripheral blood of cattle infested with Rhipicephalus microplus. *Genetics and molecular research : GMR,* 13**,** 4013-21.

DONG, X., TIAN, L., GOUIL, Q., KARIYAWASAM, H., SU, S., DE PAOLI-ISEPPI, R., PRAWER, Y. D. J., CLARK, M. B., BRESLIN, K., IMINITOFF, M., BLEWITT, M. E., LAW, C. W. & RITCHIE, M. E. 2021. The long and the short of it: unlocking nanopore long-read RNA sequencing data with short-read differential expression analysis tools. *NAR Genom Bioinform,* 3**,** lqab028.

DOS SANTOS, C. G., SOUSA, M. F., VIEIRA, J. I. G., DE MORAIS, L. R., FERNANDES, A. A. S., DE OLIVEIRA LITTIERE, T., ITAJARA OTTO, P., MACHADO, M. A., SILVA, M. V. G. B., BONAFÉ, C. M., BRAGA MAGALHÃES, A. F. & VERARDO, L. L. 2022. Candidate genes for tick resistance in cattle: a systematic review combining post-GWAS analyses with sequencing data. *Journal of Applied Animal Research,* 50**,** 460-470.

DOUBE, B. M. & WHARTON, R. H. 1980. The effect of locality, breed and previous tick experience on seasonal changes in the resistance of cattle to Boophilus microplus (Ixodoidea: Ixodidae). *Experientia,* 36**,** 1178-9.

DOYLE, J. L., BERRY, D. P., VEERKAMP, R. F., CARTHY, T. R., EVANS, R. D., WALSH, S. W. & PURFIELD, D. C. 2020. Genomic regions associated with muscularity in beef cattle differ in five contrasting cattle breeds. *Genetics Selection Evolution,* 52**,** 1-18.

EMAM, M., LIVERNOIS, A., PAIBOMESAI, M., ATALLA, H. & MALLARD, B. 2019. Genetic and Epigenetic Regulation of Immune Response and Resistance to Infectious Diseases in Domestic Ruminants. *Vet Clin North Am Food Anim Pract,* 35**,** 405-429.

FAO, F. 2018. The future of food and agriculture: alternative pathways to 2050. *Food and Agriculture Organization of the United Nations Rome*.

FARHADIAN, M., RAFAT, S. A., PANAHI, B. & MAYACK, C. 2021. Weighted gene co-expression network analysis identifies modules and functionally enriched pathways in the lactation process. *Scientific Reports,* 11**,** 2367.

FERNANDO, R. L. & GROSSMAN, M. 1989. Marker assisted selection using best linear unbiased prediction. *Genetics Selection Evolution,* 21**,** 467.

FIKRET ISIK, J. H., CHRISTIAN MALTECCA 2017. *Genetic Data Analysis for Plant and Animal Breeding*, Springer Cham.

FONSECA, P. A. S., SUÁREZ-VEGA, A. & CÁNOVAS, A. 2020. Weighted Gene Correlation Network Meta-Analysis Reveals Functional Candidate Genes Associated with High- and Sub-Fertile Reproductive Performance in Beef Cattle. *Genes (Basel),* 11.

FRANCISCHETTI, I. M., SA-NUNES, A., MANS, B. J., SANTOS, I. M. & RIBEIRO, J. M. 2009. The role of saliva in tick feeding. *Frontiers in bioscience: a journal and virtual library,* 14**,** 2051.

FRANZIN, A. M., MARUYAMA, S. R., GARCIA, G. R., OLIVEIRA, R. P., RIBEIRO, J. M. C., BISHOP, R., MAIA, A. A. M., MORÉ, D. D., FERREIRA, B. R. & SANTOS, I. K. F. D. M. 2017a. Immune and biochemical responses in skin differ between bovine hosts genetically susceptible and resistant to the cattle tick Rhipicephalus microplus. *Parasites & Vectors,* 10**,** 51.

FRANZIN, A. M., MARUYAMA, S. R., GARCIA, G. R., OLIVEIRA, R. P., RIBEIRO, J. M. C., BISHOP, R., MAIA, A. A. M., MORÉ, D. D., FERREIRA, B. R. & SANTOS, I. K. F. D. M. 2017b. Immune and biochemical responses in skin differ between bovine hosts genetically susceptible and resistant to the cattle tick Rhipicephalus microplus. *Parasites & vectors,* 10**,** 1-24.

FRAZER, K. A., MURRAY, S. S., SCHORK, N. J. & TOPOL, E. J. 2009. Human genetic variation and its contribution to complex traits. *Nature Reviews Genetics,* 10**,** 241-251.

FRISCH, J., DRINKWATER, R., HARRISON, B. & JOHNSON, S. 1997. Classification of the southern African sanga and East African shorthorned zebu. *Animal Genetics,* 28**,** 77-83.

GAO, F., CHIU, S. M., MOTAN, D. A. L., ZHANG, Z., CHEN, L., JI, H. L., TSE, H. F., FU, Q. L. & LIAN, Q. 2016. Mesenchymal stem cells and immunomodulation: current status and future prospects. *Cell Death & Disease,* 7**,** e2062-e2062.

GARCíA-GARCíA, J. C., MONTERO, C., REDONDO, M., VARGAS, M., CANALES, M., BOUE, O., RODRíGUEZ, M., JOGLAR, M., MACHADO, H. & GONZÁLEZ, I. L. 2000. Control of ticks resistant to immunization with Bm86 in cattle vaccinated with the recombinant antigen Bm95 isolated from the cattle tick, Boophilus microplus. *Vaccine,* 18**,** 2275-2287.

GARNER, J. B., DOUGLAS, M. L., WILLIAMS, S. R., WALES, W. J., MARETT, L. C., NGUYEN, T. T., REICH, C. M. & HAYES, B. J. 2016. Genomic Selection Improves Heat Tolerance in Dairy Cattle. *Sci Rep,* 6**,** 34114.

GASISOVA, A. I., ATKENOVA, A. B., AHMETZHANOVA, N. B., MURZABEKOVA, L. M. & BEKENOVA, A. C. 2017. Morphostructure of Immune System Organs in Cattle of Different Age. *Anat Histol Embryol,* 46**,** 132-142.

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. *Animal Genetics,* 38**,** 453-459.

GENG, R., LI, Z., YU, S., YUAN, C., HONG, W., WANG, Z., WANG, Q., YI, Z. & FANG, Y. 2020. Weighted gene co-expression network analysis identifies specific modules and hub genes related to subsyndromal symptomatic depression. *World J Biol Psychiatry,* 21**,** 102-110.

GHEBREHIWET, B., KANDOV, E., KISHORE, U. & PEERSCHKE, E. 2018. Is the A-Chain the Engine That Drives the Diversity of C1q Functions? Revisiting Its Unique Structure. *Frontiers in Immunology,* 9.

GIOLAI, M., PAAJANEN, P., VERWEIJ, W., WITEK, K., JONES, J. D. & CLARK, M. D. 2017. Comparative analysis of targeted long read sequencing approaches for characterization of a plant's immune receptor repertoire. *BMC genomics,* 18**,** 1-15.

GLOWACKA, W. K., ALBERTS, P., OUCHIDA, R., WANG, J. Y. & ROTIN, D. 2012. LAPTM5 protein is a positive regulator of proinflammatory signaling pathways in macrophages. *J Biol Chem,* 287**,** 27691-702.

GOFF, W. L., BASTOS, R. G., BROWN, W. C., JOHNSON, W. C. & SCHNEIDER, D. A. 2010. The bovine spleen: interactions among splenic cell populations in the innate immunologic control of hemoparasitic infections. *Vet Immunol Immunopathol,* 138**,** 1-14.

GOOSSENS, K., VAN SOOM, A., VAN ZEVEREN, A., FAVOREEL, H. & PEELMAN, L. J. 2009. Quantification of fibronectin 1 (FN1) splice variants, including two novel ones, and analysis of integrins as candidate FN1 receptors in bovine preimplantation embryos. *BMC Dev Biol,* 9**,** 1.

GREELY, H. T. 2007. The uneasy ethical and legal underpinnings of large-scale genomic biobanks. *Annu Rev Genomics Hum Genet,* 8**,** 343-64.

GROSSMANN, D., BERENGUER-ESCUDER, C., BELLET, M. E., SCHEIBNER, D., BOHLER, J., MASSART, F., RAPAPORT, D., SKUPIN, A., FOUQUIER D'HÉROUËL, A., SHARMA, M., GHELFI, J., RAKOVIĆ, A., LICHTNER, P., ANTONY, P., GLAAB, E., MAY, P., DIMMER, K. S., FITZGERALD, J. C., GRÜNEWALD, A. & KRÜGER, R. 2019. Mutations in RHOT1 Disrupt Endoplasmic Reticulum-Mitochondria Contact Sites Interfering with Calcium Homeostasis and Mitochondrial Dynamics in Parkinson's Disease. *Antioxid Redox Signal,* 31**,** 1213-1234.

GRÜNBERGER, F., FERREIRA-CERCA, S. & GROHMANN, D. 2022. Nanopore sequencing of RNA and cDNA molecules in Escherichia coli. *Rna,* 28**,** 400-417.

GUERRERO, F. D., MILLER, R. J. & DE LEÓN, A. A. P. 2012. Cattle tick vaccines: many candidate antigens, but will a commercially viable product emerge? *International Journal for Parasitology,* 42**,** 421-427.

HADLEY WICKHAM, M. A., JENNIFER BRYAN, WINSTON CHANG, LUCY D'AGOSTINO MCGOWAN, ROMAIN FRANÇOIS, GARRETT GROLEMUND, ALEX HAYES, LIONEL HENRY, JIM HESTER, MAX KUHN, THOMAS LIN PEDERSEN, EVAN MILLER, STEPHAN MILTON BACHE, KIRILL MÜLLER, JEROEN OOMS, DAVID ROBINSON, DANA PAIGE SEIDEL, VITALIE SPINU, KOHSKE TAKAHASHI, DAVIS VAUGHAN, CLAUS WILKE, KARA WOO, HIROAKI YUTANI 2019. tidyverse. Zenodo.

HALL, S. J. G. & BUNCE, R. G. H. 2019. The use of cattle Bos taurus for restoring and maintaining holarctic landscapes: Conclusions from a long-term study (1946–2017) in northern England. *Ecology and Evolution,* 9**,** 5859-5869.

HAMADA, T. 2002. Lipoid proteinosis. *Clinical and experimental dermatology,* 27**,** 624-629.

HAN, Z., NI, J., SMITS, P., UNDERHILL, C. B., XIE, B., CHEN, Y., LIU, N., TYLZANOWSKI, P., PARMELEE, D. & FENG, P. 2001. Extracellular matrix protein 1 (ECM1) has angiogenic properties and is expressed by breast tumor cells. *The FASEB Journal,* 15**,** 988-994.

HAYES, B. & GODDARD, M. 2010. Genome-wide association and genomic selection in animal breeding. *Genome,* 53**,** 876-883.

HAYES, B. J., BOWMAN, P. J., CHAMBERLAIN, A. J. & GODDARD, M. E. 2009. Invited review: Genomic selection in dairy cattle: progress and challenges. *J Dairy Sci,* 92**,** 433-43.

HENSHALL, J. 2004a. A genetic analysis of parasite resistance traits in a tropically adapted line of Bos taurus. *Australian Journal of Agricultural Research,* 55**,** 1109-1116.

HENSHALL, J. M. 2004b. A genetic analysis of parasite resistance traits in a tropically adapted line of Bos taurus. *Australian Journal of Agricultural Research,* 55**,** 1109-1116.

HIGA, L. D. O. S., GARCIA, M. V., BARROS, J. C., KOLLER, W. W. & ANDREOTTI, R. 2016. Evaluation of Rhipicephalus (Boophilus) microplus (Acari: Ixodidae) resistance to different acaricide formulations using samples from Brazilian properties. *Revista Brasileira de Parasitologia Veterinária,* 25**,** 163-171.

HIRSCHHORN, J. N. 2005. Genetic approaches to studying common diseases and complex traits. *Pediatr Res,* 57**,** 74r-77r.

HO, C.-H. & TSAI, S.-F. 2017. Functional and biochemical characterization of a T cell-associated anti-apoptotic protein, GIMAP6. *Journal of Biological Chemistry,* 292**,** 9305-9319.

HONG, E. P. & PARK, J. W. 2012. Sample size and statistical power calculation in genetic association studies. *Genomics Inform,* 10**,** 117-22.

HU, Y., FANG, L., CHEN, X., ZHONG, J. F., LI, M. & WANG, K. 2021. LIQA: long-read isoform quantification and analysis. *Genome Biol,* 22**,** 182.

HUANG DA, W., SHERMAN, B. T. & LEMPICKI, R. A. 2009a. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res,* 37**,** 1-13.

HUANG DA, W., SHERMAN, B. T. & LEMPICKI, R. A. 2009b. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc,* 4**,** 44-57.

HUANG, X., CHEN, Q., LUO, W., PAKVASA, M., ZHANG, Y., ZHENG, L., LI, S., YANG, Z., ZENG, H., LIANG, F., ZHANG, F., HU, D. A., QIN, K. H., WANG, E. J., QIN, D. S., REID, R. R., HE, T. C., ATHIVIRAHAM, A., EL DAFRAWY, M. & ZHANG, H. 2022. SATB2: A versatile transcriptional regulator of craniofacial and skeleton development, neurogenesis and tumorigenesis, and its applications in regenerative medicine. *Genes Dis,* 9**,** 95-107.

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. *Mammalian Genome,* 19**,** 226-245.

IRIE, A., YAMAMOTO, K., MIKI, Y. & MURAKAMI, M. 2017. Phosphatidylethanolamine dynamics are required for osteoclast fusion. *Scientific reports,* 7**,** 46715.

IUNG, L. H. S., PETRINI, J., RAMÍREZ-DÍAZ, J., SALVIAN, M., ROVADOSCKI, G. A., PILONETTO, F., DAURIA, B. D., MACHADO, P. F., COUTINHO, L. L., WIGGANS, G. R. & MOURÃO, G. B. 2019. Genome-wide association study for milk production traits in a Brazilian Holstein population. *Journal of Dairy Science,* 102**,** 5305-5314.

JAIN, M., OLSEN, H. E., PATEN, B. & AKESON, M. 2016. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol,* 17**,** 239.

JAIRES, G., ANDRÉ LUCIO FRANCESCHINI, S., LORENA LOPES, F., JOHN, C. C., STEPHEN, J. P., JOHN, A. P., ADALBERTO, A. P., MICHAEL, A. B. & LÍGIA MIRANDA FERREIRA, B. 2016. Quantification of brown dog tick repellents, 2-hexanone and benzaldehyde, and release from tick-resistant beagles, Canis lupus familiaris. *Journal of Chromatography B,* 1022**,** 64-69.

JENJAROENPUN, P., WONGSURAWAT, T., PEREIRA, R., PATUMCHAROENPOL, P., USSERY, D. W., NIELSEN, J. & NOOKAEW, I. 2018. Complete genomic and transcriptional landscape analysis using third-generation sequencing: a case study of Saccharomyces cerevisiae CEN. PK113-7D. *Nucleic acids research,* 46**,** e38-e38.

JOHNSTON, D., MUKIIBI, R., WATERS, S. M., MCGEE, M., SURLIS, C., MCCLURE, J. C., MCCLURE, M. C., TODD, C. G. & EARLEY, B. 2020. Genome wide association study of passive immunity and disease traits in beef-suckler and dairy calves on Irish farms. *Scientific Reports,* 10**,** 18998.

JOHNSTON, L. A. & HAYDOCK, K. P. 1969. The effect of cattle tick (Boophilus microplus)on production of Brahman-cross and British-breed cattle in northern Australia. *Aust Vet J,* 45**,** 175-9.

JONGEJAN, F. & UILENBERG, G. 1994. Ticks and control methods. *Rev Sci Tech,* 13**,** 1201-26.

JONGEJAN, F. & UILENBERG, G. 2004. The global importance of ticks. *Parasitology,* 129 Suppl**,** S3-14.

JONSSON, N. N., PIPER, E. K. & CONSTANTINOIU, C. C. 2014. Host resistance in cattle to infestation with the cattle tick Rhipicephalus microplus. *Parasite Immunology,* 36**,** 553-559.

KANEHISA, M. 2019. Toward understanding the origin and evolution of cellular organisms. *Protein Sci,* 28**,** 1947-1951.

KANEHISA, M., FURUMICHI, M., SATO, Y., KAWASHIMA, M. & ISHIGURO-WATANABE, M. 2023. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res,* 51**,** D587-d592.

KANEHISA, M. & GOTO, S. 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res,* 28**,** 27-30.

KANEHISA, M., GOTO, S., SATO, Y., KAWASHIMA, M., FURUMICHI, M. & TANABE, M. 2014. Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic acids research,* 42**,** D199-D205.

KARIUKI, S. N., GHODKE-PURANIK, Y., DORSCHNER, J. M., CHRABOT, B. S., KELLY, J. A., TSAO, B. P., KIMBERLY, R. P., ALARCÓN-RIQUELME, M. E., JACOB, C. O. & CRISWELL, L. A. 2015. Genetic analysis of the pathogenic molecular sub-phenotype interferon-alpha identifies multiple novel loci involved in systemic lupus erythematosus. *Genes & Immunity,* 16**,** 15-23.

KEMP, D. H., HALES, J. R., SCHLEGER, A. V. & FAWCETT, A. A. 1983. Comparison of cutaneous hyperemia in cattle elicited by larvae of Boophilus microplus and by prostaglandins and other mediators. *Experientia,* 39**,** 725-7.

KITSOU, C., FIKRIG, E. & PAL, U. 2021. Tick host immunity: vector immunomodulation and acquired tick resistance. *Trends Immunol,* 42**,** 554-574.

KLEIN, R. J., ZEISS, C., CHEW, E. Y., TSAI, J.-Y., SACKLER, R. S., HAYNES, C., HENNING, A. K., SANGIOVANNI, J. P., MANE, S. M. & MAYNE, S. T. 2005. Complement factor H polymorphism in age-related macular degeneration. *Science,* 308**,** 385-389.

KOBOLDT, D. C., STEINBERG, K. M., LARSON, D. E., WILSON, R. K. & MARDIS, E. R. 2013. The next-generation sequencing revolution and its impact on genomics. *Cell,* 155**,** 27-38.

KOLSET, S. O. & PEJLER, G. 2011. Serglycin: a structural and functional chameleon with wide impact on immune cells. *The journal of Immunology,* 187**,** 4927-4933.

KONGSUWAN, K., JOSH, P., COLGRAVE, M. L., BAGNALL, N. H., GOUGH, J., BURNS, B. & PEARSON, R. 2010. Activation of several key components of the epidermal differentiation pathway in cattle following infestation with the cattle tick, Rhipicephalus (Boophilus) microplus. *Int J Parasitol,* 40**,** 499-507.

KONGSUWAN, K., PIPER, E., BAGNALL, N., RYAN, K., MOOLHUIJZEN, P., BELLGARD, M., LEW, A., JACKSON, L. & JONSSON, N. N. 2008. Identification of genes involved with tick infestation in Bos taurus and Bos indicus. *Animal Genomics for Animal Health.* Karger Publishers.

KÖNIG, S., SIMIANER, H. & WILLAM, A. 2009. Economic evaluation of genomic breeding programs. *J Dairy Sci,* 92**,** 382-91.

KORPETINOU, A., SKANDALIS, S. S., LABROPOULOU, V. T., SMIRLAKI, G., NOULAS, A., KARAMANOS, N. K. & THEOCHARIS, A. D. 2014. Serglycin: at the crossroad of inflammation and malignancy. *Frontiers in oncology,* 3**,** 327.

KOUDSTAAL, D., KEMP, D. H. & KERR, J. D. 1978. Boophilus microplus: rejection of larvae from British breed cattle. *Parasitology,* 76**,** 379-86.

KRÜCKEN, J., SCHROETEL, R. M., MÜLLER, I. U., SAÏDANI, N., MARINOVSKI, P., BENTEN, W. P. M., STAMM, O. & WUNDERLICH, F. 2004. Comparative analysis of the human gimap gene cluster encoding a novel GTPase family. *Gene,* 341**,** 291-304.

LAFRAMBOISE, T. 2009. Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances. *Nucleic Acids Res,* 37**,** 4181-93.

LANGFELDER, P. & HORVATH, S. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics,* 9**,** 559.

LEE, B. Y., LEE, K. N., LEE, T., PARK, J. H., KIM, S. M., LEE, H. S., CHUNG, D. S., SHIM, H. S., LEE, H. K. & KIM, H. 2015. Bovine Genome-wide Association Study for Genetic Elements to Resist the Infection of Foot-and-mouth Disease in the Field. *Asian-Australas J Anim Sci,* 28**,** 166-70.

LEW-TABOR, A. E. & RODRIGUEZ VALLE, M. 2016. A review of reverse vaccinology approaches for the development of vaccines against ticks and tick borne diseases. *Ticks and Tick-borne Diseases,* 7**,** 573-585.

LI, H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics,* 34**,** 3094-3100.

LI, J., ZHOU, D., QIU, W., SHI, Y., YANG, J.-J., CHEN, S., WANG, Q. & PAN, H. 2018. Application of weighted gene co-expression network analysis for data from paired design. *Scientific reports,* 8**,** 622.

LI, T., FU, J., ZENG, Z., COHEN, D., LI, J., CHEN, Q., LI, B. & LIU, X. S. 2020. TIMER2.0 for analysis of tumor-infiltrating immune cells. *Nucleic Acids Res,* 48**,** W509-w514.

LI, Z., ZHANG, Y., LIU, Z., WU, X., ZHENG, Y., TAO, Z., MAO, K., WANG, J., LIN, G., TIAN, L., JI, Y., QIN, M., SUN, S., ZHU, X. & SUN, B. 2011. ECM1 controls TH2 cell egress from lymph nodes through re-expression of S1P1. *Nature Immunology,* 12**,** 178-185.

LÍGIA MIRANDA FERREIRA, B., JAIRES, G., LORENA LOPES, F., CARLA CRISTINA BRAZ, L., JOHN, A. P. & MICHAEL, A. B. 2015. Identification of non-host semiochemicals for the brown dog tick, Rhipicephalus sanguineus sensu lato (Acari: Ixodidae), from tick-resistant beagles, Canis lupus familiaris. *Ticks and Tick-borne Diseases,* 6**,** 676-682.

LIU, L., ZHOU, J., CHEN, C. J., ZHANG, J., WEN, W., TIAN, J., ZHANG, Z. & GU, Y. 2020. GWAS-Based Identification of New Loci for Milk Yield, Fat, and Protein in Holstein Cattle. *Animals (Basel),* 10.

LÓPEZ-JUÁREZ, A., HOWARD, J., ULLOM, K., HOWARD, L., GRANDE, A., PARDO, A., WACLAW, R., SUN, Y.-Y., YANG, D. & KUAN, C.-Y. 2013. Gsx2 controls region-specific activation of neural stem cells and injury-induced neurogenesis in the adult subventricular zone. *Genes & development,* 27**,** 1272-1287.

LORD, E., LUMSDEN, J., DODDS, K., HENRY, H., CRAWFORD, A., ANSARI, H., PEARCE, P., MAHER, D., STONE, R. & KAPPES, S. 1996. The linkage map of

sheep chromosome 6 compared with orthologous regions in other species. *Mammalian Genome,* 7**,** 373-376.

LOULY, C. C. B., SOARES, S. F., DA NÓBREGA SILVEIRA, D., GUIMARÃES, M. S. & BORGES, L. M. F. 2010. Differences in the behavior of Rhipicephalus sanguineus tested against resistant and susceptible dogs. *Experimental and Applied Acarology,* 51**,** 353-362.

LOW, V. L., TAY, S. T., KHO, K. L., KOH, F. X., TAN, T. K., LIM, Y. A. L., ONG, B. L., PANCHADCHARAM, C., NORMA-RASHID, Y. & SOFIAN-AZIRUN, M. 2015. Molecular characterisation of the tick Rhipicephalus microplus in Malaysia: new insights into the cryptic diversity and distinct genetic assemblages throughout the world. *Parasites & Vectors,* 8**,** 341.

LUND, M. S., ROOS, A. P., VRIES, A. G., DRUET, T., DUCROCQ, V., FRITZ, S., GUILLAUME, F., GULDBRANDTSEN, B., LIU, Z., REENTS, R., SCHROOTEN, C., SEEFRIED, F. & SU, G. 2011. A common reference population from four European Holstein populations increases reliability of genomic predictions. *Genet Sel Evol,* 43**,** 43.

MACHADO, M. A., AZEVEDO, A. L., TEODORO, R. L., PIRES, M. A., PEIXOTO, M. G., DE FREITAS, C., PRATA, M. C., FURLONG, J., DA SILVA, M. V., GUIMARÃES, S. E., REGITANO, L. C., 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.

MACHUGH, D. E., SHRIVER, M. D., LOFTUS, R. T., CUNNINGHAM, P. & BRADLEY, D. G. 1997. Microsatellite DNA variation and the evolution, domestication and phylogeography of taurine and zebu cattle (Bos taurus and Bos indicus). *Genetics,* 146**,** 1071-1086.

MACKENZIE, M. & ARGYROPOULOS, C. 2023. An Introduction to Nanopore Sequencing: Past, Present, and Future Considerations. *Micromachines (Basel),* 14.

MALLARD, B. A., EMAM, M., PAIBOMESAI, M., THOMPSON-CRISPI, K. & WAGTER-LESPERANCE, L. 2015. Genetic selection of cattle for improved immunity and health. *Jpn J Vet Res,* 63 Suppl 1**,** S37-44.

MANJUNATHACHAR, H. V., SARAVANAN, B. C., KESAVAN, M., KARTHIK, K., RATHOD, P., GOPI, M., TAMILMAHAN, P. & BALARAJU, B. L. 2014. Economic importance of ticks and their effective control strategies. *Asian Pacific Journal of Tropical Disease,* 4**,** S770-S779.

MANTILLA VALDIVIESO, E. F., ROSS, E. M., RAZA, A., NASEEM, M. N., KAMRAN, M., HAYES, B. J., JONSSON, N. N., JAMES, P. & TABOR, A. E. 2022. Transcriptional changes in the peripheral blood leukocytes from Brangus cattle before and after tick challenge with Rhipicephalus australis. *BMC Genomics,* 23**,** 454.

MANTILLA VALDIVIESO, E. F., ROSS, E. M., RAZA, A., NGUYEN, L., HAYES, B. J., JONSSON, N. N., JAMES, P. & TABOR, A. E. 2024. Expression network analysis of bovine skin infested with Rhipicephalus australis identifies pro-inflammatory genes contributing to tick susceptibility. *Scientific Reports,* 14**,** 4419.

MAPHOLI, N. O., MAIWASHE, A., MATIKA, O., RIGGIO, V., BISHOP, S., MACNEIL, M., BANGA, C., TAYLOR, J. & DZAMA, K. 2016. Genome-wide association study of tick resistance in South African Nguni cattle. *Ticks and Tick-Borne Diseases,* 7**,** 487-497.

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 and tick-borne diseases,* 5**,** 475-483.

MARIMA, J. K. 2022. *The application of genomics and transcriptomics for the characterization of the genetic diversity of tick-resistance in Angus, Brahman, Nguni, and Santa Gertrudis cattle artificially infested with Rhipicephalus microplus and Rhipicephalus decoloratus.* University of Stellenbosch.

MARIMA, J. K., NEL, C. L., MARUFU, M. C., JONSSON, N. N., DUBE, B. & DZAMA, K. 2020. A genetic and immunological comparison of tick-resistance in beef cattle following artificial infestation with Rhipicephalus ticks. *Experimental and Applied Acarology,* 80**,** 569-590.

MARITZ-OLIVIER, C., VAN ZYL, W. & STUTZER, C. 2012. A systematic, functional genomics, and reverse vaccinology approach to the identification of vaccine candidates in the cattle tick, Rhipicephalus microplus. *Ticks and tick-borne diseases,* 3**,** 179-187.

MARTIN, J. A. & WANG, Z. 2011. Next-generation transcriptome assembly. *Nat Rev Genet,* 12**,** 671-82.

MARTINEZ, M. L., MACHADO, M. A., NASCIMENTO, C. S., SILVA, M. V., TEODORO, R. L., FURLONG, J., PRATA, M. C., CAMPOS, A. L., GUIMARÃES, M. F., AZEVEDO, A. L., PIRES, M. F. & VERNEQUE, R. S. 2006. Association of BoLA-DRB3.2 alleles with tick (Boophilus microplus) resistance in cattle. *Genet Mol Res,* 5**,** 513-24.

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 and tick-borne diseases,* 2**,** 172-177.

MCHUGO, G. P., DOVER, M. J. & MACHUGH, D. E. 2019. Unlocking the origins and biology of domestic animals using ancient DNA and paleogenomics. *BMC Biology,* 17**,** 98.

MEUWISSEN, T., HAYES, B. & GODDARD, M. 2016. Genomic selection: A paradigm shift in animal breeding. *Animal frontiers,* 6**,** 6-14.

MEUWISSEN, T. H., HAYES, B. J. & GODDARD, M. E. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics,* 157**,** 1819-29.

MKIZE, N., MAIWASHE, A., DZAMA, K., DUBE, B. & MAPHOLI, N. 2021. Suitability of GWAS as a Tool to Discover SNPs Associated with Tick Resistance in Cattle: A Review. *Pathogens,* 10.

MOON, R. 2009. Chapter 16: Muscid flies (Muscidae). *Medical and veterinary entomology, 2nd ed. Academic Press, New York, NY***,** 267-287.

MORÉ, D. D., CARDOSO, F. F., MUDADU, M. A., MALAGÓ-JR, W., GULIAS-GOMES, C. C., SOLLERO, B. P., IBELLI, A. M. G., COUTINHO, L. L. & REGITANO, L. C. A. 2019. Network analysis uncovers putative genes affecting resistance to tick infestation in Braford cattle skin. *BMC Genomics,* 20**,** 998.

MOTA, R. R., SILVA, F. F., LOPES, P. S., TEMPELMAN, R. J., SOLLERO, B. P., AGUILAR, I. & CARDOSO, F. F. 2018. Analyses of reaction norms reveal new chromosome regions associated with tick resistance in cattle. *Animal,* 12**,** 205-214.

MOTA, R. R., TEMPELMAN, R. J., LOPES, P. S., AGUILAR, I., SILVA, F. F. & CARDOSO, F. F. 2016. Genotype by environment interaction for tick resistance of Hereford and Braford beef cattle using reaction norm models. *Genetics Selection Evolution,* 48**,** 1-12.

MOURSI, A. M., DAMSKY, C. H., LULL, J., ZIMMERMAN, D., DOTY, S. B., AOTA, S.-I. & GLOBUS, R. K. 1996. Fibronectin regulates calvarial osteoblast differentiation. *Journal of cell science,* 109**,** 1369-1380.

MUELLER, M. L., MCNABB, B. R., OWEN, J. R., HENNIG, S. L., LEDESMA, A. V., ANGOVE, M. L., CONLEY, A. J., ROSS, P. J. & VAN EENENNAAM, A. L. 2023. Germline ablation achieved via CRISPR/Cas9 targeting of NANOS3 in bovine zygotes. *Front Genome Ed,* 5**,** 1321243.

NAESSENS, J., NEWSON, J., MCHUGH, N., HOWARD, C., PARSONS, K. & JONES, B. 1990. Characterization of a bovine leucocyte differentiation antigen of 145,000 MW restricted to B lymphocytes. *Immunology,* 69**,** 525.

NAIK, J., HAU, C. M., TEN BLOEMENDAAL, L., MOK, K. S., HAJJI, N., WEHMAN, A. M., MEISNER, S., MUNCAN, V., PAAUW, N. J., DE VRIES, H. E., NIEUWLAND, R., PAULUSMA, C. C. & BOSMA, P. J. 2019. The P4-ATPase ATP9A is a novel determinant of exosome release. *PLoS One,* 14**,** e0213069.

NASCIMENTO, C. S., MACHADO, M. A., GUIMARÃES, S. E., GUIMARÃES, M. F., PEIXOTO, J. O., FURLONG, J., PRATA, M. C., VERNEQUE, R. S., TEODORO, R. L. & LOPES, P. S. 2010. Differential expression of genes in resistant versus susceptible Gyr x Holstein cattle challenged with the tick Rhipicephalus (Boophilus) microplus. *Genet Mol Res,* 9**,** 1974-9.

NICIURA, S. C. M., BENAVIDES, M. V., OKINO, C. H., IBELLI, A. M. G., MINHO, A. P., ESTEVES, S. N. & CHAGAS, A. C. S. 2022. Genome-Wide Association Study for Haemonchus contortus Resistance in Morada Nova Sheep. *Pathogens,* 11.

OHADI, F., RAHGOZAR, S. & GHODOUSI, E. S. 2020. Sal-Like Protein 4 Transcription Factor: A Significant Diagnostic Biomarker Involved in Childhood ALL Resistance and Relapse. *Cancer Manag Res,* 12**,** 1611-1619.

OIKONOMOPOULOS, S., WANG, Y. C., DJAMBAZIAN, H., BADESCU, D. & RAGOUSSIS, J. 2016. Benchmarking of the Oxford Nanopore MinION sequencing for quantitative and qualitative assessment of cDNA populations. *Scientific reports,* 6**,** 31602.

OTTO, P. I., GUIMARÃES, S. E. F., VERARDO, L. L., AZEVEDO, A. L. S., VANDENPLAS, J., SOARES, A. C. C., SEVILLANO, C. A., VERONEZE, R., DE FATIMA, A. P. M., DE FREITAS, C., PRATA, M. C. A., FURLONG, J., VERNEQUE, R. S., MARTINS, M. F., PANETTO, J. C. C., CARVALHO, W. A., GOBO, D. O. R., DA SILVA, M. & MACHADO, M. A. 2018. Genome-wide association studies for tick resistance in Bos taurus × Bos indicus crossbred cattle: A deeper look into this intricate mechanism. *J Dairy Sci,* 101**,** 11020-11032.

OUCHIDA, R., KUROSAKI, T. & WANG, J.-Y. 2010. A role for lysosomal-associated protein transmembrane 5 in the negative regulation of surface B cell receptor levels and B cell activation. *The Journal of Immunology,* 185**,** 294-301.

OUCHIDA, R., YAMASAKI, S., HIKIDA, M., MASUDA, K., KAWAMURA, K., WADA, A., MOCHIZUKI, S., TAGAWA, M., SAKAMOTO, A. & HATANO, M. 2008. A lysosomal protein negatively regulates surface T cell antigen receptor expression by promoting CD3ζ-chain degradation. *Immunity,* 29**,** 33-43.

OXFORD NANOPORE TECHNOLOGIES. 2024. *Oxford Nanopore Technologies* [Online]. Gosling Building, Edmund Halley Road, Oxford Science Park, OX4 4DQ, UK. Available: https://nanoporetech.com/ [Accessed].

OYAMA, N. & MERREGAERT, J. 2017. The Extracellular Matrix Protein 1 (ECM1) in Molecular-Based Skin Biology.

PAPATHEODOROU, I., MORENO, P., MANNING, J., FUENTES, A. M.-P., GEORGE, N., FEXOVA, S., FONSECA, N. A., FÜLLGRABE, A., GREEN, M. & HUANG, N. 2020. Expression Atlas update: from tissues to single cells. *Nucleic acids research,* 48**,** D77-D83.

PARK, S. D., MAGEE, D. A., MCGETTIGAN, P. A., TEASDALE, M. D., EDWARDS, C. J., LOHAN, A. J., MURPHY, A., BRAUD, M., DONOGHUE, M. T., LIU, Y., CHAMBERLAIN, A. T., RUE-ALBRECHT, K., SCHROEDER, S., SPILLANE, C., TAI, S., BRADLEY, D. G., SONSTEGARD, T. S., LOFTUS, B. J. & MACHUGH, D. E. 2015. Genome sequencing of the extinct Eurasian wild aurochs, Bos primigenius, illuminates the phylogeography and evolution of cattle. *Genome Biol,* 16**,** 234.

PATRO, R., DUGGAL, G., LOVE, M. I., IRIZARRY, R. A. & KINGSFORD, C. 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods,* 14**,** 417--419.

PAUL, D., BRIDOUX, L., REZSÖHAZY, R. & DONNAY, I. 2011. HOX genes are expressed in bovine and mouse oocytes and early embryos. *Mol Reprod Dev,* 78**,** 436-49.

PEREZ-COGOLLO, L. C., RODRIGUEZ-VIVAS, R. I., RAMIREZ-CRUZ, G. T. & MILLER, R. J. 2010. First report of the cattle tick Rhipicephalus microplus resistant to ivermectin in Mexico. *Vet Parasitol,* 168**,** 165-9.

PERIÑÁN, M. T., GÓMEZ-GARRE, P., BLAUWENDRAAT, C., MIR, P. & BANDRES-CIGA, S. 2021. The role of RHOT1 and RHOT2 genetic variation on Parkinson disease risk and onset. *Neurobiol Aging,* 97**,** 144.e1-144.e3.

PIPER, E. 2010. *Bovine immune responses to cattle tick infestation.* The University of Queensland.

PIPER, E. K., JACKSON, L. A., BAGNALL, N. H., KONGSUWAN, K. K., LEW, A. E. & JONSSON, N. N. 2008. Gene expression in the skin of Bos taurus and Bos indicus cattle infested with the cattle tick, Rhipicephalus (Boophilus) microplus. *Veterinary Immunology and Immunopathology,* 126**,** 110-119.

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**,** 431-41.

PIPER, E. K., JONSSON, N. N., GONDRO, C., LEW-TABOR, A. E., MOOLHUIJZEN, P., VANCE, M. E. & JACKSON, L. A. 2009. Immunological profiles of Bos taurus and Bos indicus cattle infested with the cattle tick, Rhipicephalus (Boophilus) microplus. *Clin Vaccine Immunol,* 16**,** 1074-86.

PIPER, E. K., JONSSON, N. N., GONDRO, C., VANCE, M. E., LEW-TABOR, A. & JACKSON, L. A. 2017. Peripheral cellular and humoral responses to infestation with the cattle tick Rhipicephalus microplus in Santa Gertrudis cattle. *Parasite Immunol,* 39.

POLLARD, M. O., GURDASANI, D., MENTZER, A. J., PORTER, T. & SANDHU, M. S. 2018. Long reads: their purpose and place. *Human Molecular Genetics,* 27**,** R234-R241.

PORTO NETO, L. R., BUNCH, R. J., HARRISON, B. E., PRAYAGA, K. C. & BARENDSE, W. 2010a. Haplotypes that include the integrin alpha 11 gene are associated with tick burden in cattle. *BMC Genet,* 11**,** 55.

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**,** 165-72.

PORTO NETO, L. R., JONSSON, N. N., INGHAM, A., BUNCH, R. J., HARRISON, B. E. & BARENDSE, W. 2012. The RIPK2 gene: a positional candidate for tick burden supported by genetic associations in cattle and immunological response of knockout mouse. *Immunogenetics,* 64**,** 379-88.

PORTO NETO, L. R., PIPER, E. K., JONSSON, N. N., BARENDSE, W. & GONDRO, C. Meta-analysis of Genome Wide Association and Gene Expression Studies to Identify Candidate Genes for Tick Burden in Cattle. 2010b. German Society for Animal Science.

PRAYAGA, K., CORBET, N., JOHNSTON, D., 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. *Animal Production Science,* 49**,** 413-425.

PRINGLE, E. S., FIRTH, M. A., CHATTHA, K. S., HODGINS, D. C. & SHEWEN, P. E. 2012. Expression of complement receptors 1 (CR1/CD35) and 2 (CR2/CD21), and co-signaling molecule CD19 in cattle. *Developmental & Comparative Immunology,* 38**,** 487-494.

PUGLIA, G. D., PRJIBELSKI, A. D., VITALE, D., BUSHMANOVA, E., SCHMID, K. J. & RACCUIA, S. A. 2020. Hybrid transcriptome sequencing approach improved assembly and gene annotation in Cynara cardunculus (L.). *BMC genomics,* 21**,** 1-17.

QIAN, Y., KANG, Z., LIU, C. & LI, X. 2010. IL-17 signaling in host defense and inflammatory diseases. *Cellular & Molecular Immunology,* 7**,** 328-333.

QUEENSLAND GOVERNMENT, B. Q. 2016. *Cattle tick fever* [Online]. Available: https://www.business.qld.gov.au/industries/farms-fishing-forestry/agriculture/biosecurity/animals/diseases/cattle-tick-fever/manage [Accessed 7 Dec 2023].

QUINLAN, A. R. & HALL, I. M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics,* 26**,** 841-842.

RAO, M. S., VAN VLEET, T. R., CIURLIONIS, R., BUCK, W. R., MITTELSTADT, S. W., BLOMME, E. A. G. & LIGUORI, M. J. 2018. Comparison of RNA-Seq and Microarray Gene Expression Platforms for the Toxicogenomic Evaluation of Liver From Short-Term Rat Toxicity Studies. *Front Genet,* 9**,** 636.

RAZA, A., SCHULZ, B. L., NOUWENS, A., JACKSON, L. A., PIPER, E. K., JAMES, P., JONSSON, N. N. & TABOR, A. E. 2021. Serum proteomes of Santa Gertrudis cattle before and after infestation with Rhipicephalus australis ticks. *Parasite Immunology,* 43**,** e12836.

RAZA, A., SCHULZ, B. L., NOUWENS, A., NASEEM, M. N., KAMRAN, M., MANTILLA VALDIVIESO, E. F., KERR, E. D., CONSTANTINOIU, C., JONSSON, N. N., JAMES, P. & TABOR, A. E. 2023. Application of quantitative proteomics to discover biomarkers for tick resistance in cattle. *Frontiers in Immunology,* 14.

REED, E., NUNEZ, S., KULP, D., QIAN, J., REILLY, M. P. & FOULKES, A. S. 2015. A guide to genome-wide association analysis and post-analytic interrogation. *Statistics in Medicine,* 34**,** 3769-3792.

REGITANO, L. & PRAYAGA, K. 2010. Breeding for Disease Resistance in Farm Animals. CABI Publishing: Oxon, UK.

RIEK, R. 1962. Studies on the reactions of animals to infestation with ticks. VI. Resistance of cattle to infestation with the tick Boophilus microplus (Canestrini). *Australian Journal of Agricultural Research,* 13**,** 532-550.

ROBBERTSE, L., RICHARDS, S. A., CLIFT, S. J., BARNARD, A. C., LEISEWITZ, A., CRAFFORD, J. E. & MARITZ-OLIVIER, C. 2018. Comparison of the differential regulation of T and B-lymphocyte subsets in the skin and lymph nodes amongst three cattle breeds as potential mediators of immune-resistance to Rhipicephalus microplus. *Ticks Tick Borne Dis,* 9**,** 976-987.

ROBBERTSE, L., RICHARDS, S. A., STUTZER, C., OLIVIER, N. A., LEISEWITZ, A. L., CRAFFORD, J. E. & MARITZ-OLIVIER, C. 2020. Temporal analysis of the bovine lymph node transcriptome during cattle tick (Rhipicephalus microplus) infestation. *Vaccine,* 38**,** 6889-6898.

ROBINSON, D. Join Tables Together on Inexact Matching [R package fuzzyjoin version 0.1.6]. 2020.

ROBINSON, M. D., MCCARTHY, D. J. & SMYTH, G. K. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics,* 26**,** 139-40.

ROCHA, C. C., DA SILVA ANDRADE, S. C., DE MELO, G. D., MOTTA, I. G., COUTINHO, L. L., GONELLA-DIAZA, A. M., BINELLI, M. & PUGLIESI, G.

2020. Early pregnancy-induced transcripts in peripheral blood immune cells in Bos indicus heifers. *Scientific Reports,* 10**,** 13733.
RODRIGUEZ-VIVAS, R. I., ALONSO-DÍAZ, M. A., RODRÍGUEZ-AREVALO, F., FRAGOSO-SANCHEZ, H., SANTAMARIA, V. M. & ROSARIO-CRUZ, R. 2006. Prevalence and potential risk factors for organophosphate and pyrethroid resistance in Boophilus microplus ticks on cattle ranches from the State of Yucatan, Mexico. *Vet Parasitol,* 136**,** 335-42.
RODRÍGUEZ-VIVAS, R. I., HODGKINSON, J. E. & TREES, A. J. 2012. Resistencia a los acaricidas en Rhipicephalus (Boophilus) microplus: situación actual y mecanismos de resistencia. *Revista Mexicana de Ciencias Pecuarias,* 3**,** 9-24.
RODRIGUEZ-VIVAS, R. I., JONSSON, N. N. & BHUSHAN, C. 2018. Strategies for the control of Rhipicephalus microplus ticks in a world of conventional acaricide and macrocyclic lactone resistance. *Parasitology Research,* 117**,** 3-29.
ROY, S. K., SHRIVASTAVA, A., SRIVASTAV, S., SHANKAR, S. & SRIVASTAVA, R. K. 2020. SATB2 is a novel biomarker and therapeutic target for cancer. *J Cell Mol Med,* 24**,** 11064-11069.
RSTUDIO TEAM 2022. RStudio: Integrated Development Environment for R. 2022.07.576 ed.: RStudio, PBC.
SABINO, M., CARMELO, V. A. O., MAZZONI, G., CAPPELLI, K., CAPOMACCIO, S., AJMONE-MARSAN, P., VERINI-SUPPLIZI, A., TRABALZA-MARINUCCI, M. & KADARMIDEEN, H. N. 2018. Gene co-expression networks in liver and muscle transcriptome reveal sex-specific gene expression in lambs fed with a mix of essential oils. *BMC Genomics,* 19**,** 236.
SAFIULINA, D., KUUM, M., CHOUBEY, V., HICKEY, M. A. & KAASIK, A. 2019. Mitochondrial transport proteins RHOT1 and RHOT2 serve as docking sites for PRKN-mediated mitophagy. *Autophagy,* 15**,** 930-931.
SALLEH, S., MAZZONI, G., LØVENDAHL, P. & KADARMIDEEN, H. N. 2018. Gene co-expression networks from RNA sequencing of dairy cattle identifies genes and pathways affecting feed efficiency. *BMC bioinformatics,* 19**,** 1-15.
SANTIAGO-LÓPEZ, L., AGUILAR-TOALÁ, J. E., HERNÁNDEZ-MENDOZA, A., VALLEJO-CORDOBA, B., LICEAGA, A. M. & GONZÁLEZ-CÓRDOVA, A. F. 2018. Invited review: Bioactive compounds produced during cheese ripening and health effects associated with aged cheese consumption. *J Dairy Sci,* 101**,** 3742-3757.
SAUNA, Z. E. & KIMCHI-SARFATY, C. 2011. Understanding the contribution of synonymous mutations to human disease. *Nature Reviews Genetics,* 12**,** 683-691.
SCHNEIDER, M. V. & ORCHARD, S. 2011. Omics technologies, data and bioinformatics principles. *Methods Mol Biol,* 719**,** 3-30.
SCHNELL, S., DÉMOLLIÈRE, C., VAN DEN BERK, P. & JACOBS, H. 2006. Gimap4 accelerates T-cell death. *Blood,* 108**,** 591-599.
SCHRUTH, D. 2023. *A Collection of Database, Data Structure, Visualization, and Utility Functions for R* [Online]. Available: https://github.com/cran/caroline [Accessed].
SCHWEFEL, D., ARASU, B. S., MARINO, S. F., LAMPRECHT, B., KÖCHERT, K., ROSENBAUM, E., EICHHORST, J., WIESNER, B., BEHLKE, J. & ROCKS, O. 2013. Structural insights into the mechanism of GTPase activation in the GIMAP family. *Structure,* 21**,** 550-559.
SCULLY, O. J., CHUA, P. J., HARVE, K. S., BAY, B. H. & YIP, G. W. 2012. Serglycin in health and diseases. *Anat Rec (Hoboken),* 295**,** 1415-20.
SEIFERT, G. W. 1971. Variations between and within breeds of cattle in resistance to field infestations of the cattle tick (Boophilus microplus). *Crop & Pasture Science,* 22**,** 159-168.
SEKI, M., KATSUMATA, E., SUZUKI, A., SEREEWATTANAWOOT, S., SAKAMOTO, Y., MIZUSHIMA-SUGANO, J., SUGANO, S., KOHNO, T., FRITH, M. C. &

TSUCHIHARA, K. 2019. Evaluation and application of RNA-Seq by MinION. *DNA Research,* 26**,** 55-65.

SHANNON, P., MARKIEL, A., OZIER, O., BALIGA, N. S., WANG, J. T., RAMAGE, D., AMIN, N., SCHWIKOWSKI, B. & IDEKER, T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research,* 13**,** 2498-2504.

SHAUL, O. 2017. How introns enhance gene expression. *Int J Biochem Cell Biol,* 91**,** 145-155.

SHI, J., XIONG, R., ZHOU, T., SU, P., ZHANG, X., QIU, X., LI, H., LI, S., YU, C., WANG, B., DING, C., SMITHGALL, T. E. & ZHENG, Y. H. 2018. HIV-1 Nef Antagonizes SERINC5 Restriction by Downregulation of SERINC5 via the Endosome/Lysosome System. *J Virol,* 92.

SHYMA, K., GUPTA, J. P. & SINGH, V. 2015. Breeding strategies for tick resistance in tropical cattle: a sustainable approach for tick control. *Journal of Parasitic Diseases,* 39**,** 1-6.

SNOWBALL, G. 1956. The effect of self-licking by cattle on infestations of cattle tick, <I>Boophilus microplus</I> (Canestrini). *Australian Journal of Agricultural Research,* 7**,** 227-232.

SODELAND, M., KENT, M., OLSEN, H., OPSAL, M., SVENDSEN, M., SEHESTED, E., HAYES, B. & LIEN, S. 2011. Quantitative trait loci for clinical mastitis on chromosomes 2, 6, 14 and 20 in Norwegian Red cattle. *Animal Genetics,* 42**,** 457-465.

SOLLERO, B. P., JUNQUEIRA, V. S., GOMES, C. C., 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. *Genetics Selection Evolution,* 49**,** 1-15.

SOPP, P. 1996. Ruminant cluster CD21. *Veterinary immunology and immunopathology,* 52**,** 249.

STARK, R., GRZELAK, M. & HADFIELD, J. 2019. RNA sequencing: the teenage years. *Nature Reviews Genetics,* 20**,** 631-656.

STELZER, G., ROSEN, N., PLASCHKES, I., ZIMMERMAN, S., TWIK, M., FISHILEVICH, S., STEIN, T. I., NUDEL, R., LIEDER, I. & MAZOR, Y. 2016. The GeneCards suite: from gene data mining to disease genome sequence analyses. *Current protocols in bioinformatics,* 54**,** 1.30. 1-1.30. 33.

STEWART, N. P., DALGLIESH, R. J. & TRATT, T. 1981. Life cycles of the cattle tick and tick fever parasites. *Queensland Agricultural Journal,* 107**,** 305-308.

SUN, B., XU, L., BI, W. & OU, W. B. 2022. SALL4 Oncogenic Function in Cancers: Mechanisms and Therapeutic Relevance. *Int J Mol Sci,* 23.

SUNG, Y. J., SCHWANDER, K., ARNETT, D. K., KARDIA, S. L., RANKINEN, T., BOUCHARD, C., BOERWINKLE, E., HUNT, S. C. & RAO, D. C. 2014. An empirical comparison of meta-analysis and mega-analysis of individual participant data for identifying gene-environment interactions. *Genet Epidemiol,* 38**,** 369-78.

SUTHERST, R. W., KERR, J. D., MAYWALD, G. F. & STEGEMAN, D. A. 1983. The effect of season and nutrition on the resistance of cattle to the tick Boophilus microplus. *Crop & Pasture Science,* 34**,** 329-339.

SZKLARCZYK, D., GABLE, A. L., LYON, D., JUNGE, A., WYDER, S., HUERTA-CEPAS, J., SIMONOVIC, M., DONCHEVA, N. T., MORRIS, J. H., BORK, P., JENSEN, L. J. & MERING, C. V. 2019. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res,* 47**,** D607-d613.

TABOR, A. E., ALI, A., REHMAN, G., ROCHA GARCIA, G., ZANGIROLAMO, A. F., MALARDO, T. & JONSSON, N. N. 2017. Cattle Tick Rhipicephalus microplus-

Host Interface: A Review of Resistant and Susceptible Host Responses. *Frontiers in Cellular and Infection Microbiology*, 7.
TAM, V., PATEL, N., TURCOTTE, M., BOSSÉ, Y., PARÉ, G. & MEYRE, D. 2019. Benefits and limitations of genome-wide association studies. *Nature Reviews Genetics,* 20**,** 467-484.
TANAKA, Y., ONO, N., SHIMA, T., TANAKA, G., KATOH, Y., NAKAYAMA, K., TAKATSU, H. & SHIN, H.-W. 2016. The phospholipid flippase ATP9A is required for the recycling pathway from the endosomes to the plasma membrane. *Molecular biology of the cell,* 27**,** 3883-3893.
TATCHELL, R. J. & BENNETT, G. F. 1969. Boophilus microplus: antihistaminic and tranquilizing drugs and cattle resistance. *Exp Parasitol,* 26**,** 369-77.
TATCHELL, R. J. & MOORHOUSE, D. E. 1970. Neutrophils: their role in the formation of a tick feeding lesion. *Science,* 167**,** 1002-3.
TECHNOLOGIES, O. N. 2008. *PCR-cDNA barcoding Kit (SQK-PCB109)* [Online]. ONT. Available: https://nanoporetech.com/ [Accessed].
THOMAS, D. B., RENFU, S. & STEPHEN, C. B. 2014. Phylogenetic analysis of mitochondrial genome sequences indicates that the cattle tick, Rhipicephalus (Boophilus) microplus, contains a cryptic species. *Molecular Phylogenetics and Evolution,* 76**,** 241-253.
TIEZZI, F., PARKER-GADDIS, K. L., COLE, J. B., CLAY, J. S. & MALTECCA, C. 2015. A genome-wide association study for clinical mastitis in first parity US Holstein cows using single-step approach and genomic matrix re-weighting procedure. *PLoS One,* 10**,** e0114919.
TILGNER, H., RAHA, D., HABEGGER, L., MOHIUDDIN, M., GERSTEIN, M. & SNYDER, M. 2013. Accurate identification and analysis of human mRNA isoforms using deep long read sequencing. *G3: Genes, Genomes, Genetics,* 3**,** 387-397.
TIZARD, I. A. N. R. *IMUNOLOGIA VETERINARIA: UMA INTRODUÇAO*, ELSEVIER (MEDICINA).
TODD, D. L., WOOLLIAMS, J. A. & ROUGHSEDGE, T. 2011. Gene flow in a national cross-breeding beef population. *Animal,* 5**,** 1874-86.
TRÍBULO, P., RABAGLINO, M. B., BO, M. B., CARVALHEIRA, L. D. R., BISHOP, J. V., HANSEN, T. R. & HANSEN, P. J. 2019. Dickkopf-related protein 1 is a progestomedin acting on the bovine embryo during the morula-to-blastocyst transition to program trophoblast elongation. *Scientific Reports,* 9**,** 11816.
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. *Animal Production Science,* 50**,** 235-245.
UNTALAN, P. M., PRUETT, J. H. & STEELMAN, C. D. 2007. Association of the bovine leukocyte antigen major histocompatibility complex class II DRB3*4401 allele with host resistance to the Lone Star tick, Amblyomma americanum. *Vet Parasitol,* 145**,** 190-5.
USAMI, Y., WU, Y. & GÖTTLINGER, H. G. 2015. SERINC3 and SERINC5 restrict HIV-1 infectivity and are counteracted by Nef. *Nature,* 526**,** 218-23.
UTECH, K., WHARTON, R. & KERR, J. 1978a. Resistance to Boophilus microplus (Canestrini) in different breeds of cattle. *Australian Journal of Agricultural Research,* 29**,** 885-895.
UTECH, K. B. W., SEIFERT, G. W. & WHARTON, R. H. 1978b. Breeding Australian Illawarra Shorthorn cattle for resistance to Boophilus microplus. I. Factors affecting resistance. *Crop & Pasture Science,* 29**,** 411-422.
VAN EENENNAAM, A. L. & YOUNG, A. E. 2014. Prevalence and impacts of genetically engineered feedstuffs on livestock populations. *J Anim Sci,* 92**,** 4255-78.

WAGLAND, B. 1978. Host resistance to cattle tick (Boophilus microplus) in Brahman (Bos indicus) cattle. III.* Growth on previously unexposed animals. *Australian Journal of Agricultural Research,* 29**,** 401-409.

WANG, L., YU, J., NI, J., XU, X.-M., WANG, J., NING, H., PEI, X.-F., CHEN, J., YANG, S. & UNDERHILL, C. B. 2003. Extracellular matrix protein 1 (ECM1) is over-expressed in malignant epithelial tumors. *Cancer letters,* 200**,** 57-67.

WANG, Y., REVERTER, A., KEMP, D., MCWILLIAM, S., INGHAM, A., DAVIS, C., MOORE, R. & LEHNERT, S. 2007. Gene expression profiling of Hereford Shorthorn cattle following challenge with Boophilus microplus tick larvae. *Australian Journal of Experimental Agriculture,* 47**,** 1397-1407.

WANG, Y., ZHAO, Y., BOLLAS, A., WANG, Y. & AU, K. F. 2021. Nanopore sequencing technology, bioinformatics and applications. *Nature Biotechnology,* 39**,** 1348-1365.

WHARTON, R. H., UTECH, K. B. W. & TURNER, H. G. 1970. Resistance to the cattle tick, Boophilus microplus in a herd of Australian Illawarra Shorthorn cattle: its assessment and heritability. *Crop & Pasture Science,* 21**,** 163-181.

WIKEL, S. K. 1988. Immunological control of hematophagous arthropod vectors: utilization of novel antigens. *Veterinary Parasitology,* 29**,** 235-264.

WILLADSEN, P., BIRD, P., COBON, G. S. & HUNGERFORD, J. 1995. Commercialisation of a recombinant vaccine against Boophilus microplus. *Parasitology,* 110 Suppl**,** S43-50.

WILLADSEN, P. & KEMP, D. H. 1988. Vaccination with 'concealed' antigens for tick control. *Parasitology Today,* 4**,** 196-198.

WILLADSEN, P., RIDING, G., MCKENNA, R., KEMP, D., TELLAM, R., NIELSEN, J., LAHNSTEIN, J., COBON, G. & GOUGH, J. 1989. Immunologic control of a parasitic arthropod. Identification of a protective antigen from Boophilus microplus. *Journal of immunology (Baltimore, Md.: 1950),* 143**,** 1346-1351.

YAMADA, K. M. 1991. Adhesive recognition sequences. *Journal of Biological Chemistry,* 266**,** 12809-12812.

YUAN, J., PALIOURA, S., SALAZAR, J. C., SU, D., O'DONOGHUE, P., HOHN, M. J., CARDOSO, A. M., WHITMAN, W. B. & SÖLL, D. 2006. RNA-dependent conversion of phosphoserine forms selenocysteine in eukaryotes and archaea. *Proc Natl Acad Sci U S A,* 103**,** 18923-7.

ZELTZ, C. & GULLBERG, D. 2016. Integrin Alpha11 (ITGA11). *In:* CHOI, S. (ed.) *Encyclopedia of Signaling Molecules.* New York, NY: Springer New York.

ZHANG, B. & HORVATH, S. 2005. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol,* 4**,** Article17.

ZHANG, C., XIE, L. & LIN, Z. 2023. Homeobox-D 1 and FTO form a transcriptional–epigenetic feedback loop to promote head and neck cancer proliferation. *Cell Biology International,* 47**,** 1987-1998.

ZHANG, X., YIN, M. & ZHANG, L. J. 2019. Keratin 6, 16 and 17-Critical Barrier Alarmin Molecules in Skin Wounds and Psoriasis. *Cells,* 8.

ZHENG, W., LIN, Q., ISSAH, M. A., LIAO, Z. & SHEN, J. 2021. Identification of PLA2G7 as a novel biomarker of diffuse large B cell lymphoma. *BMC Cancer,* 21**,** 927.

ZHONG, J., SHI, S., PENG, W., LIU, B., YANG, B., NIU, W., ZHANG, B., QIN, C., ZHONG, D., CUI, H., ZHANG, Z. & SUN, X. 2022. Weighted Gene Co-Expression Network Analysis (WGCNA) Reveals the Functions of Syndecan-1 to Regulate Immune Infiltration by Influenced T Cells in Glioma. *Front Genet,* 13**,** 792443.

ZIEGLER-HEITBROCK, H. W. & ULEVITCH, R. J. 1993. CD14: cell surface receptor and differentiation marker. *Immunol Today,* 14**,** 121-5.

ZIMIN, A. V., DELCHER, A. L., FLOREA, L., KELLEY, D. R., SCHATZ, M. C., PUIU, D., HANRAHAN, F., PERTEA, G., VAN TASSELL, C. P. & SONSTEGARD, T.

S. 2009. A whole-genome assembly of the domestic cow, Bos taurus. *Genome biology,* 10**,** 1-10.