

**Identification of candidate signature genes and key regulators
associated with trypanotolerance in the Sheko breed**

Dissertation submitted

To obtain the Doctor of philosophy (Ph.D) degree
at the Faculty of Agricultural Sciences,
Georg-August-University of Göttingen, Germany

Presented by

Yonatan Ayalew Mekonnen
Born in Dessie, Ethiopia

Göttingen, December 2019

D7

- Reference 1: Prof. Dr. Armin Otto Schmitt
Breeding Informatics Group Göttingen
Department of Animal Sciences
Georg-August-University Göttingen, Germany
- Reference 2: Prof. Dr. Henner Simianer
Animal Breeding and Genetics Group
Department of Animal Sciences
Georg-August-University Göttingen, Germany
- Reference 3: Prof. Dr. Olivier Hanotte
Cells, Organisms and Molecular Genetics Group
School of Life Sciences
University of Nottingham, Nottingham, United Kingdom
LiveGene, International Livestock Research Institute (ILRI), Addis Ababa, Ethiopia

Date of dissertation: 31 January 2020

Abstract

African Animal Trypanosomiasis (AAT) is caused by a protozoan parasite that affects the health of livestock. Livestock production in Ethiopia is severely hampered by AAT and various controlling measures were not successful to eradicate the disease. AAT affects the indigenous breeds in varying degrees. However, the Sheko breed shows better trypanotolerance than other breeds. The tolerance attributes of Sheko are believed to be associated with its taurine genetic background but the genetic controls of these tolerance attributes of Sheko are not well understood. In order to investigate the level of taurine background in the genome, the genome of Sheko is compared with that of 11 other African breeds. The result shows that Sheko has an admixed genome composed of taurine and indicine ancestries. To identify selective sweeps in the Sheko genome, three methods were applied: (i) The integrated haplotype score (*iHS*), (ii) the standardized log ratio of integrated site specific extended haplotype homozygosity (EHH) between populations (*Rsb*), and (iii) the composite likelihood ratio (CLR) method. The combined results of these methods reveal 99 genomic regions harboring 364 signature genes in Sheko. Out of the signature genes, 15 genes are selected based on their biological importance described in the literature. In addition, 13 over-represented pathways and 10 master regulators are identified in Sheko using the TRANSPATH database in the geneXplain platform. Most of the pathways are related to oxidative stress responses indicating a possible selection response against the induction of oxidative stress following trypanosomiasis infection in Sheko. Moreover, the trypanotolerance tendencies of the Nuer, Benshangul, and Gindeberet breeds are assessed by comparing the candidate genomic regions, genes, hub genes, overrepresented pathways, and master regulators identified in each breed with Sheko and among themselves. In addition, the identified genes and genomic regions are compared with the trypanotolerant QTL regions in N'Dama, and genes as well as genomic regions of Muturu. The common genomic regions and genes in Nuer, Benshangul, and Gindeberet that are shared in common with Sheko, N'Dama, and Muturu are identified. Furthermore, the hub genes, overrepresented pathways, and master regulators in Nuer, Benshangul, and Gindeberet breeds which are in common with Sheko are identified. These results indicate that the Nuer, Benshangul, and Gindeberet breeds have undergone similar evolutionary responses against trypanosomiasis. The findings reported in this thesis show that the master regulator Caspase which is identified in Sheko, Nuer, and Benshangul is a key protease that plays a major role in the emergence of adaptive immunity in harmony with the other master regulators. In this thesis, I present for the first time the importance of master regulators involved in trypanotolerance not only for the breeds included in this thesis but also in the context of cattle genomics. These results suggest that designing and implementing genetic intervention strategies is necessary to improve the performance of susceptible animals. Moreover, the identification of master regulator Caspase suggests potential candidate therapeutic targets for the development of new drugs for trypanosomiasis treatment.

Zusammenfassung

Die Afrikanische Trypanosomiasis wird durch ein parasitisches Protozoon verursacht, das die Gesundheit von Nutztieren beeinträchtigt. Die Tierproduktion in Äthiopien wird durch diese Krankheit erheblich gestört. Diverse Maßnahmen zu ihrer Eindämmung blieben erfolglos. Die Afrikanische Trypanosomiasis befällt einheimische Rinderrassen in unterschiedlichem Ausmaß. Die Rasse Sheko weist jedoch ein höheres Toleranzniveau als andere Rassen auf. Die Toleranzeigenschaften bei Sheko werden gemeinhin mit seinem taurinen Genomanteil in Verbindung gebracht, sind jedoch auf genetischer Ebene noch nicht verstanden. Um den taurinen Anteil des Shekogenoms zu untersuchen, wird es mit elf anderen afrikanischen Rassen verglichen. Es zeigte sich, dass das Shekogenom sowohl taurine als auch indicine Wurzeln hat. Um selektive Sweeps im Shekogenom aufzudecken, werden drei Methoden angewandt: (i) der integrierte Haplotypen-Score (*iHS*), (ii) das standardisierte logarithmierte Verhältnis der integrierten positionsspezifischen erweiterten Haplotypenhomozygotität (EHH) zwischen Populationen (*Rsb*) und (iii) die zusammengesetzte Likelihood-Verhältnis-Methode (CLR). Die zusammengeführten Ergebnisse dieser drei Methoden umfassen 99 genomische Regionen mit 364 sogenannten Signatur-Genen in Sheko. Unter diesen Genen wurden aufgrund ihrer in Publikationen dokumentierten biologischen Bedeutung 15 Gene ausgewählt. Zusätzlich wurden 13 überrepräsentierte Pathways und zehn Master-Regulatoren basierend auf Einträgen in der TRANSPATH-Datenbank der geneXplain-Plattform ermittelt. Die meisten dieser Pathways sind mit Reaktionen auf oxidativen Stress verknüpft, was eine mögliche Reaktion auf oxidativen Stress aufgrund der Trypanosomiasis-Infektion bei Sheko nahelegt. Weiterhin wurde die ansatzweise feststellbare Trypanotoleranz bei den Rassen Nuer, Benshangul und Gindeberet untersucht, indem die genomischen Kandidatenregionen, Gene, Schlüsselgene, überrepräsentierte Pathways und Master-Regulatoren in jeder Rasse mit denen von Sheko und untereinander verglichen. Zusätzlich wurden die identifizierten Gene und genomischen Regionen mit QTLs für Trypanotoleranz bei N'Dama und Muturu verglichen. Die gemeinsamen genomischen Regionen und Gene in Nuer, Benshangul und Gindeberet einerseits und in Sheko, N'Dama und Muturu andererseits wurden identifiziert. Die Schlüsselgene, überrepräsentierten Pathways und Master-Regulatoren, die Nuer, Benshangul und Gindeberet mit Sheko gemeinsam haben, wurden identifiziert. Die Ergebnisse legen nahe, dass Nuer, Benshangul und Gindeberet durch Trypanosomiasis evolutionär ähnlich geformt wurden. Die Ergebnisse in dieser Arbeit zeigen, dass der Master-Regulator Caspase, der in Sheko, Nuer und Benshangul gefunden wurde, eine Schlüssel-Protease ist, die zusammen mit anderen Master-Regulatoren eine wichtige Rolle beim Aufkommen einer adaptiven Immunität spielt. In dieser Arbeit wurde zum ersten Mal die Wichtigkeit von Master-Regulatoren bei der Trypanotoleranz nicht nur der hier behandelten Rassen, sondern allgemein des Rinds, aufgezeigt. Diese Ergebnisse legen nahe, dass genetische Interventionsstrategien notwendig sind, um die Leistung anfälliger Tiere zu steigern. Weiterhin zeigt die Identifikation des Master-Regulators Caspase potentielle therapeutische Targets für die Entwicklung neuer Wirkstoffe zur Behandlung von Trypanosomiasis auf.

Acknowledgements

First and for most I would like to thank the almighty God for his mercy and giving me the strength to accomplish this PhD thesis. I would like to thank his mother Saint Marry for keeping me in her intercession with God. Then, I would like to thank my main supervisor Professor Armin Schmitt for giving me the chance in his group to work on this interesting topic. You have given me unreserved support during my stay in your group. Your inspiring scientific discussions, sincerity, visions, and motivation guided me to see the bigger picture of science. Your advice at each step of the project was very helpful. Thank you a lot!

I would like to thank my second supervisor Professor Henner Simianer. I appreciate your role as a facilitator to enroll in this PhD position and for your continuous support as a supervisor. Your guidance and valuable insights were helpful to improve this PhD research.

Further, I would like to extend my sincere gratitude to Professor Olivier Hanotte. This work would not have been possible without your support in all stages of my PhD research. Thank you for your warm welcome during our visit to ILRI.

I am indebted to forward my special thanks to Dr. Mehmet Gültas for providing guidance, support, and feedback during my PhD research carrier. Especially, your advises and constructive comments were helpful for shaping my thesis. Thank you, Mehmet!

In addition, I would like to thank Ms. Siebert Monika. You have been helping me in all issues whenever I needed your help. I also thank the breeding informatics group in general for accompanying me throughout these periods. Selina, Faisal, Abirami, Hendrik, Martin, and Felix thank you for proofreading this thesis.

I would like to thank my father, Ayalew Mekonnen and my mother Aselefech Haile. My siblings and family members, Meron, Yishak, Mersha, Misaye, Tenagne, Beteal, Alemtsehay, Dereje, and Endalkachew thank you for your continued care, support, and encouragement. Your love, support and encouragement were the basis of my strength to accomplish this research.

Last, but not least, to my wife, Dr. Bethelhem Legesse (Betty), I am so lucky to have you in my life. You have been extremely patient and supportive throughout this project. I can not thank you enough for your countless scarifies to get me reach at the final stages of this research. My son Bereket, and my daughter Rediet, thank you for providing necessary breaks from science and for creating a refreshing atmosphere throughout the entire process. This thesis is dedicated to them. Thank you!

Contents

1. Introduction	1
1.1. Structure of the thesis	3
1.2. Impact	4
2. Biological background	5
2.1. DNA, Gene, Genome and Mutation	5
2.2. Haplotype	5
2.3. Biotic Stress	6
2.4. Molecular Mechanisms of Adaptation	8
2.5. Signature of Selection	8
2.6. Genome Wide Association Study	8
2.7. Bioinformatic Databases and Tools	10
2.7.1. TRANSPATH	11
2.7.2. GeneMANIA plugin in Cytoscape platform	12
2.7.3. BEAGLE	13
2.7.4. ADMIXTURE	13
2.7.5. PLINK	14
3. Theoretical background	15
3.1. Extended Haplotype Homozygosity	15
3.1.1. Integrated Haplotype Score	15
3.1.2. Relative Integrated EHHS between populations	16
3.2. Composite Likelihood Ratio	16
4. Materials and Methods	19
4.1. Data, SNP Genotyping and Quality Control	19
4.2. Genetic Background of the Cattle Population	20
4.3. Breed Differentiation, Genetic Relationship and Structure	20
4.4. Analysis of Signatures of Positive Selection	21
4.4.1. Extended Haplotype Homozygosity Based Methods	21
4.4.2. Spatial Distribution of Allele Frequency Based Method	22
4.5. Functional Annotation of Selected Candidate Regions	22
4.6. Comparative Identification of Trypanotolerance Tendency	22
4.7. Comparative Identification of Signature of Selection	23
4.8. Summary of the Analysis of Signatures of Positive Selection	23

5. Results	27
5.1. The Genetic Relationship and Structure of Cattle Populations	27
5.2. Identification of Candidate Signature Genes Associated with Trypanotolerance .	28
5.3. Functional Annotation of Candidate Signature Genes	36
5.4. Identification of Overrepresented Pathways in the Candidate Signature Gene Sets	36
5.5. Identification of Master Regulators based on Candidate Signature Genes	39
5.6. Identification of Common Candidate Signature Genes Using GWAS and Signature of Selection Detecting Methods	40
5.7. Comparative identification of trypanotolerance tendency	43
5.7.1. Comparative Identification of Signature Genes Associated with Trypanotolerance	43
5.7.2. Treemap Comparisons Between Breeds	44
5.7.3. Comparison with N'Dama and Muturu for Signature of Positive Selection against Trypanosomiasis	45
5.7.4. Comparative Identification of Hub Genes Involved in Trypanotolerance .	46
5.7.5. Comparative Identification of Overrepresented Pathway Involved in Trypanotolerance	46
5.7.6. Comparative Identification of Master Regulators	55
6. Discussion	59
6.1. Population Structure and Genetic Relationship	59
6.1.1. Identification of Candidate Signature Genes Associated with Trypanotolerance	60
6.2. Identification of Overrepresented Pathways in the Candidate Signature Gene Sets	61
6.3. Identification of Master Regulators based on Candidate Signature Genes	62
6.4. Comparative Study for the Identification of Trypanotolerance Tendency	63
7. Conclusion	67
7.1. Summary	67
7.2. Outlook	69
Bibliography	71
A. Appendix	98
A.1. Identification of Candidate Signature Genes and Key Regulators Associated with Trypanotolerance in the Sheko Breed	98
A.2. Genetic Diversity and Population Structure of Six Ethiopian Cattle Breeds from Different Geographical Regions Using High Density Single Nucleotide Polymorphism Markers	119
A.3. Evolutionary Responses Associated with Trypanotolerance in Three Ethiopian Breeds – A Comparative Genomics Approach	125

List of Figures

2.1. Biotic stress-response model	7
2.2. Schematic illustration of background selection and selective sweep	9
2.3. Schematic illustration of case-control genome wide association study	10
2.4. TRANSPATH database integrated in the geneXplain platform	12
2.5. Interface of the GeneMANIA plugin in the Cytoscape platform	13
4.1. Analysis work-flow	25
5.1. PCA plots of the first two principal components	30
5.2. Admixture bar plots	31
5.3. Cross validation error in dependence to K	32
5.4. Venn diagrams of the overlapping genomic regions and candidate genes	32
5.5. Manhattan plots of genome-wide <i>iHS</i> , <i>Rsb</i> , and CLR analyses	33
5.6. Gene Ontology treemap for the 260 functionally enriched genes	41
5.7. The master regulatory networks identified in Sheko	42
5.8. UpSet plot of the overlapping candidate genes and genomic regions identified by <i>iHS</i> , CLR and <i>Rsb</i>	44
5.9. Gene Ontology treemap comparisons for the functionally enriched genes identified in Nuer breed	45
5.10. Gene Ontology treemap comparisons for the functionally enriched genes identified in Gindeberet breed	48
5.11. Gene Ontology treemap comparisons for the functionally enriched genes identified in Benshangul breed	49
5.12. Gene networks identified in Benshangul	50
5.13. Gene networks identified in Gindeberet	51
5.14. Gene networks identified in Nuer	52
5.15. Gene networks identified in Sheko	53
5.16. The master regulatory networks identified in Benshangul	56
5.17. The master regulatory networks identified in Gindeberet	57
5.18. The master regulatory networks identified in Nuer	58

List of Tables

4.1. Cattle breeds included in the study	20
5.1. Proportion of admixture within each of the analyzed breeds	28
5.2. Summary of major candidate signature regions identified by CLR, <i>iHS</i> , and <i>Rsb</i> analyses	34
5.3. Overrepresented pathways for the identified candidate signature genes	37
5.4. Overrepresented pathway analyses of the three breeds	54

1. Introduction

In most parts of the world, livestock production plays a critical role in rural economies. Especially, in African countries livestock are the basis for food security and income for smallholder farmers and pastoralist [1, 2, 3]. However, livestock production in Africa is severely affected by diseases among which trypanosomiasis is the most prevalent livestock disease [4]. Trypanosomiasis is caused by uni-cellular protozoan parasites which affects the health of humans and livestock. In Africa, this disease is referred to as African Animal Trypanosomiasis (AAT) [1, 3]. AAT is the major livestock production constraint especially in sub-Saharan African countries. It is mainly caused by *Trypanosoma congolense*, *Trypanosoma vivax*, and *Trypanosoma brucei brucei* [5, 2, 6, 7]. Particularly, *T. congolense* is the most frequent cause of livestock disease in this region [8]. The disease is transmitted from infected animals to healthy animals by tsetse fly as a vector [9]. The infected animal shows symptoms such as anemia [8, 10], nervous symptoms [11, 12], reduced productivity, infertility, abortion [13], listlessness and emaciation [14, 15, 16, 17]. If not treated, it can lead to death [1, 12, 13]. AAT severely affects the draft power as well as meat and milk production of the animals [18]. Therefore, AAT is considered as a major challenge constraining the path towards ensuring food security and combating poverty in this region [19]. Hence, this disease has a major economic impact that accounts for an annual loss of US\$ 5 billion in sub-Saharan countries [1, 12].

Ethiopia is one of the sub-Saharan African countries located in the eastern part of the tsetse belt. The tsetse fly distribution in the country spans from the south western to the north western regions covering 22,000 km² between longitude 33° and 38° East and latitude 5° and 12° North along river basins [20, 21]. About 14 million cattle, 7 million horses, 1.8 million camels, and 14 million small ruminants are kept in the infection zone [22]. In Ethiopia, Sheko is a cattle breed that shows better trypanotolerance attributes than other breeds such as Abigar and Horro [23]. Sheko is found in the southern region of the Bench Maji Zone, the adjoining areas of Keffa and Shaka and is considered as an endangered breed due to extensive interbreeding with local indicine and sanga breeds [24]. Sheko cattle are kept in the tsetse infested regions likely explaining their degree of trypanotolerance [25, 26].

Until now, a number of methods have been applied to control the spread of this disease such as trypanocidal drugs, insect traps and insecticides [12, 27, 28]. But none of these controlling measures have been successful to eradicate the disease. The current situation is deteriorating because the parasite became trypanocidal drug resistant due to inappropriate drug usage. Moreover, pharmaceutical companies are less attracted to invest in new drug discovery and development due to high cost [1, 29, 30, 31].

In general, to control the spread of trypanosomiasis disease, Lutje *et al.* [32] suggested an effective cross breeding strategy between trypanotolerant and susceptible cattle, together with vector control. Accordingly, Hanotte *et al.* [25] performed cross breeding between the west African trypanotolerant N'Dama and the east African trypanosusceptible Boran breeds to produce an F₂ population that shows heterosis. This leads to the assumption that an F₂ cross between trypanotolerant and susceptible breeds could produce a trypanotolerant synthetic breed whose performance exceeds that of either parents. Consequently, marker assisted selection, from the F₂ breed, would be the most promising strategy to produce a breed that combines high production and trypanotolerance [17, 25]. However, this requires the identification of genes and genomic regions controlling trypanotolerance. Despite the fact that there have been some efforts to identify genes and genomic regions controlling trypanotolerance in N'Dama [25, 33], reports on the genetic controls and mechanisms of trypanotolerance in cattle which are necessary for the implementation of marker assisted selection strategies are limited. Therefore, the aim of this thesis is to reveal the genetic controls and regulatory mechanisms that are involved in trypanotolerance in cattle breeds especially focusing on the Sheko breed since the trypanotolerance mechanisms of Sheko have not been well studied.

In this thesis, genotyping data is used to analyze the tolerance attributes of the Sheko breed at the molecular level in detail. The rationale is that natural or artificial selection targets the genome in response to environmental pressures or stresses as shaping adaptation and evolution. This implies that if the new allele of a mutation is beneficial (increases the fitness of their carriers) under certain environmental pressure or stress, then the frequency of these alleles will rapidly increase in the population [34]. Under positive selection, strong and long range linkage disequilibrium (LD) and unexpectedly high local haplotype homozygosity might occur in the genome [35, 36].

Likewise, trypanosomiasis is considered as an environmental pressure which plays a major role to create selection signatures in the genome and which is thus leading to breed formation [1, 2, 7]. These signs or traces of selection in the genome could be detected by using a "bottom-up" or a "from genotype to phenotype" approach [37]. In this thesis, I focus on traces or signs of positive selection in the genome of Sheko against trypanosomiasis using the "bottom-up" approach. In response to trypanosomiasis as the environmental pressure, the genome of Sheko could undergo changes at the molecular level. With the aim to identify the mechanism of Sheko tolerance, I use extended haplotype homozygosity (*iHS* and *Rsb*) and spatial distribution of allele frequency (CLR) based methods to identify genes that are associated with selection pressure in the Sheko breed. Combining methods for the detection of selection signature regions has been suggested as a means of increasing the power of the study compared to single analysis [38, 39].

In addition, other cattle breeds that are kept in the tsetse infested regions for several generations might display some degree of trypanotolerance due to adaptive responses against this particular environmental pressure [25, 26]. The Nuer, Benshangul, and Gindeberet breeds are also kept in the tsetse infested region in Ethiopia with trypanosome challenge [40, 41]. Therefore, these breeds might also have developed a certain level of trypanotolerance attributes. However, their level of trypanotolerance is not documented. In order to assess the trypanotolerance attributes of these breeds and to understand the underlying mechanisms of trypanotolerance, I perform comparative

genome wide signature gene identification using the total gene and genomic regions that are identified by *iHS*, CLR, and *Rsb* methods. Moreover, the genomic regions and genes identified by signature of selection detecting methods (*iHS*, CLR, and *Rsb*) are compared with the results of a genome wide association study (GWAS) to assess the signature genes commonly detected by these two approaches.

Furthermore, to understand the interactions between the genes and to gain more insights into the regulatory mechanisms of trypanotolerance, the combined gene sets of each method for each breed are used in the analysis pipeline to construct: i) Gene networks to identify hub genes using the GeneMANIA prediction server in the Cytoscape platform [42]; ii) overrepresented pathways and master regulators using the TRANSPATH database in the geneXplain platform [43]. Then, the identified hub genes, overrepresented pathways, and master regulators are compared among breeds.

Moreover, knowledge about the genetic diversity and the population structure of indigenous cattle is crucial to precisely define breeds, to design sustainable genetic improvement programs, and to develop utilization and conservation strategies [44]. Thus far, the Benshangul and Gindeberet populations have not yet been registered as breeds in the Domestic Animal Genetic Resources Information System (DAGRIS) database (<http://dagris.info/countries/192/breeds>). To the best of our knowledge, there is no research publication on the genotypic characterization of the indigenous cattle populations of Benshangul and Gindeberet. This thesis investigates the genetic diversity and the population structure of the indigenous cattle populations from Ethiopia.

1.1. Structure of the thesis

This thesis is organized as follows. In the second Chapter, I provide a definition of biological terminologies that are used in this thesis. Then, I provide a brief overview of the evolutionary processes by focusing on environmental stress responses. After introducing the basic biological concepts, I present bioinformatics tools and databases used in this thesis. In Chapter 3, I first introduce the theory of extended haplotype homozygosity (EHH) in relation to the signature of selection detection. Then, I describe the integrated haplotype score (*iHS*) and the standardized log ratio of integrated site specific EHH values between populations (*Rsb*). Next, I present an overview of the composite likelihood ratio method (CLR). Afterwards, I present methods applied in this thesis in Chapter 4. First, I describe the data and quality control measures used to filter the data set. Then, I describe the genetic background of the cattle populations used in this thesis. Following this, I present PCA and admixture methods that are used to reveal breed differentiation, genetic relationship, and population structure. In the last sections of this Chapter, I present the applications of the methods used to detect the signature of selection (selective sweep) (i.e., *iHS*, *Rsb*, and CLR), the functional properties and molecular mechanisms involved in the trypanotolerance attributes and tendencies of the breeds included in this thesis. In Chapter 5, I provide the findings of the analyses. In Chapter 6, I discuss the results reported in this thesis. In the last Chapter, I summarize the results reported in this project and provide a future perspective of the research reported in this thesis.

1.2. Impact

Journal articles:

The identified candidate signature genes, master regulators and overrepresented pathways related to trypanotolerance in Sheko are published in *Frontier in Genetics*. In addition, the genetic diversity and population structure of the indigenous Ethiopian breeds included in this thesis is under review in *Livestock Science*. Moreover, the identification of trypanotolerance tendencies in Nuer, Benshangul, and Gindeberet breeds is under preparation:

- [1] **Yonatan A. Mekonnen**, Mehmet Gültas, Kefena Effa, Olivier Hanotte, Armin O. Schmitt (2019). Identification of Candidate Signature Genes and Key Regulators Associated with Trypanotolerance in the Sheko Breed. *Front. Genet* **10**:1095. doi: 10.3389/f-gene.2019.01095 (Appendix A.1).
- [2] Selam Meseret, **Yonatan A. Mekonnen**, Bertram Brenig, Ekkehard Schütz, Olivier Hanotte, Mehmet Gültas, and Armin O. Schmitt (2020). Genetic Diversity and Population Structure of Six Ethiopian Cattle Populations from Different Geographical Regions Using High-Density Single Nucleotide Polymorphisms. *Livestock Science*: 103979. doi: 10.1016/j.livsci.2020.103979 (Appendix A.2).
- [3] **Yonatan A. Mekonnen**, Mehmet Gültas, Olivier Hanotte, Armin O. Schmitt (2020). Evolutionary Responses Associated with Trypanotolerance in Three Ethiopian Breeds – A Comparative Genomics Approach (under preparation) (Appendix A.3).

Conferences and Workshops

I have attended and presented topics included in this thesis in the following conferences and workshops.

- Bioinformatics poster day: poster presentation, Göttingen, Germany (May, 2017)
- Annual meeting of the DGfZ, Stuttgart, Germany (September, 2017)
- Bioinformatics workshop at ILRI, Addis Ababa, Ethiopia (February, 2018)
- CiBreed workshop, Göttingen, Germany (September, 2019)
- International Congress on Domestic Animal Breeding Genetics and Husbandry, Prague, Czech Republic (September, 2019)

Project

- 1000 Bull Genomes Project

2. Biological background

In this chapter, I provide an overview about the evolutionary processes that are involved in response to environmental stress. Thereby, I will start with the introduction of DNA, gene, genome, and mutation. Furthermore, I will give an overview of haplotypes, molecular mechanisms of adaptation, biotic stress, Genome Wide Association Study (GWAS), and signature of selection. In the final part of this chapter, I will briefly present the bioinformatics databases and tools that are used and applied to carry out the analyses of the genetic relationship of cattle populations and molecular regulatory processes involved in trypanotolerance.

2.1. DNA, Gene, Genome and Mutation

Deoxyribonucleic acid (DNA) is one of the most studied cellular component composed of two nucleotide chains that are coiled around each other to form a structure known as double helix which was discovered by James Watson and Francis Crick in 1953. The DNA is composed of sugar molecules (deoxyribose), phosphate groups and one of the four nucleobases (i.e., adenine (A), guanine (G), cytosine (C), and thymine (T)). As a building block of a helical structure, the sugar molecule (deoxyribose) is joined by covalent bonds of the phosphate groups and forms a sugar-phosphate backbone of the helical structure. The nitrogen bases are paired together by hydrogen bonds according to the rules of base pairings (bp) in which A pairs with T and C pairs with G and form the inside of the helical structure.

The part of DNA that contains information of the physical development, growth, and production of an organism is called gene [45]. Therefore, for the normal functioning of the organisms, the nucleus of the cells consists of a complete set of DNA sequences, including all genes which is known as genome. However, a change in the nucleotide sequence of the genome could occur and alter the characteristics of the organism. A change in those nucleotide sequence is called mutation [46].

2.2. Haplotype

A haplotype is initially defined as a group of alleles or genes within an organism that are inherited together from the same parent [47]. However, this term is also extended to a group of conserved sequences which are inherited together for several generations. In addition, the term is used to refer to the inheritance of a cluster of linked alleles that always occur together [48]. In this study, I use the latter definition of haplotype in which a pattern of variations in a single position in the DNA sequence is used to study the association between disease or tolerance traits of the organism.

2.3. Biotic Stress

Biotic stress is an environmental stress caused by the damages of living organisms on another organism. Usually, the stress causing organisms are bacteria, fungi, viruses, parasites, and insects [49]. The effect of biotic stress factors varies depending on the geographic origin, climate type, host species, and resistance [50, 51, 52]. Moreover, the exposure of these stress factors over a certain period of time leads to different types of stress responses and adaptations. A variety of stress responses are involved in gene expression, morphological and physiological changes in favor of the survival of the organism [52]. When biotic stress factors (pathogens) enter into the body of another host organism, they induce the formation of reactive oxygen species (ROS) and activate stress-associated signaling pathways which play an important role in linking sensors and genetic responses [53] (Figure 2.1).

Reactive oxygen species such as super oxides and hydrogen peroxides are known to cause oxidative stress (hypoxia). Through a variety of signaling cascades, hypoxia triggers innate and adaptive immunity responses through the activation of hypoxia induced factor (HIF) which plays a key role in the induction of stress tolerance in animals [54, 55, 56, 57].

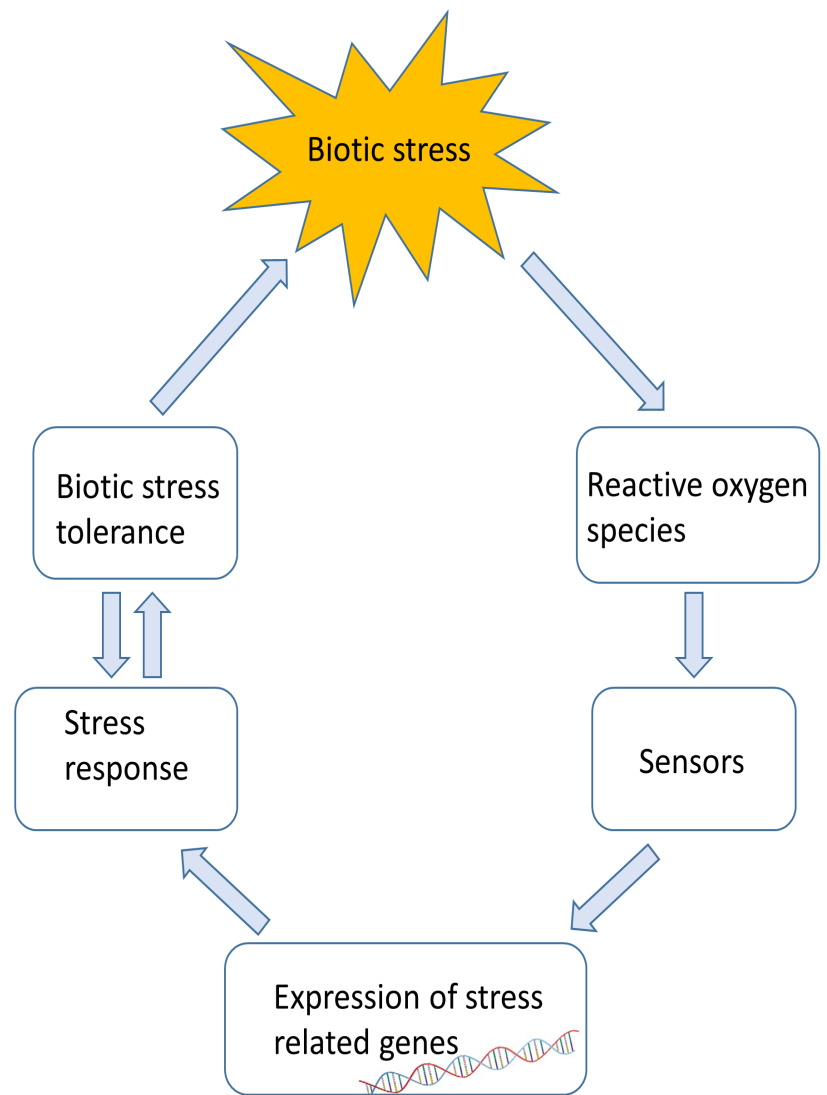


Figure 2.1.: Biotic stress-response model.

2.4. Molecular Mechanisms of Adaptation

Mutations in the genome play a crucial role in shaping the evolutionary response to a changing environment. Depending on whether the mutations are deleterious or adaptive, the organisms' fitness is determined by the type of mutation in response to the changing environment [58]. This indicates that adaptive evolutionary processes have a genetic basis which is involved in diverse molecular mechanisms such as regulation in gene networks and various biochemical pathways [59]. Therefore, the genome of the evolving organism could undergo an evolutionary change in response to the environmental stress. This implies that if the mutations of the alleles are beneficial (increase the fitness of their carriers) under certain environmental pressure or stress, natural selection favours these mutations to rapidly increase the frequency of those alleles in the population [34]. If the mutations are deleterious, then they will be removed (i.e., background selection) or substituted by beneficial alleles (Figure 2.2). The change in the allele frequencies under the neutral evolution model could occur due to a phenomenon known as genetic drift [60]. This means that substituting the deleterious allele by the beneficial allele could produce a "hitchhiking" effect or selective sweep on the frequency of neutral alleles at linked loci which results in the reduction of the genetic variation around the beneficial mutation [61, 34].

2.5. Signature of Selection

A signature of selection is defined as the genomic region composed of sequence variants in higher frequency that are functionally important [62, 63]. As presented in Section 2.4, the frequency of beneficial alleles will likely rapidly increase in the population. This means that strong and long range linkage disequilibrium (LD) and unexpectedly high local haplotype homozygosity might occur in the genome over a period of time [35, 36]. The distinctive patterns of the DNA, flanking up and downstream sequences of the beneficial variant, are due to the "hitchhiking" effect on the frequency of neutral linked loci [64]. Therefore, the genes in these genomic regions could share a common pattern of mutations or expressions and are called signature genes [65, 66]. These phenomena are the basis for the detection of selection signatures due to a certain environmental pressure.

2.6. Genome Wide Association Study

A genome wide association study (GWAS) is a hypothesis-free (non-candidate-driven) observational study which is intended to identify associations between genetic regions and phenotype. The rationale is that only tightly linked markers are present at a higher frequency in a population with traits (cases) compared to a population without the traits (controls) (Figure 2.3). This means that the GWAS analysis scans the entire genome for SNPs that occurs with higher frequency in diseased individuals in contrast to healthy individuals [67]. However, the identified SNPs might not be a causative mutation for the trait variations between case and control groups. This could be due to a stochastic noise that might be coursed by LD in the sample, especially in small popula-

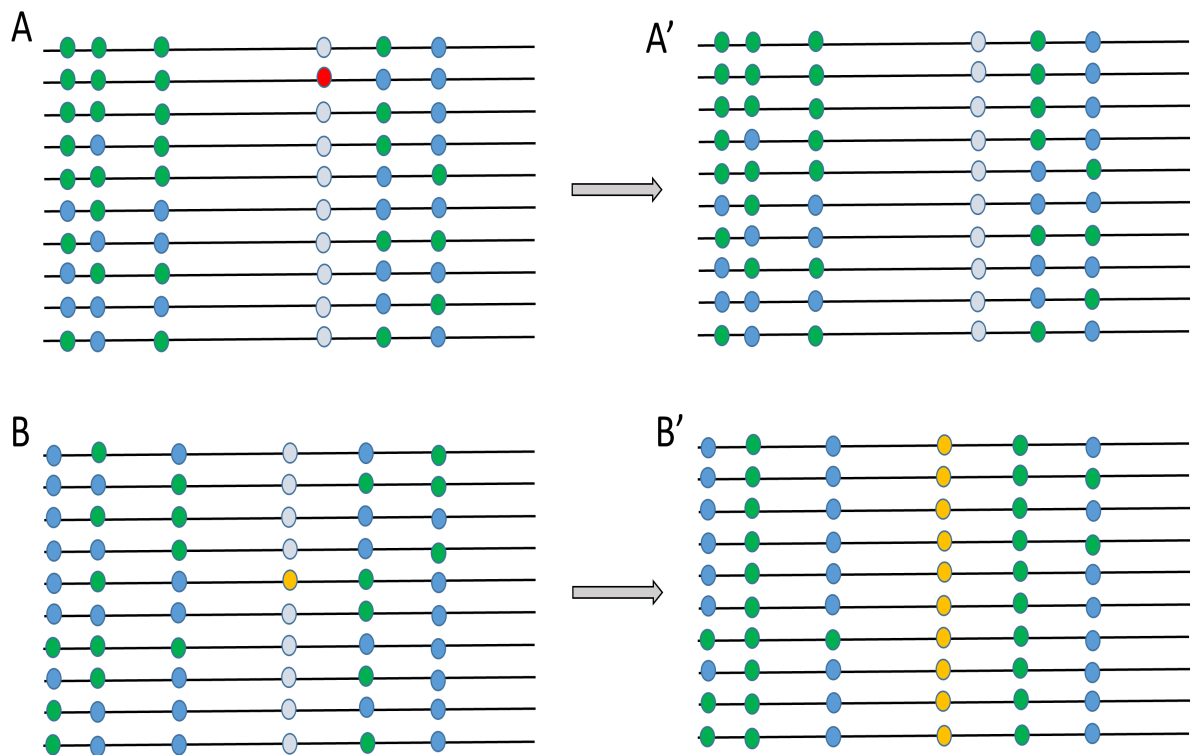


Figure 2.2.: A schematic illustration of background selection (A and A'), and selective sweep (B and B'). Each line represents a region in a single chromosome in which the red circle indicates a deleterious mutations, yellow circles indicate beneficial mutations, green circles indicate the reference allele, blue circles indicate the alternative allele, grey circles indicate the alternative alleles of the beneficial/deleterious alleles, and straight horizontal lines indicate the non-polymorphic sites.

tions or there might be a correlation between loci and the trait by chance [68]. Moreover, GWAS can result in spurious associations due to population stratification (i.e. if the ancestry between case and controls systematically differs) [69].

The genomic control method was developed by Devlin and Roeder [70] to control the confounding effects in GWAS caused by population stratification. However, this approach lack power since the variance inflation is not the same across all SNPs [69, 71]. The most widely used method to overcome confounding effects in GWAS is through the use of PCs as covariates in the regression model before testing [70, 72]. However, few PCs are used to capture the population structure which may not be adequate to effectively control cryptic relatedness and stratification in the population [73, 74].

The mixed linear model (MLM) based methods have become a popular approach to handle confounding effects more efficiently compared to PCs based methods in GWAS [73]. The MLM approach is basically developed to compute genome wide association statistics that account for

the estimate of phenotypic variance contributed by the genetic relationship matrix using a random effect model [75]. However, the identified SNP and other variants need additional studies such as differential expression of targeted genes or gene knockout experiments to validate the association between the genetic polymorphisms and the trait of interest [76].

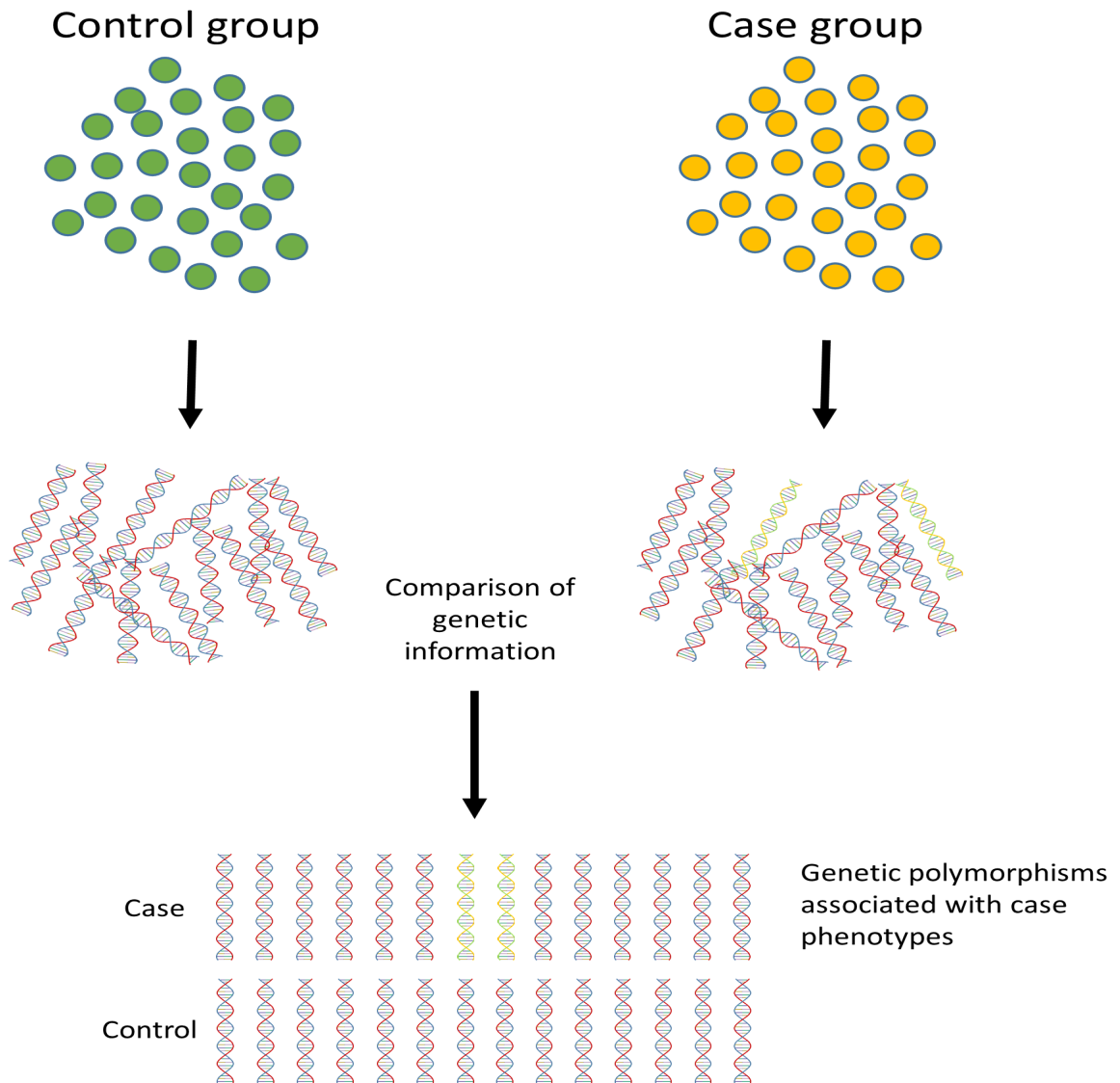


Figure 2.3.: A schematic illustration of case-control genome wide association study (GWAS) investigating differences of genetic variants between disease (illustrated as yellow circles) and control groups (illustrated as green circles).

2.7. Bioinformatic Databases and Tools

In this section, I will explain the bioinformatic databases and tools which are used for this thesis. First, I will start with the signaling pathway database TRANSPATH, and the gene function predic-

tion tool GeneMANIA plugin in the Cytoscape platform. Afterwards, I will give a brief overview about the bioinformatic tools BEAGLE, ADMIXTURE, and Plink.

2.7.1. TRANSPATH

TRANSPATH is one of the few signaling pathway databases in which a hierarchy of signaling cascades are used to unravel gene regulatory networks and molecules involved in biological processes [77] (Figure 2.4). In the current TRANSPATH database, signal transductions collected from more than 298,000 molecules and 80,000 genes that are available in the scientific literature are included (<http://genexplain.com/transpath/>). Lists of genes or proteins are used as input for TRANSPATH analyses and to perform searches for signaling molecules, pathways and their reactions in TRANSPATH database, at least two entries are required (the direction of the analysis (upstream or downstream pathway directions) and the number of permitted reaction steps (maximum distance options)) [77, 78]. The downstream pathway is defined as the identification of metabolic and regulatory pathways enhanced/provoked by the induced genes. However, the downstream analysis is limited to provide the causes of the observed effect of the induced gene expressions. Whereas the upstream pathway analysis provides the signaling pathways that activate the genes involved in metabolic and regulatory pathways [79].

The image shows a screenshot of the geneXplain website. At the top, there is a dark blue header with the geneXplain logo on the left, a 'Log in' button, and a search bar on the right. Below the header is a navigation menu with links for Home, Products, Services, Free, Research, News, About us, and Contact. The main content area is titled 'TRANSPATH'. On the left, there is a vertical sidebar with buttons for 'TRANSPATH', 'Structure', 'TGFbeta network', 'Key features', 'Benefits', and 'Pathway analysis'. The central part of the page features the TRANSPATH logo and a promotional banner that reads 'TRANSPATH® release 2019.2 is out now!'. Below the banner, there is text describing the database as one of the largest pathway databases available, optimally suited for geneXplain's proprietary Upstream Analysis. It also mentions that its contents can be used for pathway analysis with the geneXplain platform, requiring separate licensing. To the right of this text is a vertical sidebar with buttons for 'New release TRANSPATH 2019.2', 'Free trial', 'Price request', 'TRANSPATH applications', 'Information downloads', and 'Publications'.

Figure 2.4.: TRANSPATH database integrated in the geneXplain platform ([http://genexplain-platform.com/bioulmlweb/\(08/2019\)](http://genexplain-platform.com/bioulmlweb/(08/2019))).

2.7.2. GeneMANIA plugin in Cytoscape platform

GeneMANIA is a prediction server in which query gene lists are used to find functionally similar genes [42]. The BioGRID [80], IRefIndex [81], GEO [82], I2D [83], and Pathway Commons [84] databases are used to prioritize genes regarding their functional assays [42]. GeneMANIA uses these databases to search for interactions between genes (e.g., co-expression, physical interaction, predicted functional relationships, genetic interactions, co-localization, shared proteins, and common pathways). Based on the query of gene lists, GeneMANIA finds genes that are closely connected and constructs networks among the query genes [42]. For the desktop use and fast prediction of gene functions, GeneMANIA prediction server is mirrored by Cytoscape [42] (Figure 2.5). The Cytoscape is designed to analyze and visualize networks, and the GeneMANIA-Cytoscape plugin allows biologists to apply queries without restricting the number of genes as long as the memory capacity of their machine is sufficient to construct those networks [85].

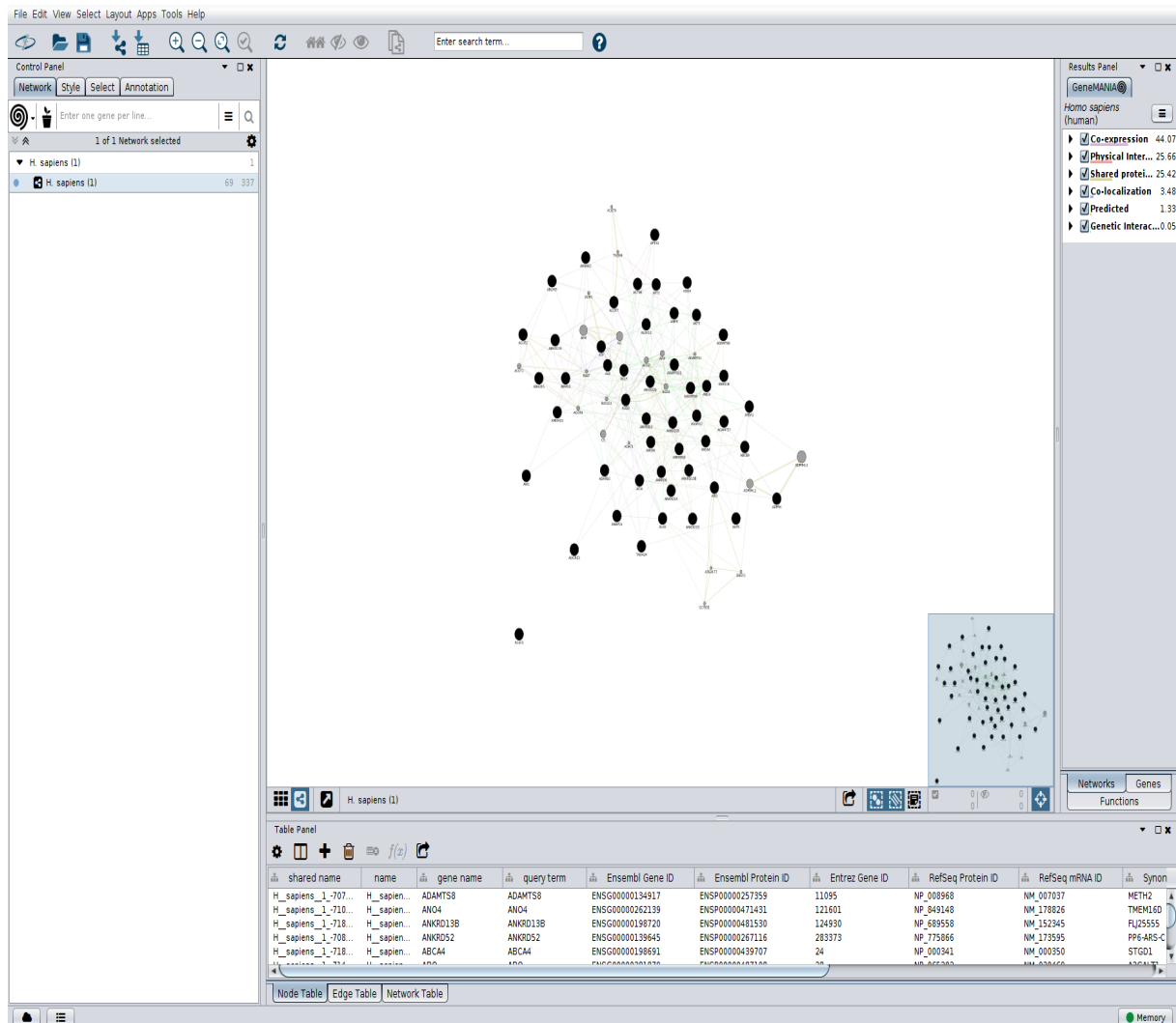


Figure 2.5.: Interface of the GeneMANIA plugin in the Cytoscape platform.

2.7.3. BEAGLE

Most of the software that are used in statistical analyses for the identification of genetic variants associated with evolutionary responses (signature of selection) require a complete data (without missing alleles) for their model parameter estimations [86]. However, with the availability of genotyping platforms for high throughput SNPs, missing data has become an indispensable issue. To overcome the issue of missing allele, BEAGLE software is developed to impute the ungenotyped markers. Therefore, BEAGLE imputes the missing allele either using a reference genome or without using a reference genome (i.e., using the entire data set as a genomic background) [87].

2.7.4. ADMIXTURE

ADMIXTURE is a software which is used to estimate the ancestries from the genotype data of the current population [88]. ADMIXTURE uses a model-based estimation approach (Markov Chain Monte Carlo (MCMC) algorithm) to estimate admixture coefficients and frequencies of ancestral alleles in a higher computational speed than *structure* [89] and EIGENSTRAT [90]. The

admixture analysis performs the estimation of ancestry in two levels: 1) Local ancestry estimation; and 2) global ancestry estimation. The former considers the individual genome as segments of chromosomes and assigns the origin of each segment based on the segment boundaries while the latter estimates the proportion of ancestry from multiple populations using the entire genome of the individuals in each population [88].

2.7.5. PLINK

In the past, it has been proven that genetic variations are beyond single Mendelian mutations. Rather, the molecular basis of complex diseases are associated with many genes with small effect (polygenic effect) which requires to include the whole genome sequence. With the advent of high-throughput sequencing, there was an increasing demand for robust algorithms to handle such a large dataset. PLINK is developed by Shaun Purcell [91] which was initially aimed to conduct whole genome association studies (WGAS) with computational efficiency. Currently, PLINK can carry out a wide range of tasks such as data management, summary statistics, quality controls, population stratification detection, and identity-by-descent (IBS) analyses [91].

3. Theoretical background

In this chapter, the theory for the detection of signatures of selection will be presented. I will first introduce the theory of extended haplotype homozygosity (EHH). Then, I will present the integrated haplotype score (*iHS*) and the standardized log ratio of integrated site specific EHH values between populations (*Rsb*). Finally, I will give a theoretical overview of the composite likelihood ratio method (CLR).

3.1. Extended Haplotype Homozygosity

Extended haplotype homozygosity (EHH) was developed by Sabeti et al. in 2002 [92]. The initial aim of the EHH method was to detect positive selection in humans by searching for a long range linkage disequilibrium (LD) and unusually highly frequent haplotypes in the genome [92]. Afterwards, it has been applied to many other species including cattle [36]. Hence, EHH method is used to calculate the probability of the SNPs carrying a core haplotype (i.e., a set of closely linked SNPs in which recombination does not take place) are homozygous from the core region to distance x . Therefore, the EHH based methods integrated haplotype score (*iHS*) and the standardized log ratio of integrated site specific EHH values between populations (*Rsb*) will elucidate the genomic regions that are under recent positive selection [92, 93].

3.1.1. Integrated Haplotype Score

The measure of EHH in comparison to ancestral and derived alleles at a given SNP is calculated as integrated haplotype score (*iHS*) [93]. The *iHS* test starts with a core haplotype identification [92, 94]. Then, the observed decay of LD as a function of the distance from the core haplotypes is computed as the integrated EHH (*iHH*). The *iHH* is defined as the area under the EHH curve against the distance from the core haplotype to the last haplotype carrying the core SNP [93] (**Note:** If the ancestral core SNP is considered, the *iHH* is considered as iHH_A , and if the derived core SNP is considered, then the *iHH* is considered as iHH_D). Therefore, the *iHS* is given as:

$$iHS(\text{unstandardized}) = \ln \left(\frac{iHH_A}{iHH_D} \right), \quad (3.1.1)$$

to allow the direct comparisons among different SNPs regardless of their allele frequencies, the *iHS* values are standardized using their mean and standard deviation [93, 35].

$$iHS(\text{standardized}) = \frac{\ln \left(\frac{iHH_A}{iHH_D} \right) - E_p \left[\ln \left(\frac{iHH_A}{iHH_D} \right) \right]}{SD_p \left[\ln \left(\frac{iHH_A}{iHH_D} \right) \right]}, \quad (3.1.2)$$

the empirical distribution of the SNPs with a frequency p of derived allele (the same as the frequency of the core SNP) is used to estimate the expectation $E_p \left[\ln \left(\frac{iHH_A}{iHH_D} \right) \right]$ and standard deviation $SD_p \left[\ln \left(\frac{iHH_A}{iHH_D} \right) \right]$ of the unstandardized iHS [93].

3.1.2. Relative Integrated EHHS between populations

The decay of EHH starting from the individual SNP site i of a population as a function of distance j regardless of the status of the allele (without considering ancestral and derived alleles) is given as:

$$EHHS_{i,j} = \frac{E(H_{O_{i,j}})}{E(H_{O_i})}, \quad (3.1.3)$$

where $EHHS_{i,j}$ is the decay of EHH of an individual SNP site i to SNP site j , $E(H_{O_{i,j}})$ is a haplotype homozygosity from SNP site i to SNP site j , and $E(H_{O_i})$ is a haplotype homozygosity at SNP site i . Therefore, integrated EHHS (iES) is defined as the area under the EHHS curve against the distance from the core haplotype to the last haplotype carrying the core SNP. The iES is used to summarize the decay of EHHS for a single SNP site in a population. Hence, the relative integrated site-specific EHH between populations (Rsb) can be given as:

$$Rsb(\text{unstandardized}) = \ln \left(\frac{iES_{pop1}}{iES_{pop2}} \right), \quad (3.1.4)$$

where $pop1$ refers to the study population and $pop2$ refers to the reference population. Due to a slower decay of EHH in one population compared to the other, there might be extreme values of Rsb . Therefore, unlike iHS , the standardization of Rsb values uses the median instead of the mean, since the median is less sensitive to extreme values [95]. Therefore, the standardized Rsb is given as:

$$Rsb(\text{standardized}) = \frac{\ln \left(\frac{iES_{pop1}}{iES_{pop2}} \right) - med \left[\ln \left(\frac{iES_{pop1}}{iES_{pop2}} \right) \right]}{SD \left[\ln \left(\frac{iES_{pop1}}{iES_{pop2}} \right) \right]}, \quad (3.1.5)$$

where $med \left[\ln \left(\frac{iES_{pop1}}{iES_{pop2}} \right) \right]$ is the median and $SD \left[\ln \left(\frac{iES_{pop1}}{iES_{pop2}} \right) \right]$ is the standard deviation of the unstandardized Rsb . The median and standard deviation of the unstandardized Rsb are estimated from the empirical distribution of the total set of SNPs [95].

3.2. Composite Likelihood Ratio

The composite likelihood ratio (CLR) test is an LD based selective sweep searching method using information from the spatial distribution of allele frequencies [96]. This means that the CLR test identifies selective sweeps using the patterns of allele frequencies that differ from the total pattern of SNPs included in the study. Therefore, CLR is used to identify skewed patterns of the allele frequency spectrum towards excess of rare alleles and high frequency alternative alleles due to the

hitchhiking effect [64, 97, 98]. The composite likelihood function for SNP k is calculated as:

$$CL(p) \equiv \prod_{i=1}^k px_i = \prod_{i=1}^k p_j^{k_j}, \quad (3.2.1)$$

where p is the probability of observing the derived allele, $CL(p)$ is a composite likelihood function of p , x_i is the frequency of the derived allele at locus i , and p_j is the probability that a derived allele has frequency j in the k SNPs under consideration. Therefore, the composite likelihood ratio test statistic T is given as:

$$T = 2\{\log CL(\hat{p}_{v \leftrightarrow b}; v \leftrightarrow b) - \log CL(\hat{p}; v \leftrightarrow b)\}, \quad (3.2.2)$$

where $\log CL(\hat{p}_{v \leftrightarrow b}; v \leftrightarrow b)$ is the log likelihood of the data under the alternative model, $\log CL(\hat{p}; v \leftrightarrow b)$ is the log likelihood of the data under the null model, and \hat{p} is the estimate of the maximum composite likelihood of p from SNP v to SNP b [98, 99].

4. Materials and Methods

In this chapter, I will provide the materials and methods used in this thesis. First, I will describe the data used in this study and the quality controls implemented to filter the data. Second, I will briefly explore the genetic backgrounds and origin of breeds included in the study. Third, I will present statistical methods applied to detect population structure, genetic relationships, and signatures of selection. Fourth, I will present bio-informatics tools and databases used to identify functionally enriched genes, hub genes, overrepresented pathways, and master regulators. Fifth, I will provide a description of comparative approaches to identify the tendency of trypanotolerance in Benshangul, Nuer, and Gindeberet breeds. Part of the following sections are previously published in Mekonnen *et al.* [100].

4.1. Data, SNP Genotyping and Quality Control

The data used in this study is genetic information (DNA samples) of cattle breeds. sDNA was extracted from 67 blood and tissue samples according to the QIAGEN DNA extraction protocol [101]. 19 samples from Gindeberet, 12 from Sheko, 13 from Nuer, 12 from Benshangul and 11 from Fogera breeds were collected. All samples were taken randomly from unrelated animals based on the information given by livestock keepers at the time of sampling. All samples were genotyped for 777,962 SNPs using the Illumina BovineHD Genotyping Bead chip. In addition, the genotyping data of two west African breeds (24 N'Dama and 8 Muturu), and five east African breeds (92 East African Shorthorn Zebu (EASZ), 25 Ankole, 16 Karamojong, 23 Nganda, and 12 Serere) were obtained from the International Livestock Research Institute (ILRI, Addis Ababa, Ethiopia; [102]). The total sample size for the down stream analysis consisted of 265 samples and 715,712 SNPs. For quality control (QC), Plink1.9 [91] was used on 735,293 autosomal SNPs. SNPs with minor allele frequency (MAF) of less than 1% were excluded (19,581 SNPs). Minimum genotyping call rate (<95%) and maximum identity-by-state (IBS) ($\geq 95\%$) were also used as filtering criteria. Two Benshangul samples failed the genotyping call rate criterion and were excluded from further analyses but no pair of samples was excluded due to the IBS filtering criterion. BEAGLE 4 [87] was used for inferring haplotype phasing and imputing the missing alleles. The default setting was used to perform the imputation. The imputation was performed by fitting 83 sliding windows across the autosomes in which on average 8600 markers were included. With in each window twelve iterations were executed. Since our samples consist of indigenous African breeds, the total of 264 (n-1) animals included in this study are used as a background to impute the missing alleles in the context of indigenous African cattle genomes [i.e., without using the reference genome (UMD3.1)].

4.2. Genetic Background of the Cattle Population

In the eastern part of Africa, the mixture of African taurine and indicine cattle populations is common which reflects the immigration waves of these two different ancestral aurochs in the region [102, 103, 104]. With respect to these two ancestral populations, the N'Dama and Muturu breeds are considered as African taurine whereas the Fogera, EASZ, Ankole, Karamojong, and Serere breeds are referred to as African zebu [102]. The Nuer and Ankole breeds are classified as African sanga [24] while the Nganda breed is assigned to African zenga [102]. The sanga and zenga cattle are crossbreeds between the indigenous humpless cattle and zebu. The latter have higher zebu genetic introgression than the former [105]. Interestingly, the Sheko breed is considered as the last oddment of the primordial *Bos taurus* cattle in eastern Africa. However, some animals in the present population of Sheko display small humps which indicates the genetic introgression of zebu cattle [24]. Today, there is no research publication or documentation available on the genetic background of the Benshangul and Gindeberet breeds which are included in this study. The breed type and origin of the cattle samples included in this study are presented in Table 4.1.

Table 4.1.: Cattle breeds included in the study. Modified from Mekonnen *et al.* [100].

Breed name	*Breed category	Breed origin	No. of animals	Trypanosusceptible
N'Dama	African taurine	Guinea	24	No [25]
Muturu	African taurine	Nigeria	8	No [106]
Ankole	Sanga	Uganda	25	Yes [107]
Karamojong	African zebu	Uganda	16	Yes [108]
Serere	African zebu	Uganda	12	Yes [109]
Nganda	Zenga	Uganda	23	Yes [110]
EASZ	African zebu	Kenya	92	Yes [111, 112]
Sheko	African taurine and zebu	Ethiopia	12	No [23]
Nuer	Sanga	Ethiopia	13	Not available
Gindeberet	Not available	Ethiopia	19	Not available
Benshangul	Not available	Ethiopia	10	Not available
Fogera	African zebu	Ethiopia	11	Yes [113]

*Breed category according to DAGRIS (2009).

4.3. Breed Differentiation, Genetic Relationship and Structure

In order to understand the genomic structure of Sheko in comparison to other breeds, I considered a total of 12 indigenous African breeds genotyped with the Illumina BovineHD Genotyping BeadChip. To assess the within and between population genetic structure and admixture, PCA and admixture analyses were conducted. PCA was performed using Plink 1.9 to estimate the eigenvectors of the variance-standardized relationship matrix of all samples. In order to refine the genetic structure of the indigenous Ethiopian cattle breeds, separate PCA calculations were made for samples that were collected in Ethiopia (Sheko, Benshangul, Gindeberet, Fogera, and Nuer). Admixture analysis was performed using the ADMIXTURE 1.3 software with cross-validation and 200 bootstraps for the hypothetical number of ancestries K ($2 \leq K \leq 7$). Both PCA and ad-

mixture analyses were used to determine the level of admixture and genetic differentiation of the populations. Furthermore, admixture analysis was used to determine the level of indicine and taurine ancestries of each breed at the genome-wide level. In particular, PCA and admixture analyses were performed to show and quantify the taurine background of Sheko.

4.4. Analysis of Signatures of Positive Selection

In general, methods for the detection of selection signatures are based on the spatial distribution of allele frequencies and the property of segregating haplotypes in the population [114]. As suggested by Ma *et al.* [38] and Vatsiou *et al.* [39], combining these methods would help to reach a higher power than would be possible with a single analysis. In this thesis, I used extended haplotype homozygosity (EHH) and spatial distribution of allele frequency based methods to identify signatures of positive selection in the genome of the Sheko breed. This denotes that integrated haplotype score (*iHS*) and Composite Likelihood Ratio (CLR) analyses were performed on Sheko ($n = 12$) while the ratio of site-specific EHH (EHHS) between populations (*Rsb*) analysis were performed between Sheko ($n = 12$) and combined trypanosusceptible reference cattle populations (179) [(EASZ ($n = 92$) [111, 112], Ankole ($n = 25$) [107], Karamojong ($n = 16$) [108], Nganda ($n = 23$) [110], Serere ($n = 12$) [109] and Fogera ($n = 11$) [113]]).

4.4.1. Extended Haplotype Homozygosity Based Methods

Rsb and *iHS* are linkage disequilibrium (LD) based approaches which are implemented in the R package *rehh*. Both *Rsb* and *iHS* are used to identify genome-wide signatures of selection [35]. The *Rsb* analysis was performed between the study population and the combined group of trypanosusceptible breeds. For each group, the integrated site-specific EHH of each SNP (iES) was calculated. To calculate *Rsb* values, the standardized log-ratio between iES of the two groups was used. The *iHS* values were calculated as the natural log ratio of integrated EHH (iHH) between reference and alternative alleles for each SNP [26, 35].

The bovine reference genome (UMD3.1) is used as the reference allele while the allele of the study population is considered as the alternative allele. The *iHS* values were standardized based on the calculated mean and standard deviation values. This allows direct comparisons among different SNPs regardless of their allele frequencies [35]. For the standardization of *Rsb* values, median and standard deviation values were used. One-tailed Z-tests for *Rsb* and two-tailed Z-tests for *iHS* were applied on the standardized and normally distributed *Rsb* and *iHS* values to identify statistically significant SNPs that are under positive selection. For one-tailed Z-tests, $P = 1 - \Phi(Rsb)$, whereas $P = 1 - 2|\Phi(iHS) - 0.5|$ was used for the two sided tests with Φ being the Gaussian cumulative density function. For both *Rsb* and *iHS* *P*-values, the significance threshold of $\alpha = 10^{-4}$ was applied following the study of Bahbahani *et al.* [26] and Tijjani *et al.* [106] to identify candidate regions.

4.4.2. Spatial Distribution of Allele Frequency Based Method

The CLR test is an LD based selective sweep searching approach using the information from the spatial distribution of allele frequencies [96]. CLR is used to identify skewed patterns of the allele frequency spectrum towards excess of rare alleles and high frequency alternative alleles due to the hitchhiking effect [64, 97, 98]. The P -values were calculated by the rank of the genome wide scan of CLR values. As suggested by Wilches *et al.* [115], the 95th quantile of the distribution of the top CLR P -values was used to identify a significance threshold of $\alpha = 10^{-5}$. For CLR analysis, the Sweepfinder2 [116] software was used for each chromosome with a window size of 50kb including on average 226 SNPs per window. Sweepfinder2 estimates CLR values in the context of background selection to identify sweeps [116, 117].

4.5. Functional Annotation of Selected Candidate Regions

Genomic regions are designated as a group of three or more significant SNPs which are separated by a maximum of 500kb on both sides of each of the consecutive SNPs following the study of Cheruiyot *et al.* [118]. Genes found within 25 kb around the most significant SNPs were considered as candidate genes [26]. Protein-coding and RNA genes found within the candidate regions were retrieved using the BioMart tool [119]. The R package *Enrichr* [120] was used to determine the candidate signature genes that are functionally enriched in Gene Ontology (GO) terms with respect to the whole bovine reference genome background ($\alpha = 0.05$). These functionally enriched candidate signature genes were used to produce a treemap (a visualization of large hierarchical collections of data in the form of nested rectangles that make up the map [121]) which shows clusters of functional terms based on the biological functions of the candidate signature genes. Since trypanosomiasis is considered as an environmental pressure, the enriched signature genes might work together in a network to govern the genes that are involved in the evolving breed against trypanosomiasis. In order to identify the gene networks, I applied the GeneMANIA prediction server in the Cytoscape platform [42, 122]. For the network construction analysis, co-expression, physical interaction, predicted functional relationships, genetic interactions, and co-localization between genes were considered.

To gain more insight into the functional properties and molecular mechanisms involved in trypanotolerance, overrepresented pathways were analyzed using the TRANSPATH database [77] of the geneXplain platform. Furthermore, to understand the regulatory mechanisms of the candidate signature genes and the signaling cascades in the regulatory hierarchy involved in trypanotolerance, the identification of master regulators was conducted 10 steps upstream in the regulatory hierarchy using the TRANSPATH database.

4.6. Comparative Identification of Trypanotolerance Tendency

For generations, cattle breeds such as Benshangul, Gindeberet, and Nuer have been kept in the infected regions without proper prevention and medication. However, the tolerance level of these

breeds is not well investigated. To explore the tendency of trypanotolerance attributes in these breeds, *iHS*, *Rsb*, and CLR analyses were conducted for Nuer, Benshangul, and Gindeberet breeds following the same procedure in sections 4.4.1 and 4.4.2. Furthermore, signature genes, hub genes, overrepresented pathways, and master regulators are identified for these breeds following the same procedure in section 4.5. The identified hub genes, overrepresented pathways and master regulators of of Nuer, Benshangul, and Gindeberet breeds are compared with those of Sheko and among themselves. In addition, the identified signature genes are compared with the genomic regions and genes of N'Dama and Muturu breeds that are described in the literature.

4.7. Comparative Identification of Signature of Selection Using GWAS and Signature of Selection Detecting Methods

Case-control association analysis is performed using GEMMA [73] to identify genes and genomic regions that are significantly associated with the phenotype (i.e., trypanosomiasis). For GEMMA analysis, trypanosusceptible breeds are used as case group (EASZ [111, 112], Ankole [107], Karamojong [108], Nganda [110], Serere [109], and Fogera [113] breeds) and the trypanotolerant breed (Sheko [23]) as control group. Separate case-control association analysis was also performed using GEMMA for the combined trypanotolerant control group (Sheko [23], N'Dama [25], and Muturu [106] breeds) with the same case group as above. To compare the genes and genomic regions identified by the association analyses and the signature of selection detecting methods, *iHS*, CLR, and *Rsb* analyses were performed for the combined trypanotolerant group. Then, the genes and genomic regions identified by GEMMA and the signature of selection detecting methods (*iHS*, CLR, and *Rsb*) are compared to detect genomic regions and genes captured by both approaches.

4.8. Summary of the Analysis of Signatures of Positive Selection

Our work-flow can be divided into four major steps as described below (see also Figure 4.1): 1) I analyzed the genetic relationship and structure of Sheko and 11 other indigenous African breeds using Plink 1.9 and the ADMIXTURE 1.3 software. 2) The identified candidate signature genes were then used in the analysis pipeline comprising the following five sub-steps: i) First, I identified genomic regions and signature genes under positive selection towards trypanotolerance in Sheko using *iHS*, CLR, and *Rsb* analyses. As an intermediate result, I present the 15 genes resulting from a literature survey; ii) in the second step, I applied enrichment analysis in gene ontology (GO) terms in the combined gene sets of the three methods and made clusters of enriched GO terms in the form of a treemap using the geneXplain platform; iii) I then identified overrepresented pathways based upon the significant genes found in (ii) using the TRANSPATH database in the geneXplain platform; iv) I identified the master regulators 10 steps upstream in the regulatory hierarchy using the significant genes found in (ii) using the TRANSPATH database in the geneXplain platform; v) I identified top three hub genes using the significant genes found in (ii) using GeneMANIA prediction server in Cytoscape platform. 3) I identified genomic regions and

signatures genes using *iHS*, CLR, and *Rsb* analyses for Nuer, Benshangul, and Gindeberet breeds.

4) The identified signature genes for Nuer, Gindeberet, and Benshangul breeds were then used in the analysis pipeline comprising the following five sub-steps: i) The significant signature genes of Benshangul, Nuer and Gindeberet are compared with the 15 most significant genes of Sheko that are identified in the second step (i); ii) the identified signature genes and genomic regions are compared with genes and genomic regions of N'Dama and Muturu that are reported in the literature; iii) I identified the overrepresented pathways and the master regulators for Nuer, Benshangul, and Gindeberet following the same procedure as described in the second step [(iii) and (iv)]; iv) I identified the top three hub genes for each breed following the same procedure as described in the second step (v); iv) the identified top three hub genes, overrepresented pathways, and master regulators of the three breeds are compared with Sheko and among themselves.

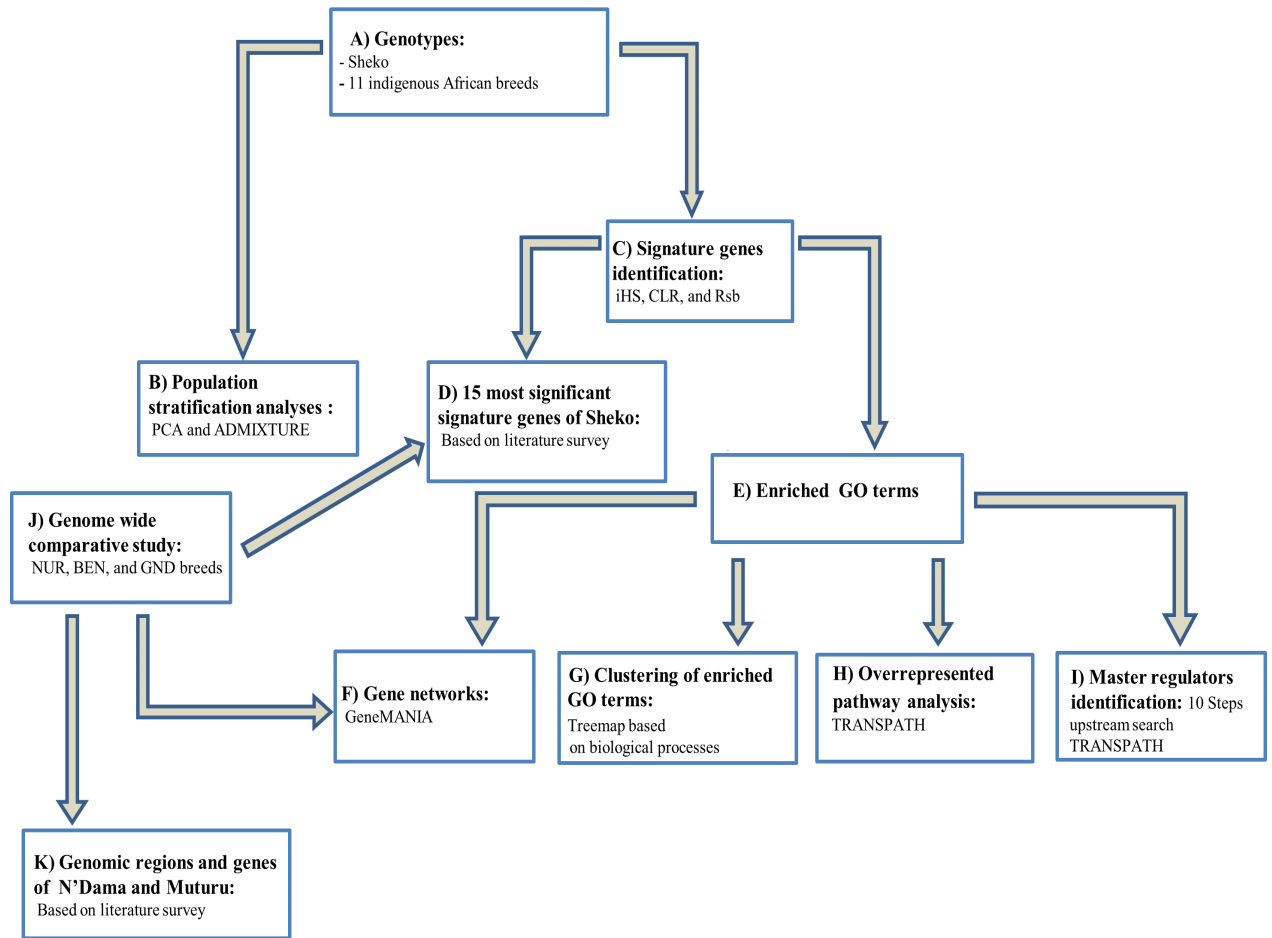


Figure 4.1.: Analysis work-flow to identify candidate genes and key regulators that are associated with trypanotolerance. (A) The genotypes of the Sheko and 11 other indigenous African breeds are obtained and quality control filtering is performed. (B) The genomic structure of Sheko in comparison to 11 other indigenous African breeds is analyzed using PCA and ADMIXTURE. (C) The identification of signature genes is performed using *iHS*, CLR, and *Rsb* analyses. (D) Among 364 signature genes in Sheko, the 15 most significant genes that are associated with trypanotolerance attributes are identified and disclosed. (E) Significantly functionally enriched terms (GO terms) are identified for the signature genes. Functionally enriched signature genes are analyzed to identify; (F) the top three hub genes for Sheko are identified using GeneMANIA in the Cytoscape platform, (G) a treemap is produced based on the biological processes, (H) overrepresented pathways, and (I) a master regulatory network is generated up to 10 steps upstream in the regulatory hierarchy. The treemap, overrepresented pathway, and master regulator analyses were performed in the geneXplain platform. (J) The identified signature genes of Benshangul, Nuer, and Gindeberet are compared with the 15 signature genes of Sheko that are identified in (D), with genomic regions, and genes of N'Dama and Muturu which are reported in the literature (K). Moreover, the identified Tree maps, hub genes, overrepresented pathways, and master regulators of Nuer, Benshangul, and Gindeberet breeds are compared with Sheko (F, G, H, and I) and among themselves. NUR, Nuer; BEN, Benshangul; GND, Gindeberet. Modified from Mekonnen *et al.* [100]

5. Results

In this chapter, I will first present the results of the population structure and genetic relationship of breeds. Then, I will present the results of the signature of selection analysis mainly focusing on the Sheko breed in relation to the trypanotolerance. Next, I will provide the results of the pathway analysis and transcriptional machineries involved in the regulatory mechanisms of trypanotolerance in the Sheko breed. Then, I will explore the trypanotolerance tendencies of Nuer, Benshangul, and Gindeberet breeds in comparison to the genes and genomic regions of Sheko reported in this thesis. Furthermore, the trypanotolerance tendencies of the Nuer, Benshangul, and Gindeberet breeds will be assessed by making a comparison with genes and genomic regions of N'Dama and Muturu breeds that are reported in the literature. Finally, I will present a comparative analysis on GWAS and signature of selection detecting methods (*iHS*, CLR, and *Rsb*) to detect the common genomic regions and genes identified by both approaches. Part of the following sections are previously published in Mekonnen *et al.* [100].

5.1. The Genetic Relationship and Structure of Cattle Populations

In order to understand the genetic structure of Sheko in comparison with eleven other African breeds, principal component analysis (PCA) was used. The result shows that the first two principal components (PCs), which explain 30.3% and 4.6% of the total variation, distinguishes the African taurine (N'Dama and Muturu) from the African indicine breeds (Benshangul, Serere, Karamojong, EASZ, Fogera, and Gindeberet) (Figure 5.1 A). Moreover, the Sheko, Nganda, Ankole, and Nuer are positioned between the African taurine and the African indicine clusters. I also conducted PCA exclusively for indigenous Ethiopian breeds. The result shows that the Sheko and Nuer form separate groups while the indicine type breeds (Benshangul, Fogera, and Gindeberet) form a cluster in both PCs (Figure 5.1 B).

For the further understanding of the degree of admixture in the populations, the ADMIXTURE 1.3 [88] software was used for $K = 2$ to 7 hypothetical ancestral populations (Figure 5.2). I start from 2 hypothetical ancestral populations with the aim to determine the degree of indicine and taurine genetic background in the cattle breeds. Since the CV errors from $K = 3$ to $K = 6$ have not exceeded the cross-validation (CV) errors of $K = 2$ (see Figure 5.3), I extend the hypothetical population up to $K = 7$ which has the highest CV error (Figure 5.3).

At $K = 2$, the two ancestries of taurine and indicine are revealed. The genomes of Ankole, Nganda, Nuer, and Sheko are mainly of indicine origin but have substantial taurine admixture. At $K = 3$, Ankole, Nuer and Sheko show genetic heterogeneity with a considerable level of taurine admixture. EASZ, Karamojong, Serere, Benshangul, Fogera, and Gindeberet also show minor levels

of taurine admixture whereas Nganda reveals a high level of within breed genetic differentiation. Moreover, with the increment of the value of K, Sheko and Nuer show a higher level of genetic heterogeneity than the other east African breeds. Furthermore, at K = 6 and K = 7, the African taurine breeds N'Dama and Muturu show separate genetic backgrounds. In general, Sheko shows the highest level of African taurine genomic contribution for all values of K among East African breeds. The proportion of admixture in each of the analyzed breeds are presented in Table 5.1 for K = 7.

Table 5.1.: Proportion of admixture within each of the analyzed breeds for K = 7. Adapted from Mekonnen *et al.* [100].

Breed	Number of inferred cluster						
	1	2	3	4	5	6	7
Muturu	0.000	0.988	0.000	0.011	0.000	0.000	0.000
N'Dama	0.000	0.003	0.000	0.996	0.000	0.000	0.000
EASZ	0.001	0.002	0.007	0.003	0.168	0.770	0.047
Ankole	0.001	0.002	0.008	0.004	0.943	0.035	0.005
Karamojong	0.001	0.003	0.026	0.005	0.142	0.730	0.088
Nganda	0.471	0.003	0.322	0.002	0.097	0.098	0.005
Serere	0.005	0.001	0.144	0.003	0.010	0.742	0.092
Nuer	0.001	0.041	0.015	0.070	0.123	0.178	0.572
Benshangul	0.001	0.002	0.010	0.005	0.042	0.133	0.806
Fogera	0.001	0.001	0.000	0.002	0.027	0.136	0.832
Gindeberet	0.001	0.000	0.001	0.001	0.010	0.161	0.824
Sheko	0.013	0.049	0.009	0.083	0.202	0.000	0.643

The result of the admixture proportion obtained for K = 7 shows that about 99% of Muturu and 100% of N'Dama are assigned to cluster 2 and cluster 4, respectively. The genome of EASZ is assigned to cluster 5 (17%) and, with a high genomic proportion, to cluster 6 (77%). Similarly, Karamojong is assigned to cluster 5 (14%) and cluster 6 (73%). A high proportion of Ankole is assigned to cluster 5 (94%) with 3% of its genome being assigned to cluster 6. Only Nganda is assigned to cluster 1 with a high genome proportion (47%) and cluster 3 (32%). The genome of Serere is assigned to cluster 3 (14%) and cluster 6 (74%). All of the Ethiopian breeds (Nuer (57%), Benshangul (80%), Fogera (83%), Gindeberet (82%), and Sheko (64%)) are assigned to cluster 7 with high genome proportions. Among these Ethiopian breeds, 17% of Nuer, 13% of Benshangul, 14% of Fogera, and 16% of Gindeberet breeds belong to cluster 6 whilst 20% of Sheko is assigned to cluster 5. Additional analyses on the genetic diversity and population structure of these Ethiopian breeds are found in Appendix A.2.

5.2. Identification of Candidate Signature Genes Associated with Trypanotolerance

A total of 20, 14, and 65 genomic regions harbouring 109, 64, and 202 candidate signature genes were identified by *iHS*, CLR, and *Rsb* analyses in 22, 10, and 27 autosomes in Sheko, respectively (Figure 5.4 and Supplementary Tables 2-4 in Appendix A.1). Among the 364 unique candidate

signature genes identified by *iHS*, CLR, and *Rsb* analyses, 260 disposed of enriched GO terms ($\alpha = 0.05$) (Supplementary Tables 5-7 in Appendix A.1). Moreover, 96, 323, and 463 intergenic variants were identified in gene desert regions by *iHS*, CLR, and *Rsb* analyses in all candidate regions, respectively (Supplementary Tables 8-10 in Appendix A.1).

Mainly focusing on the top ten candidate signature genes of each of the three methods, I performed a literature survey and identified 15 (4 genes identified by *iHS*, 3 genes identified by CLR, 7 genes identified by *Rsb*, and 1 gene identified by both *iHS* and CLR) candidate signature genes that are associated with trypanotolerant attributes which have been reported in previous studies (Table 5.2). Polymorphisms in or nearby the *MIGA1*, *CDAN1*, *HSPA9* and *PCSK6* genes in the genome of Sheko might be associated with anemia. The *MIGA1* gene is associated with iron deficiency anemia and immunity [123, 124]. This gene also plays a major role in the development and proliferation of lymphocyte since defective T- and B-cell activation is caused by inadequate iron uptake [124, 125]. The encoded protein of this gene is primarily bound by the transferrin receptor (TfR) and was identified as a component of TfR [123]. When the cellular iron level is low, increment and stabilization of iron are performed by iron regulatory proteins binding to the 3'UTR (untranslated regions) of TfR [124]. TfR also plays a major role in the development and proliferation of lymphocyte since defective T- and B-cell activation is caused by inadequate iron uptake [125].

Another candidate signature gene related to anemia is *CDAN1*. Polymorphisms in this gene are associated with congenital dyserythropoietic anemia type 1 [126, 127]. Moreover, the hsp70 protein family and the heat shock 70kDa protein 9 (*HSPA9*) gene play a role as a downstream mediator of erythropoietin signaling and contribute to normal erythropoiesis [54, 128, 129, 130]. The mutation in this gene is associated with sideroblastic anemia [131], while the *PCSK6* gene is involved in iron homeostasis and hence related to iron deficiency anemia [132].

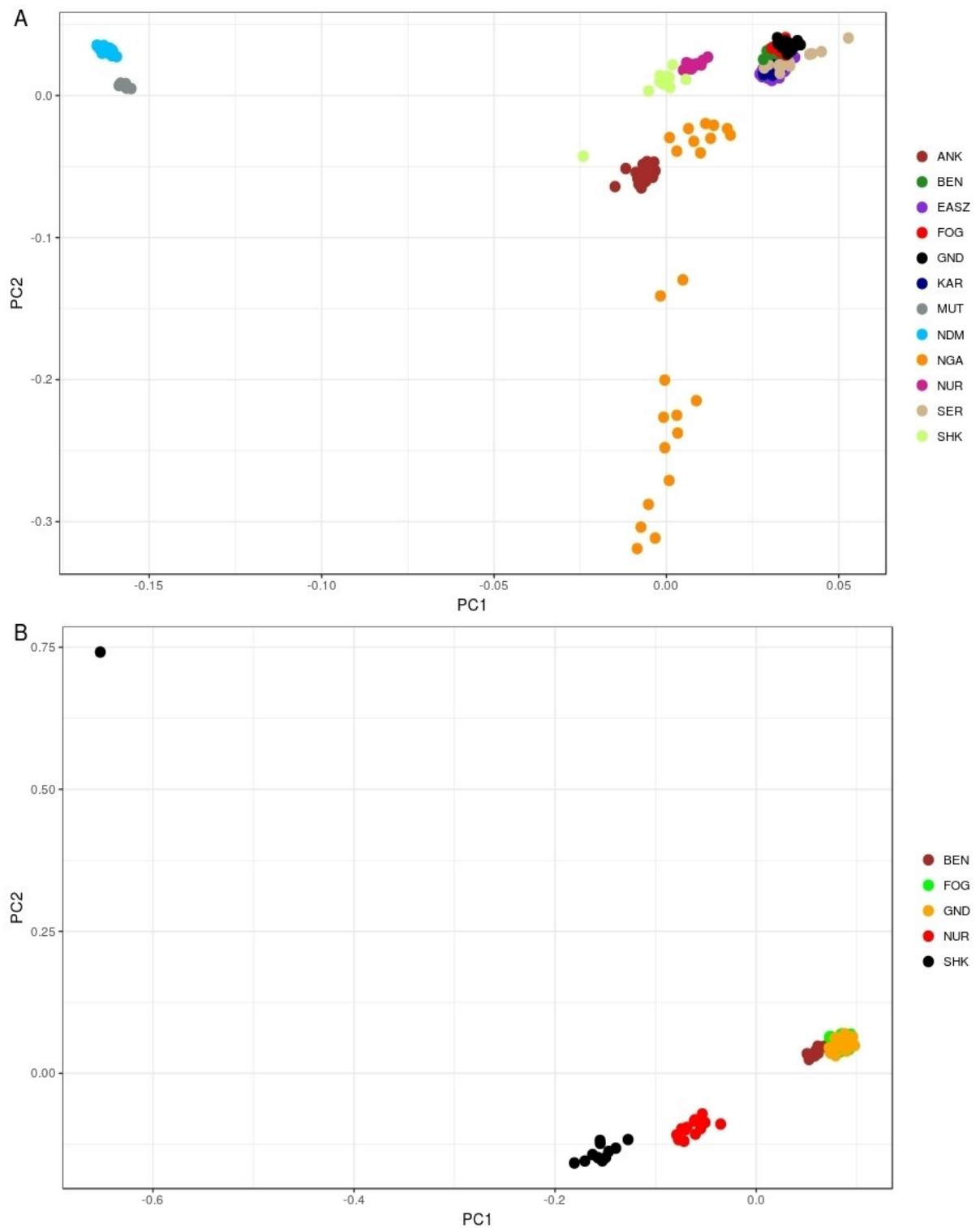


Figure 5.1.: PCA plots of the first two principal components showing the genetic relationship between cattle breeds. (A) PCA plot for all cattle breeds included in this study, and (B) PCA plot for the Ethiopian cattle breeds. ANK: Ankole, BEN: Benshangul, FOG: Fogera, GND: Gindeberet, KAR: Karamojong, MUT: Muturu, NDM: N'Dama, NGA: Nganda, NUR: Nuer, SER: Serere, SHK: Sheko. Adapted from Mekonnen *et al.* [100].

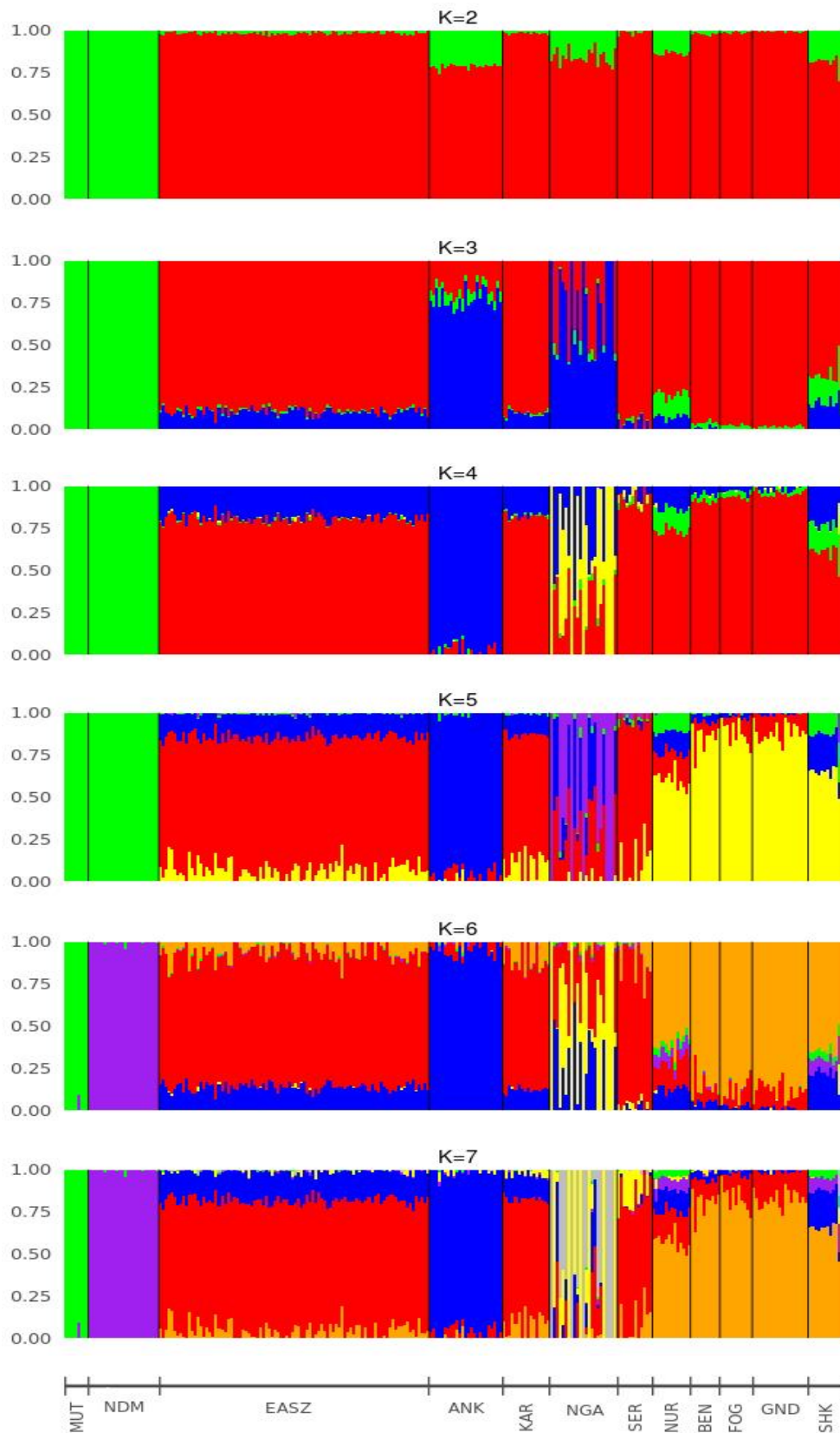


Figure 5.2.: Admixture bar plots of each individual assuming different numbers of ancestral breeds (K=2 to K=7). ANK: Ankole, BEN: Benshangul, FOG: Fogera, GND: Gindeberet, KAR: Karamojong, MUT: Muturu, NDM: N'Dama, NGA: Nganda, NUR: Nuer, SER: Serere, SHK: Sheko [100]. Adapted from Mekonnen *et al.* [100].

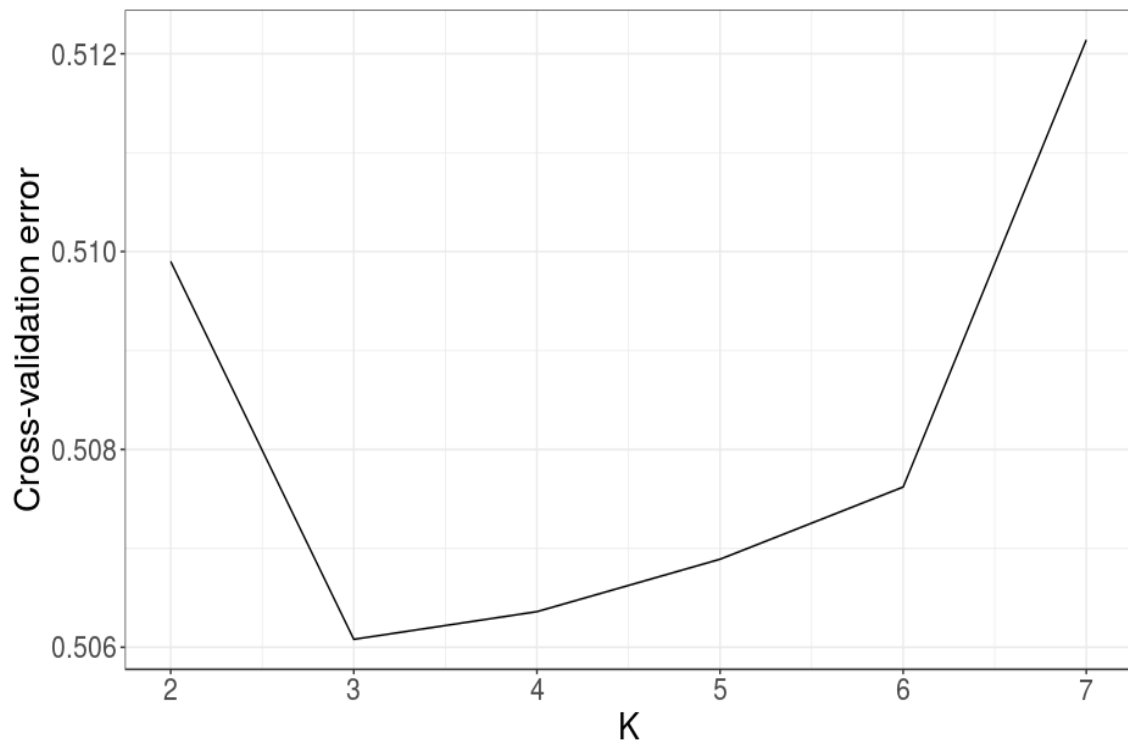


Figure 5.3.: Cross validation error in dependence to K (number of hypothetical ancestral populations) for the ADMIXTURE analyses. Adapted from Mekonnen *et al.* [100].

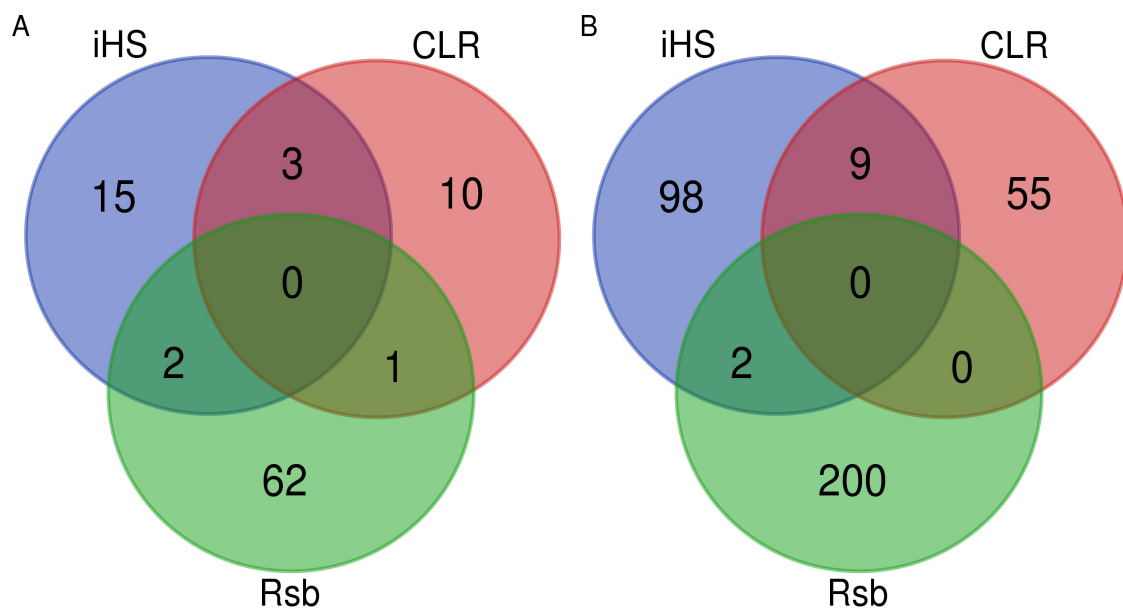


Figure 5.4.: Venn diagrams of the overlapping (A) genomic regions and (B) candidate genes identified by *iHS*, *CLR*, and *Rsb* analyses. Adapted from Mekonnen *et al.* [100].

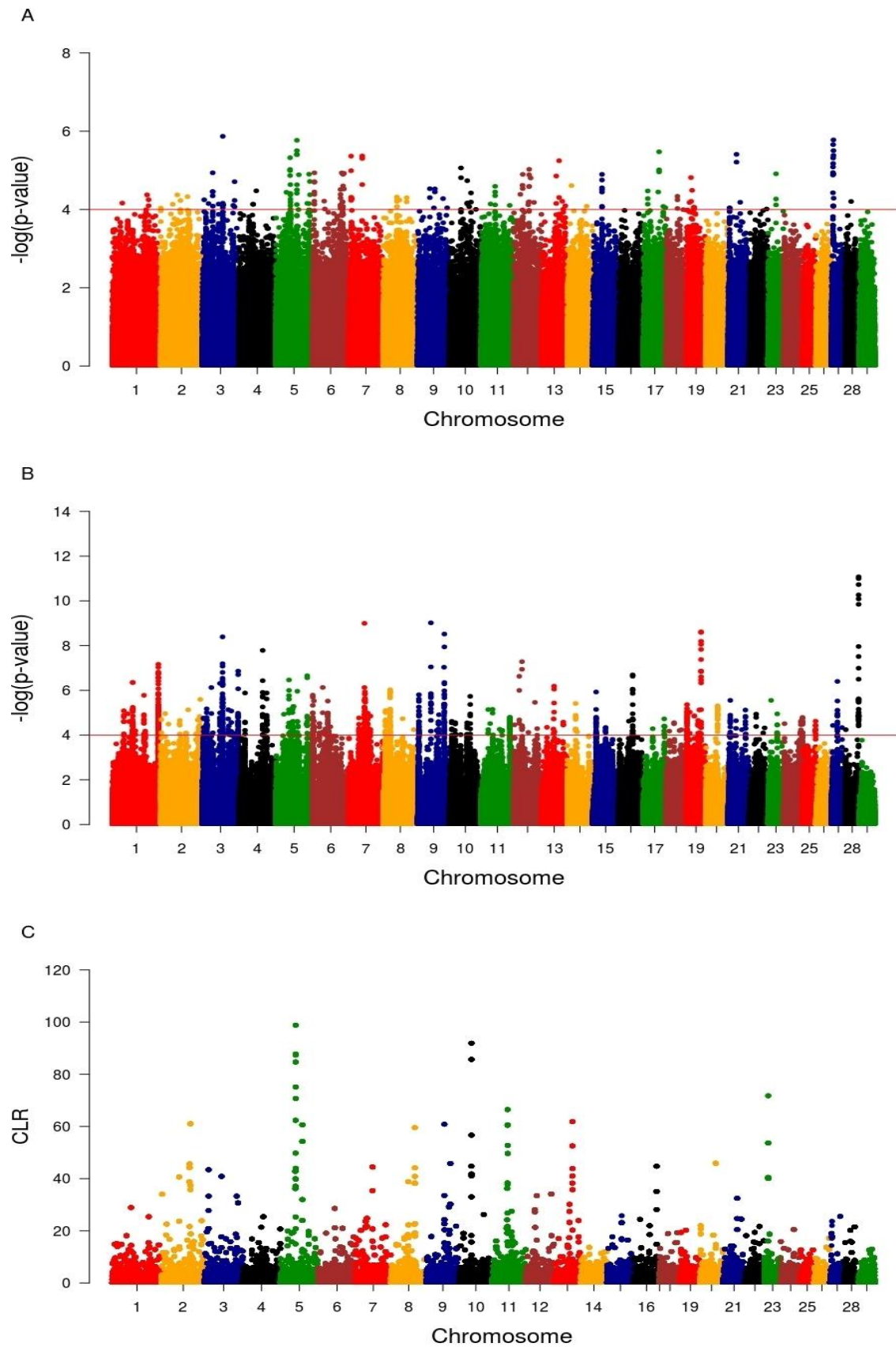


Figure 5.5.: Manhattan plots of genome-wide *iHS* (A), *Rsb* (B), and CLR (C) analyses. The x-axis shows the autosomal chromosomes and the y-axis shows -log transformed *P*-values (A and B) and CLR values (C). Adapted from Mekonnen *et al.* [100].

Table 5.2.: Summary of major candidate signature regions identified by CLR, *iHS*, and *Rsb* analyses. Adapted from Mekonnen *et al.* [100].

Genes	Method	CHR	Association	Position (UMD3.1) Start - End
MIGA1	Rsb	3	anemia, immune tolerance, and neurological dysfunction [123, 124, 125]	6706504-67137909
CDAN1	CLR	10	anemia [126, 127]	38138863-38151656
HSPA9	Rsb	7	anemia [54, 128, 129, 130, 131]	51506219-51521515
PCSK6	<i>iHS</i>	21	anemia [132]	29553201-29673109
SPAG11B	<i>iHS</i>	27	immune tolerance [133, 134]	4920083-4942958
RAET1G	Rsb	9	immune tolerance [135, 136, 137]	88232044-88402262
PPP1R14C	Rsb	9	immune tolerance, anemia, and neurological dysfunction [138, 139, 140, 141, 142]	88384683-88500749
TTC3	Rsb	1	immune tolerance and neurological dysfunction [143, 144, 145]	151034217-151141015
ERN1	Rsb	19	immune tolerance and neurological dysfunction [146, 147, 148, 149, 150, 151]	48924511-48971838
CAPG	CLR	11	immune tolerance and neurological dysfunction [146, 147, 148, 149, 150, 151, 152]	49423731-49438680
TTBK2	CLR	10	neurological dysfunction [153, 154]	38159317-38248606
POLR3B	<i>iHS</i>	5	neurological dysfunction [155, 156]	70062608-70178439
GNAS	<i>iHS</i> and CLR	13	neurological dysfunction [11, 12, 157]	58010287-58049012
CHAT	Rsb	28	listlessness [158]	44143245-44187239
AP1M1	<i>iHS</i>	7	listlessness [159]	7820650-7850254

The SPAG11B, RAET1G, PPP1R14C, and TTC3 genes are involved in immune tolerance in Sheko. The PPP1R14C gene could play an important role as a regulator of PP1 which is a competitive inhibitor of ATP binding of Src tyrosine kinase family members [138, 140]. The inhibition of Src kinase is associated with the termination of stem cell factor induced proliferation of hemopoietic cells [139]. It was also reported that Src kinases are involved as a primary activator of AKT (serine/threonine kinase family). AKT plays a critical role in adaptive immunity through the inhibition of regulatory T-cells (T_{reg} cells), which could play a key role in maintaining the immune tolerance [140, 141, 142]. AKT plays a critical role in adaptive immunity through the inhibition of T_{reg} cells, which play a key role in maintaining immune tolerance [141, 142].

However, T_{reg} cells are also involved in the host susceptibility to infection. The T_{reg} cells are matured in the thymus and released into the periphery to suppress immunity during infection [142]. This process negatively regulates adaptive immunity and increases the host susceptibility to infection. Therefore, T_{reg} cells are negatively regulated by S1P1 through the downstream mediation of the AKT-mTOR pathway. The negative regulation of T_{reg} cells by the S1P1-AKT-mTOR pathway thus promotes adaptive immunity responses [140]. In addition, activated AKT is a mediator of neuronal cell survival. In this regard, p53-mediated upregulation of pro-apoptotic protein Bax is blocked by AKT to promote neuronal cell survival [160]. Moreover, the TTC3 gene is also involved in the regulation of AKT signaling and is related to immune tolerance and neuronal cell survival [143, 144, 145].

Moreover, the candidate signature gene RAET1G is one of the few genes that could encode a ligand recognized by NKG2D proteins in response to stress and infections [135, 136, 137]. The expression of NKG2D on the cell surface of NK, dendritic and T-cells is modulated by cytokines such as IL2, IL7, IL12, IL15, IL21, IFN β , and TGF β which are involved in the host defence mechanisms [161, 162, 163, 164]. For instance, the IFN- β mediates host defence through toll-like receptor (TLR) signaling pathways against trypanosoma infection [165]. When infection occurred by pathogens, mononuclear phagocytes and dendritic cells produce interleukin-12 (IL-12) and IL-23. Natural killer cells then induce IFN- β and IFN- γ by these cytokines. In the presence of these IFNs, naive T-cell differentiates into T-helper 1 cells which play an important role in the adaptive immune system [166, 167]. Furthermore, the isoforms of the SPAG11B gene encode defensin-like peptides which are expressed by phagocytic cells [133]. These structurally diverse peptides make multimeric forms during infection and disrupt the membrane of the pathogen [134]. They are also involved in the recruitment of T- and dendritic cells to facilitate the adaptive immunity [133].

Furthermore, polymorphisms in the POLR3B, MIGA1, TTC3, ERN1, CAPG, GNAS, and TTBK2 genes might be associated with the neuronal dysfunctions (abnormalities), thyroid and parathyroid gland dysfunctions. The endoplasmic reticulum to nucleus signaling 1 (ERN1) and capping protein gelsolin-like (CAPG) genes are involved in the regulation of hypoxia (a state of a cell with inadequate or reduced oxygen availability) [146]. The reduction of the hypoxic response element in the spinal cord results in the progressive degradation of the motor neuron [147, 150]. Therefore, mutations in the ERN1 and CAPG genes are associated with neurological dysfunction [149, 150]. The ERN1 and CAPG genes might also be involved in the innate immune response since hypoxia triggers innate immune responses through the activation of the hypoxia induced factor α 1 (HIF-1 α) [147, 148, 151].

Another reported candidate signature gene related to neurological dysfunction is the TTBK2 gene. A mutation in the TTBK2 gene is associated with spinocerebellar ataxia which is a genetic syndrome causing progressive degeneration of the cerebellum and the spinal cord [153, 154]. Moreover, a mutation in the POLR3B gene is associated with hypomyelinating leukodystrophy which is characterized by a deficiency in myelin deposition of the white matter of the brain [155, 156]. In addition, the POLR3B gene is also involved in positive regulation of the interferon-beta production and the innate immune response (GO:0032728, GO:0045089). A mutation in the GNAS gene is associated with pseudohypoparathyroidism which is characterized by a low level of calcium and a high phosphate level in the blood [157]. Most importantly, the AP1M1 gene is a member of the adapter protein complex which is involved in thyroid abnormalities [158]. Due to the thyroid gland dysfunction (hypothyroidism), the nerves are unable to conduct electrical impulses properly. This leads to general weakness, lethargy, and listlessness [168]. The CHAT gene encodes a protein that catalyzes the synthesis of the neurotransmitter acetylcholine [169]. The mutation in the CHAT gene is associated with myasthenia gravis which is an autoimmune disease characterized by load-dependent muscle weakness [159].

Among the 15 identified candidate signature genes (Table 5.2), the MIGA1, RAETG, and PPP1R1AC genes are not significantly functionally enriched ($\alpha = 0.05$). Moreover, the identified

signature regions of the three methods were compared with trypanotolerant quantitative trait locus (QTL) regions which were reported by Hanotte *et al.* [25]. Among the 55 trypanotolerant QTL, which were identified by crossing trypanotolerant N'Dama and susceptible Boran, 6 regions were overlapping with trypanotolerant QTL in N'Dama (Supplementary Table 12 in Appendix A.1). Furthermore, among the identified candidate signature genes in Table 5.2, the AP1M1 and GNAS genes are found in these overlapping regions.

5.3. Functional Annotation of Candidate Signature Genes

In order to characterize the biological functions of functionally enriched candidate genes, a treemap was produced using the geneXplain platform [77]. The treemap shows the clusters of 30 functional terms. Most of these terms are associated with cellular transport, metabolic processes, and biological regulation (Figure 5.6). Among the 30 enriched terms, two GO-terms are T-cell chemotaxis and cell-cell adhesion which play a critical role in the immune system [170, 171, 172]. T-cell chemotaxis (chemoattractant cytokines) is a process that requires the movement of T-cells in response to a certain signal or external stimulus. The movement or circulation of immune cells in the blood and lymph as non-adherent cells and in tissues as adherent cells is critical for patrolling the body against infectious organisms effectively [170]. For instance, β defensin is chemotactic for chemokine receptors of macrophages, natural killer cells, immature dendritic cells, and memory T-cells. Therefore, the recruitment of these cells to the site of a microbial invasion provides a link between innate and adaptive immunity [133]. In the presence of infectious organisms (foreign antigens), the immune cells aggregate at the site of the infection and through their adhesion receptors, they adhere to cells bearing a foreign antigen and trigger destruction [170].

5.4. Identification of Overrepresented Pathways in the Candidate Signature Gene Sets

I performed pathway analysis using the TRANSPATH database of the geneXplain platform. The TRANSPATH pathway analysis identified 15 genes out of 260 functionally enriched genes that are involved in 13 overrepresented TRANSPATH pathways (Table 5.3). Among these genes, the immunoproteasome PSMD7 gene is involved in most of the overrepresented pathways. This gene is involved in the processes of presenting antigens by the major histocompatibility complex (MHC) class I proteins to CD8+ T lymphocytes [173, 174, 175]. The sufficient induction of CD8+ during infection leads to pathogen elimination. This gene plays a critical role in the development of adaptive immunity or tolerance [176].

In general, most of the overrepresented pathways (PDGFB \rightarrow STATs, stress associated pathways, IMP \rightarrow ADP, ARIP1 \rightarrow atrophin 1, p38 pathway, IL-3 signaling, oxygen independent HIF-1alpha degradation, and Cul3 $-/$ Nrf2) are activated by cellular stresses and antigens while others [E2F network, G2/M phase (cyclin B:Cdk1), S phase (Cdk2), Plk1 cell cycle regulation, and Aurora-B cell cycle regulation] are involved in cell cycle processes.

Table 5.3.: Overrepresented pathways for the identified candidate signature genes. The names of the pathways are provided by the TRANSPATH database of the geneXplain platform. Adapted from Mekonnen *et al.* [100].

Pathway	P-value	Genes
PDGF B → STATs	0.003	STAT3, STAT5A
stress-associated pathways	0.007	MBP, MEF2A, PSMD7, RAF1, RBX1, STAT3
E2F network	0.008	AKT3, CDC25C, PPP2R5A, PSMD7, RAF1, RBX1
G2/M phase (cyclin B:Cdk1)	0.015	AKT3, CDC25C, PSMD7, RBX1
IMP → ADP	0.025	AK5, AMPD3
ARIP1 → atrophin1	0.034	AKT3, APBA1
p38 pathway	0.039	MBP, MEF2A, STAT3
Plk1 cell cycle regulation	0,039	CDC25C, PSMD7, RBX1
IL-3 signaling	0.043	MBP, RAF1, STAT5A
Aurora-B cell cycle regulation	0.045	CENPE, PSMD7, RBX1
oxygen independent HIF-1alpha degradation	0.045	PSMD7, RBX1, UBE2R2
Cul3 —/ Nrf2	0.047	PSMD7, RBX1
S phase (Cdk2)	0.048	CDC25C, RAF1, RBX1

The first two pathways in Table 5.3 (PDGFB → STATs and stress associated pathways) are related to the immune system and anemia. Especially, in stress associated pathways, I find MBP, RAF1, MEF2A, and STAT3 genes that are involved in the immune and nervous systems. In the MBP gene, there are eight different mRNAs due to the alternative splicing of exons [177]. Three of the eight splice variants are expressed in the brain, macrophages and hemolymphopoietic tissues such as the spleen, bone marrow, and thymus [177]. This gene is also involved in the interleukin (IL)-3 signaling pathway. IL-3 is a T-cell-derived hematopoiesis stimulating cytokine which is involved in the production, differentiation, and functioning of granulocytes and macrophages [178, 179]. The serine/threonine kinase proto-oncogene RAF1 is also related with the stress associated pathway and is involved in inducing adaptive immunity by regulating the expression of cytokines that are important for the differentiation of T-helper cells [180].

Moreover, STATs family members are also involved in the activation of various cytokines and in the promotion of cell survival by inducing the expression of antiapoptotic BCL2L1/BCL-X(L) genes [181, 182, 183]. For instance, STAT3 activation by trypanosomes was associated with the survival of cardiomyocytes during infection [184, 185]. The other gene involved in defence response is MEF2A which is associated with promoting antimicrobial peptide expression during infection [186]. This gene is also involved in neuronal cell survival and loss of function [187, 188]. Neurotoxins induce ubiquitination of MEF2A in response to toxic stress which leads to the loss of neuronal viability [189]. Furthermore, the increment of platelet-derived growth factor (PDGF)-B related signaling is associated with induced chemokine secretion, which is a mediator of innate and adaptive immune responses [190, 191].

The E2F network as well as the Cdk1 and Cdk2 related pathways are associated with cell cycle processes [192, 193]. The tumor suppressor retinoblastoma (Rb) is the inhibitor of E2Fs. When Rb binds to E2Fs, it prevents E2F mediated activation of transcriptional genes. In quiescent cells,

E2F is required for cell differentiation through a series of signal transduction cascades, including Cdks activation and phosphorylation. The Aurora-B and Plk1 pathways are involved in the activation and phosphorylation of Cdks, respectively. As a result of these and several other signaling cascades, E2Fs are activated while inactivating Rb. The activated E2F mediates quiescent cells for S phase entry and cell cycle progression [192, 194, 195, 196, 197]. The serine/threonine kinase family isoforms of the AKT gene is also involved in the E2F and Cdk1 pathways. This gene is activated in the host cells during trypanosome infection [198, 199]. The host kinase AKT promotes infected host cell survival and restricts the growth of intracellular parasites [200]. AKT3 is also a key mediator of down stream signaling pathways of activated receptor tyrosine kinases (RTKs) which play a role in STAT3 activation [183, 198]. The different isoforms of the kinase AKT regulate the development of immunity and autoimmunity. AKT is predominantly expressed in the innate immune cells [164]. The isoforms of AKT are primarily involved in regulating inflammatory responses although it has been reported that AKT also modulates adaptive immune responses [140]. Moreover, the AKT related pathway Atrophin-1 plays a role in erythroid and lymphoid cell differentiation and in E3 ubiquitin ligase atrophin-1 interacting protein 4 (ITCH) signaling cascades. Atrophin-1 is involved in the regulation of immune responses through Notch-mediated signaling pathways [201, 202, 203]. It is also associated with spinocerebellar degeneration caused by extended CAG repeats encoding several glutamine units (polyglutamine tract) in the atrophin-1 protein [204].

Furthermore, the hypoxia inducible factor (HIF) and the nuclear factor-erythroid 2-related factor 2 (NRF2) pathways are related to immune-mediated hemolysis [205]. During hypoxia, HIF facilitates a high production of red blood cells (erythropoiesis) in order to overcome the shortage of oxygen [206]. The other pathway, NRF2, regulates the expression of an antioxidant responsive element (ARE) driven gene and plays a critical role in the ARE-driven cellular protection [207].

Further important pathways are p38, $\text{IMP} \rightarrow \text{ADP}$, and the aurora B-cell cycle regulation pathways that are involved in the host defence mechanism [194, 208, 209, 210, 211]. The p38 pathway is a MAPK-related pathway which is activated by various physical and chemical stresses such as hypoxia and various cytokines. The activation of the p38 pathway is critical for normal immunity and inflammatory responses [208]. Moreover, the AK5 and AMPD3 genes are involved in the $\text{IMP} \rightarrow \text{ADP}$ pathway and play a central role in the regulation of inflammation and red blood cell homeostasis [209, 210]. AK5 is associated with double-positive thymocyte and auto-immunity regulation in the brain and pancreatic tissues [212], while the AMPD3 gene is involved in the regulation of the energy state of red blood cells during oxidative stress (hypoxia) [209]. In addition to that, the aurora B-cell cycle regulation pathway is involved in the progression of T-lymphocytes which play a critical role in the development of innate and adaptive immunity [194, 211]. The HIF and NRF2 related pathways are directly associated with the induction of host innate and adaptive immunity under oxidative stress [54, 55, 56, 57, 213].

5.5. Identification of Master Regulators based on Candidate Signature Genes

To gain more insight into the regulatory mechanisms of the identified candidate signature genes, I performed a master regulatory network analysis using the TRANSPATH database in the geneXplain platform. Applying the maximum radius of 10 steps upstream in the regulatory hierarchy, I identified ten master regulators (Figure 5.7). The master regulator Caspase, which is a family of protease enzymes, is associated mainly with regulating the reduction of the load of intracellular parasites, induction of nitric oxide (NO) production, increasing the level of CD4 and CD8+ T-cells, secretion of IFN γ , and control of trypanosome infection by macrophages [214]. This master regulator is involved in programmed cell death such as pyroptosis and necroptosis. These types of programmed cell deaths play a role for protecting an organism against oxidative stress (stress signals) and pathogenic attack [215]. In addition, Caspase also plays a role in the normal erythroid differentiation in the terminal stages [216].

Most of the regulatory molecules (Syk, Lck, Lyn, Jak1, Jak2, and Jak3) are protein tyrosine kinases while others (VHR and PTP1B) are protein tyrosine phosphatases and activated kinase (PAK1). These master regulators are mainly associated with innate and adaptive immune responses and are critical for the functioning of the nervous and immune systems. For instance, the activation of the regulatory molecule Syk requires the regulatory molecule Lck to phosphorylate immunoreceptor tyrosine-based activation motifs (ITAMs). Then, the phosphorylated ITAMs modulate T-cell proliferation and differentiation by recruiting Syk protein tyrosine kinases [217, 218]. In addition, coupling of the other master molecules JAK1 and JAK3 occurs on the cell surface receptor of IFN γ , followed by phosphorylation of the IFN γ receptor 1. This process leads to the activation of the STAT1 protein. The STAT1 protein binds to the target element of the IFN γ inducible gene in the nucleus and facilitates the transcription of the target regions during immunity responses [167, 166]. Another reported regulator molecule VHR is also involved in the phosphorylation of STAT proteins and in the T-lymphocyte physiology [219, 220]. Moreover, the regulatory molecule JAK2 plays a critical role in the maintenance of hematopoiesis.

Furthermore, a related master molecule, the protein tyrosine phosphatase 1B (PTP1B), is reported to modulate the activation of macrophages and plays a key role in mediating the central dendritic cell function of bridging innate and adaptive immunity [221, 222]. The kinase family regulator molecule Lyn is also involved in the regulation of innate and adaptive immune responses [223]. Lyn is also known for mediating the production of type I interferon (IFN-I) which is involved in host defence mechanisms against invading pathogens [224, 225, 226]. The related kinase regulatory molecule PAK1 is highly expressed in most leukocytes that are involved in immune responses. PAK1 also plays an important role in the activation of MAP-kinase pathways which are involved in all aspects of immune responses, from innate immunity to the activation of adaptive immune responses [227, 228, 229, 230, 231, 232].

The stress induced protein kinases could also induce or aggravate auto-immunity by phosphorylating self-antigens to be recognized by auto-antibodies [233, 234]. However, Caspase-mediated apoptosis plays an important role in arresting the development of auto-immunity by eliminating

auto-reactive and pro-inflammatory cells [235]. Moreover, the activation of Caspase and JAK2 is essential for the processes of erythroid differentiation and for the maintenance of hematopoiesis [216]. On the other hand, the inhibition of Caspase dependent mechanisms contributes to cell survival [236].

5.6. Identification of Common Candidate Signature Genes Using GWAS and Signature of Selection Detecting Methods

GWAS and signature of selection detecting methods identify the three genes MAPT, STXBP5L, and ALCAM that are commonly captured by both approaches. The MAPT gene is identified in the trypanotolerant group (Sheko, N'Dama, and Muturu) whereas both STXBP5L and ALCAM genes are identified in Sheko. The mutation in the MAPT gene is associated with neurological disorders such as dementia [237] and Parkinson's diseases [238]. The STXBP5L gene is orthologous to the mouse gene STXBP5 [239] which is involved in the secretion of platelet and in the normal hemostasis [240]. Moreover, the ALCAM gene is involved in cell-cell adhesion and cell migration processes [241].



Figure 5.6.: Gene Ontology treemap for the 260 functionally enriched ($P < 0.05$) genes. The size of the boxes corresponds to the $-\log_{10} P$ -value of the GO-term. The boxes are grouped together based on the upper-hierarchy GO-term which is written in bold letters. Adapted from Mekonnen *et al.* [100].

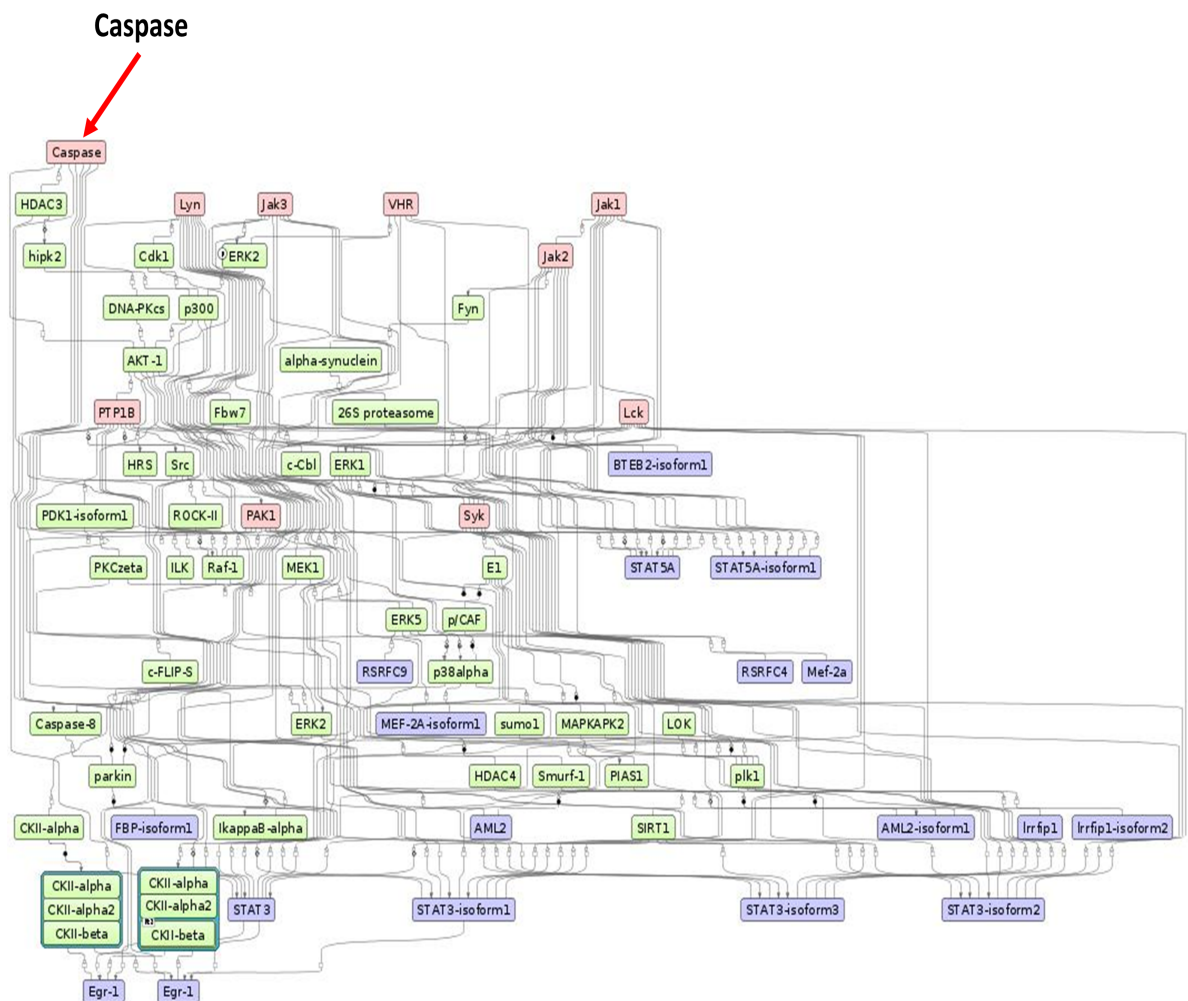


Figure 5.7.: The master regulatory networks identified in Sheko (Caspase, Lyn, Jak1, Jak2, Jak3, VHR, PTP1B, PAK1, Lck and Syk). Red, blue and green indicate master regulators, regulated proteins and connecting molecules, respectively. Adapted from Mekonnen *et al.* [100].

5.7. Comparative identification of trypanotolerance tendency in Nuer, Benshangul, and Gindeberet breeds

5.7.1. Comparative Identification of Signature Genes Associated with Trypanotolerance

The combined gene sets of *iHS*, *CLR*, and *Rsb* analyses contain 220, 449, 1022, and 439 candidate genes with a signature of positive selection in 22, 27, 29, and 29 autosomes in Sheko ([100]), Benshangul, Gindeberet, and Nuer, respectively (Figure 5.8 A). The comparison of the genomic regions and signature genes among these breeds using the combined gene sets of *iHS*, *CLR*, and *Rsb* analyses reveal that Benshangul and Gindeberet share the biggest overlaps in both genomic regions and genes while Sheko and Benshangul share few genes (Figure 5.8). To explore the trypanotolerance tendencies of Benshangul, Nuer, and Gindeberet breeds, the identified signature genes of these breeds are compared with those of Sheko. Sheko shares 15 genes (i.e., referring to shared haplotypes) with Benshangul, 38 genes with Gindeberet, and 34 genes with Nuer. On these overlapping gene sets, I perform a literature survey and identify 6 genes in Benshangul (*STXBP3*, *GNAS*, *KCNIP4*, *DHX33*, *C1QBP*, and *AK5*), 3 genes in Gindeberet (*JMJD1C*, *EGR1*, and *SLC24A3*), and 2 genes (*CD86* and *L3MBTL3*) in Nuer are associated with trypanotolerance attributes. The *STXBP3*, *GNAS*, and *KCNIP4* genes are involved in neuronal dysfunctions. Especially, the *KCNIP4* gene encodes a potassium ion transporter protein which is directly involved in neuronal functions [242]. The polymorphisms in *KCNIP4* gene are associated with neuro-developmental disorders such as Attention-deficit/hyperactivity disorder (ADHD) [243]. Moreover, the *STXBP3* gene is involved in neurological dysfunction such as hearing loss [244], whereas the *GNAS* gene is associated with pseudohypoparathyroidism (low level of calcium and a high phosphate level in the blood) [157].

Another commonly identified gene in both Benshangul and Sheko is *DHX33*. This gene is involved in the regulation of RNA induced NLRP3 inflammasome activation [245]. Inflammasomes facilitate the activation of pro-inflammatory caspase-1 which catalyzes the maturation of pro-IL-1 β and pro-IL-18. The activated IL-1 β is involved in antigen presenting processes in host defence and in the production of CD4+ and CD8+ T-cells. These T-cells produce IFN- γ which activates the phagocytic cells to kill invading parasites [246]. Moreover, the *C1QBP* is reported as a direct target gene of the *ZNF32* gene which plays a critical role in the cellular resistance to oxidative stress [247]. In addition, the *AK5* gene is involved in the regulation of thymocyte maturation and auto-immunity [212, 248].

Furthermore, the *JMJD1C* and *EGR1* genes which are commonly identified in Gindeberet and Sheko show upregulation in hypoxic conditions that are induced by the hypoxia-inducible factor (HIF)-1 [249, 250]. The *SLC24A3* gene is involved in cellular calcium ion homeostasis and its expression is induced by chronic hypoxia [251]. Moreover, the *ARG1* gene is involved in the defence response against protozoans (GO:0042832). This gene is also involved in parasite-induced tissue damage repairing processes [252, 253].

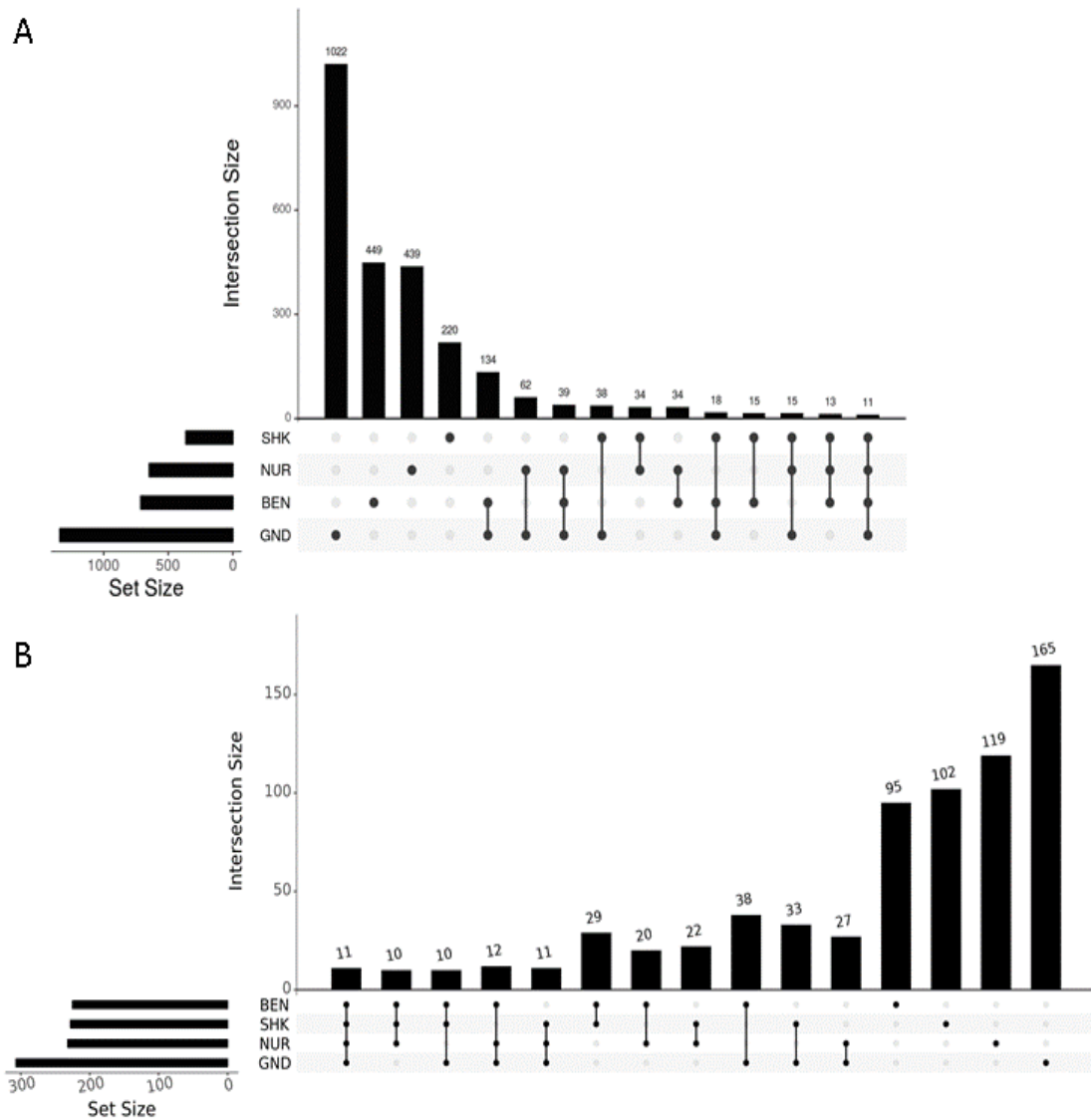


Figure 5.8.: UpSet plot of the overlapping candidate genes (A) and genomic regions (B) identified by *iHS*, *CLR* and *Rsb*.

Furthermore, the CD86 gene in Nuer and Sheko plays an important role in T-cell activation and co-stimulation of naive T-cells [254] while the L3MBTL3 gene is involved in erythropoiesis.

Sheko, Nuer, Benshangul, and Gindeberet breeds share FTMT, RAET1G, and PPP1R14C genes. The FTMT gene encodes a protein that plays a protective role against oxidative stress by storing excess iron. However, the accumulation of iron as ferric iron leads to iron deficiency in refractory anemia with ring sideroblasts (RARS) [255, 256, 257]. The function of the RAET1G and PPP1R14C genes are presented in section 5.2.

5.7.2. Treemap Comparisons Between Breeds

The biological functions for the combined signature gene sets from each Benshangul, Gindeberet, and Nuer breeds are clustered into a treemap of 30 functional terms (Figures 5.9 - 5.11). Most of

these terms are involved in cellular transport, metabolic process, and regulation. Among the 30 enriched functional terms, one of the GO-terms in Sheko (see Figure 5.6) and Nuer (Figure 5.9) is T cell chemotaxis which plays a critical role in immune responses [171, 172] (see section 5.3).

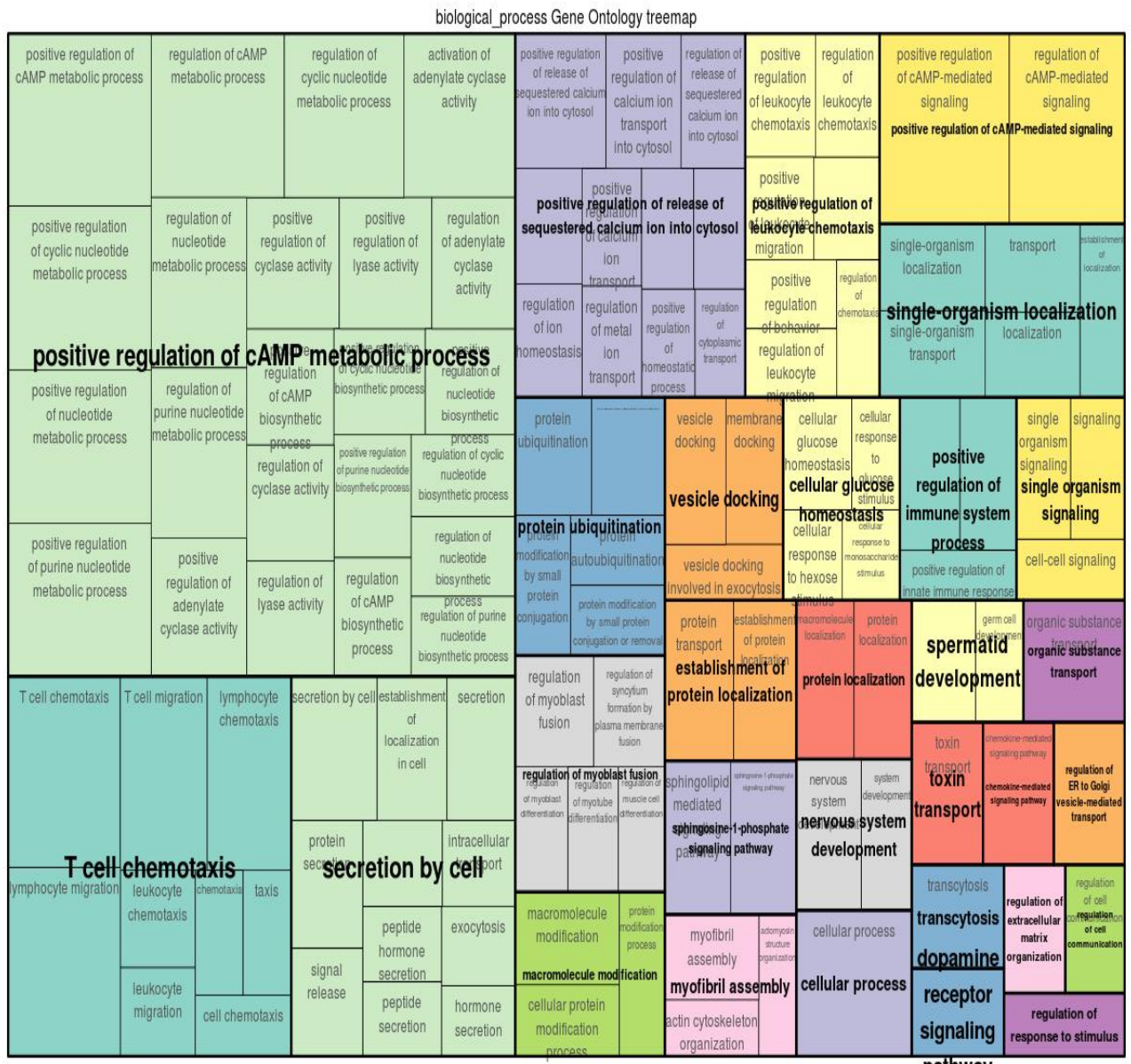


Figure 5.9.: Gene Ontology treemap for the functionally enriched ($P < 0.05$) genes identified in Nuer breed. The size of the boxes corresponds to the $-\log_{10} P$ -value of the GO-term. The boxes are grouped together based on the upper-hierarchy GO-term which is written in bold letters.

5.7.3. Comparison with N'Dama and Muturu for Signature of Positive Selection against Trypanosomiasis

Among the 55 trypanotolerant QTL identified in N'Dama [25], 23 regions in Benshangul, 21 regions in Nuer, and 26 regions in Gindeberet overlap. Moreover, among the genes that are associated with anemia in N'Dama [17, 258], RPS26 gene in Benshangul and Gindeberet, GNA14 gene

in Gindeberet and Sheko [100], LCT and MCM6 genes in Gindeberet, FGF20 gene in Sheko [100], and MADCAM1 gene in Benshangul are commonly identified. Moreover, the ENSBTAG00000047764 and ST6GALNAC5 genes in Benshangul, Gindeberet, Nuer, and Sheko [100] are commonly identified in the Muturu breed which is reported in the literature [106]. Furthermore, Benshangul and Nuer share the JAZF1, PPP2R2A, and BNIP3L genes while only Benshangul and Sheko [100] share the AK5 gene with Muturu [106].

The ENSBTAG00000047764 gene is orthologous to the human gene ABCC4 which is involved in cellular detoxification [259] while the ST6GALNAC5 and JAZF1 genes are involved in the glycolipid biosynthesis [260, 261]. The PPP2R2A gene is expressed as the sub-unit of the regulatory protein phosphatases 2 (PP2A) which is entangled in a variety of cellular processes involved in the (negative) control of cell growth and division [262]. Moreover, the BNIP3L gene is involved in apoptosis and cellular responses to hypoxia [263, 264].

5.7.4. Comparative Identification of Hub Genes Involved in Trypanotolerance

I performed a gene network search analysis using GeneMANIA and identified the top three hub genes that have the highest connectivity with the nodes as shown in Figures 5.12 - 5.15. Among the top three hub genes, the KCNIP4 gene is shared by all breeds except Nuer. The *Foxp1* hub gene is found in both Sheko and Benshangul, while the *FYN* hub gene is shared by Benshangul and Nuer. Moreover, the *PSMD7* hub gene in Sheko, *CTNNA2* and *RDX* hub genes in Gindeberet, *AAK1* and *SYT1* hub genes in Nuer are unique for each breed. The biological function of the most commonly shared hub gene *KCNIP4* is described in section 5.7.1.

Moreover, the hub gene *FOXP1* is involved in the transcriptional regulatory networks of B lymphopoiesis, cardiomyocyte proliferation, motor neuron, and regulatory T-cell functions [265, 266, 267, 268]. Further, the hub gene *FYN* is activated by hypoxia [269] and interacts with *Nrf2* by phosphorylating tyrosine 568 of *Nrf2* which is involved in the activation of defence genes during stressful conditions [270]. In addition, the hub gene *FYN* plays a critical role in the initiation of T-cell receptor signaling pathways [271].

Furthermore, mutations in the hub genes *RDX* and *CTNNA2* are associated with neurological dysfunctions such as hearing loss [272, 273]. The *RDX* hub gene is also involved in leukocyte extravasation signaling processes [274]. The other hub gene, *AAK1*, which is identified in Nuer is involved in hypoxia induced cellular processes [275], whereas the hub gene *SYT1* functions as calcium sensor in the cell that triggers cellular response to a calcium ion [276]. The function of the hub gene *PSMD7* identified in Sheko is described in section 5.4.

5.7.5. Comparative Identification of Overrepresented Pathway Involved in Trypanotolerance

The TRANSPATH pathway analysis reveals 29, 18, and 32 genes involved in 20, 12, and 22 overrepresented TRANSPATH pathways in Benshangul, Nuer, and Gindeberet, respectively (Table 5.4). Sheko (see Table 5.2) and Gindeberet share two overrepresented pathways involved in PDGFB → STATs and ARIP1 → atrophin1 signaling cascades. Gindeberet and Benshangul share

two overrepresented pathways involved in E1 → ERBB3 and p53 related signaling cascades. The PDGF B → STATs pathway is the topmost overrepresented pathway in Sheko which is involved in the induction of chemokine secretion and plays an important role in innate and adaptive immunity responses [191] as described in section 5.4. The other overrepresented pathway shared by Sheko and Gindeberet is ARIP1 → atrophin1 which is involved in erythroid and lymphoid differentiation. Moreover, atrophin1 related pathways play a role in immune responses [201, 202, 203] and are also associated with spinocerebellar degeneration (see section 5.4).

The E1 → ERBB3 pathway, shared by Gindeberet and Benshangul, is the second most overrepresented pathway in Benshangul which is involved in the ubiquitin-proteasome system and plays a key role in the degradation of cellular proteins [277]. The topmost overrepresented pathway in Benshangul is a platelet-activating factor (PAF) which is related to host defence system [278]. The other overrepresented pathway commonly shared by Gindeberet and Benshangul is p53 → cytochrome C which is involved in the induction of apoptosis by Caspase activation [279]. The two topmost overrepresented pathways in Gindeberet are N-cadherin and parkin related pathways which are involved in immunity and neurological dysfunctions [170, 280, 281]. The N-cadherin related pathway is involved in cell migration and cell-cell contact [281] which play an important role to patrol the body against pathogens [170]. The parkin related pathway is related to parkin protein. This protein is associated with a neurological disorder known as autosomal recessive juvenile parkinsonism (AR-JP) [280].

There were no overlapping overrepresented pathways between Nuer and the other three breeds. The first overrepresented pathway in Nuer is Syk → RhoA which is involved in the engulfment of the pathogen by macrophages [282], whereas the second overrepresented pathway is acetyl-CoA → palmitic acid which is involved in lipid metabolism [283].



Figure 5.10.: Gene Ontology treemap for the functionally enriched ($P < 0.05$) genes identified in Gindeberet breed. The size of the boxes corresponds to the $-\log_{10} P$ -value of the GO-term. The boxes are grouped together based on the upper-hierarchy GO-term which is written in bold letters.



Figure 5.11.: Gene Ontology treemap for the functionally enriched ($P < 0.05$) genes identified in Benshangul breed. The size of the boxes corresponds to the $-\log_{10} P$ -value of the GO-term. The boxes are grouped together based on the upper-hierarchy GO-term which is written in bold letters.

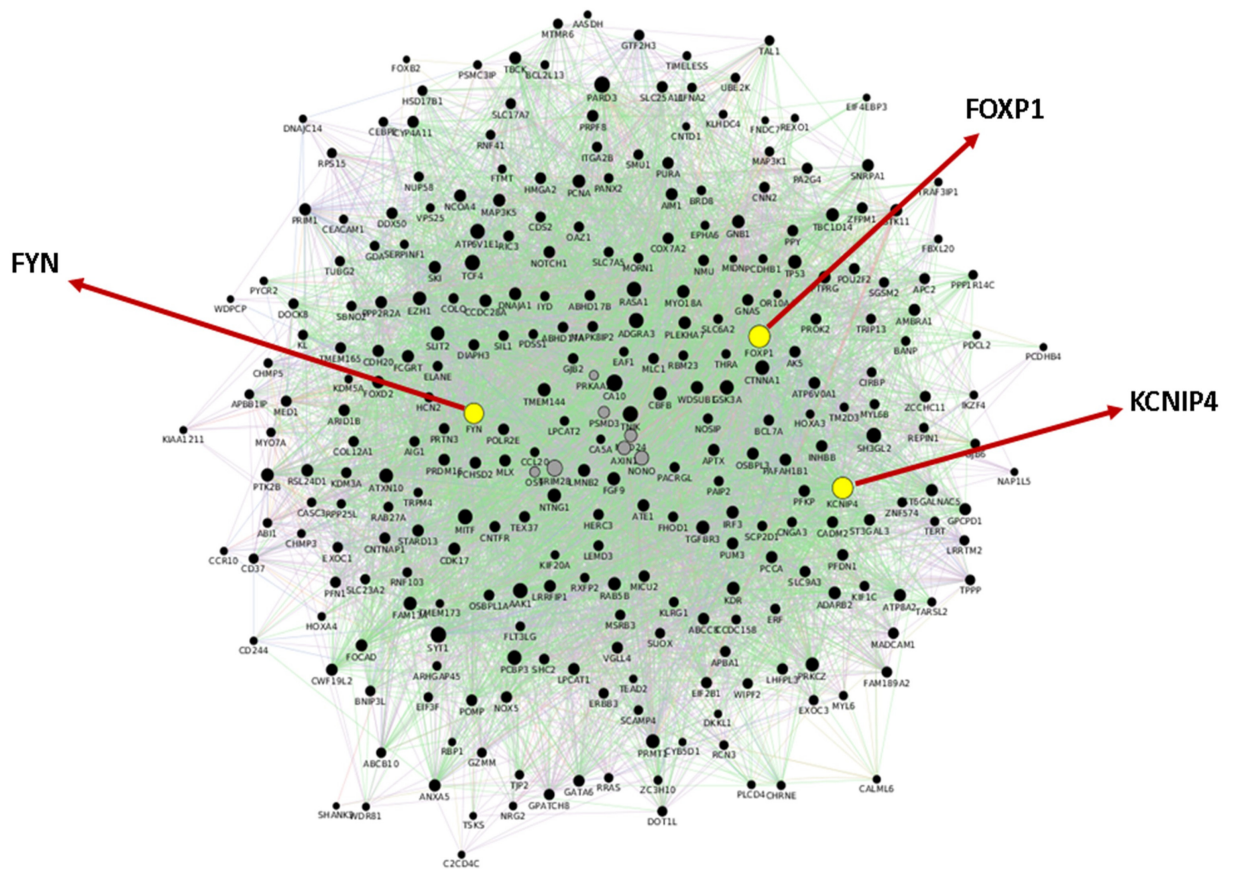


Figure 5.12.: Gene networks identified in Benshangul. The highlighted circles indicate hub genes.

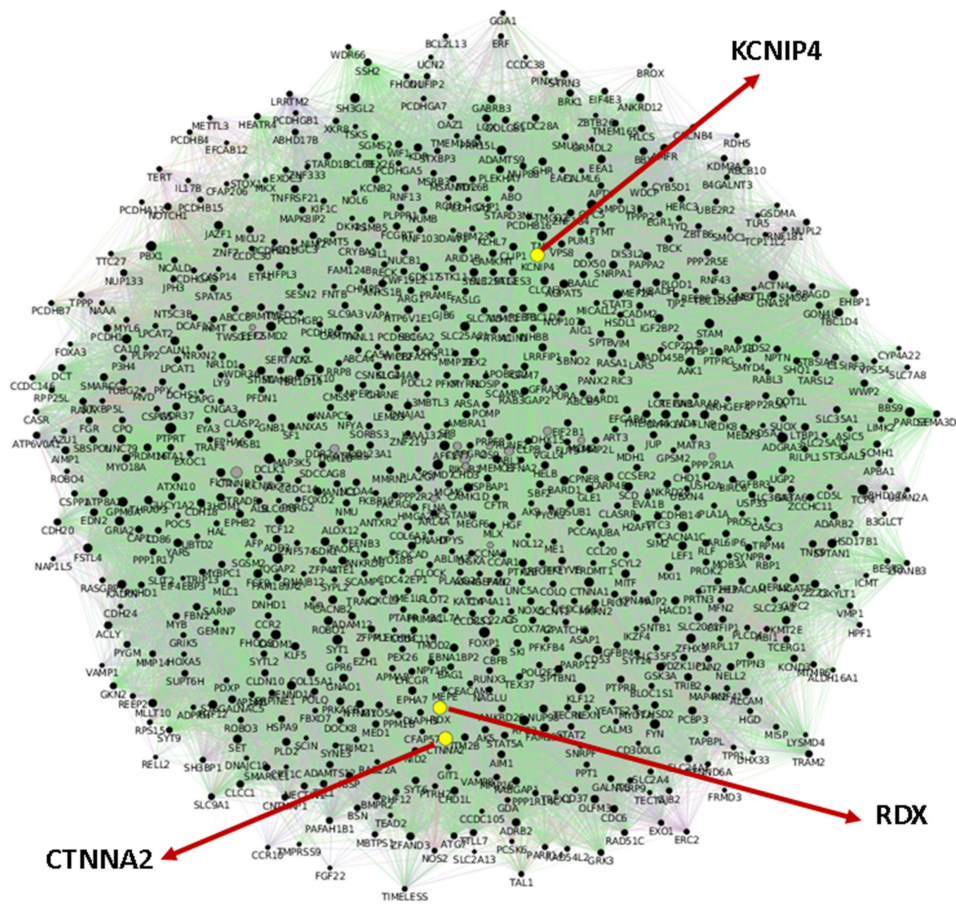


Figure 5.13.: Gene networks identified in Gindeberet. The highlighted circles indicate hub genes.

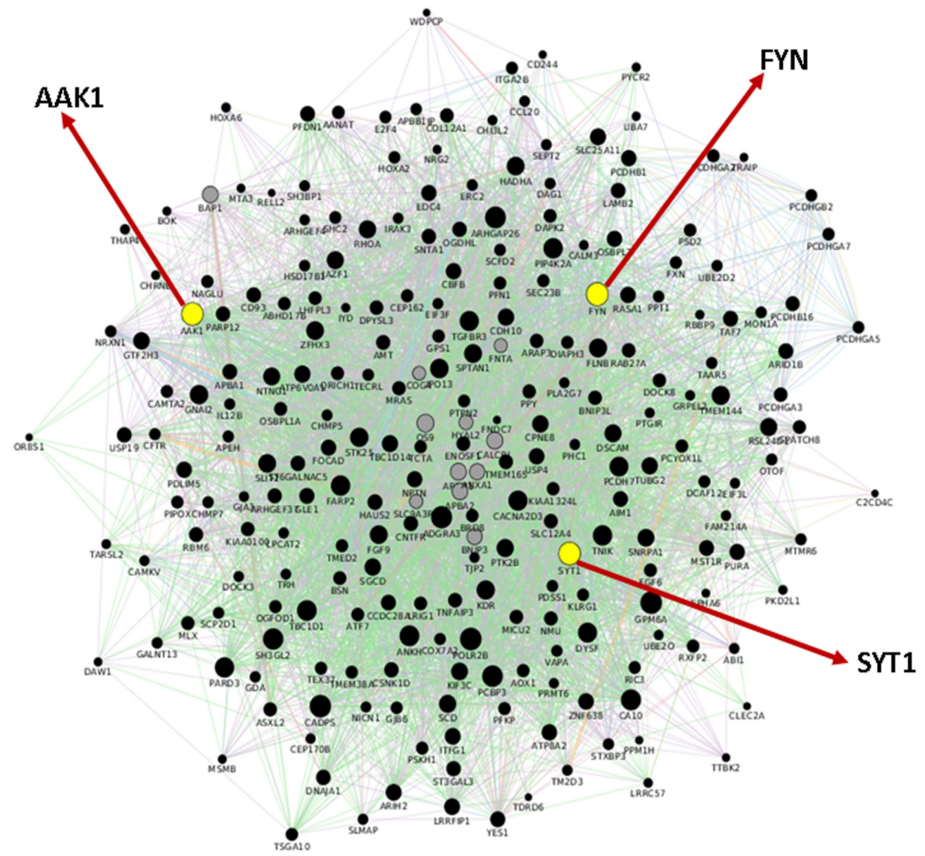


Figure 5.14.: Gene networks identified in Nuer. The highlighted circles indicate hub genes.

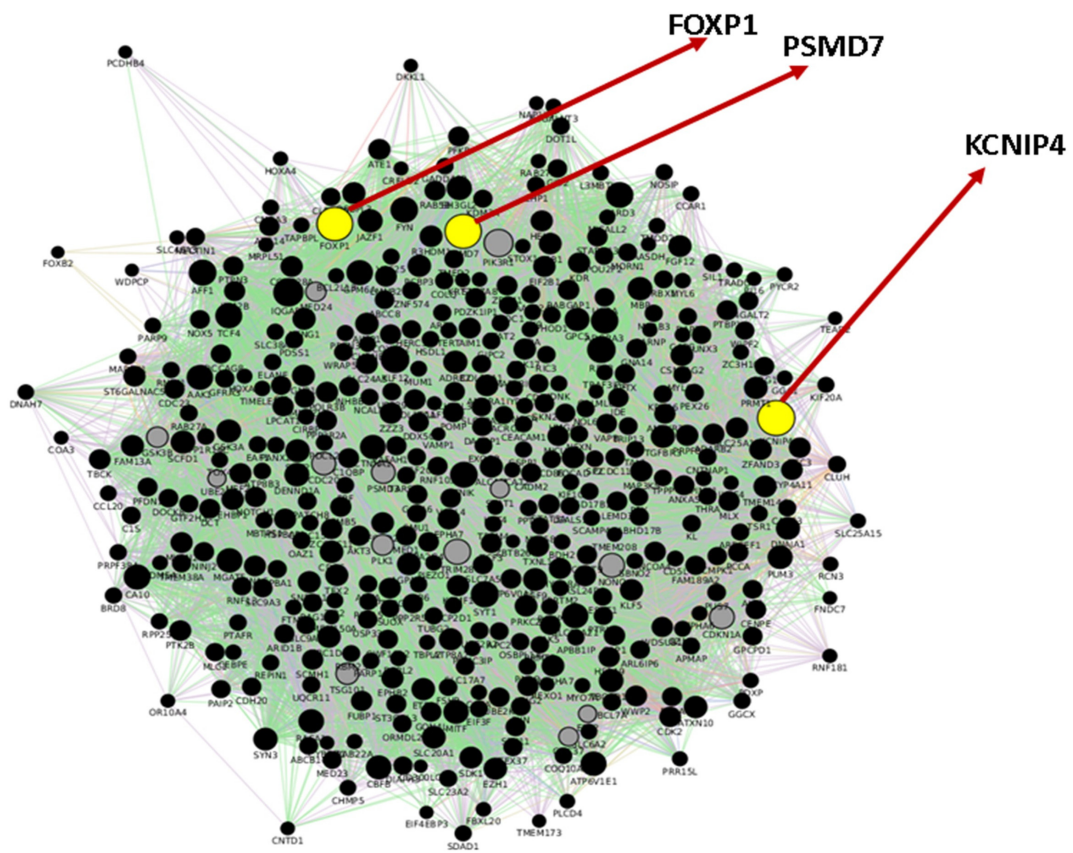


Figure 5.15.: Gene networks identified in Sheko. The highlighted circles indicate hub genes.

Table 5.4.: Overrepresented pathway analyses for Benshangul, Gindeberet, and Nuer breeds.

Pathways	P-value	Genes
Gindeberet		
N-cadherin:plakoglobin, homophilic ligation	0.003	CTNNA1, CTNNA2, JUP
NOS —/ parkin	0.003	NOS2, PRKN
E1 —/ ErbB3	0.008	ERBB3, RNF41, UBE2D2
H2O2 → Pyk2	0.009	PLD2, PTK2B
9-cis-retinol → 9-cis-retinoic acid	0.009	HSD17B6, RDH5
Nrdp1 —/ apollon	0.013	BIRC6, PSMA7, PSMD3, RNF41, UBE2D2
Cdk2 → p53	0.018	CDK2, TP53
ARIP1 → atrophin1	0.024	AKT1, AKT3, APBA1, CTNNA1
N-cadherin network	0.027	CTNNA1, CTNNA2, JUP
citrate cycle	0.028	ACLY, CS, DLST, MDH1
CH000000331	0.028	CTNNA1, JUP
OSM → STAT3	0.028	OSMR, STAT3
EGF → STAT3	0.028	PTK2B, STAT3
AKT-1 —/ JNK1	0.028	AKT1, MAP3K5
p53 → cytochrome C	0.028	AKT1, AKT3
uridine → beta-alanine	0.028	DPYD, DPYS
2'-deoxythymidine → deoxythymidine 5'-monophosphate	0.028	DPYD, DPYS
PDGF B → STATs	0.041	STAT3, STAT5A
Bad → 14-3-3	0.041	AKT1, PRKACB
N-cadherin:beta-catenin, homophilic ligation	0.041	CTNNA1, CTNNA2
VE-cadherin, ligation	0.041	CTNNA1, CTNNA2
Nuer		
Syk → RhoA	0.002	RHOA, SYK
acetyl-CoA → palmitic acid	0.002	FASN, PPT1
fatty acid synthesis	0.007	FASN, PPT1
PMCA4 —/ nNOS	0.007	DAG1, SNTA1
fatty acid chain elongation	0.016	FASN, HADHA
acyl-CoA, malonyl-CoA → fatty acyl-CoA	0.016	FASN, HADHA
glucose-1-p → UDP-D-galactose	0.020	B4GALT1, B4GALT2
lactose metabolism	0.020	B4GALT1, B4GALT2
cytidine 5'-phosphate → RNA-P-C	0.025	POLR1B, POLR1E, POLR2B, POLR3D
histidine metabolism	0.026	AOC1, AOX1
biosynthesis of saturated and n - 9 series of MUFA and PUFA	0.026	FASN, SCD
L-methionine → dimethylglycine	0.032	BHMT2, DNMT3B
Benshangul		
1-alkyl-glycerol 3-phosphate → platelet-activating factor	0.001	LPCAT1, LPCAT2
E1 —/ ErbB3	0.002	ERBB3, RNF41, UBE2D2
Cdk2 → p53	0.006	CDK2, TP53
p53 → cytochrome C	0.011	BID, TP53
kennedy pathway	0.014	CDS2, LPCAT1, LPCAT2, PLPP2
CH000000249	0.016	MBD3, TP53
CH000000250	0.016	MBD3, TP53
sn-glycerol 3-phosphate → lysophosphatidylserine	0.019	LPCAT1, LPCAT2, PLPP2
Fer → beta-cateninTyr142	0.022	CTNNA1, FYN
dsRNA → p50:RelA	0.026	CDC34, MAP3K1, UBE2D2
TNF-alpha → c-Jun	0.029	MAP3K5, TRADD
alpha-D-Ribose 5-phosphate → inosine 5'-phosphate	0.029	PAICS, PPAT

Table 5.4 – continued from previous page

Pathways	<i>P</i> -value	Genes
synthesis of purine ribonucleotides	0.029	PAICS, PPAT
platelet-activating factor → fatty aldehyde	0.029	PAFAH1B, PLA2G7
TNF —MEKK1→ c-Jun	0.029	MAP3K1, TRADD
CH000000332	0.036	CYBA, MAP3K1
LKB1 —AMPKalpha-2, AMPKbeta-2→ PPARalpha	0.036	PRKCZ, STK11
metabolism of androgens	0.043	HSD17B1, HSD17B6,SRD5A3
fMLP → NADPH oxidase	0.044	CYBA, PRKCZ
EP2 → VEGFA	0.048	CYBA, GNAS, GNB1

5.7.6. Comparative Identification of Master Regulators

The trypanotolerance attributes are further elucidated by the identification of 10 master regulators in Nuer, Gindeberet, and Benshangul breeds using the maximum radius of 10 steps upstream in the regulatory hierarchy using the TRANSPATH database. The master regulator Caspase family proteins are shared mostly by Sheko (Figure 5.7), Benshangul (Figure 5.16), and Gindeberet (Figure 5.17). The function of the master regulator protein Caspase is presented in section 5.5.

The top master regulator ITCH is only found in Nuer (Figure 5.18) which is involved in tolerance and immune responses including T-cell activation and T-helper cell differentiation [284]. In addition, only Sheko and Nuer share the master regulator PAK1 which is involved in the activation of MAP-kinase pathway and plays an important role in innate and adaptive immunity [227, 228, 232, 285]. Moreover, all the master regulators of Benshangul and Gindeberet are identical.

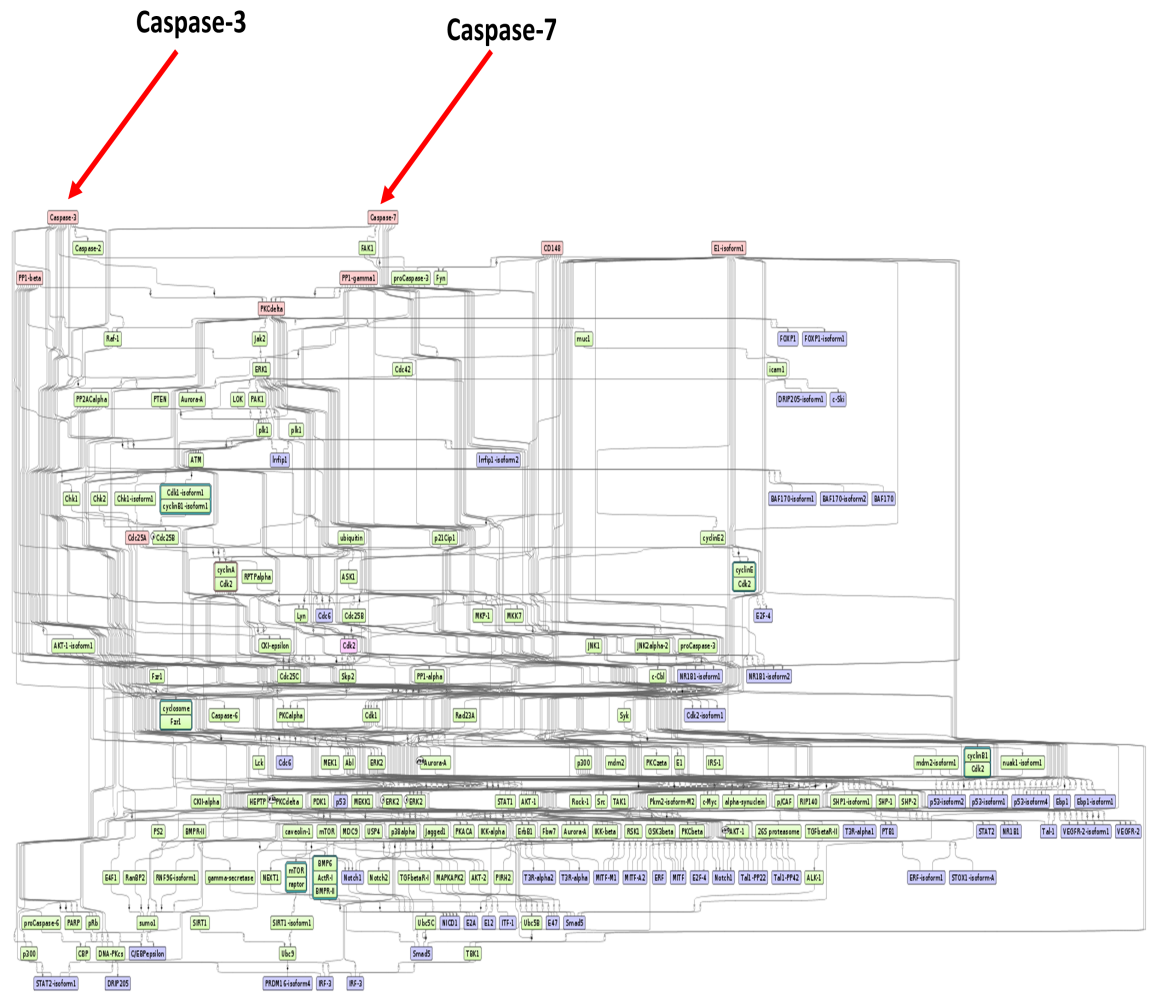


Figure 5.16.: The master regulatory networks identified in Beshangul. The colors red, blue, and green indicate master regulators, regulated proteins, and connecting molecules, respectively. The red/pink coloration of the master regulators represents the intensity of the degree of overrepresentation for the respective master regulator.

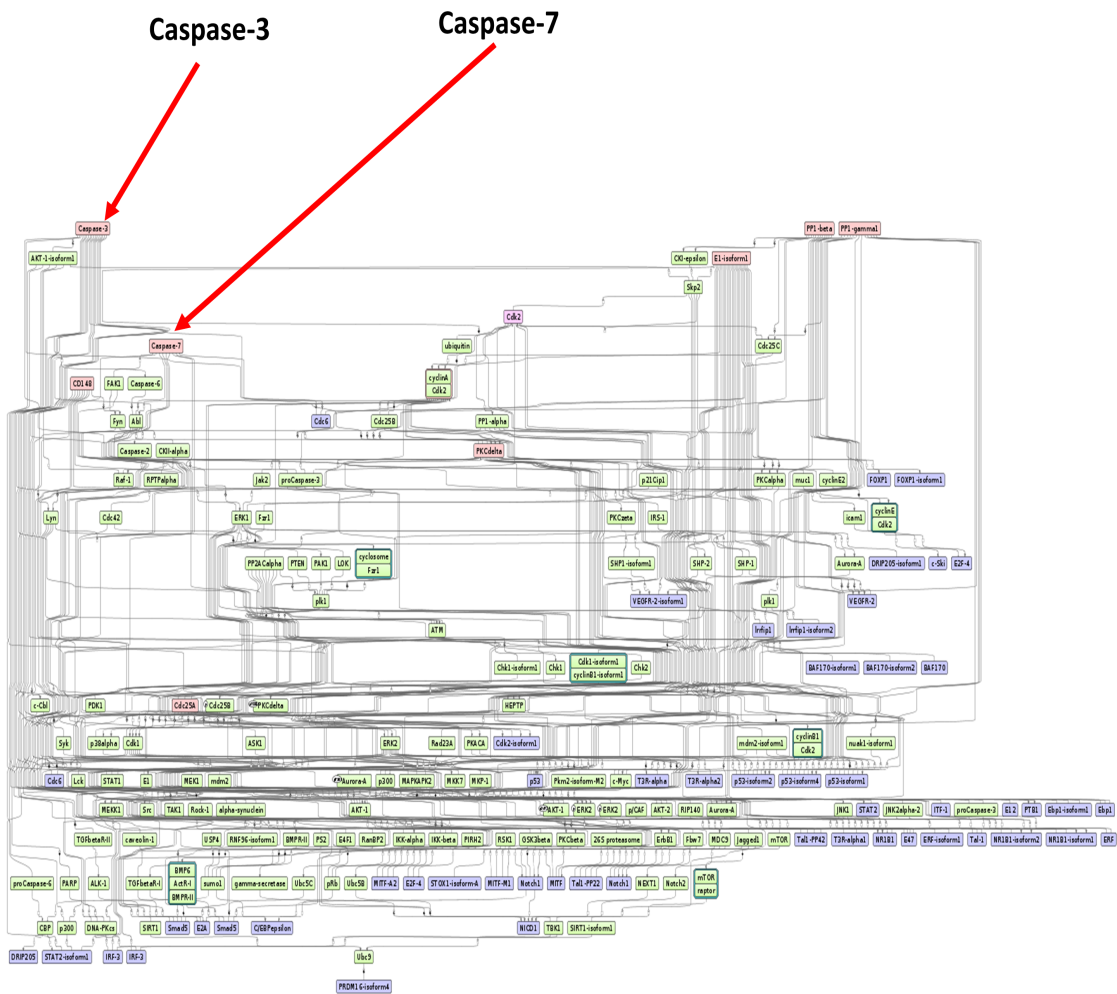


Figure 5.17.: The master regulatory networks identified in Gindeberet. The colors red, blue, and green indicate master regulators, regulated proteins, and connecting molecules, respectively. The red/pink coloration of the master regulators represents the intensity of the degree of overrepresentation for the respective master regulator.

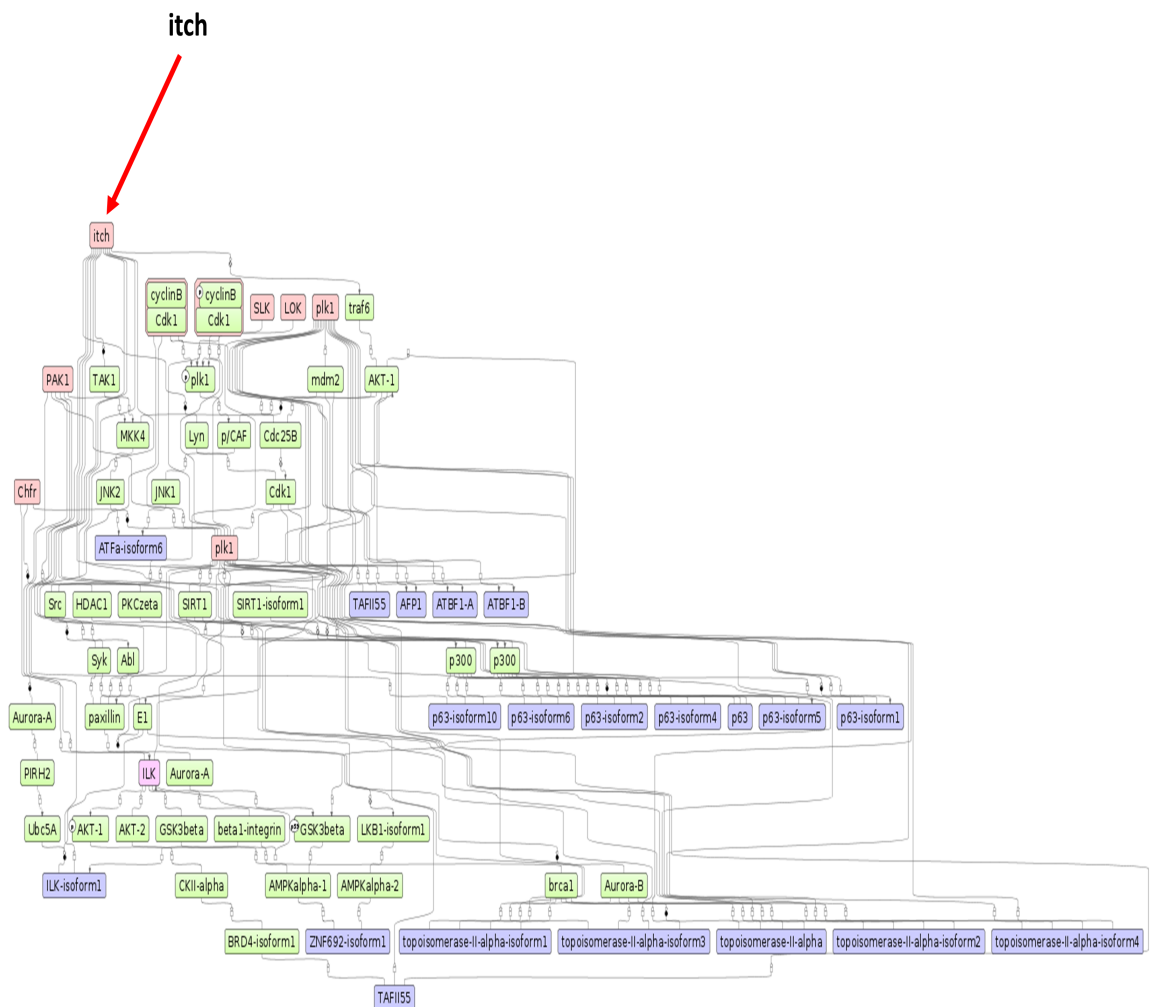


Figure 5.18.: The master regulatory networks identified in Nuer. The colors red, blue, and green indicate master regulators, regulated proteins, and connecting molecules, respectively. The red/pink coloration of the master regulators represents the intensity of the degree of overrepresentation for the respective master regulator.

6. Discussion

In this chapter, I will discuss the results reported in this thesis. First, I will discuss the results concerning population structure and genetic relationship between breeds. Second, I will elaborate on the results of the signature of selection analysis mainly focusing on the Sheko breed in relation to the trypanotolerance attribute. Third, I will discuss the results of pathways and regulatory mechanisms of trypanotolerance in the Sheko breed. Fourth, I will discuss the results of trypanotolerance tendencies of Nuer, Benshangul, and Gindeberet breeds. Finally, I will discuss the results of the comparative analysis of GWAS and the signature of selection detecting methods. Part of the following sections are previously published in Mekonnen *et al.* [100].

6.1. Population Structure and Genetic Relationship

The Sheko, Nganda, Ankole, and Nuer breeds are close to the indicine cluster and thereby support the admixture of more indicine than taurine type genomes in these breeds. The PCA result also shows the highest level of genetic heterogeneity in the Nganda breed which might be caused by ongoing cross breeding of Nganda with exotic breeds to enhance their productivity [286].

Consistent with the previous findings and the origins of the genetic backgrounds of the cattle breeds worldwide [26, 287], $K = 2$ in the admixture analysis highlights best the ancient divergence between indicine and taurine cattle. However, the three optimal genetic clusters suggested by the minimal CV error (Figure 5.3) reflect the common genetic background unique to east Africa besides taurine and indicine ancestral genetic admixture. In agreement with our study, Bahbahani *et al.* [26] reported east African genetic background unique to east African cattle breeds. Moreover, the admixture plots show two individuals of Sheko with a high level of taurine admixture. One of these individuals with higher taurine admixture is also detected by the PCA (Figure 5.1 B, upper left corner). This could be due to the recent crossbreeding of Sheko with European dairy breeds. There were similar observations in Butana, and it was speculated that farmers might have been involved in the crossbreeding with European dairy breeds in order to increase milk production [26]. I believe that the introgression of the European dairy breeds into the genome of indigenous breeds such as Sheko and Butana might distort their adaptive evolutionary responses against their natural environmental stresses. In this regard, future studies should assess the impact of European dairy breeds on the genome of the indigenous African breeds with respect to their natural adaptation and tolerance attributes.

It is believed that the taurine background of the Sheko is linked to its trypanotolerance characteristics [23, 288]. The taurine admixture of Sheko is likely a legacy of the first taurine occurrence on the African continent [103, 104]. A study on mtDNA indicates that all African cattle breeds ana-

lyzed so far carried taurine mtDNA haplotypes which suggest that these waves of indicine arrival into Africa were male-mediated [289, 290].

6.1.1. Identification of Candidate Signature Genes Associated with Trypanotolerance

The identified candidate signature genes in Table 5.2 might play a major role in the natural tolerance attributes of Sheko against trypanosomiasis. Notably, polymorphisms in or nearby the MIGA1, CDAN1, HSPA9 and PCSK6 genes in the genome of Sheko might be associated with the evolutionary response against the shortage of red blood cells due to trypanosome infection. In agreement with our findings, it has been reported by several studies that trypanotolerant N'Dama can better cope with control anemia, a process mediated by hematopoietic cell differentiation, than trypanosusceptible breeds [8, 291].

In previous studies, trypanotolerant animals were reported to switch from innate immune response to adaptive immune response with the induction of active macrophages (M2) following trypanosome infection [292, 293]. For instance, humoral response differences between trypanosusceptible (Boran) and trypanotolerant (N'Dama) cattle corresponding to the amount of antibody (Ab) titers have been observed. There is a difference in trypanosome-specific antiparasite Ab secreting cells in spleen and B cell activation between trypanotolerant and trypanosusceptible cattle [294, 295, 296]. In agreement with this, I identified the SPAG11B, RAET1G, PPP1R14C, and TTC3 genes which are involved in immune tolerance in Sheko. Moreover, trypanosome parasites are also known for their ability to manipulate the host immune responses. One of the mechanisms of innate immune evasion by these parasites is the reduction of HIF-1 α by indolepyruvate. Therefore, the reduction of hypoxic response elements in the spinal cord results in the progressive degradation of the motor neurons [147]. Therefore, the mutation in the ERN1 and CAPG genes, in particular, would be related to the host innate immune evasion of the parasite. Furthermore, the mutations or the differential expression of these genes are critical for the immune tolerance of Sheko to combat anemia and neurological dysfunction caused by trypanosome infection.

Trypanosomiasis is also reported to affect the nervous system of the animal. Fatihu *et al.* [297] and Allam *et al.* [298] reported causes of thyroid and parathyroid gland dysfunction following trypanosome infection in cattle. The dysfunctioning of the thyroid and parathyroid glands often result in neurological complications or cerebral pathology [168, 169]. Therefore, mutations in the POLR3B, MIGA1, TTC3, ERN1, CAPG, GNAS, and TTBK2 genes might be associated with the response to the trypanosome parasite in the brain white matter, cerebral fluid, thyroid, and parathyroid glands.

Moreover, during trypanosome infection, listlessness and emaciation are some of the clinical signs of the infection [14, 16, 17]. These clinical signs might be associated with the destruction of the thyroid gland by trypanosome parasites in cattle [297]. The candidate signature genes AP1M1 and CHAT are related to these clinical signs. Allam *et al.* [298] reported a similar profile during trypanosome infection in cattle that could be associated with neurological dysfunction such as

muscle spasm [11, 12, 157]. Therefore, the polymorphisms in AP1M1 and CHAT genes might be associated with the destruction of the thyroid gland by trypanosome infection.

Our findings show strong selective sweeps (Figure 5.5 A, B, and C) in the genomic regions around the selected signature genes of Table 5.2 (Supplementary Table 11 in Appendix A.1). This might indicate that the mutations in these genes have reached fixation or are near fixation. Therefore, the identified candidate signature genes in Table 5.2 might play a major role in the natural tolerance attributes of Sheko against trypanosomiasis. Moreover, the comparison of candidate signature genes identified by the *iHS*, CLR, and *Rsb* methods show more overlaps between *iHS* and CLR than between *Rsb* and *iHS* or *Rsb* and CLR analyses (Figure 5.4 A and B), in agreement with *Rsb* being a powerful method to detect selection signatures when the selected allele has reached fixation [26, 95, 299].

Furthermore, among the 15 identified candidate signature genes (Table 5.2), the MIGA1, RAETG, and PPP1R1AC genes are not significantly functionally enriched ($\alpha = 0.05$). This might indicate that these candidate signature genes in Sheko could be specific to the environmental pressure in the region such as trypanosomiasis. Interestingly, among the identified candidate signature genes in Table 5.2, the AP1M1 and GNAS genes are found in the trypanotolerant QTL regions of N'Dama [25]. The overlapping regions and genes of Sheko and N'Dama might indicate selection against the same environmental pressures at the same genes in these two breeds.

6.2. Identification of Overrepresented Pathways in the Candidate Signature Gene Sets

Pathway analysis has become a powerful tool in order to reveal the molecular mechanisms of disease tolerance. The rationale of pathway analysis lies in the detection of overrepresentation of biologically defined pathways based upon the functionally enriched candidate selected genes. The result of the pathway analysis indicates that most of the overrepresented pathways reported in the Sheko breed are mainly associated with host defense mechanisms against pathogens and anemia.

The previous study indicates that trypanosome infection could trigger auto-immunity by depleting thymocytes [300]. As a result, immature T-lymphocytes are released from the thymic central tolerance and differentiate into mature T-helper cells in the lymph nodes. This process would induce auto-immunity against self-antigens. Moreover, during trypanosome infection, the red blood cell membrane might be damaged by parasite enzymes such as proteases or phospholipases. This could expose epitopes which are not recognized as self-antigens and would trigger immune-mediated hemolysis due to antibody response against these self-antigens [301]. This could be controlled by suppressing the development of auto-reactive immune cells through ubiquitination which is a degradative tag to be recognized by a proteasome complex such as PSMD7 [302, 303]. Furthermore, some of the identified candidate signature genes are also associated with protein ubiquitination processes which might indicate that these genes are also involved in the functions described above (see Supplementary Table 2-4 in Appendix A.1). It has been reported that immunoproteasome subunits are key determinants of the CD8+ T-cell level and quality involved in

host resistance to trypanosomes infection [304]. To the best of our knowledge, our study is the first to show the potential of a molecular mechanism for controlling auto-reactive immune cells caused by trypanosomiasis in cattle (Appendix A.1). In agreement with our finding, Kiertstein *et al.* [305] reported that a trypanotolerant mouse strain showed overexpression of several genes encoding proteases.

In general, our findings of the search for signature genes appear to be well substantiated by the results of the overrepresented pathway analyses. Particularly, stress-associated, HIF, and NRF2 related pathways are involved in oxidative stress responses. Moreover, knockout mice for NRF2 show regenerative immune-mediated hemolytic anemia which indicates that this pathway is involved in erythrocyte maintenance during oxidative stress [205]. In addition, knock-out mice for PDGF-B, E2F8 (i.e., E2F gene family) and Rb develop severe anemia [306, 307]. Interestingly, trypanosome infection induces the production of superoxide, hydrogen peroxide, peroxy radicals, and hydroxyl radicals which are known to cause oxidative stress followed by tissue damage and hemolysis (rupturing of the red blood cells) [308]. Under oxidative stress (hypoxia), erythrocytes are important mobile oxidative sinks (antioxidant) for themselves, other cells and tissues. However, these properties of the red blood cells during oxidative stress contribute to its susceptibility towards hemolysis. The death of the red blood cells leads to anemia [309, 310] which is the most prominent and consistent clinical sign of trypanosome infection [17, 192, 193, 306]. In order to overcome the shortage of oxygen, stress-associated, HIF, and NRF2 related pathways play a critical role in the production of red blood cells in which hemoglobin acts as oxygen repository for red blood cells and other cells [206, 309, 310].

None of the most significant candidate signature genes (Table 5.2) was contained in the overrepresented pathway gene list (Table 5.3). This indicates that the candidate signature genes might be involved in the evolutionary gear particularly towards trypanotolerance in Sheko. For instance, candidate signature genes involved in the regulation of hypoxia (ERN1 and CAPG) are not identified in the overrepresented hypoxia related pathways. This might indicate that these candidate signature genes might be specific to oxidative stress tolerance attributes in Sheko. Hence, trypanotolerance of Sheko could be controlled by some selected genes with major effect and cohorts of genes with minor effects.

6.3. Identification of Master Regulators based on Candidate Signature Genes

The result of the master regulatory network analysis indicates that the identified proteins and master regulatory molecules are a large family of signaling enzymes expressed in various immune cells and regulate immune cell differentiation, cytokine production, and immune responses. Therefore, to maintain the tolerance against a pathogen, the regulation of these signaling pathways is critical [311, 312]. I believe that the candidate signature genes involved in anemia, neurological dysfunction, listlessness, and immune tolerance might be governed by the top master regulator Caspase in harmony with other regulatory molecules. In general, our study provides a first report on the top master regulators for trypanotolerance of Sheko and the overall analysis framework might

be helpful to understand the underlying mechanisms of different cattle diseases in future works (Appendix A.1).

6.4. Comparative Study for the Identification of Trypanotolerance Tendency in Nuer, Benshangul, and Gindeberet Breeds

In this thesis, the signature genes, hub genes, pathways, and master regulators of Gindeberet, Nuer, and Benshangul breeds are compared with the Sheko breed to detect the tendency of trypanotolerance in those breeds. Trypanosomiasis is considered as a major environmental pressure in the region where Gindeberet, Nuer, and Benshangul breeds are kept. Comparing the signature genes of these breeds with Sheko and among themselves shows a similar or distinct pattern of adaption towards trypanotolerance (see Section 5.7). The common signature genes shared by Nuer, Gindeberet, and Benshangul with those of Sheko are related to trypanotolerance attributes that are involved in neurological dysfunction, anemia, and immunity. Since these trypanotolerant attributes are the most common symptoms of the disease [10, 11, 12, 313], polymorphisms in STXBP3, GNAS, KCNIP4, DHX33, C1QBP, AK5, JMJD1C, EGR1, SLC24A3, CD86, and L3MBTL3 genes might be related with an evolutionary response against the effect of the parasite.

Specifically, the STXBP3, C1QBP, DHX33, GNAS, KCNIP4, and AK5 genes which are commonly identified in Benshangul and Sheko are associated with trypanotolerance attributes related to immune tolerance and neurological dysfunctions. Moreover, the JMJD1C, EGR1, SLC24A3 and ARG1 genes in Gindeberet which are shared by Sheko are involved in trypanotolerance attributes related to immune responses. The CD86 and L3MBTL3 genes shared by Nuer and Sheko are involved in trypanotolerance attributes related to immune responses and anemia, respectively. The knockout experiment on the L3MBTL3 gene results severe anemia in mice [314]. Hence, the identified signature genes involved in immune responses, neurological dysfunctions, and anemia in Benshangul, Nuer, and Gindeberet in common with Sheko might be engaged in a similar pattern of adaptation (similar selection responses) against the same environmental pressure related to trypanosomiasis. Consistent with our results, trypanotolerant breeds such as N'Dama show a shift from innate to adaptive responses during trypanosome infection [292, 293] (see Section 6.1.1).

Intriguingly, Sheko and Nuer shared the GO term known as T-cell chemotaxis. This indicates that there is a positive selection response to diseases in these breeds. In agreement with our result, mice infected with trypanosomes show the differentially expressed genes that are involved in immune responses such as T-chemotaxis, T-cell maturation, and presenting antigens by MHC class II [315]. Likewise, T-cell mediated migration of thymocyte towards chemokines was observed following trypanosome infection in human [316].

Our result is further substantiated by the identification of the hub genes involved in the gene networks. In response to an environmental pressure or stress, the expression of genes is regulated in a coordinated manner [317]. Likewise, the identified signature genes could also work in such a coordinated manner and there might be a few genes (hub genes) that are involved in the regulation of the expression of other genes in response to a given environmental pressure. Interestingly, the

gene PSMD7 which is involved in most of the overrepresented pathways (see Section 6.2), is also identified as a hub gene in Section 5.7.4. Hence, the PSMD7 gene might be obliged in harmonizing proteins (proteases) to degrade auto-reactive immune cells caused by trypanosomiasis [302, 303]. In addition, most of the hub genes are involved in the immune system, neurological dysfunctions, and hypoxia. Remarkably, the hub genes CTNNA2 (in Gindeberet) and KCNIP4 (in Gindeberet, Benshangul, and Nuer) are also reported in the known trypanotolerant N'Dama [258] which are associated with neurological dysfunctions [272, 273]. The CTNNA2 gene is also reported in N'Dama as a major hub gene that plays a role in orchestrating the immune cells in response to the presence of a pathogen [258].

Moreover, trypanosomes show the ability to manipulate the immune response of the host by reducing the HIF-1 α [147]. The reduction of HIF-1 α negatively affects the cellular responses such as high production of red blood cells [206, 205] to overcome the shortage of oxygen. In addition, progressive degradation of the motor neurons in the spinal cord is associated with the reduction of hypoxic response elements in mice [147]. Therefore, the hub genes RDX, AAK1, SYT1, and FOXP1 which are related to neurological dysfunctions and hypoxia induced cellular processes might be involved in the regulation of adaptive evolutionary responses against the host immune response escaping mechanisms of trypanosomes.

Furthermore, the identification of overrepresented pathways reveals the underlying molecular mechanisms of tolerance against trypanosomiasis. Notably, the common pathways identified in Gindeberet, Sheko, and Benshangul are involved in immune responses and oxidative stresses. Not only the common pathways between these breeds but also the first two overrepresented pathways of each breed are related to immunity responses and oxidative stresses. Consistent with our results, trypanosome infection is known to induce oxidative stress (see Section 6.2). Therefore, these results strengthen the hypothesis that Nuer, Gindeberet, and Benshangul are evolving or are in the processes of rapid evolutionary change in response to the common environmental stress (e.g. trypanosomiasis).

It has been reported in previous studies that trypanosomes utilize lipid nutrients of the host for growth and differentiation [318]. Trypanotolerant breeds such as N'Dama show a low level of plasma cholesterol and triglycerides compared to trypanosusceptible zebu breeds such as White Fulani [319]. This entails that trypanotolerant cattle has a superior ability of lipid clearance and hormonal control of lipid synthesis in order to limit the lipid nutrients which is required by trypanosomes [319]. Interestingly, the second most overrepresented pathway in Nuer, the acetyl-CoA \rightarrow palmitic acid pathway, which is involved in lipid metabolism, might play a significant role similar to N'Dama control of lipid synthesis. Remarkably, polymorphisms in the ST6GALNAC5 gene in Benshangul, Gindeberet, Nuer, Sheko, and Muturu, as well as the JAZF1 gene in Benshangul and Nuer are involved in lipid metabolism [260, 261]. Polymorphisms in these genes might be related to lipid clearance trypanotolerance attributes mentioned above. In agreement with this result, the overexpression of the JAZF1 gene is associated with suppression of lipid accumulation (lipid clearance) in mice [261].

For the further understanding of the regulatory mechanisms of trypanotolerance, key master reg-

ulators are identified that are potentially involved in the trypanotolerance attributes in Nuer, Benshangul, and Gindeberet breeds. Strikingly, the top master regulator Caspase family proteins are commonly identified in the Sheko, Gindeberet, and Benshangul breeds. Caspase might be involved in governing the evolutionary responses against trypanosomiasis (see Section 6.3). Consistent with our findings Paroli *et al.* [246] reported that Caspase-1 deficient mice show an increment on the level of parasitemia compared to the wild type during trypanosome infection. The knockout mice for Caspase-1 show IFN- γ , CD4+, and CD8+ cell reduction which plays a major role in the induction of adaptive immunity [246]. Therefore, Caspase protein families might be involved in orchestrating the expression of the signature genes which are involved in trypanotolerance. Complementary to Caspase, the master regulator ITCH might play an important role in adaptive immunity in the Nuer breed [284, 320]. Moreover, most of the master regulators of Sheko and Nuer are kinase families which might have similar biological functions and show ubiquitous expressions in various tissues. Interestingly, the Benshangul and Gindeberet master regulators are all identical and most of these regulators are phosphatase family proteins which might indicate that these two breeds have undergone a similar environmental pressure for generations.

In addition, the trypanotolerance tendencies of the Nuer, Benshangul, and Gindeberet breeds are supported by the overlapping QTL regions of trypanotolerant N'Dama. Gindeberet shares the largest trypanotolerant QTL regions with N'Dama which provides additional evidence that this breed has a certain level of trypanotolerance. However, the small number of overlapping trypanotolerant QTL regions identified between Sheko and N'Dama compared to the Nuer, Benshangul, and Gindeberet breeds indicate that the Sheko and N'Dama breeds could share regions with similar evolutionary changes while other regions are unique for each individual breed's trypanotolerance attributes. In agreement with this, Noyes *et al.* [17] reported unique genomic regions under positive selection in the N'Dama that are associated with trypanotolerance in comparison to the Sheko breed.

Furthermore, the ALCAM (in the Sheko) and MAPT [in the trypanotolerant group (Sheko, N'Dama, and Muturu)] genes are commonly captured by the GWAS analysis and signature of selection detecting methods (*iHS*, CLR, and *Rsb*). These genes are involved in immune responses and neurological dysfunctions which could be associated with trypanotolerant attributes as described above. To this end, the gene STXBP5 which is commonly identified by GWAS and signature of selection detecting methods in Sheko could be associated with the effect of the trypanosome parasite on blood coagulation [240, 321, 322]. Consistent with our result, there were reports associated with the depletion of coagulation factors by trypanosome parasite in human [321] and cattle [322]. In general, small overlapping regions are detected between the results of GWAS and the signature of selection detecting methods. In accordance with our results, Igoshin *et al.* [323] captured a few overlapping regions between the result of scans for selective sweeps and GWAS in search of genomic regions associated with cold-stress in Siberian cattle. These regions contain two genes (MSANTD4 and GRIA4) that are associated with cold-stress resistance. Therefore, combining GWAS and signature of selection detecting methods provides an additional level of confidence to the overlapping candidate regions which are responsible for the trypanotolerance attributes of these breeds.

7. Conclusion

In this chapter of my thesis, I will present a summary of my study by highlighting the main results. Moreover, I will provide the contributions of my work to the literature and future studies. Part of the following sections are previously published in Mekonnen *et al.* [100].

7.1. Summary

In sub-Saharan African countries, livestock production is severely affected by trypanosomiasis which is the most prevalent livestock disease. This disease is caused by a protozoan parasite known as trypanosome which is transmitted by the tsetse fly as a vector. The three sub-species of trypanosomes (*Trypanosoma congolense*, *Trypanosoma vivax*, and *Trypanosoma brucei brucei*) frequently cause infection in cattle. Even though there have been numerous measures to control the disease transmission, none of them were successful to eradicate the disease. Nowadays the parasite has become drug resistant due to the inappropriate usage of the drug to treat the disease. However, there are trypanotolerant breeds that could perform better in the regions where trypanosomiasis is highly prevalent. In Ethiopia, Sheko is known as trypanotolerant compared to other breeds such as Abigara and Horro. Nevertheless, the genetic controls that are involved in the trypanotolerance attributes of Sheko have not been studied well. Hence, this thesis closely investigates the trypanotolerance mechanisms of the Sheko breed. To explore the trypanotolerance attribute of the Sheko breed at the molecular level, I consider trypanosomiasis as an environmental selection pressure. Therefore, signs or traces of selection in the genome are detected by using a “bottom-up” or “from genotype to phenotype” approach in this thesis. Selection detecting methods (*iHS*, CLR, *Rsb*) are used to investigate the trypanotolerance mechanisms of Sheko.

In this thesis, I developed four basic analysis frameworks using genotyping data for the identification of the molecular mechanisms of trypanotolerance in Sheko and trypanotolerance tendencies in the Nuer, Benshangul, and Gindeberet breeds. The first analysis framework consists of PCA and admixture analyses to reveal the level of taurine background in the genome of Sheko. The PCA result shows that Sheko is positioned between African taurine type breeds and east African zebu type breeds. This suggests that Sheko is a mixture of African taurine and zebu ancestral populations. The admixture result further strengthens our confidence in Sheko by illustrating the level of taurine and zebu genomic admixture. Both the PCA and admixture analyses reveal a considerable level of taurine introgression in the genome of Sheko in comparison to the other east African breeds which can be considered as the potential of trypanotolerance in this breed. However, the majority of the Sheko genome indicates zebu origin (see Section 5.1).

The second analysis framework consists of the identification of genomic regions and genes that are associated with trypanotolerance in Sheko. The candidate signature genes and genomic regions that control trypanotolerance in Sheko are identified using *iHS*, CLR, and *Rsb* methods. The identified genomic regions and genes in this thesis (see Section 5.2) could be used as input for the development of genetic intervention strategies such as marker assisted selection which is considered as a promising strategy to eradicate trypanosomiasis in cattle.

Another promising strategy to eradicate the disease is the development of new drugs for trypanosomiasis treatment. Therefore, the third analysis framework includes the identification of overrepresented pathways, hub genes, and master regulators that are associated with trypanotolerance in Sheko (see Sections 5.4 and 5.5). Especially, the master regulator search result reveals that the top master regulator Caspase is a potential candidate drug target in which the proper regulation of this protease is critical to maintain trypanotolerance in Sheko. Usually, regulatory molecules and proteins are used as therapeutic drug targets. Therefore, Caspase could be used as a drug target for the development of new drugs to treat the disease.

The fourth analysis framework is the identification of trypanotolerance tendency for Nuer, Benshangul, and Gindeberet breeds. These breeds are kept in the infection zones for generations without proper prevention and medication. However, there is no documentation about their trypanotolerance attributes. In order to identify the trypanotolerance tendencies of Nuer, Gindeberet, and Benshangul breeds, the candidate signature genes and genomic regions identified in each breed are compared with Sheko and among themselves. The common genes and genomic regions identified in Sheko and those of the Nuer, Gindeberet, and Benshangul breeds indicate that these breeds have some degree of trypanotolerance (see Section 5.7.1). Moreover, the identified signature genes in the Sheko, Nuer, Benshangul, and Gindeberet breeds are compared with genes and genomic regions of the N'Dama and Muturu breeds which are associated with trypanotolerance in the literature (see Section 5.7.3). The identified common QTL regions with N'Dama, and genes as well as genomic regions of Muturu confirm that Nuer, Benshangul, and Gindeberet breeds have trypanotolerance tendencies. The trypanotolerance tendencies of the Nuer, Benshangul, and Gindeberet breeds are further elucidated by a comparison with Sheko using overrepresented pathways, hub genes, and master regulators that are identified for each breed (see Sections 5.7.4 - 5.7.6). The common pathways, hub genes, and master regulators shared by these breeds indicate that the Sheko, Nuer, Benshangul, and Gindeberet breeds have undergone a similar evolutionary response against the same environmental pressure related to trypanosomiasis. Especially, the identification of the top master regulator Caspase family proteins in the Sheko, Benshangul, and Gindeberet breeds indicate that the evolutionary response against trypanosomiasis might be primarily modulated by this master regulator in these cattle breeds.

In general, all four analysis frameworks deciphered the trypanotolerance attribute of Sheko and trypanotolerance tendencies of Nuer, Benshangul, and Gindeberet breeds. Moreover, I provide new insights to understand the underlying mechanisms of trypanotolerance in cattle populations included in this thesis. Furthermore, the findings reported in this thesis could be used for the development of genetic intervention strategies as well as new drugs for trypanosomiasis treat-

ment. Therefore, the overall analysis frameworks can be used to unravel different disease tolerance mechanisms in cattle.

7.2. Outlook

For generations, African Animal Trypanosomiasis has been the major selection pressure in the region. In this thesis, the candidate causative genes, pathways, and master regulators which are associated with the adaptation of the Sheko breed to its natural environmental pressure are identified. Most of the identified candidate signature genes, overrepresented pathways, and master regulator molecules were involved in immune tolerance, neurological dysfunction and anemia. This entails that the genome of Sheko was targeted by these environmental pressures which are associated with trypanosomiasis. Furthermore, the Nuer, Gindeberet, and Benshangul breeds might have undergone an evolutionary change against the same environmental pressure. Comparatively, I have identified traces of these changes as candidate signature genes, overrepresented pathways, and master regulators which are associated with adaptation of these breeds to their natural environmental pressure. These breeds have common as well as unique tolerance mechanisms against trypanosome infection.

In general, this study contributes to the existing literature in three ways: 1) The genetic controls of Sheko against trypanosomiasis have not been well studied and this study examines the genomic signatures in response to trypanosomiasis in detail; 2) this study presents pathways and master regulators which could help to understand the upstream biological processes involved in trypanotolerance. Particularly, this study for the first time identifies the master regulators involved in the regulatory mechanisms of trypanotolerance in relation to signatures of selection not only for the breeds included in this thesis but also in the context of cattle genomics, which can be used for the development of effective new drugs; 3) The trypanotolerance tendencies of Nuer, Benshangul, and Gindeberet breeds have not been documented. In this thesis, the trypanotolerance tendencies of these breeds are disclosed using a comparative genomics approach. Therefore, this study helps as an input for designing and implementing genetic intervention strategies to improve the performance of susceptible as well as animals which are relatively tolerant towards higher trypanotolerance.

The improvement of the cattle health contributes to increase the production of milk and meat. The improvement of the cattle health enhances the draft power of the animal which is associated with increasing crop production. This implies that, increasing animal and crop production significantly contributes to eradicate poverty in the area. Therefore, designing and implementing cross breeding with different level of tolerance and mechanism would produce a synthetic breed that would combine the tolerance attributes of the two parental populations. Hence, this study helps as an input for designing and implementing genetic intervention strategies to improve the performance of these breeds towards higher trypanotolerance. On the top of that, our findings reveal the importance of pathway and master regulators analyses which can be used as targets for the development of effective drugs for trypanosome infection. However, additional studies such as differential expressions

of targeted genes and regulatory molecules may be required to further confirm the validity of the results reported in this thesis.

Bibliography

- [1] Kristjanson PM, Swallow BM, Rowlands G, Kruska R, De Leeuw P: **Measuring the costs of African animal trypanosomosis, the potential benefits of control and returns to research.** *Agricultural systems* 1999, **59**:79–98.
- [2] Abebe G: **Current situation of Trypanosomiasis: In review article on: Trypanosomosis in Ethiopia.** *Ethiop. J Biol Sci.* 2005, **4**:75–121.
- [3] Shaw AP, Cecchi G, Wint GR, Mattioli RC, Robinson TP: **Mapping the economic benefits to livestock keepers from intervening against bovine trypanosomosis in Eastern Africa.** *Prev. Vet. Med.* 2014, **113**(2):197–210.
- [4] Franco JR, Simarro PP, Diarra A, Jannin JG: **Epidemiology of human African trypanosomiasis.** *Clinical epidemiology* 2014, **6**:257.
- [5] Hoare CA: *The trypanosomes of mammals. A zoological monograph.* Oxford and Edinburgh: UKs: Publications BS 1972.
- [6] Batista JS, Rodrigues CM, Garcia HA, Bezerra FS, Olinda RG, Teixeira MM, Soto-Blanco B: **Association of Trypanosoma vivax in extracellular sites with central nervous system lesions and changes in cerebrospinal fluid in experimentally infected goats.** *Vet. Res.* 2011, **42**:63.
- [7] Yaro M, Munyard KA, Stear MJ, Groth DM: **Combating African Animal Trypanosomiasis (AAT) in livestock: The potential role of trypanotolerance.** *Vet. Parasitol.* 2016, **225**:43–52.
- [8] Naessens J: **Bovine trypanotolerance: A natural ability to prevent severe anaemia and haemophagocytic syndrome?** *Int. J. Parasitol.* 2006, **36**(5):521–528.
- [9] Welburn SC, Molyneux DH, Maudlin I: **Beyond Tsetse—Implications for Research and Control of Human African Trypanosomiasis Epidemics.** *Trends Parasitol.* 2016, **32**(3):230–241.
- [10] Murray M, Trail JC, D'Ieteren GD: **Trypanotolerance in cattle and prospects for the control of trypanosomiasis by selective breeding.** *Rev. - Off. Int. Epizoot.* 1990, **9**(2):369–386.
- [11] Tuntasuvan D, Sarataphan N, Nishikawa H: **Cerebral trypanosomiasis in native cattle.** *Vet. Parasitol.* 1997, **73**(3-4):357–363.

- [12] Giordani F, Morrison LJ, Rowan TG, DE Koning HP, Barrett MP: **The animal trypanosomiasis and their chemotherapy: a review.** *Parasitology* 2016, **143**(14):1862–1889.
- [13] Barrett AD, Stanberry LR: *Vaccines for biodefense and emerging and neglected diseases.* Academic Press 2009.
- [14] Nantulya VM: **Immunological approaches to the control of animal trypanosomiasis.** *Parasitol. Today (Regul. Ed.)* 1986, **2**(6):168–173.
- [15] Batista JS, Riet-Correa F, Teixeira MM, Madruga CR, Simoes SD, Maia TF: **Trypanosomiasis by *Trypanosoma vivax* in cattle in the Brazilian semiarid: Description of an outbreak and lesions in the nervous system.** *Vet. Parasitol.* 2007, **143**(2):174–181.
- [16] Steverding D: **The history of African trypanosomiasis.** *Parasit Vectors* 2008, **1**:3.
- [17] Noyes H, Brass A, Obara I, Anderson S, Archibald AL, Bradley DG, Fisher P, Freeman A, Gibson J, Gicheru M, Hall L, Hanotte O, Hulme H, McKeever D, Murray C, Oh SJ, Tate C, Smith K, Tapio M, Wambugu J, Williams DJ, Agaba M, Kemp SJ: **Genetic and expression analysis of cattle identifies candidate genes in pathways responding to *Trypanosoma congolense* infection.** *Proc. Natl. Acad. Sci. U.S.A.* 2011, **108**(22):9304–9309.
- [18] Chanie M, Adula D, Bogale B: **Socio-economic assessment of the impacts of trypanosomiasis on cattle in Girja District, Southern Oromia Region, Southern Ethiopia.** *Acta Parasitologica Globalis* 2013, **4**(3):80–85.
- [19] Meyer A, Holt HR, Oumarou F, Chilongo K, Gilbert W, Fauron A, Mumba C, Guitian J: **Integrated cost-benefit analysis of tsetse control and herd productivity to inform control programs for animal African trypanosomiasis.** *Parasites & vectors* 2018, **11**:154.
- [20] NTTICC: **National Tsetse and Trypanosomosis Investigation and Control Center. Report for the period 7th June 2003 to 6th July 2004. Bedele, Ethiopia. 21-24 2004.**
- [21] Andrew AH: *Bovine Medicine Diseases and Husbandry of Cattle.* 9600 Garsington Road, Oxford OX42DQ, UKs: Blackwell Science Ltd 2004.
- [22] MoARD: **Ministry Of Agriculture and Rural Development of the Government of Ethiopia (MoARD). Tsetse and trypanosomiasis prevention and control strategies. Paper presented on Farming In Tsetse Controlled Areas (FITCA), Ethiopia final workshop. Adama, Ethiopia 2004.**
- [23] Lemecha H, Mulatu W, Hussein I, Rege E, Tekle T, Abdicho S, Ayalew W: **Response of four indigenous cattle breeds to natural tsetse and trypanosomosis challenge in the Ghibe valley of Ethiopia.** *Vet. Parasitol.* 2006, **141**(1-2):165–176.
- [24] DAGRIS: **Domestic Animal Genetic Resources Information System (DAGRIS). In: Rege, J., Hanotte, O., Mamo Y., Asrat B. and Dessie T. Addis Ababa: International Livestock Research Institute (ILRI) 2007.**

- [25] Hanotte O, Ronin Y, Agaba M, Nilsson P, Gelhaus A, Horstmann R, Sugimoto Y, Kemp S, Gibson J, Korol A, Soller M, Teale A: **Mapping of quantitative trait loci controlling trypanotolerance in a cross of tolerant West African N'Dama and susceptible East African Boran cattle.** *Proc. Natl. Acad. Sci. U.S.A.* 2003, **100**(13):7443–7448.
- [26] Bahbahani H, Salim B, Almathen F, Al Enezi F, Mwacharo JM, Hanotte O: **Signatures of positive selection in African Butana and Kenana dairy zebu cattle.** *PLoS ONE* 2018, **13**:e0190446.
- [27] Slingenbergh J: **Tsetse control and agricultural development in Ethiopia.** *World Anim Rev.* 1992, **70-71**:30–36.
- [28] Leak SG, Peregrine AS, Mulatu W, Rowlands GJ, D'Ieteren G: **Use of insecticide-impregnated targets for the control of tsetse flies (*Glossina* spp.) and trypanosomiasis occurring in cattle in an area of south-west Ethiopia with a high prevalence of drug-resistant trypanosomes.** *Trop. Med. Int. Health* 1996, **1**(5):599–609.
- [29] Mulugeta W, Wilkes J, Mulatu W, Majiwa PA, Masake R, Peregrine AS: **Long-term occurrence of *Trypanosoma congolense* resistant to diminazene, isometamidium and homidium in cattle at Ghibe, Ethiopia.** *Acta Trop.* 1997, **64**(3-4):205–217.
- [30] Codjia V, Mulatu W, Majiwa PA, Leak SG, Rowlands GJ, Authie E, d'Ieteren GD, Peregrine AS: **Epidemiology of bovine trypanosomiasis in the Ghibe valley, southwest Ethiopia. 3. Occurrence of populations of *Trypanosoma congolense* resistant to diminazene, isometamidium and homidium.** *Acta Trop.* 1993, **53**(2):151–163.
- [31] Naula C, Burchmore R: **A plethora of targets, a paucity of drugs: progress towards the development of novel chemotherapies for human African trypanosomiasis.** *Expert Rev Anti Infect Ther* 2003, **1**:157–165.
- [32] Lutje V, Taylor KA, Kennedy D, Authie E, Boulange A, Gettinby G: ***Trypanosoma congolense*: a comparison of T-cell-mediated responses in lymph nodes of trypanotolerant and trypanosusceptible cattle during primary infection.** *Exp. Parasitol.* 1996, **84**(3):320–329.
- [33] Kim J, Hanotte O, Mwai OA, Dessie T, Bashir S, Diallo B, Agaba M, Kim K, Kwak W, Sung S, et al.: **The genome landscape of indigenous African cattle.** *Genome biology* 2017, **18**:34.
- [34] Charlesworth B: **A hitch-hiking guide to the genome: a commentary on 'The hitch-hiking effect of a favourable gene' by John Maynard Smith and John Haigh.** *Genet. Res.* 2007, **89**(5-6):389–390.
- [35] Gautier M, Vitalis R: **rehh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure.** *Bioinformatics* 2012, **28**(8):1176–1177.

- [36] Bomba L, Nicolazzi EL, Milanese M, Negrini R, Mancini G, Biscarini F, Stella A, Valentini A, Ajmone-Marsan P: **Relative extended haplotype homozygosity signals across breeds reveal dairy and beef specific signatures of selection.** *Genetics Selection Evolution* 2015, **47**:25.
- [37] McGuire SE, McGuire AL: **Don't throw the baby out with the bathwater: enabling a bottom-up approach in genome-wide association studies.** *Genome research* 2008, **18**(11):1683–1685.
- [38] Ma Y, Ding X, Qanbari S, Weigend S, Zhang Q, Simianer H: **Properties of different selection signature statistics and a new strategy for combining them.** *Heredity (Edinb)* 2015, **115**(5):426–436.
- [39] Vatsiou AI, Bazin E, Gaggiotti OE: **Detection of selective sweeps in structured populations: a comparison of recent methods.** *Molecular ecology* 2016, **25**:89–103.
- [40] Tasew S, Duguma R: **Cattle anaemia and trypanosomiasis in western Oromia State, Ethiopia.** *evue Méd. Vét.* 2012, **163**(12):581–588.
- [41] Leta S, Alemayehu G, Seyoum Z, Bezie M: **Prevalence of bovine trypanosomosis in Ethiopia: a meta-analysis.** *Parasit Vectors* 2016, **9**:139.
- [42] Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, Franz M, Grouios C, Kazi F, Lopes CT, et al.: **The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function.** *Nucleic acids research* 2010, **38**(suppl_2):W214–W220.
- [43] Wlochowitz D, Haubrock M, Arackal J, Bleckmann A, Wolff A, Beissbarth T, Wingender E, Gultas M: **Computational Identification of Key Regulators in Two Different Colorectal Cancer Cell Lines.** *Front Genet* 2016, **7**:42.
- [44] Groeneveld L, Lenstra J, Eding H, Toro M, Scherf B, Pilling D, Negrini R, Finlay E, Jianlin H, Groeneveld E, et al.: **Genetic diversity in farm animals—a review.** *Animal genetics* 2010, **41**:6–31.
- [45] Portin P, Wilkins A: **The evolving definition of the term “gene”.** *Genetics* 2017, **205**(4):1353–1364.
- [46] Rittner D, McCabe TL: *Encyclopedia of biology.* Infobase Publishing 2004.
- [47] Cox CB, Moore PD, Ladle R: *Biogeography: an ecological and evolutionary approach.* John Wiley & Sons 2016.
- [48] Consortium IH, et al.: **The international HapMap project.** *Nature* 2003, **426**(6968):789.
- [49] Flynn P: **Biotic vs. abiotic—distinguishing disease problems from environmental stresses.** *ISU Entomology* 2003.

- [50] Rejeb I, Pastor V, Mauch-Mani B: **Plant responses to simultaneous biotic and abiotic stress: molecular mechanisms.** *Plants* 2014, **3**(4):458–475.
- [51] Suzuki N, Rivero RM, Shulaev V, Blumwald E, Mittler R: **Abiotic and biotic stress combinations.** *New Phytologist* 2014, **203**:32–43.
- [52] Sun S, Zhou J: **Molecular mechanisms underlying stress response and adaptation.** *Thoracic cancer* 2018, **9**(2):218–227.
- [53] Scandalios J: **Oxidative stress: molecular perception and transduction of signals triggering antioxidant gene defenses.** *Brazilian journal of medical and biological research* 2005, **38**(7):995–1014.
- [54] Singh B, Soltys BJ, Wu ZC, Patel HV, Freeman KB, Gupta RS: **Cloning and some novel characteristics of mitochondrial Hsp70 from Chinese hamster cells.** *Exp. Cell Res.* 1997, **234**(2):205–216.
- [55] Cramer T, Yamanishi Y, Clausen BE, Forster I, Pawlinski R, Mackman N, Haase VH, Jaenisch R, Corr M, Nizet V, Firestein GS, Gerber HP, Ferrara N, Johnson RS: **HIF-1 α is essential for myeloid cell-mediated inflammation.** *Cell* 2003, **112**(5):645–657.
- [56] Jantsch J, Wiese M, Schodel J, Castiglione K, Glasner J, Kolbe S, Mole D, Schleicher U, Eckardt KU, Hensel M, Lang R, Bogdan C, Schnare M, Willam C: **Toll-like receptor activation and hypoxia use distinct signaling pathways to stabilize hypoxia-inducible factor 1 α (HIF1A) and result in differential HIF1A-dependent gene expression.** *J. Leukoc. Biol.* 2011, **90**(3):551–562.
- [57] McNamee EN, Korn Johnson D, Homann D, Clambey ET: **Hypoxia and hypoxia-inducible factors as regulators of T cell development, differentiation, and function.** *Immunol. Res.* 2013, **55**(1-3):58–70.
- [58] Bleuven C, Landry CR: **Molecular and cellular bases of adaptation to a changing environment in microorganisms.** *Proceedings of the Royal Society B: Biological Sciences* 2016, **283**(1841):20161458.
- [59] Byers KJ, Xu S, Schlüter PM: **Molecular mechanisms of adaptation and speciation: why do we need an integrative approach?** *Molecular ecology* 2017, **26**:277–290.
- [60] Rédei GP: *Encyclopedia of genetics, genomics, proteomics, and informatics.* Springer Science & Business Media 2008.
- [61] Smith JM, Haigh J: **The hitch-hiking effect of a favourable gene.** *Genet. Res.* 2007, **89**(5-6):391–403.
- [62] O'Brien AMP, Utsunomiya YT, Mészáros G, Bickhart DM, Liu GE, Van Tassell CP, Sonstegard TS, Da Silva MV, Garcia JF, Sölkner J: **Assessing signatures of selection through variation in linkage disequilibrium between taurine and indicine cattle.** *Genetics Selection Evolution* 2014, **46**:19.

- [63] Nielsen R: **Molecular signatures of natural selection.** *Annu. Rev. Genet.* 2005, **39**:197–218.
- [64] Qanbari S, Pausch H, Jansen S, Somel M, Strom TM, Fries R, Nielsen R, Simianer H: **Classic selective sweeps revealed by massive sequencing in cattle.** *PLoS Genet.* 2014, **10**(2):e1004148.
- [65] Itadani H, Mizuarai S, Kotani H: **Can systems biology understand pathway activation? Gene expression signatures as surrogate markers for understanding the complexity of pathway activation.** *Current genomics* 2008, **9**(5):349–360.
- [66] Cantini L, Calzone L, Martignetti L, Rydenfelt M, Blüthgen N, Barillot E, Zinovyev A: **Classification of gene signatures for their information value and functional redundancy.** *NPJ systems biology and applications* 2017, **4**:2.
- [67] Manolio TA: **Genomewide association studies and assessment of the risk of disease.** *New England journal of medicine* 2010, **363**(2):166–176.
- [68] Platt A, Vilhjálmsson BJ, Nordborg M: **Conditions under which genome-wide association studies will be positively misleading.** *Genetics* 2010, **186**(3):1045–1052.
- [69] Shmulewitz D, Zhang J, Greenberg DA: **Case-control association studies in mixed populations: correcting using genomic control.** *Human heredity* 2004, **58**(3-4):145–153.
- [70] Devlin B, Roeder K: **Genomic control for association studies.** *Biometrics* 1999, **55**(4):997–1004.
- [71] Lee S, Wright FA, Zou F: **Control of population stratification by correlation-selected principal components.** *Biometrics* 2011, **67**(3):967–974.
- [72] Golan D, Rosset S: **Mixed models for case-control genome-wide association studies: major challenges and partial solutions.** *Borgan Ø, Breslow N, Chatterjee N, et al. (1st edn). Handbook of Statistical Methods for Case-Control Studies. Boca Raton, FL: Chapman and Hall/CRC* 2018, :495–514.
- [73] Hellwege JN, Keaton JM, Giri A, Gao X, Velez Edwards DR, Edwards TL: **Population stratification in genetic association studies.** *Current protocols in human genetics* 2017, **95**:1–22.
- [74] Price AL, Zaitlen NA, Reich D, Patterson N: **New approaches to population stratification in genome-wide association studies.** *Nature Reviews Genetics* 2010, **11**(7):459.
- [75] Yang J, Zaitlen NA, Goddard ME, Visscher PM, Price AL: **Advantages and pitfalls in the application of mixed-model association methods.** *Nature genetics* 2014, **46**(2):100.
- [76] Gallagher MD, Chen-Plotkin AS: **The post-GWAS era: from association to function.** *The American Journal of Human Genetics* 2018, **102**(5):717–730.

- [77] Krull M, Pistor S, Voss N, Kel A, Reuter I, Kronenberg D, Michael H, Schwarzer K, Potapov A, Choi C, et al.: **TRANSPATH®: an information resource for storing and visualizing signaling pathways and their pathological aberrations.** *Nucleic acids research* 2006, **34**(suppl_1):D546–D551.
- [78] Krull M, Voss N, Choi C, Pistor S, Potapov A, Wingender E: **TRANSPATH®: an integrated database on signal transduction and a tool for array analysis.** *Nucleic acids research* 2003, **31**:97–100.
- [79] Koschmann J, Bhar A, Stegmaier P, Kel A, Wingender E: **“Upstream analysis”: an integrated promoter-pathway analysis approach to causal interpretation of microarray data.** *Microarrays* 2015, **4**(2):270–286.
- [80] Stark C, Breitkreutz BJ, Chatr-Aryamontri A, Boucher L, Oughtred R, Livstone MS, Nixon J, Van Auken K, Wang X, Shi X, et al.: **The BioGRID interaction database: 2011 update.** *Nucleic acids research* 2010, **39**(suppl_1):D698–D704.
- [81] Razick S, Magklaras G, Donaldson IM: **iRefIndex: a consolidated protein interaction database with provenance.** *BMC bioinformatics* 2008, **9**:405.
- [82] Barrett T, Suzek TO, Troup DB, Wilhite SE, Ngau WC, Ledoux P, Rudnev D, Lash AE, Fujibuchi W, Edgar R: **NCBI GEO: mining millions of expression profiles—database and tools.** *Nucleic acids research* 2005, **33**(suppl_1):D562–D566.
- [83] Brown KR, Jurisica I: **Online predicted human interaction database.** *Bioinformatics* 2005, **21**(9):2076–2082.
- [84] Cerami EG, Gross BE, Demir E, Rodchenkov I, Babur Ö, Anwar N, Schultz N, Bader GD, Sander C: **Pathway Commons, a web resource for biological pathway data.** *Nucleic acids research* 2010, **39**(suppl_1):D685–D690.
- [85] Montojo J, Zuberi K, Rodriguez H, Kazi F, Wright G, Donaldson SL, Morris Q, Bader GD: **GeneMANIA Cytoscape plugin: fast gene function predictions on the desktop.** *Bioinformatics* 2010, **26**(22):2927–2928.
- [86] Sun YV, Kardia SL: **Imputing missing genotypic data of single-nucleotide polymorphisms using neural networks.** *European Journal of Human Genetics* 2008, **16**(4):487.
- [87] Browning SR, Browning BL: **Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering.** *Am. J. Hum. Genet.* 2007, **81**(5):1084–1097.
- [88] Alexander D, Novembre J, Lange K: **Fast model-based estimation of ancestry in unrelated individuals.** *Genome Research* 2009, **19**:1655–1664.
- [89] Pritchard JK, Wen W, Falush D: **Documentation for STRUCTURE software: Version 2** 2003.

- [90] Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D: **Principal components analysis corrects for stratification in genome-wide association studies.** *Nature genetics* 2006, **38**(8):904.
- [91] Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC: **PLINK: a toolset for whole-genome association and population-based linkage analysis.** *American Journal of Human Genetics* 2007, **81**:559–575.
- [92] Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, Schaffner SF, Gabriel SB, Platko JV, Patterson NJ, McDonald GJ, Ackerman HC, Campbell SJ, Altshuler D, Cooper R, Kwiatkowski D, Ward R, Lander ES: **Detecting recent positive selection in the human genome from haplotype structure.** *Nature* 2002, **419**(6909):832–837.
- [93] Voight BF, Kudravalli S, Wen X, Pritchard JK: **A map of recent positive selection in the human genome.** *PLoS biology* 2006, **4**(3):e72.
- [94] Skipper M: **Human genetics: Tracking positive selection.** *Nature Reviews Genetics* 2002, **3**(11):824.
- [95] Tang K, Thornton KR, Stoneking M: **A new approach for using genome scans to detect recent positive selection in the human genome.** *PLoS biology* 2007, **5**(7):e171.
- [96] Charlesworth B: **The effects of deleterious mutations on evolution at linked sites.** *Genetics* 2012, **190**:5–22.
- [97] Kim Y, Stephan W: **Detecting a local signature of genetic hitchhiking along a recombining chromosome.** *Genetics* 2002, **160**(2):765–777.
- [98] Nielsen R, Williamson S, Kim Y, Hubisz MJ, Clark AG, Bustamante C: **Genomic scans for selective sweeps using SNP data.** *Genome Res.* 2005, **15**(11):1566–1575.
- [99] Pavlidis P, Hutter S, Stephan W: **A population genomic approach to map recent positive selection in model species.** *Molecular Ecology* 2008, **17**(16):3585–3598.
- [100] Mekonnen YA, Gültas M, Effa K, Hanotte O, Schmitt AO: **Identification of Candidate Signature Genes and Key Regulators Associated With Trypanotolerance in the Sheko Breed.** *Frontiers in Genetics* 2019, **10**.
- [101] Qiagen A: **DNeasy© blood and tissue handbook.** *Qiagen AG, Hombrechtikon, Switzerland* 2006.
- [102] Bahbahani H, Tijjani A, Mukasa C, Wragg D, Almathen F, Nash O, Akpa GN, Mbole-Kariuki M, Malla S, Woolhouse M, Sonstegard T, Van Tassell C, Blythe M, Huson H, Hanotte O: **Signatures of Selection for Environmental Adaptation and Zebu x Taurine Hybrid Fitness in East African Shorthorn Zebu.** *Front Genet* 2017, **8**:68.

- [103] Hanotte O, Tawah CL, Bradley DG, Okomo M, Verjee Y, Ochieng J, Rege JE: **Geographic distribution and frequency of a taurine *Bos taurus* and an indicine *Bos indicus* Y specific allele amongst sub-saharan African cattle breeds.** *Mol. Ecol.* 2000, **9**(4):387–396.
- [104] Salim B, Taha KM, Hanotte O, Mwacharo JM: **Historical demographic profiles and genetic variation of the East African Butana and Kenana indigenous dairy zebu cattle.** *Anim. Genet.* 2014, **45**(6):782–790.
- [105] Rege J: **The state of African cattle genetic resources I. Classification framework and identification of threatened and extinct breeds.** *Animal Genetic Resources/Resources génétiques animales/Recursos genéticos animales* 1999, **25**:1–25.
- [106] Tijjani A, Utsunomiya YT, Ezekwe A, Nash O, Hanotte OH: **Genome Sequence Analysis Reveals Selection Signatures in Endangered Trypano-tolerant West African Muturu Cattle.** *Frontiers in Genetics* 2019, **10**:442.
- [107] Magona JW, Walubengo J, Odimim JJ: **Differences in susceptibility to trypanosome infection between Nkedi Zebu and Ankole cattle, under field conditions in Uganda.** *Ann Trop Med Parasitol* 2004, **98**(8):785–792.
- [108] Muhanguzi D, Mugenyi A, Bigirwa G, Kamusiime M, Kitibwa A, Akurut GG, Ochwo S, Amanyire W, Okech SG, Hattendorf J, Tweyongyere R: **African animal trypanosomiasis as a constraint to livestock health and production in Karamoja region: a detailed qualitative and quantitative assessment.** *BMC Vet. Res.* 2017, **13**:355.
- [109] Ocaido M, Otim CP, Okuna NM, Erume J, Ssekitto C, Wafula RZO, Kakaire D, Walubengo J, Monrad J: **Socio-economic and livestock disease survey of agro-pastoral communities in Serere County, Soroti District, Uganda.** *Livestock Research for Rural Development* 2005, **17**.
- [110] FAO: **THE UGANDA COUNTRY REPORT AS PART OF THE STATE OF THE WORLD'S ANIMAL GENETIC RESOURCES (SoW-AnGR) REPORT** 2004.
- [111] Muhanguzi D, Picozzi K, Hatendorf J, Thrusfield M, Welburn SC, Kabasa JD, Waiswa C: **Improvements on restricted insecticide application protocol for control of Human and Animal African Trypanosomiasis in eastern Uganda.** *PLoS Negl Trop Dis* 2014, **8**(10):e3284.
- [112] Van Wyk IC, Goddard A, de C Bronsvoort BM, Coetzer JA, Handel IG, Hanotte O, Jennings A, Lesosky M, Kiara H, Thumbi SM, Toye P, Woolhouse MW, Penzhorn BL: **The impact of co-infections on the haematological profile of East African Short-horn Zebu calves.** *Parasitology* 2014, **141**(3):374–388.
- [113] Sinshaw A, Abebe G, Desquesnes M, Yoni W: **Biting flies and *Trypanosoma vivax* infection in three highland districts bordering lake Tana, Ethiopia.** *Vet. Parasitol.* 2006, **142**(1-2):35–46.

- [114] Hayes BJ, Pryce J, Chamberlain AJ, Bowman PJ, Goddard ME: **Genetic architecture of complex traits and accuracy of genomic prediction: coat colour, milk-fat percentage, and type in Holstein cattle as contrasting model traits.** *PLoS Genet.* 2010, **6**(9):e1001139.
- [115] Wilches R, Voigt S, Duchon P, Laurent S, Stephan W: **Fine-mapping and selective sweep analysis of QTL for cold tolerance in *Drosophila melanogaster*.** *G3 (Bethesda)* 2014, **4**(9):1635–1645.
- [116] DeGiorgio M, Huber CD, Hubisz MJ, Hellmann I, Nielsen R: **SweepFinder2: increased sensitivity, robustness and flexibility.** *Bioinformatics* 2016, **32**(12):1895–1897.
- [117] Huber CD, DeGiorgio M, Hellmann I, Nielsen R: **Detecting recent selective sweeps while controlling for mutation rate and background selection.** *Mol. Ecol.* 2016, **25**:142–156.
- [118] Cheruiyot EK, Bett RC, Amimo JO, Yi Z, Mrode R, Mujibi FD: **Signatures of selection in admixed dairy cattle in Tanzania.** *Frontiers in genetics* 2018, **9**:607.
- [119] Kinsella RJ, Kahari A, Haider S, Zamora J, Proctor G, Spudich G, Almeida-King J, Staines D, Derwent P, Kerhornou A, Kersey P, Flicek P: **Ensembl BioMart: a hub for data retrieval across taxonomic space.** *Database (Oxford)* 2011, **2011**:bar030.
- [120] Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, McDermott MG, Monteiro CD, Gundersen GW, Ma'ayan A: **Enrichr: a comprehensive gene set enrichment analysis web server 2016 update.** *Nucleic Acids Res.* 2016, **44**(W1):W90–97.
- [121] Jungmeister WA, Turo D: **Adapting treemaps to stock portfolio visualization.** Tech. rep. 1992.
- [122] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T: **Cytoscape: a software environment for integrated models of biomolecular interaction networks.** *Genome research* 2003, **13**(11):2498–2504.
- [123] Moura IC, Centelles MN, Arcos-Fajardo M, Malheiros DM, Collawn JF, Cooper MD, Monteiro RC: **Identification of the transferrin receptor as a novel immunoglobulin (Ig)A1 receptor and its enhanced expression on mesangial cells in IgA nephropathy.** *J. Exp. Med.* 2001, **194**(4):417–425.
- [124] Rouault TA: **The role of iron regulatory proteins in mammalian iron homeostasis and disease.** *Nat. Chem. Biol.* 2006, **2**(8):406–414.
- [125] Jabara HH, Boyden SE, Chou J, Ramesh N, Massaad MJ, Benson H, Bainter W, Fraulino D, Rahimov F, Sieff C, Liu ZJ, Alshemmari SH, Al-Ramadi BK, Al-Dhekri H, Arnaut R, Abu-Shukair M, Vatsayan A, Silver E, Ahuja S, Davies EG, Sola-Visner M, Ohsumi TK, Andrews NC, Notarangelo LD, Fleming MD, Al-Herz W, Kunkel LM, Geha RS: A

- missense mutation in TFRC, encoding transferrin receptor 1, causes combined immunodeficiency.** *Nat. Genet.* 2016, **48**:74–78.
- [126] Dgany O, Avidan N, Delaunay J, Krasnov T, Shalmon L, Shalev H, Eidelitz-Markus T, Kapelushnik J, Cattan D, Pariente A, Tulliez M, Cretien A, Schischmanoff PO, Iolascon A, Fibach E, Koren A, Rossler J, Le Merrer M, Yaniv I, Zaizov R, Ben-Asher E, Olender T, Lancet D, Beckmann JS, Tamary H: **Congenital dyserythropoietic anemia type I is caused by mutations in codanin-1.** *Am. J. Hum. Genet.* 2002, **71**(6):1467–1474.
- [127] Renella R, Roberts NA, Brown JM, De Gobbi M, Bird LE, Hassanali T, Sharpe JA, Sloane-Stanley J, Ferguson DJ, Cordell J, Buckle VJ, Higgs DR, Wood WG: **Codanin-1 mutations in congenital dyserythropoietic anemia type 1 affect HP1alpha localization in erythroblasts.** *Blood* 2011, **117**(25):6928–6938.
- [128] Ran Q, Wadhwa R, Kawai R, Kaul SC, Sifers RN, Bick RJ, Smith JR, Pereira-Smith OM: **Extramitochondrial localization of mortalin/mthsp70/PBP74/GRP75.** *Biochem. Biophys. Res. Commun.* 2000, **275**:174–179.
- [129] Ohtsuka R, Abe Y, Fujii T, Yamamoto M, Nishimura J, Takayanagi R, Muta K: **Mortalin is a novel mediator of erythropoietin signaling.** *Eur. J. Haematol.* 2007, **79**(2):114–125.
- [130] Chen TH, Kambal A, Krysiak K, Walshauer MA, Raju G, Tibbitts JF, Walter MJ: **Knock-down of Hspa9, a del(5q31.2) gene, results in a decrease in hematopoietic progenitors in mice.** *Blood* 2011, **117**(5):1530–1539.
- [131] Schmitz-Abe K, Ciesielski SJ, Schmidt PJ, Campagna DR, Rahimov F, Schilke BA, Cuijpers M, Rieneck K, Lausen B, Linenberger ML, Sendamarai AK, Guo C, Hofmann I, Newburger PE, Matthews D, Shimamura A, Snijders PJ, Towne MC, Niemeyer CM, Watson HG, Dziegiel MH, Heeney MM, May A, Bottomley SS, Swinkels DW, Markianos K, Craig EA, Fleming MD: **Congenital sideroblastic anemia due to mutations in the mitochondrial HSP70 homologue HSPA9.** *Blood* 2015, **126**(25):2734–2738.
- [132] Guillemot J, Seidah NG: **PACE4 (PCSK6): another proprotein convertase link to iron homeostasis?** *Haematologica* 2015, **100**(9):e377.
- [133] Yang D, Chertov O, Bykovskaia SN, Chen Q, Buffo MJ, Shogan J, Anderson M, Schroder JM, Wang JM, Howard OM, Oppenheim JJ: **Beta-defensins: linking innate and adaptive immunity through dendritic and T cell CCR6.** *Science* 1999, **286**(5439):525–528.
- [134] Ganz T: **Defensins: antimicrobial peptides of innate immunity.** *Nat. Rev. Immunol.* 2003, **3**(9):710–720.
- [135] Eagle RA, Trowsdale J: **Promiscuity and the single receptor: NKG2D.** *Nat. Rev. Immunol.* 2007, **7**(9):737–744.
- [136] Tomasec P, Wang EC, Groh V, Spies T, McSharry BP, Aicheler RJ, Stanton RJ, Wilkinson GW: **Adenovirus vector delivery stimulates natural killer cell recognition.** *J. Gen. Virol.* 2007, **88**(Pt 4):1103–1108.

- [137] Lanier LL: **NKG2D Receptor and Its Ligands in Host Defense.** *Cancer Immunol Res* 2015, **3**(6):575–582.
- [138] Hanke JH, Gardner JP, Dow RL, Changelian PS, Brissette WH, Weringer EJ, Pollok BA, Connelly PA: **Discovery of a novel, potent, and Src family-selective tyrosine kinase inhibitor. Study of Lck- and FynT-dependent T cell activation.** *J. Biol. Chem.* 1996, **271**(2):695–701.
- [139] Linnekin D, DeBerry CS, Mou S: **Lyn associates with the juxtamembrane region of c-Kit and is activated by stem cell factor in hematopoietic cell lines and normal progenitor cells.** *J. Biol. Chem.* 1997, **272**(43):27450–27455.
- [140] Liu QR, Zhang PW, Zhen Q, Walther D, Wang XB, Uhl GR: **KEPI, a PKC-dependent protein phosphatase 1 inhibitor regulated by morphine.** *J. Biol. Chem.* 2002, **277**(15):13312–13320.
- [141] Haynes MP, Li L, Sinha D, Russell KS, Hisamoto K, Baron R, Collinge M, Sessa WC, Bender JR: **Src kinase mediates phosphatidylinositol 3-kinase/Akt-dependent rapid endothelial nitric-oxide synthase activation by estrogen.** *J. Biol. Chem.* 2003, **278**(4):2118–2123.
- [142] Vignali DA, Collison LW, Workman CJ: **How regulatory T cells work.** *Nat. Rev. Immunol.* 2008, **8**(7):523–532.
- [143] Chen HK, Fernandez-Funez P, Acevedo SF, Lam YC, Kaytor MD, Fernandez MH, Aitken A, Skoulakis EM, Orr HT, Botas J, Zoghbi HY: **Interaction of Akt-phosphorylated ataxin-1 with 14-3-3 mediates neurodegeneration in spinocerebellar ataxia type 1.** *Cell* 2003, **113**(4):457–468.
- [144] Liu G, Burns S, Huang G, Boyd K, Proia RL, Flavell RA, Chi H: **The receptor S1P1 overrides regulatory T cell-mediated immune suppression through Akt-mTOR.** *Nat. Immunol.* 2009, **10**(7):769–777.
- [145] Pulst SM: **Genetics of neurodegenerative diseases.** *Neurol Genet* 2016, **2**:e52.
- [146] Leach RM, Treacher DF: **Oxygen transport-2. Tissue hypoxia.** *BMJ* 1998, **317**(7169):1370–1373.
- [147] Oosthuysen B, Moons L, Storkebaum E, Beck H, Nuyens D, Brusselmans K, Van Dorpe J, Hellings P, Gorselink M, Heymans S, Theilmeier G, Dewerchin M, Laudénbach V, Vermeylen P, Raat H, Acker T, Vleminckx V, Van Den Bosch L, Cashman N, Fujisawa H, Drost MR, Sciôt R, Bruyninckx F, Hicklin DJ, Ince C, Gressens P, Lupu F, Plate KH, Robberecht W, Herbert JM, Collen D, Carmeliet P: **Deletion of the hypoxia-response element in the vascular endothelial growth factor promoter causes motor neuron degeneration.** *Nat. Genet.* 2001, **28**(2):131–138.

- [148] Rius J, Guma M, Schachtrup C, Akassoglou K, Zinkernagel AS, Nizet V, Johnson RS, Haddad GG, Karin M: **NF-kappaB links innate immunity to the hypoxic response through transcriptional regulation of HIF-1alpha**. *Nature* 2008, **453**(7196):807–811.
- [149] Liao SH, Zhao XY, Han YH, Zhang J, Wang LS, Xia L, Zhao KW, Zheng Y, Guo M, Chen GQ: **Proteomics-based identification of two novel direct targets of hypoxia-inducible factor-1 and their potential roles in migration/invasion of cancer cells**. *Proteomics* 2009, **9**(15):3901–3912.
- [150] Minchenko OH, Tsymbal DO, Minchenko DO, Kovalevska OV, Karbovskiy LL, Bikfalvi A: **Inhibition of ERN1 signaling enzyme affects hypoxic regulation of the expression of E2F8, EPAS1, HOXC6, ATF3, TBX3 and FOXF1 genes in U87 glioma cells**. *Ukr Biochem J* 2015, **87**(2):76–87.
- [151] Singh Y, Garden OA, Lang F, Cobb BS: **MicroRNAs regulate T-cell production of interleukin-9 and identify hypoxia-inducible factor-2I α as an important regulator of T helper 9 and regulatory T-cell differentiation**. *Immunology* 2016, **149**:74–86.
- [152] Zhang R, Zhou L, Li Q, Liu J, Yao W, Wan H: **Up-regulation of two actin-associated proteins prompts pulmonary artery smooth muscle cell migration under hypoxia**. *American journal of respiratory cell and molecular biology* 2009, **41**(4):467–475.
- [153] Jackson PK: **TTBK2 kinase: linking primary cilia and cerebellar ataxias**. *Cell* 2012, **151**(4):697–699.
- [154] Matilla-Duenas A: **The ever expanding spinocerebellar ataxias**. Editorial. *Cerebellum* 2012, **11**(4):821–827.
- [155] Schiffmann R, van der Knaap MS: **Invited article: an MRI-based approach to the diagnosis of white matter disorders**. *Neurology* 2009, **72**(8):750–759.
- [156] Daoud H, Tetreault M, Gibson W, Guerrero K, Cohen A, Gburek-Augustat J, Synofzik M, Brais B, Stevens CA, Sanchez-Carpintero R, Goizet C, Naidu S, Vanderver A, Bernard G: **Mutations in POLR3A and POLR3B are a major cause of hypomyelinating leukodystrophies with or without dental abnormalities and/or hypogonadotropic hypogonadism**. *J. Med. Genet.* 2013, **50**(3):194–197.
- [157] Bastepe M: **The GNAS locus and pseudohypoparathyroidism**. *Adv. Exp. Med. Biol.* 2008, **626**:27–40.
- [158] Johnson KR, Gagnon LH, Chang B: **A hypomorphic mutation of the gamma-1 adaptin gene (Ap1g1) causes inner ear, retina, thyroid, and testes abnormalities in mice**. *Mamm. Genome* 2016, **27**(5-6):200–212.
- [159] Molenaar PC, Newsom-Davis J, Polak RL, Vincent A: **Eaton-Lambert syndrome: acetylcholine and choline acetyltransferase in skeletal muscle**. *Neurology* 1982, **32**(9):1061–1065.

- [160] Brunet A, Bonni A, Zigmond MJ, Lin MZ, Juo P, Hu LS, Anderson MJ, Arden KC, Blenis J, Greenberg ME: **Akt promotes cell survival by phosphorylating and inhibiting a Forkhead transcription factor.** *Cell* 1999, **96**(6):857–868.
- [161] Roberts AI, Lee L, Schwarz E, Groh V, Spies T, Ebert EC, Jabri B: **NKG2D receptors induced by IL-15 costimulate CD28-negative effector CTL in the tissue microenvironment.** *J. Immunol.* 2001, **167**(10):5527–5530.
- [162] Maasho K, Opoku-Anane J, Marusina AI, Coligan JE, Borrego F: **NKG2D is a costimulatory receptor for human naive CD8+ T cells.** *J. Immunol.* 2005, **174**(8):4480–4484.
- [163] Park YP, Choi SC, Kiesler P, Gil-Krzewska A, Borrego F, Weck J, Krzewski K, Coligan JE: **Complex regulation of human NKG2D-DAP10 cell surface expression: opposing roles of the γ c cytokines and TGF- β 1.** *Blood* 2011, **118**(11):3019–3027.
- [164] Zhang Y, Wang X, Yang H, Liu H, Lu Y, Han L, Liu G: **Kinase AKT controls innate immune cell development and function.** *Immunology* 2013, **140**(2):143–152.
- [165] Koga R, Hamano S, Kuwata H, Atarashi K, Ogawa M, Hisaeda H, Yamamoto M, Akira S, Himeno K, Matsumoto M, Takeda K: **TLR-dependent induction of IFN-beta mediates host defense against Trypanosoma cruzi.** *J. Immunol.* 2006, **177**(10):7059–7066.
- [166] Rosenzweig SD, Holland SM: **Defects in the interferon-gamma and interleukin-12 pathways.** *Immunol. Rev.* 2005, **203**:38–47.
- [167] Casanova JL, Abel L: **Human genetics of infectious diseases: a unified theory.** *EMBO J.* 2007, **26**(4):915–922.
- [168] Jaggy A, Oliver JE, Ferguson DC, Mahaffey EA, Glaus T: **Neurological manifestations of hypothyroidism: a retrospective study of 29 dogs.** *J. Vet. Intern. Med.* 1994, **8**(5):328–336.
- [169] Wu D, Hersh LB: **Choline acetyltransferase: celebrating its fiftieth year.** *J. Neurochem.* 1994, **62**(5):1653–1663.
- [170] Springer TA: **Adhesion receptors of the immune system.** *Nature* 1990, **346**(6283):425.
- [171] Gerard C, Rollins BJ: **Chemokines and disease.** *Nat. Immunol.* 2001, **2**(2):108–115.
- [172] Bach TL, Chen QM, Kerr WT, Wang Y, Lian L, Choi JK, Wu D, Kazanietz MG, Koretzky GA, Zigmond S, Abrams CS: **Phospholipase cbeta is critical for T cell chemotaxis.** *J. Immunol.* 2007, **179**(4):2223–2227.
- [173] Morrot A, Zavala F: **Effector and memory CD8+ T cells as seen in immunity to malaria.** *Immunol. Rev.* 2004, **201**:291–303.
- [174] Goldszmid RS, Sher A: **Processing and presentation of antigens derived from intracellular protozoan parasites.** *Curr. Opin. Immunol.* 2010, **22**:118–123.

- [175] Jordan KA, Hunter CA: **Regulation of CD8+ T cell responses to infection with parasitic protozoa.** *Exp. Parasitol.* 2010, **126**(3):318–325.
- [176] Doolan DL, Hoffman SL: **IL-12 and NK cells are required for antigen-specific adaptive immunity against malaria initiated by CD8+ T cells in the Plasmodium yoelii model.** *J. Immunol.* 1999, **163**(2):884–892.
- [177] Zelenika D, Grima B, Pessac B: **A new family of transcripts of the myelin basic protein gene: expression in brain and in immune system.** *J. Neurochem.* 1993, **60**(4):1574–1577.
- [178] Ymer S, Tucker WQ, Sanderson CJ, Hapel AJ, Campbell HD, Young IG: **Constitutive synthesis of interleukin-3 by leukaemia cell line WEHI-3B is due to retroviral insertion near the gene.** *Nature* 1985, **317**(6034):255–258.
- [179] Dorssers L, Burger H, Bot F, Delwel R, Geurts van Kessel AH, Lowenberg B, Wage-maker G: **Characterization of a human multilineage-colony-stimulating factor cDNA clone identified by a conserved noncoding sequence in mouse interleukin-3.** *Gene* 1987, **55**:115–124.
- [180] Gringhuis SI, den Dunnen J, Litjens M, van der Vlist M, Wevers B, Bruijns SC, Geijtenbeek TB: **Dectin-1 directs T helper cell differentiation by controlling noncanonical NF- κ B activation through Raf-1 and Syk.** *Nat. Immunol.* 2009, **10**(2):203–213.
- [181] Benito A, Silva M, Grillot D, Nunez G, Fernandez-Luna JL: **Apoptosis induced by erythroid differentiation of human leukemia cell lines is inhibited by Bcl-XL.** *Blood* 1996, **87**(9):3837–3843.
- [182] Packham G, White EL, Eischen CM, Yang H, Parganas E, Ihle JN, Grillot DA, Zambetti GP, Nunez G, Cleveland JL: **Selective regulation of Bcl-XL by a Jak kinase-dependent pathway is bypassed in murine hematopoietic malignancies.** *Genes Dev.* 1998, **12**(16):2475–2487.
- [183] Yuan ZL, Guan YJ, Wang L, Wei W, Kane AB, Chin YE: **Central role of the threonine residue within the p+1 loop of receptor tyrosine kinase in STAT3 constitutive phosphorylation in metastatic cancer cells.** *Mol. Cell. Biol.* 2004, **24**(21):9390–9400.
- [184] Ponce NE, Cano RC, Carrera-Silva EA, Lima AP, Gea S, Aoki MP: **Toll-like receptor-2 and interleukin-6 mediate cardiomyocyte protection from apoptosis during Trypanosoma cruzi murine infection.** *Med. Microbiol. Immunol.* 2012, **201**(2):145–155.
- [185] Stahl P, Ruppert V, Meyer T, Schmidt J, Campos MA, Gazzinelli RT, Maisch B, Schwarz RT, Debierre-Grockiego F: **Trypomastigotes and amastigotes of Trypanosoma cruzi induce apoptosis and STAT3 activation in cardiomyocytes in vitro.** *Apoptosis* 2013, **18**(6):653–663.
- [186] Clark RI, Tan SWS, Péans CB, Roostalu U, Vivancos V, Bronda K, Pilátová M, Fu J, Walker DW, Berdeaux R, Geissmann F, Dionne MS: **MEF2 Is an In Vivo Immune-Metabolic Switch.** *Cell* 2013, **155**:435–447.

- [187] Gong X, Tang X, Wiedmann M, Wang X, Peng J, Zheng D, Blair LA, Marshall J, Mao Z: **Cdk5-mediated inhibition of the protective effects of transcription factor MEF2 in neurotoxicity-induced apoptosis.** *Neuron* 2003, **38**:33–46.
- [188] She H, Yang Q, Shepherd K, Smith Y, Miller G, Testa C, Mao Z: **Direct regulation of complex I by mitochondrial MEF2D is disrupted in a mouse model of Parkinson disease and in human patients.** *J. Clin. Invest.* 2011, **121**(3):930–940.
- [189] She H, Yang Q, Mao Z: **Neurotoxin-induced selective ubiquitination and regulation of MEF2A isoform in neuronal stress response.** *J. Neurochem.* 2012, **122**(6):1203–1210.
- [190] He C, Medley SC, Hu T, Hinsdale ME, Lupu F, Virmani R, Olson LE: **PDGFRI² signalling regulates local inflammation and synergizes with hypercholesterolaemia to promote atherosclerosis.** *Nat Commun* 2015, **6**:7770.
- [191] Kim CH, Broxmeyer HE: **Chemokines: signal lamps for trafficking of T and B cells for development and effector function.** *J Leukoc Biol* 1999, **65**:6–15.
- [192] Dimova DK, Dyson NJ: **The E2F transcriptional network: old acquaintances with new faces.** *Oncogene* 2005, **24**(17):2810–2826.
- [193] Hu H, Sun SC: **Ubiquitin signaling in immune responses.** *Cell Res.* 2016, **26**(4):457–483.
- [194] Song J, Salek-Ardakani S, So T, Croft M: **The kinases aurora B and mTOR regulate the G1-S cell cycle progression of T lymphocytes.** *Nat. Immunol.* 2007, **8**:64–73.
- [195] Dyson N: **The regulation of E2F by pRB-family proteins.** *Genes Dev.* 1998, **12**(15):2245–2262.
- [196] Nevins JR: **Toward an understanding of the functional complexity of the E2F and retinoblastoma families.** *Cell Growth Differ.* 1998, **9**(8):585–593.
- [197] Trimarchi JM, Lees JA: **Sibling rivalry in the E2F family.** *Nat. Rev. Mol. Cell Biol.* 2002, **3**:11–20.
- [198] Chuenkova MV, PereiraPerrin M: **Trypanosoma cruzi targets Akt in host cells as an intracellular antiapoptotic strategy.** *Sci Signal* 2009, **2**(97):ra74.
- [199] Woolsey AM, Sunwoo L, Petersen CA, Brachmann SM, Cantley LC, Burleigh BA: **Novel PI 3-kinase-dependent mechanisms of trypanosome invasion and vacuole maturation.** *J. Cell. Sci.* 2003, **116**(Pt 17):3611–3622.
- [200] Caradonna KL, Engel JC, Jacobi D, Lee CH, Burleigh BA: **Host metabolism regulates intracellular growth of Trypanosoma cruzi.** *Cell Host Microbe* 2013, **13**:108–117.
- [201] Qiu L, Joazeiro C, Fang N, Wang HY, Elly C, Altman Y, Fang D, Hunter T, Liu YC: **Recognition and ubiquitination of Notch by Itch, a hect-type E3 ubiquitin ligase.** *J. Biol. Chem.* 2000, **275**(46):35734–35737.

- [202] You F, Sun H, Zhou X, Sun W, Liang S, Zhai Z, Jiang Z: **PCBP2 mediates degradation of the adaptor MAVS via the HECT ubiquitin ligase AIP4.** *Nat. Immunol.* 2009, **10**(12):1300–1308.
- [203] Aki D, Zhang W, Liu YC: **The E3 ligase Itch in immune regulation and beyond.** *Immunol. Rev.* 2015, **266**:6–26.
- [204] Kanazawa I: **Molecular pathology of dentatorubral-pallidoluysian atrophy.** *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 1999, **354**(1386):1069–1074.
- [205] Lee JM, Chan K, Kan YW, Johnson JA: **Targeted disruption of Nrf2 causes regenerative immune-mediated hemolytic anemia.** *Proc. Natl. Acad. Sci. U.S.A.* 2004, **101**(26):9751–9756.
- [206] Silva B, Faustino P: **An overview of molecular basis of iron metabolism regulation and the associated pathologies.** *Biochim. Biophys. Acta* 2015, **1852**(7):1347–1359.
- [207] Cho HY, Jedlicka AE, Reddy SP, Kensler TW, Yamamoto M, Zhang LY, Kleeberger SR: **Role of NRF2 in protection against hyperoxic lung injury in mice.** *Am. J. Respir. Cell Mol. Biol.* 2002, **26**(2):175–182.
- [208] Roux PP, Blenis J: **ERK and p38 MAPK-activated protein kinases: a family of protein kinases with diverse biological functions.** *Microbiol. Mol. Biol. Rev.* 2004, **68**(2):320–344.
- [209] Tavazzi B, Di Pierro D, Amorini AM, Fazzina G, Tuttobene M, Giardina B, Lazzarino G: **Energy metabolism and lipid peroxidation of human erythrocytes as a function of increased oxidative stress.** *Eur. J. Biochem.* 2000, **267**(3):684–689.
- [210] Mabley JG, Szabo C: **Inflammatory disease and sunlight: the vitamin D–poly (ADP-ribose) polymerase connection.** *Future Rheumatol* 2008, **3**(2):169–181.
- [211] Paul WE, Steinman R, Beutler B, Hoffmann J: **Bridging innate and adaptive immunity.** *Cell* 2011, **147**(6):1212–1215.
- [212] Stanojevic V, Habener JF, Holz GG, Leech CA: **Cytosolic adenylate kinases regulate K-ATP channel activity in human beta-cells.** *Biochem. Biophys. Res. Commun.* 2008, **368**(3):614–619.
- [213] Battino M, Giampieri F, Pistollato F, Sureda A, de Oliveira MR, Pittala V, Fallarino F, Nabavi SF, Atanasov AG, Nabavi SM: **Nrf2 as regulator of innate immunity: A molecular Swiss army knife!** *Biotechnol. Adv.* 2018, **36**(2):358–370.
- [214] Goncalves VM, Matteucci KC, Buzzo CL, Miollo BH, Ferrante D, Torrecilhas AC, Rodrigues MM, Alvarez JM, Bortoluci KR: **NLRP3 controls Trypanosoma cruzi infection through a caspase-1-dependent IL-1R-independent NO production.** *PLoS Negl Trop Dis* 2013, **7**(10):e2469.

- [215] Shalini S, Dorstyn L, Dawar S, Kumar S: **Old, new and emerging functions of caspases.** *Cell death and differentiation* 2015, **22**(4):526.
- [216] Zermati Y, Garrido C, Amsellem S, Fishelson S, Bouscary D, Valensi F, Varet B, Solary E, Hermine O: **Caspase activation is required for terminal erythroid differentiation.** *J. Exp. Med.* 2001, **193**(2):247–254.
- [217] Acuto O, Di Bartolo V, Michel F: **Tailoring T-cell receptor signals by proximal negative feedback mechanisms.** *Nat. Rev. Immunol.* 2008, **8**(9):699–712.
- [218] Au-Yeung BB, Deindl S, Hsu LY, Palacios EH, Levin SE, Kuriyan J, Weiss A: **The structure, regulation, and function of ZAP-70.** *Immunol. Rev.* 2009, **228**:41–57.
- [219] Alonso A, Saxena M, Williams S, Mustelin T: **Inhibitory role for dual specificity phosphatase VHR in T cell antigen receptor and CD28-induced Erk and Jnk activation.** *J. Biol. Chem.* 2001, **276**(7):4766–4771.
- [220] Hoyt R, Zhu W, Cerignoli F, Alonso A, Mustelin T, David M: **Cutting edge: selective tyrosine dephosphorylation of interferon-activated nuclear STAT5 by the VHR phosphatase.** *J. Immunol.* 2007, **179**(6):3402–3406.
- [221] Heinonen KM, Dube N, Bourdeau A, Lapp WS, Tremblay ML: **Protein tyrosine phosphatase 1B negatively regulates macrophage development through CSF-1 signaling.** *Proc. Natl. Acad. Sci. U.S.A.* 2006, **103**(8):2776–2781.
- [222] Martin-Granados C, Prescott AR, Le Sommer S, Klaska IP, Yu T, Muckersie E, Giuraniuc CV, Grant L, Delibegovic M, Forrester JV: **A key role for PTP1B in dendritic cell maturation, migration, and T cell activation.** *J Mol Cell Biol* 2015, **7**(6):517–528.
- [223] Ingley E: **Functions of the Lyn tyrosine kinase in health and disease.** *Cell Commun. Signal* 2012, **10**:21.
- [224] Kawai T, Akira S: **Antiviral signaling through pattern recognition receptors.** *J. Biochem.* 2007, **141**(2):137–145.
- [225] Blasius AL, Beutler B: **Intracellular toll-like receptors.** *Immunity* 2010, **32**(3):305–315.
- [226] McNab F, Mayer-Barber K, Sher A, Wack A, O'Garra A: **Type I interferons in infectious disease.** *Nat. Rev. Immunol.* 2015, **15**(2):87–103.
- [227] Yi T, Cleveland JL, Ihle JN: **Identification of novel protein tyrosine phosphatases of hematopoietic cells by polymerase chain reaction amplification.** *Blood* 1991, **78**(9):2222–2228.
- [228] Adachi M, Sekiya M, Arimura Y, Takekawa M, Itoh F, Hinoda Y, Imai K, Yachi A: **Protein-tyrosine phosphatase expression in pre-B cell NALM-6.** *Cancer Res.* 1992, **52**(3):737–740.

- [229] Zhang S, Han J, Sells MA, Chernoff J, Knaus UG, Ulevitch RJ, Bokoch GM: **Rho family GTPases regulate p38 mitogen-activated protein kinase through the downstream mediator Pak1.** *J. Biol. Chem.* 1995, **270**(41):23934–23936.
- [230] Dong C, Davis RJ, Flavell RA: **MAP kinases in the immune response.** *Annu. Rev. Immunol.* 2002, **20**:55–72.
- [231] Wang D, Sai J, Carter G, Sachpatzidis A, Lolis E, Richmond A: **PAK1 kinase is required for CXCL1-induced chemotaxis.** *Biochemistry* 2002, **41**(22):7100–7107.
- [232] Traves PG, Pardo V, Pimentel-Santillana M, Gonzalez-Rodriguez A, Mojena M, Rico D, Montenegro Y, Cales C, Martin-Sanz P, Valverde AM, Bosca L: **Pivotal role of protein tyrosine phosphatase 1B (PTP1B) in the macrophage response to pro-inflammatory and anti-inflammatory challenge.** *Cell Death Dis* 2014, **5**:e1125.
- [233] Utz PJ, Hottel M, Schur PH, Anderson P: **Proteins phosphorylated during stress-induced apoptosis are common targets for autoantibody production in patients with systemic lupus erythematosus.** *Journal of Experimental Medicine* 1997, **185**(5):843–854.
- [234] Patterson H, Nibbs R, McInnes I, Siebert S: **Protein kinase inhibitors in the treatment of inflammatory and autoimmune diseases.** *Clinical & Experimental Immunology* 2014, **176**:1–10.
- [235] Eguchi K: **Apoptosis in autoimmune diseases.** *Internal medicine* 2001, **40**(4):275–284.
- [236] Lamkanfi M, Festjens N, Declercq W, Vanden Berghe T, Vandenabeele P: **Caspases in cell survival, proliferation and differentiation.** *Cell Death Differ.* 2007, **14**:44–55.
- [237] Rovelet-Lecrux A, Hannequin D, Guillin O, Legallic S, Jurici S, Wallon D, Frebourg T, Champion D: **Frontotemporal dementia phenotype associated with MAPT gene duplication.** *Journal of Alzheimer's Disease* 2010, **21**(3):897–902.
- [238] Zabetian CP, Hutter CM, Factor SA, Nutt JG, Higgins DS, Griffith A, Roberts JW, Leis BC, Kay DM, Yearout D, et al.: **Association analysis of MAPT H1 haplotype and subhaplotypes in Parkinson's disease.** *Annals of neurology* 2007, **62**(2):137–144.
- [239] Herrero J, Muffato M, Beal K, Fitzgerald S, Gordon L, Pignatelli M, Vilella AJ, Searle SM, Amode R, Brent S, et al.: **Ensembl comparative genomics resources.** *Database* 2016, **2016**.
- [240] Ye S, Huang Y, Joshi S, Zhang J, Yang F, Zhang G, Smyth SS, Li Z, Takai Y, Whiteheart SW: **Platelet secretion and hemostasis require syntaxin-binding protein STXBP5.** *The Journal of clinical investigation* 2014, **124**(10):4517–4528.
- [241] Swart GW: **Activated leukocyte cell adhesion molecule (CD166/ALCAM): developmental and mechanistic aspects of cell clustering and cell migration.** *European journal of cell biology* 2002, **81**(6):313–321.

- [242] Castillo-Rodríguez RA, Dávila-Borja VM, Juárez-Méndez S: **Data mining of pediatric medulloblastoma microarray expression reveals a novel potential subdivision of the Group 4 molecular subgroup.** *Oncology letters* 2018, **15**(5):6241–6250.
- [243] Weißflog L, Scholz CJ, Jacob CP, Nguyen TT, Zamzow K, Groß-Lesch S, Renner TJ, Romanos M, Rujescu D, Walitza S, et al.: **KCNIP4 as a candidate gene for personality disorders and adult ADHD.** *European Neuropsychopharmacology* 2013, **23**(6):436–447.
- [244] Kelsen JR, Ouahed J, Spessott WA, Kooshesh K, Sanmillan ML, Dawany N, Sullivan KE, Hamilton K, Slowik V, Nejentsev S, et al.: **25 MUTATIONS IN STXBP3 CONTRIBUTE TO VERY EARLY ONSET OF IBD, IMMUNODEFICIENCY AND HEARING LOSS.** *Gastroenterology* 2018, **154**:S40–S41.
- [245] Mitoma H, Hanabuchi S, Kim T, Bao M, Zhang Z, Sugimoto N, Liu YJ: **The DHX33 RNA helicase senses cytosolic RNA and activates the NLRP3 inflammasome.** *Immunity* 2013, **39**:123–135.
- [246] Paroli AF, Gonzalez PV, Díaz-Luján C, Onofrio LI, Arocena A, Cano RC, Carrera-Silva EA, Gea S: **NLRP3 inflammasome and caspase-1/11 pathway orchestrate different outcomes in the host protection against Trypanosoma cruzi Acute infection.** *Frontiers in immunology* 2018, **9**.
- [247] Li K, Gao B, Li J, Chen H, Li Y, Wei Y, Gong D, Gao J, Zhang J, Tan W, et al.: **ZNF32 protects against oxidative stress-induced apoptosis by modulating C1QBP transcription.** *Oncotarget* 2015, **6**(35):38107.
- [248] Sheridan JM: **Generation of a compartmentalised thymus organoid in vitro using fetal thymic epithelial progenitor cells** 2007.
- [249] Sperandio S, Fortin J, Sasik R, Robitaille L, Corbeil J, de Belle I: **The transcription factor Egr1 regulates the HIF-1 α gene during hypoxia.** *Molecular Carcinogenesis: Published in cooperation with the University of Texas MD Anderson Cancer Center* 2009, **48**:38–44.
- [250] Hancock RL, Dunne K, Walport LJ, Flashman E, Kawamura A: **Epigenetic regulation by histone demethylases in hypoxia.** *Epigenomics* 2015, **7**(5):791–811.
- [251] Gaudet P, Livstone MS, Lewis SE, Thomas PD: **Phylogenetic-based propagation of functional annotations within the Gene Ontology consortium.** *Brief. Bioinformatics* 2011, **12**(5):449–462.
- [252] Gordon S: **Alternative activation of macrophages.** *Nature reviews immunology* 2003, **3**:23.
- [253] El Kasmi KC, Qualls JE, Pesce JT, Smith AM, Thompson RW, Henao-Tamayo M, Basaraba RJ, König T, Schleicher U, Koo MS, et al.: **Toll-like receptor-induced arginase 1 in macrophages thwarts effective immunity against intracellular pathogens.** *Nature immunology* 2008, **9**(12):1399.

- [254] Short JJ, Vasu C, Holterman MJ, Curiel DT, Pereboev A: **Members of adenovirus species B utilize CD80 and CD86 as cellular attachment receptors.** *Virus research* 2006, **122**(1-2):144–153.
- [255] Campanella A, Rovelli E, Santambrogio P, Cozzi A, Taroni F, Levi S: **Mitochondrial ferritin limits oxidative damage regulating mitochondrial iron availability: hypothesis for a protective role in Friedreich ataxia.** *Human molecular genetics* 2008, **18**:1–11.
- [256] Cocco E, Porrini V, Derosas M, Nardi V, Biasiotto G, Maccarinelli F, Zanella I: **Protective effect of mitochondrial ferritin on cytosolic iron dysregulation induced by doxorubicin in HeLa cells.** *Molecular biology reports* 2013, **40**(12):6757–6764.
- [257] Nikpour M, Scharenberg C, Liu A, Conte S, Karimi M, Mortera-Blanco T, Gai V, Fernandez-Mercado M, Papaemmanuil E, Högstrand K, et al.: **The transporter ABCB7 is a mediator of the phenotype of acquired refractory anemia with ring sideroblasts.** *Leukemia* 2013, **27**(4):889.
- [258] Kim SJ, Ka S, Ha JW, Kim J, Yoo D, Kim K, Lee HK, Lim D, Cho S, Hanotte O, et al.: **Cattle genome-wide analysis reveals genetic signatures in trypanotolerant N'Dama.** *BMC genomics* 2017, **18**:371.
- [259] Bai J, Lai L, Yeo HC, Goh BC, Tan TM: **Multidrug resistance protein 4 (MRP4/ABCC4) mediates efflux of bimane-glutathione.** *The international journal of biochemistry & cell biology* 2004, **36**(2):247–257.
- [260] Ikehara Y, Shimizu N, Kono M, Nishihara S, Nakanishi H, Kitamura T, Narimatsu H, Tsuji S, Tatematsu M: **A novel glycosyltransferase with a polyglutamine repeat; a new candidate for GD1 α synthase (ST6GalNAc V) 1.** *FEBS letters* 1999, **463**(1-2):92–96.
- [261] Jang WY, Bae KB, Kim SH, Yu DH, Kim HJ, Ji YR, Park SJ, Park SJ, Kang MC, Jeong JI, et al.: **Overexpression of Jazf1 reduces body weight gain and regulates lipid metabolism in high fat diet.** *Biochemical and biophysical research communications* 2014, **444**(3):296–301.
- [262] Janssens V, Goris J, Van Hoof C: **PP2A: the expected tumor suppressor.** *Current opinion in genetics & development* 2005, **15**:34–41.
- [263] Bellot G, Garcia-Medina R, Gounon P, Chiche J, Roux D, Pouyssegur J, Mazure NM: **Hypoxia-induced autophagy is mediated through hypoxia-inducible factor induction of BNIP3 and BNIP3L via their BH3 domains.** *Molecular and cellular biology* 2009, **29**(10):2570–2581.
- [264] Kitamura N, Nakamura Y, Miyamoto Y, Miyamoto T, Kabu K, Yoshida M, Futamura M, Ichinose S, Arakawa H: **Mieap, a p53-inducible protein, controls mitochondrial quality by repairing or eliminating unhealthy mitochondria.** *PloS one* 2011, **6**:e16060.

- [265] Hu H, Wang B, Borde M, Nardone J, Maika S, Allred L, Tucker PW, Rao A: **Foxp1 is an essential transcriptional regulator of B cell development.** *Nature immunology* 2006, **7**(8):819.
- [266] Wang Y, Morrisey E: **Regulation of cardiomyocyte proliferation by Foxp1.** *Cell Cycle* 2010, **9**(21):4251–4252.
- [267] Adams KL, Rouso DL, Umbach JA, Novitch BG: **Foxp1-mediated programming of limb-innervating motor neurons from mouse and human embryonic stem cells.** *Nat Commun* 2015, **6**:6778.
- [268] Konopacki C, Pritykin Y, Rubtsov Y, Leslie CS, Rudensky AY: **Transcription factor Foxp1 regulates Foxp3 chromatin binding and coordinates regulatory T cell function.** *Nature immunology* 2019, **20**(2):232.
- [269] Hu H, Takano N, Xiang L, Gilkes DM, Luo W, Semenza GL: **Hypoxia-inducible factors enhance glutamate signaling in cancer cells.** *Oncotarget* 2014, **5**(19):8853–8868.
- [270] Kaspar JW, Jaiswal AK: **Tyrosine phosphorylation controls nuclear export of Fyn, allowing Nrf2 activation of cytoprotective gene expression.** *FASEB J.* 2011, **25**(3):1076–1087.
- [271] Salmond RJ, Filby A, Qureshi I, Caserta S, Zamoyska R: **T-cell receptor proximal signaling via the Src-family kinases, Lck and Fyn, influences T-cell activation, differentiation, and tolerance.** *Immunological reviews* 2009, **228**:9–22.
- [272] Flex E, Mangino M, Mazzoli M, Martini A, Migliosi V, Colosimo A, Mingarelli R, Pizzuti A, Dallapiccola B: **Mapping of a new autosomal dominant non-syndromic hearing loss locus (DFNA43) to chromosome 2p12.** *J. Med. Genet.* 2003, **40**(4):278–281.
- [273] Khan SY, Ahmed ZM, Shabbir MI, Kitajiri S, Kalsoom S, Tasneem S, Shayiq S, Ramesh A, Srisailpathy S, Khan SN, Smith RJ, Riazuddin S, Friedman TB, Riazuddin S: **Mutations of the RDX gene cause nonsyndromic hearing loss at the DFNB24 locus.** *Hum. Mutat.* 2007, **28**(5):417–423.
- [274] Kanaan Z, Qadan M, Eichenberger MR, Galandiuk S: **The actin-cytoskeleton pathway and its potential role in inflammatory bowel disease-associated human colorectal cancer.** *Genet Test Mol Biomarkers* 2010, **14**(3):347–353.
- [275] Chen Z, Krmar RT, Dada L, Efendiev R, Leibiger IB, Pedemonte CH, Katz AI, Sznajder JI, Bertorello AM: **Phosphorylation of Adaptor Protein-2 μ 2 Is Essential for Na⁺, K⁺-ATPase Endocytosis in Response to Either G Protein-Coupled Receptor or Reactive Oxygen Species.** *American journal of respiratory cell and molecular biology* 2006, **35**:127–132.
- [276] Yoo JC, yeon Lim T, Park JS, Hah YS, Park N, Hong SG, Park JY, Yoon TJ: **SYT14L, especially its C2 domain, is involved in regulating melanocyte differentiation.** *Journal of dermatological science* 2013, **72**(3):246–251.

- [277] Nobuyuki I, Sumiko K, Yukiko N, Takeshi S, Dai A, Fumio H, Fumiaki Y: **Cloning and sequence of a functionally active cDNA encoding the mouse ubiquitin-activating enzyme E1.** *Gene* 1992, **118**(2):279–282.
- [278] Zimmerman GA, McIntyre TM, Prescott SM, Stafforini DM: **The platelet-activating factor signaling system and its regulators in syndromes of inflammation and thrombosis.** *Critical care medicine* 2002, **30**(5):S294–S301.
- [279] Schuler M, Bossy-Wetzel E, Goldstein JC, Fitzgerald P, Green DR: **p53 induces apoptosis by caspase activation through mitochondrial cytochrome c release.** *Journal of Biological Chemistry* 2000, **275**(10):7337–7342.
- [280] Kitada T, Asakawa S, Hattori N, Matsumine H, Yamamura Y, Minoshima S, Yokochi M, Mizuno Y, Shimizu N: **Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism.** *nature* 1998, **392**(6676):605.
- [281] Derycke LD, Bracke ME: **N-cadherin in the spotlight of cell-cell adhesion, differentiation, embryogenesis, invasion and signalling.** *International Journal of Developmental Biology* 2004, **48**(5-6):463–476.
- [282] Tohyama Y, Yamamura H: **Complement-mediated phagocytosis-the role of Syk.** *IUBMB life* 2006, **58**(5-6):304–308.
- [283] Li LO, Klett EL, Coleman RA: **Acyl-CoA synthesis, lipid metabolism and lipotoxicity.** *Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids* 2010, **1801**(3):246–251.
- [284] Xiao N, Eto D, Elly C, Peng G, Crotty S, Liu YC: **The E3 ubiquitin ligase Itch is required for the differentiation of follicular helper T cells.** *Nature immunology* 2014, **15**(7):657.
- [285] Zhang C, Zhang J, Niu J, Zhou Z, Zhang J, Tian Z: **Interleukin-12 improves cytotoxicity of natural killer cells via upregulated expression of NKG2D.** *Hum. Immunol.* 2008, **69**(8):490–500.
- [286] Mwai O, Hanotte O, Kwon YJ, Cho S: **African indigenous cattle: unique genetic resources in a rapidly changing world.** *Asian-Australasian journal of animal sciences* 2015, **28**(7):911.
- [287] Mbole-Kariuki MN, Sonstegard T, Orth A, Thumbi SM, Bronsvort BM, Kiara H, Toye P, Conradie I, Jennings A, Coetzer K, Woolhouse ME, Hanotte O, Tapio M: **Genome-wide analysis reveals the ancient and recent admixture history of East African Shorthorn Zebu from Western Kenya.** *Heredity (Edinb)* 2014, **113**(4):297–305.
- [288] Gibbs RA, Taylor JF, Van Tassell CP, Barendse W, Eversole KA, Gill CA, Green RD, Hamernik DL, Kappes SM, Lien S, Matukumalli LK, McEwan JC, Nazareth LV, Schnabel RD, Weinstock GM, Wheeler DA, Ajmone-Marsan P, Boettcher PJ, Caetano AR, Garcia JF, Hanotte O, Mariani P, Skow LC, Sonstegard TS, Williams JL, Diallo B, Hailemariam L,

- Martinez ML, Morris CA, Silva LO, Spelman RJ, Mulatu W, Zhao K, Abbey CA, Agaba M, Araujo FR, Bunch RJ, Burton J, Gorni C, Olivier H, Harrison BE, Luff B, Machado MA, Mwakaya J, Plastow G, Sim W, Smith T, Thomas MB, Valentini A, Williams P, Womack J, Woolliams JA, Liu Y, Qin X, Worley KC, Gao C, Jiang H, Moore SS, Ren Y, Song XZ, Bustamante CD, Hernandez RD, Muzny DM, Patil S, San Lucas A, Fu Q, Kent MP, Vega R, Matukumalli A, McWilliam S, Sclep G, Bryc K, Choi J, Gao H, Grefenstette JJ, Murdoch B, Stella A, Villa-Angulo R, Wright M, Aerts J, Jann O, Negrini R, Goddard ME, Hayes BJ, Bradley DG, Barbosa da Silva M, Lau LP, Liu GE, Lynn DJ, Panzitta F, Dodds KG: **Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds.** *Science* 2009, **324**(5926):528–532.
- [289] Bradley DG, MacHugh DE, Cunningham P, Loftus RT: **Mitochondrial diversity and the origins of African and European cattle.** *Proc. Natl. Acad. Sci. U.S.A.* 1996, **93**(10):5131–5135.
- [290] Bonfiglio S, Ginja C, De Gaetano A, Achilli A, Olivieri A, Colli L, Tesfaye K, Agha SH, Gama LT, Cattonaro F, Penedo MC, Ajmone-Marsan P, Torroni A, Ferretti L: **Origin and spread of *Bos taurus*: new clues from mitochondrial genomes belonging to haplogroup T1.** *PLoS ONE* 2012, **7**(6):e38601.
- [291] Berthier D, Brenière SF, Bras-Gonçalves R, Lemesre JL, Jamonneau V, Solano P, Lejon V, Thévenon S, Bucheton B: **Tolerance to trypanosomatids: a threat, or a key for disease elimination?** *Trends in parasitology* 2016, **32**(2):157–168.
- [292] Stijlemans B, Vankrunkelsven A, Caljon G, Bockstal V, Guilliams M, Bosschaerts T, Beschin A, Raes G, Magez S, De Baetselier P: **The central role of macrophages in trypanosomiasis-associated anemia: rationale for therapeutical approaches.** *Endocr Metab Immune Disord Drug Targets* 2010, **10**:71–82.
- [293] Bosschaerts T, Morias Y, Stijlemans B, Herin M, Porta C, Sica A, Mantovani A, De Baetselier P, Beschin A: **IL-10 limits production of pathogenic TNF by M1 myeloid cells through induction of nuclear NF- κ B p50 member in *Trypanosoma congolense* infection-resistant C57BL/6 mice.** *Eur. J. Immunol.* 2011, **41**(11):3270–3280.
- [294] La Greca F, Haynes C, Stijlemans B, De Trez C, Magez S: **Antibody-mediated control of *Trypanosoma vivax* infection fails in the absence of tumour necrosis factor.** *Parasite Immunol.* 2014, **36**(6):271–276.
- [295] Mamoudou A, Njanloga A, Hayatou A, Suh PF, Achukwi MD: **Animal trypanosomiasis in clinically healthy cattle of north Cameroon: epidemiological implications.** *Parasit Vectors* 2016, **9**:206.
- [296] Morrison LJ, Vezza L, Rowan T, Hope JC: **Animal African Trypanosomiasis: Time to Increase Focus on Clinically Relevant Parasite and Host Species.** *Trends Parasitol.* 2016, **32**(8):599–607.

- [297] Fatihu M, Adamu S, Ibrahim N, Euvie L, Esievo K: **The effect of experimental Trypanosoma vivax infection on the thyroid gland in Zebu bulls.** *Vet. arhiv* 2009, **79**(5):429–437.
- [298] Allam L, Ogwu D, Agbede R, Sackey A: **Hematological and serum biochemical changes in gilts experimentally infected with xperimentally infected with Trypanosoma brucei.** *Vet. arhiv* 2011, **81**(5):597–609.
- [299] Oleksyk TK, Smith MW, O'Brien SJ: **Genome-wide scans for footprints of natural selection.** *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 2010, **365**(1537):185–205.
- [300] Flávia Nardy A, Freire-de Lima CG, Morrot A: **Immune evasion strategies of Trypanosoma cruzi.** *Journal of immunology research* 2015, **2015**.
- [301] Taylor KA: **Immune responses of cattle to African trypanosomes: protective or pathogenic?** *Int. J. Parasitol.* 1998, **28**(2):219–240.
- [302] Lodish H, Berk A, Matsudaira P, C K, Krieger M, Scott M, Zipursky L, Darnell J: *Molecular Cell Biology*. 41 Madison Avenue, New York, USA: W. H. Freeman and Company, 5 edition 2004.
- [303] Zinngrebe J, Montinaro A, Peltzer N, Walczak H: **Ubiquitin in the immune system.** *EMBO Rep.* 2014, **15**:28–45.
- [304] Ersching J, Vasconcelos JR, Ferreira CP, Caetano BC, Machado AV, Bruna-Romero O, Baron MA, Ferreira LR, Cunha-Neto E, Rock KL, Gazzinelli RT, Rodrigues MM: **The Combined Deficiency of Immunoproteasome Subunits Affects Both the Magnitude and Quality of Pathogen- and Genetic Vaccination-Induced CD8+ T Cell Responses to the Human Protozoan Parasite Trypanosoma cruzi.** *PLoS Pathog.* 2016, **12**(4):e1005593.
- [305] Kierstein S, Noyes H, Naessens J, Nakamura Y, Pritchard C, Gibson J, Kemp S, Brass A: **Gene expression profiling in a mouse model for African trypanosomiasis.** *Genes Immun.* 2006, **7**(8):667–679.
- [306] Kaminski WE, Lindahl P, Lin NL, Broudy VC, Crosby JR, Hellstrom M, Swolin B, Bowen-Pope DF, Martin PJ, Ross R, Betsholtz C, Raines EW: **Basis of hematopoietic defects in platelet-derived growth factor (PDGF)-B and PDGF beta-receptor null mice.** *Blood* 2001, **97**(7):1990–1998.
- [307] Hu T, Ghazaryan S, Sy C, Wiedmeyer C, Chang V, Wu L: **Concomitant inactivation of Rb and E2f8 in hematopoietic stem cells synergizes to induce severe anemia.** *Blood* 2012, **119**(19):4532–4542.
- [308] Saleh MA, Al-Salahy MB, Sanousi SA: **Oxidative stress in blood of camels (Camelus dromedaries) naturally infected with Trypanosoma evansi.** *Vet. Parasitol.* 2009, **162**(3-4):192–199.

- [309] Chan K, Han XD, Kan YW: **An important function of Nrf2 in combating oxidative stress: detoxification of acetaminophen.** *Proc. Natl. Acad. Sci. U.S.A.* 2001, **98**(8):4611–4616.
- [310] Sangokoya C, Telen MJ, Chi JT: **microRNA miR-144 modulates oxidative stress tolerance and associates with anemia severity in sickle cell disease.** *Blood* 2010, **116**(20):4338–4348.
- [311] Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S: **The protein kinase complement of the human genome.** *Science* 2002, **298**(5600):1912–1934.
- [312] Salmond RJ, Filby A, Qureshi I, Caserta S, Zamoyska R: **T-cell receptor proximal signaling via the Src-family kinases, Lck and Fyn, influences T-cell activation, differentiation, and tolerance.** *Immunol. Rev.* 2009, **228**:9–22.
- [313] Naessens J: **Bovine trypanotolerance: A natural ability to prevent severe anaemia and haemophagocytic syndrome?** *Int. J. Parasitol.* 2006, **36**(5):521–528.
- [314] Arai S, Miyazaki T: **Impaired maturation of myeloid progenitors in mice lacking novel Polycomb group protein MBT-1.** *The EMBO journal* 2005, **24**(10):1863–1873.
- [315] Amin DN, Ngoyi DM, Nkwachi GM, Palomba M, Rottenberg M, Büscher P, Kristensson K, Masocha W: **Identification of stage biomarkers for human African trypanosomiasis.** *The American journal of tropical medicine and hygiene* 2010, **82**(6):983–990.
- [316] Mendes-da Cruz DA, Silva JS, Cotta-de Almeida V, Savino W: **Altered thymocyte migration during experimental acute Trypanosoma cruzi infection: combined role of fibronectin and the chemokines CXCL12 and CCL4.** *Eur. J. Immunol.* 2006, **36**(6):1486–1493.
- [317] Martin E, Sung MH: **Challenges of Decoding Transcription Factor Dynamics in Terms of Gene Regulation.** *Cells* 2018, **7**(9):132.
- [318] Katunga-Rwakishaya E, MURRAY M, HOLMES PH: **Haematological erythrokinetic and blood lipid changes in sheep infected with Trypanosoma congolense.** *International Scientific Council for Trypanosomiasis Research and Control (OAU/ISCTRC) 21st Meeting. IUBMB life* 1991, (116):257.
- [319] Ogunsanmi A, Taiwo V, Onawumi B, Mbagwu H, Okoronkwo C, et al.: **Correlation of physiological plasma lipid levels with resistance of cattle to trypanosomiasis.** *Veterinarski arhiv* 2000, **70**(5):251–258.
- [320] Fang D, Elly C, Gao B, Fang N, Altman Y, Joazeiro C, Hunter T, Copeland N, Jenkins N, Liu YC: **Dysregulation of T lymphocyte function in itchy mice: a role for Itch in T H 2 differentiation.** *Nature immunology* 2002, **3**(3):281.
- [321] Barrett-Connor E, Ugoretz RJ, Braude AI: **Disseminated intravascular coagulation in trypanosomiasis.** *Archives of Internal Medicine* 1973, **131**(4):574–577.

-
- [322] Forsberg C, Valli V, Gentry P, Donworth R: **The Pathogenesis of Trypanosoma congolense Infection in Calves: IV. The Kinetics of Blood Coagulation.** *Veterinary pathology* 1979, **16**(2):229–242.
- [323] Igoshin AV, Yurchenko AA, Belonogova NM, Petrovsky DV, Aitnazarov RB, Soloshenko VA, Yudin NS, Larkin DM: **Genome-wide association study and scan for signatures of selection point to candidate genes for body temperature maintenance under the cold stress in Siberian cattle populations.** *BMC genetics* 2019, **20**:26.

A. Appendix

A.1. Identification of Candidate Signature Genes and Key Regulators Associated with Trypanotolerance in the Sheko Breed



Identification of Candidate Signature Genes and Key Regulators Associated With Trypanotolerance in the Sheko Breed

Yonatan Ayalew Mekonnen¹, Mehmet Gültas^{1,2}, Kefena Effa³, Olivier Hanotte^{4,5} and Armin O. Schmitt^{1,2*}

¹ Breeding Informatics Group, Department of Animal Sciences, University of Göttingen, Göttingen, Germany, ² Center for Integrated Breeding Research (CiBreed), University of Göttingen, Göttingen, Germany, ³ Animal Biosciences, National Program Coordinator for African Dairy Genetic Gain, International Livestock Research Institute (ILRI), Addis Ababa, Ethiopia, ⁴ Cells, Organisms and Molecular Genetics, School of Life Sciences, University of Nottingham, Nottingham, United Kingdom, ⁵ LiveGene, International Livestock Research Institute (ILRI), Addis Ababa, Ethiopia

OPEN ACCESS

Edited by:

Gábor Mészáros,
University of Natural Resources and
Life Sciences Vienna, Austria

Reviewed by:

Nina Moravčíková,
Slovak University of Agriculture,
Slovakia
Kieran G. Meade,
The Irish Agriculture and Food
Development Authority, Ireland
John B. Cole,
United States Department of
Agriculture (USDA), United States

*Correspondence:

Armin O. Schmitt
armin.schmitt@uni-goettingen.de

Specialty section:

This article was submitted to
Livestock Genomics,
a section of the journal
Frontiers in Genetics

Received: 03 June 2019

Accepted: 11 October 2019

Published: 14 November 2019

Citation:

Mekonnen YA, Gültas M, Effa K,
Hanotte O and Schmitt AO (2019)
Identification of Candidate Signature
Genes and Key Regulators
Associated With Trypanotolerance
in the Sheko Breed.
Front. Genet. 10:1095.
doi: 10.3389/fgene.2019.01095

African animal trypanosomiasis (AAT) is caused by a protozoan parasite that affects the health of livestock. Livestock production in Ethiopia is severely hampered by AAT and various controlling measures were not successful to eradicate the disease. AAT affects the indigenous breeds in varying degrees. However, the Sheko breed shows better trypanotolerance than other breeds. The tolerance attributes of Sheko are believed to be associated with its taurine genetic background but the genetic controls of these tolerance attributes of Sheko are not well understood. In order to investigate the level of taurine background in the genome, we compare the genome of Sheko with that of 11 other African breeds. We find that Sheko has an admixed genome composed of taurine and indicine ancestries. We apply three methods: (i) The integrated haplotype score (*iHS*), (ii) the standardized log ratio of integrated site specific extended haplotype homozygosity between populations (*Rsb*), and (iii) the composite likelihood ratio (CLR) method to discover selective sweeps in the Sheko genome. We identify 99 genomic regions harboring 364 signature genes in Sheko. Out of the signature genes, 15 genes are selected based on their biological importance described in the literature. We also identify 13 overrepresented pathways and 10 master regulators in Sheko using the TRANSPATH database in the geneXplain platform. Most of the pathways are related with oxidative stress responses indicating a possible selection response against the induction of oxidative stress following trypanosomiasis infection in Sheko. Furthermore, we present for the first time the importance of master regulators involved in trypanotolerance not only for the Sheko breed but also in the context of cattle genomics. Our finding shows that the master regulator Caspase is a key protease which plays a major role for the emergence of adaptive immunity in harmony with the other master regulators. These results suggest that designing and implementing genetic intervention strategies is necessary to improve

the performance of susceptible animals. Moreover, the master regulatory analysis suggests potential candidate therapeutic targets for the development of new drugs for trypanosomiasis treatment.

Keywords: trypanosomiasis, trypanotolerant, selection signature, candidate signature genes, master regulators, overrepresented pathways

INTRODUCTION

Trypanosomiasis is a disease caused by uni-cellular protozoan parasites which affects the health of humans and livestock. In Africa, this disease is referred to as African animal trypanosomiasis (AAT) (Kristjanson et al., 1999; Shaw et al., 2014). AAT is the major livestock production constraint especially in sub-Saharan African countries. It is mainly caused by *Trypanosoma congolense*, *Trypanosoma vivax*, and *Trypanosoma brucei brucei* (Hoare, 1972; Abebe, 2005; Batista et al., 2011; Yaro et al., 2016). Particularly, *T. congolense* is the most frequent cause of livestock disease in this region (Naessens, 2006). The disease is transmitted from infected animals to healthy animals by tsetse fly as a vector (Welburn et al., 2016). The infected animal shows symptoms such as anemia (Murray et al., 1990; Naessens, 2006), neurological symptoms (Tuntasuvan et al., 1997; Giordani et al., 2016), reduced productivity, infertility, abortion (Barrett and Stanberry, 2009), listlessness, and emaciation (Nantulya, 1986; Batista et al., 2007; Steverding, 2008; Noyes et al., 2011). If not treated, it can lead to death (Kristjanson et al., 1999; Barrett and Stanberry, 2009; Giordani et al., 2016). Hence, this disease has a major economic impact that accounts for an estimated annual loss of US\$ 5 billion in sub-Saharan countries (Kristjanson et al., 1999; Giordani et al., 2016).

Ethiopia is located in the eastern part of the tsetse belt. The tsetse fly distribution in the country spans from the south western to the north western regions covering 22,000 km² between longitude 38° and 38° East and latitude 5° and 12° North along river basins (Andrew, 2004; NTTICC, 2004). About 14 million cattle, 7 million horses, 1.8 million camels, and 14 million small ruminants are kept in the infection zone (MoARD, 2004). AAT severely affects the draft power as well as meat and milk production of the animals (Chanie et al., 2013). Therefore, AAT is considered as a major challenge constraining the path toward ensuring food security and combating poverty in this region (Meyer et al., 2018).

Until now, a number of methods have been applied to control the spread of this disease such as trypanocidal drugs, insect traps, and insecticides (Slingenbergh, 1992; Leak et al., 1996; Giordani et al., 2016). But none of these controlling measures has been successful to eradicate the disease. The current situation is deteriorating because of the trypanocidal drug resistance due to inappropriate drug usage. Moreover, pharmaceutical companies are less attracted to invest in new drug discovery and development due to high cost (Codjia et al., 1993; Mulugeta et al., 1997; Kristjanson et al., 1999; Naula and Burchmore, 2003). To control the spread of this disease, Lutje et al. (1996) have suggested a cross breeding

strategy between trypanotolerant and trypanosusceptible cattle, together with vector control. Accordingly, Hanotte et al. (2003) performed crossbreeding between the trypanotolerant N'Dama and trypanosusceptible Boran breeds to produce an F₂ population that shows heterosis. This led to the assumption that an F₂ cross between trypanotolerant and susceptible breeds could produce a trypanotolerant synthetic breed whose performance would exceed that of either parent. Consequently, marker assisted selection from the F₂ breed would be the most promising strategy to produce a breed that combines high production and trypanotolerance (Hanotte et al., 2003; Noyes et al., 2011).

In Ethiopia, Sheko shows better trypanotolerance attributes than other breeds such as Abigar and Horro (Lemecha et al., 2006). Sheko is found in the southern region of the Bench Maji Zone, the adjoining areas of Keffa and Shaka and is considered as an endangered breed due to extensive interbreeding with local indicine and sanga breeds (DAGRIS, 2007). Sheko cattle are kept in the tsetse infested regions likely explaining their degree of trypanotolerance (Hanotte et al., 2003; Bahbahani et al., 2018). In order to address the tolerance attributes of the Sheko breed at the molecular level, this study analyzes the genotyping data of the breed to explore the genome for candidate signature genes. The rationale is that natural or artificial selection targets the genome in response to environmental pressures or stresses as shaping adaptation and evolution. This implies that if the new allele of a mutation is beneficial (increases the fitness of their carriers) under certain environmental pressure or stress, then the frequency of these alleles will rapidly increase in the population (Charlesworth, 2007). Under positive selection, strong and long range linkage disequilibrium (LD) and unexpectedly high local haplotype homozygosity might occur in the genome (Gautier and Vitalis, 2012; Bomba et al., 2015).

Likewise, trypanosomiasis is considered as an environmental pressure which plays a major role to create selection signatures in the genome and which is thus leading to breed formation (Kristjanson et al., 1999; Abebe, 2005; Yaro et al., 2016). These signs or traces of selection in the genome could be detected by using a “bottom-up” or a “from genotype to phenotype” approach (McGuire and McGuire, 2008). This study provides traces or signs of positive selection in the genome of Sheko against trypanosomiasis using the “bottom-up” approach. In response to trypanosomiasis as the environmental pressure, the genome of Sheko could undergo changes at the molecular level. With the aim to identify the molecular mechanism of Sheko tolerance, we use extended haplotype homozygosity (EHH; *iHS* and *Rsb*) and spatial distribution of allele frequency [composite

likelihood ratio (CLR)] based methods to identify genes that are associated with this selection pressure in the Sheko breed. Combining methods for the detection of selection signature regions has been suggested as a means of increasing the power of the study compared to single analysis (e.g. Ma et al., 2015; Vatsiou et al., 2016).

Summary of the Analysis Workflow

Our workflow can be divided into two major steps as described below (see also **Figure 1**): 1) We analyzed the genetic relationship and structure of Sheko and 11 other indigenous African breeds using Plink 1.9 and the ADMIXTURE 1.3 software. 2) The identified candidate signature genes were then used in the analysis pipeline comprising the following four sub-steps: i) First, we identified genomic regions and signature genes under positive selection toward trypanotolerance in Sheko using *iHS*, CLR and *Rsb* analyses. As an intermediate result, we present the 15 genes resulting from a literature survey; ii) in the second step, we applied enrichment analysis in gene ontology (GO) terms in the combined gene sets of the three methods and made clusters of enriched GO terms in the form of a treemap using the geneXplain platform; iii) we then identified overrepresented pathways based upon the significant genes found in (ii) using the TRANSPATH database in the geneXplain platform; iv) finally, we identified the master regulators 10 steps upstream in the regulatory hierarchy using the significant genes found in (ii) using the TRANSPATH database in the geneXplain platform.

RESULT AND DISCUSSION

The Genetic Relationship and Structure of Cattle Populations

In order to understand the genetic structure of Sheko in comparison with 11 other African breeds, principal component analysis (PCA) was used. The result shows that the first two principal components (PCs), which explain 30.3% and 4.6% of the total variation, distinguishes the African taurine (N'Dama and Muturu) from the African indicine breeds [Benshangul, Serere, Karamojong, East African Shorthorn Zebu (EASZ), Fogera, and Gindeberet] (**Figure 2A**). Moreover, the Sheko, Nganda, Ankole, and Nuer are positioned between the African taurine and the African indicine clusters. These breeds are close to the indicine cluster and thereby support the admixture of more indicine than taurine type genomes in these breeds. The PCA result also shows the highest level of genetic heterogeneity in the Nganda breed which might be caused by ongoing crossbreeding of Nganda with exotic breeds to enhance their productivity (Mwai et al., 2015). We also conducted PCA exclusively for indigenous Ethiopian breeds. The result shows that the Sheko and Nuer form separate groups while the indicine type breeds (Benshangul, Fogera and Gindeberet) form a cluster in both PCs (**Figure 2B**).

For the further understanding of the degree of admixture in the populations, the ADMIXTURE 1.3 (Alexander et al., 2009) software was used for $K = 2$ to 7 hypothetical ancestral populations (**Figure 3**). We start from two hypothetical ancestral populations with the aim to determine the degree of indicine and taurine genetic background in the cattle breeds.

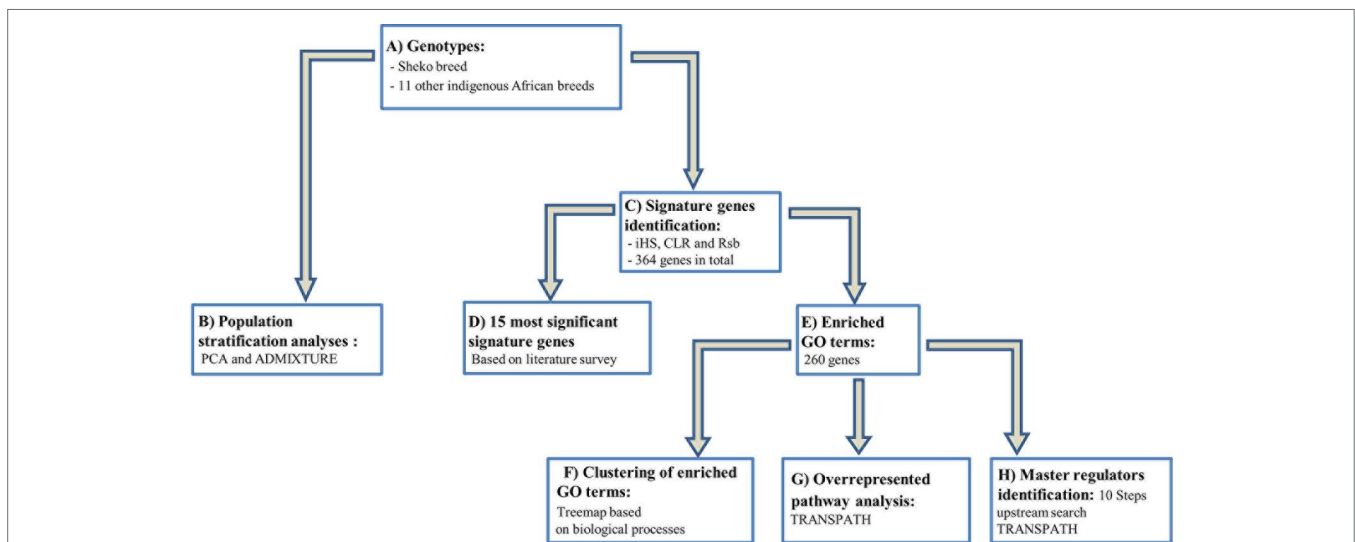


FIGURE 1 | Workflow for the study to identify candidate genes and key regulators that are associated with trypanotolerance in Sheko breed. **(A)** The genotypes of the Sheko and 11 other indigenous African breeds are obtained and quality control filtering is performed. **(B)** The genomic structure of Sheko in comparison to 11 other indigenous African breeds is analyzed using principal component analysis (PCA) and ADMIXTURE. **(C)** The identification of 364 signature genes is performed by *iHS*, CLR, and *Rsb* analyses. **(D)** Among 364 genes, the 15 most significant signature genes that are associated with trypanotolerance attributes are identified and disclosed. **(E)** Significantly functionally enriched terms [gene ontology (GO) terms] are identified for the 364 signature genes. 260 genes are identified as significantly enriched. **(F)** Using the functionally enriched 260 genes, a treemap is produced based on the biological processes. **(G)** Functionally enriched signature genes (260 genes) are analyzed to identify overrepresented pathways. **(H)** A master regulator network is generated up to 10 steps upstream using functionally enriched signature genes. The treemap, overrepresented pathway, and master regulator analyses were performed in the geneXplain platform.

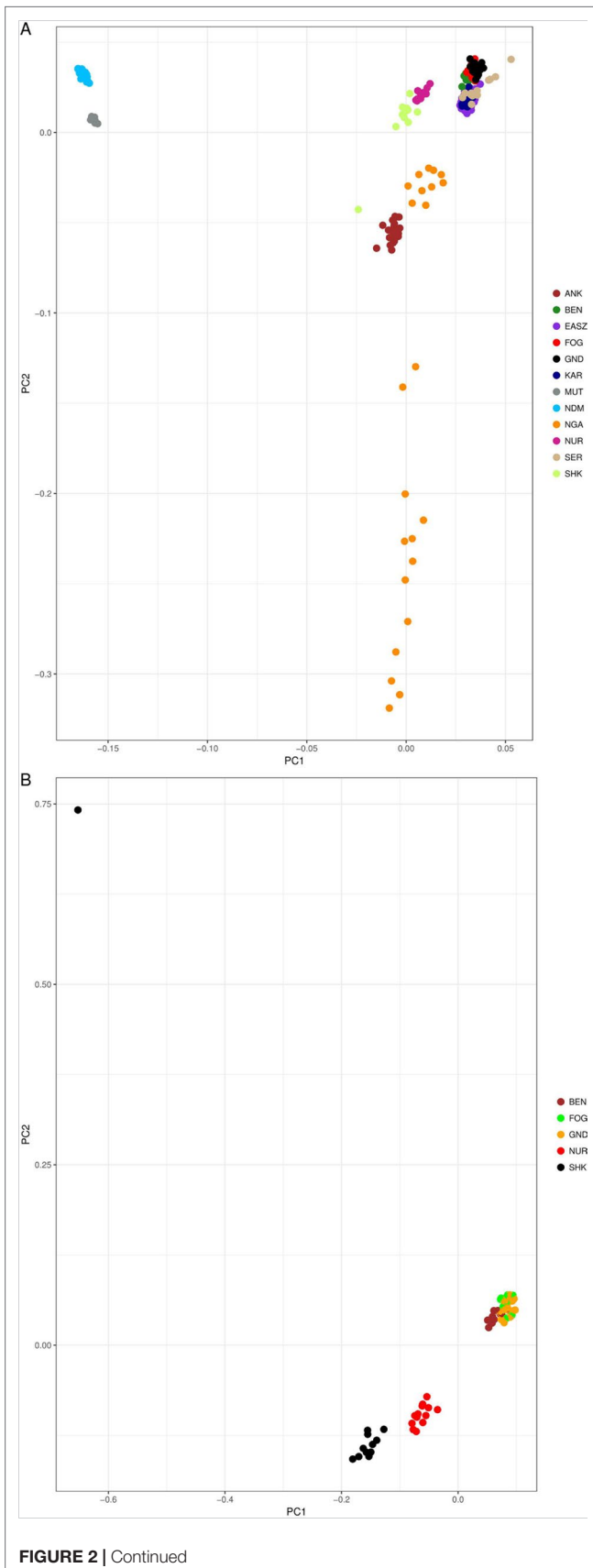
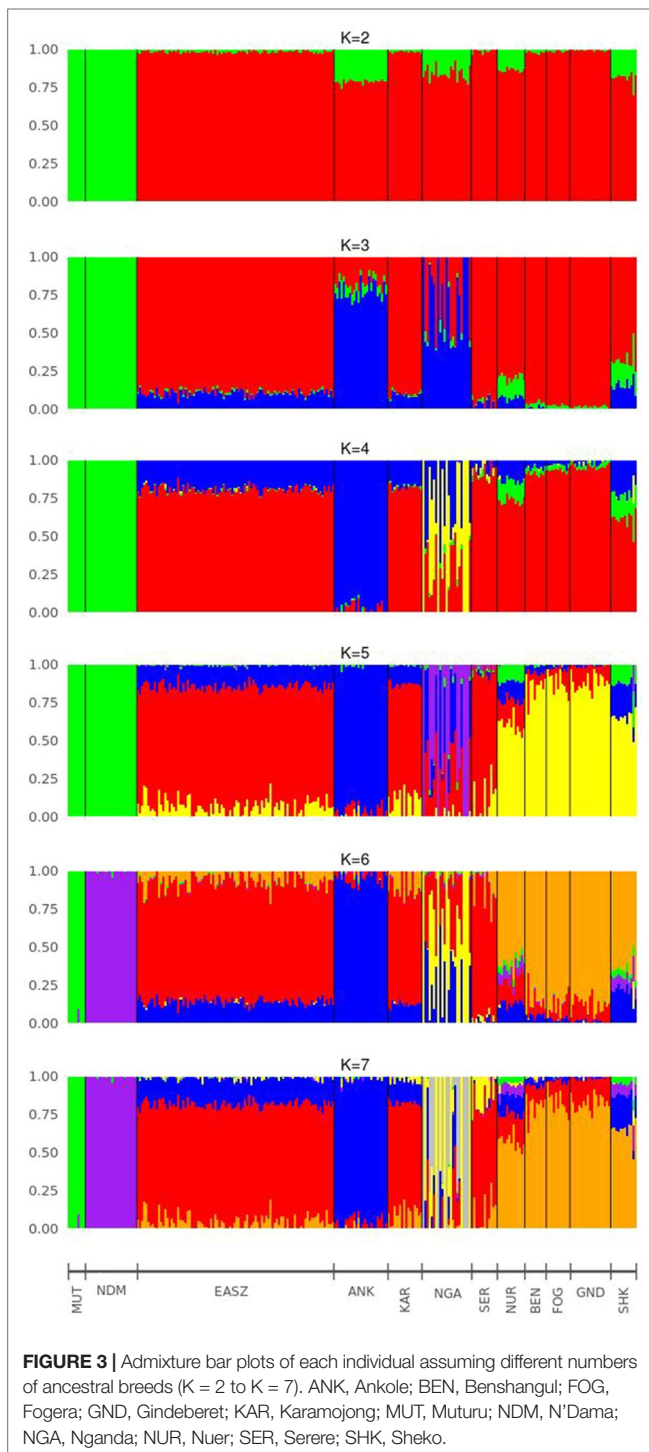


FIGURE 2 | PCA plots of the first two principal components showing the genetic relationship between cattle breeds. **(A)** PCA plot for all cattle breeds included in this study, and **(B)** PCA plot for the Ethiopian cattle breeds. ANK, Ankole; BEN, Benshangul; FOG, Fogera; GND, Gindeberet; KAR, Karamojong; MUT, Muturu; NDM, N'Dama; NGA, Nganda; NUR, Nuer; SER, Serere; SHK, Sheko.

Since the CV errors from $K = 3$ to $K = 6$ have not exceeded the cross-validation (CV) errors of $K = 2$, we extend the hypothetical population up to $K = 7$ which has the highest CV error (**Supplementary Figure 1**). At $K = 2$, the two ancestries taurine and indicine are revealed. The genomes of Ankole, Nganda, Nuer, and Sheko are mainly of indicine origin but have substantial taurine admixture, a result supporting our interpretation of the first PC of **Figure 2A**, that African taurine are separated from the East African indicine breeds and the mixed taurine-indicine type populations. At $K = 3$, Ankole, Nuer and Sheko show genetic heterogeneity with a considerable level of taurine admixture. EASZ, Karamojong, Serere, Benshangul, Fogera, and Gindeberet also show minor levels of taurine admixture whereas Nganda reveals a high level of within breed genetic differentiation. This is also in agreement with the second PC coordinate analysis in showing genetic heterogeneity within the cattle breeds (**Figure 2A**). Moreover, with the increment of the value of K , Sheko and Nuer show a higher level of genetic heterogeneity than the other east African breeds. Furthermore, at $K = 6$ and $K = 7$, the African taurine breeds N'Dama and Muturu show separate genetic backgrounds. In general, Sheko shows the highest level of African taurine genomic contribution for all values of K among East African breeds. The proportions of admixture in each of the analyzed breeds are presented for $K = 7$ in **Supplementary Table 1**.

Consistent with the previous findings and the origins of the genetic backgrounds of the cattle breeds worldwide (Mbole-Kariuki et al., 2014; Bahbahani et al., 2018), $K = 2$ highlights the ancient divergence between indicine and taurine cattle. However, the three optimal genetic clusters suggested by the minimal CV error (**Supplementary Figure 1**) reflect the common genetic background unique to East Africa besides taurine and indicine ancestral genetic admixture. In agreement with our study, Bahbahani et al. (2018) reported east African genetic background unique to East African cattle breeds. Moreover, the admixture plots show two individuals of Sheko with a high level of taurine introgression. One of these individuals with higher taurine introgression is also detected by the PCA (**Figure 1B**, upper left corner). This could be due to the recent crossbreeding of Sheko with European dairy breeds. There were similar observations in Butana, and it was speculated that farmers might have been involved in the crossbreeding with European dairy breeds in order to increase milk production (Bahbahani et al., 2018). We believe that the introgression of the European dairy breeds into the genome of indigenous breeds such as Sheko and Butana might distort their adaptive evolutionary responses against their natural environmental stresses. In this regard, future studies should assess the impact of European dairy breeds on the genome of



the indigenous African breeds with respect to their natural adaptation and tolerance attributes.

It is believed that the taurine background of the Sheko is linked to its trypanotolerance characteristics (Lemecha et al., 2006; Gibbs et al., 2009). This taurine admixture is likely a legacy of the first taurine occurrence on the African continent (Hanotte et al., 2000; Salim et al., 2014). A study on mtDNA indicates that all African cattle breeds analyzed so far carried taurine mtDNA haplotypes

which suggests that these waves of indicine arrival into Africa were male-mediated (Bradley et al., 1996; Bonfiglio et al., 2012).

Identification of Candidate Signature Genes Associated With Trypanotolerance

A total of 20, 14, and 65 genomic regions harboring 109, 64, and 202 candidate signature genes were identified by *iHS*, CLR, and *Rsb* analyses in 22, 10, and 27 autosomes in Sheko, respectively (Figure 4 and Supplementary Tables 2–4). Among the 364 unique candidate signature genes identified by *iHS*, CLR, and *Rsb* analyses, 260 disposed of enriched GO terms ($\alpha = 0.05$) (Supplementary Tables 5–7). Moreover, 96, 323, and 463 intergenic variants were identified in gene desert regions by *iHS*, CLR, and *Rsb* analyses in all candidate regions, respectively (Supplementary Tables 8–10).

Mainly focusing on the top 10 candidate signature genes of each of the three methods, we performed a literature survey and identified 15 (4 genes identified by *iHS*, 3 genes identified by CLR, 7 genes identified by *Rsb*, and 1 gene identified by both *iHS* and CLR) candidate signature genes that are associated with trypanotolerant attributes which have been reported in previous studies (Table 1). Notably, polymorphisms in or nearby the *MIGA1*, *CDAN1*, *HSPA9*, and *PCSK6* genes in the genome of Sheko might be associated with the evolutionary response against anemia. The *MIGA1* gene is associated with iron deficiency anemia and immunity (Moura et al., 2001; Rouault, 2006). This gene also plays a major role for the development and proliferation of lymphocyte since defective T- and B-cell activation is caused by inadequate iron uptake (Rouault, 2006; Jabara et al., 2016). Another interesting candidate signature gene related with anemia is *CDAN1*. Polymorphisms in this gene are associated with congenital dyserythropoietic anemia type 1 (Dgany et al., 2002; Renella et al., 2011). Moreover, the *hsp70* protein family and the heat shock 70kDa protein 9 (*HSPA9*) gene play a role as a downstream mediator of erythropoietin signaling and contribute to normal erythropoiesis (Singh et al., 1997; Ran et al., 2000; Ohtsuka et al., 2007; Chen et al., 2011). The mutation in this gene is associated with sideroblastic anemia (Schmitz-Abe et al., 2015), while the *PCSK6* gene is involved in iron homeostasis and hence related with iron deficiency anemia (Guillemot and Seidah, 2015). In agreement with our findings, it has been reported by several studies that trypanotolerant N'Dama do better control anemia, a process mediated by hematopoietic cells differentiation, than trypanosusceptible breeds (Berthier et al., 2016; Naessens, 2006).

In previous studies, trypanotolerant animals were reported to switch from innate immune response to adaptive immune response with the induction of active macrophages (M2) following trypanosome infection (Stijlemans et al., 2010; Bosschaerts et al., 2011). For instance, humoral response differences between trypanosusceptible (Boran) and trypanotolerant (N'Dama) cattle corresponding to the amount of antibody (Ab) titers have been observed. There is a difference in trypanosome-specific antiparasite Ab secreting cells in spleen and B cell activation between trypanotolerant and trypanosusceptible cattle (La Greca et al., 2014; Mamoudou et al., 2016; Morrison et al., 2016). In agreement with this, we identified the *SPAG11B*, *RAET1G*, *PPP1R14C*, and

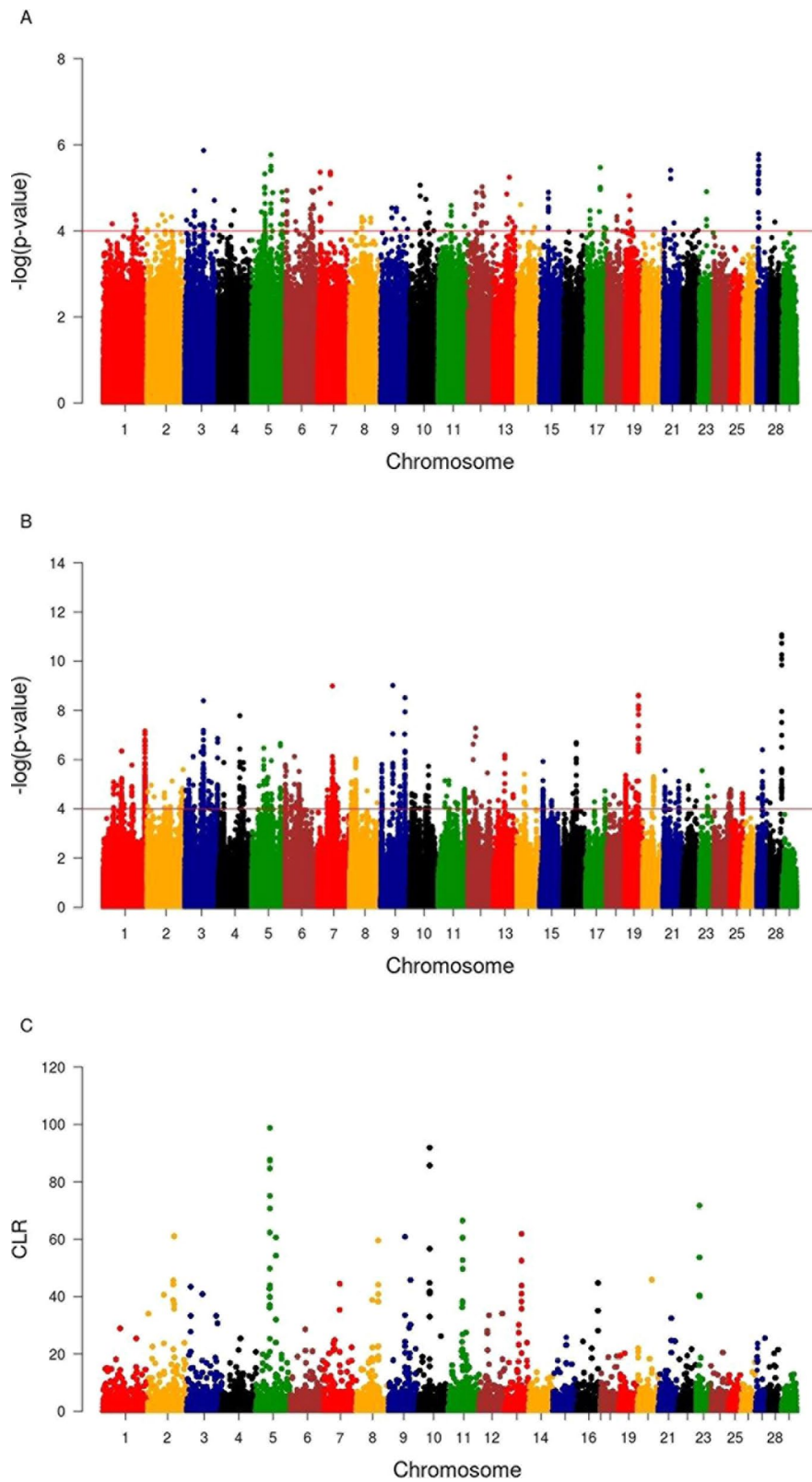


FIGURE 4 | Manhattan plots of genome-wide *iHS* (A), *Rsb* (B), and CLR (C) analyses. The x-axis shows the autosomal chromosomes and the y-axis shows $-\log$ transformed *P*-values (A and B) and CLR values (C).

TABLE 1 | Summary of major candidate signature regions identified by CLR, *iHS*, and *Rsb* analyses.

Genes	Method	CHR	Association	Position (UMD3.1) Start-End (bp)
MIGA1	Rsb	3	Anemia, immune tolerance and neurological dysfunction (Moura et al., 2001; Rouault, 2006; Jabara et al., 2016)	6706504–67137909
CDAN1	CLR	10	Anemia (Dgany et al., 2002; Renella et al., 2011)	38138863–38151656
HSPA9	Rsb	7	Anemia (Singh et al., 1997; Ran et al., 2000; Ohtsuka et al., 2007; Chen et al., 2011; Schmitz-Abe et al., 2015)	51506219–51521515
PCSK6	iHS	21	Anemia (Guillemot and Seidah, 2015)	29553201–29673109
SPAG11B	iHS	27	Immune tolerance (Yang et al., 1999; Ganz, 2003)	4920083–4942958
RAET1G	Rsb	9	Immune tolerance (Eagle and Trowsdale, 2007; Tomasec et al., 2007; Lanier, 2015)	88232044–88408862
PPP1R14C	Rsb	9	Immune tolerance and neurological dysfunction (Hanke et al., 1996; Linnekin et al., 1997; Liu et al., 2002; Haynes et al., 2003; Vignali et al., 2008)	88384683–88500749
TTC3	Rsb	1	Immune tolerance and neurological dysfunction (Chen et al., 2003; Liu et al., 2009; Pulst, 2016)	151034217–151141015
ERN1	Rsb	19	Immune tolerance and neurological dysfunction (Leach and Treacher, 1998; Oosthuysen et al., 2001; Rius et al., 2008; Liao et al., 2009; Minchenko et al., 2015; Singh et al., 2016)	48924511–48971838
CAPG	CLR	11	Immune tolerance and neurological dysfunction (Leach and Treacher, 1998; Oosthuysen et al., 2001; Rius et al., 2008; Liao et al., 2009; Zhang et al., 2009; Minchenko et al., 2015; Singh et al., 2016)	49423731–49438680
TTBK2	CRL	10	Neurological dysfunction (Jackson, 2012; Matilla-Duenas, 2012)	38159317–38248606
POLR3B	iHS	5	Neurological dysfunction (Schiffmann and van der Knaap, 2009; Daoud et al., 2013)	70062608–70178439
GNAS	iHS and CLR	13	Neurological dysfunction (Tuntasuvan et al., 1997; Bastepe, 2008; Giordani et al., 2016)	58010287–58049012
CHAT	Rsb	28	Listlessness (Johnson et al., 2016)	44143245–44187239
AP1M1	iHS	7	Listlessness (Molenaar et al., 1982)	7820650–7850254

TTC3 genes which are involved in immune tolerance in Sheko. Interestingly, the PPP1R14C gene could play an important role in the tolerance mechanisms of Sheko with PP1, a competitive inhibitor of ATP binding of Src tyrosine kinase family members (Hanke et al., 1996; Liu et al., 2002). The inhibition of Src kinase is associated with the termination of stem cell factor induced proliferation of hemopoietic cells (Linnekin et al., 1997). It was also reported that Src kinases are involved as a primary activator of AKT (serine/threonine kinase family). AKT plays a critical role in adaptive immunity through the inhibition of regulatory T-cells (T_{reg} cells), which could play a key role in maintaining the immune tolerance (Liu et al., 2002; Haynes et al., 2003; Vignali et al., 2008). In addition, activated AKT is a mediator of neuronal cell survival (Liu et al., 2002; Chen et al., 2003; Pulst, 2016).

Furthermore, the TTC3 gene is also involved in the regulation of AKT signaling and is related with immune tolerance and neuronal cell survival (Chen et al., 2003; Liu et al., 2009; Pulst, 2016). Therefore, the mutation in the PPP1R14C gene is associated with three tolerance attributes (immune tolerance, neurological dysfunction, and anemia). Remarkably, the candidate signature gene RAET1G is one of the few genes that could encode a ligand recognized by NKG2D proteins in response to stress and infections (Eagle and Trowsdale, 2007; Tomasec et al., 2007; Lanier, 2015). Furthermore, the isoforms of the SPAG11B gene encode defensin-like peptides which are expressed by phagocytic cells (Yang et al., 1999). These structurally diverse peptides make multimeric forms during infection and disrupt the membrane of the pathogen (Ganz, 2003). They are also involved in the recruitment of T- and dendritic cells to facilitate the adaptive immunity (Yang et al., 1999). Therefore, the mutations or the differential expression of these genes are critical for the immune tolerance of Sheko to combat anemia and neurological dysfunction caused by trypanosome infection.

Trypanosomiasis is also reported to affect the nervous system of the animal. Fatihu et al. (2009) and Allam et al. (2011) reported causes of thyroid and parathyroid gland dysfunction following trypanosome infection in cattle. The dysfunctioning of thyroid and parathyroid glands often result in neurological complications or cerebral pathology (Jaggy et al., 1994; Wu and Hersh, 1994). Therefore, mutations in the POLR3B, MIGA1, TTC3, ERN1, CAPG, GNAS, and TTBK2 genes might be associated with the response to the presence of the parasite in the brain white matter, cerebral fluid, thyroid, and parathyroid glands. The endoplasmic reticulum to nucleus signaling 1 (ERN1) and capping protein gelsolin-like (CAPG) genes are involved in the regulation of hypoxia (a state of a cell with inadequate or reduced oxygen availability) (Leach and Treacher, 1998). The reduction of the hypoxic response element in the spinal cord results in the progressive degradation of the motor neuron (Oosthuysen et al., 2001; Minchenko et al., 2015). Therefore, mutations in the ERN1 and CAPG genes are associated with neurological dysfunction (Liao et al., 2009; Minchenko et al., 2015). The ERN1 and CAPG genes might also be involved in the innate immune response since hypoxia triggers innate immunity responses through the activation of the hypoxia induced factor $\alpha 1$ (HIF-1 α) (Oosthuysen et al., 2001; Rius et al., 2008; Singh et al., 2016).

Trypanosome parasites are also known for their ability to manipulate the host immune responses. One of the mechanisms of innate immune evasion by these parasites is the reduction of HIF-1 α by indolepyruvate. Therefore, the reduction of hypoxic response elements in the spinal cord results in the progressive degradation of the motor neuron (Oosthuysen et al., 2001). Therefore, the mutation in the ERN1 and CAPG genes in particular would be related to the host innate immune evasion of the parasite. Another reported candidate signature gene related with neurological dysfunction is the TTBK2 gene. A mutation

in the *TTBK2* gene is associated with spinocerebellar ataxia which is a genetic syndrome causing progressive degeneration of the cerebellum and the spinal cord (Jackson, 2012; Matilla-Duenas, 2012). Moreover, a mutation in the *POLR3B* gene is associated with hypomyelinating leukodystrophy which is characterized by a deficiency in myelin deposition of the white matter of the brain (Schiffmann and van der Knaap, 2009; Daoud et al., 2013). In addition, the *POLR3B* gene is also involved in positive regulation of the interferon-beta production and the innate immune response (GO:0032728, GO:0045089).

Strikingly, a mutation in the *GNAS* gene is associated with pseudohypoparathyroidism which is characterized by a low level of calcium and a high phosphate level in the blood (Bastepe, 2008). Allam et al. (2011) reported a similar profile during trypanosome infection in cattle that could be associated with neurological dysfunction such as muscle spasm (Tuntasuvan et al., 1997; Bastepe, 2008; Giordani et al., 2016). Furthermore, during trypanosome infection, listlessness and emaciation are some of the clinical signs of the infection (Nantulya, 1986; Steverding, 2008; Noyes et al., 2011). These clinical signs might be associated with the destruction of the thyroid gland by trypanosome parasites in cattle (Fatihu et al., 2009). The candidate signature genes *AP1M1* and *CHAT* are related with these clinical signs. Most importantly, the *AP1M1* gene is a member of the adapter protein complex which is involved in thyroid abnormalities (Johnson et al., 2016). Due to the thyroid gland dysfunction (hypothyroidism), the nerves are unable to conduct electrical impulses properly. This leads to general weakness, lethargy, and listlessness (Jaggy et al., 1994). The mutation in the *CHAT* gene is associated with myasthenia gravis which is an autoimmune disease characterized by load dependent muscle weakness (Molenaar et al., 1982).

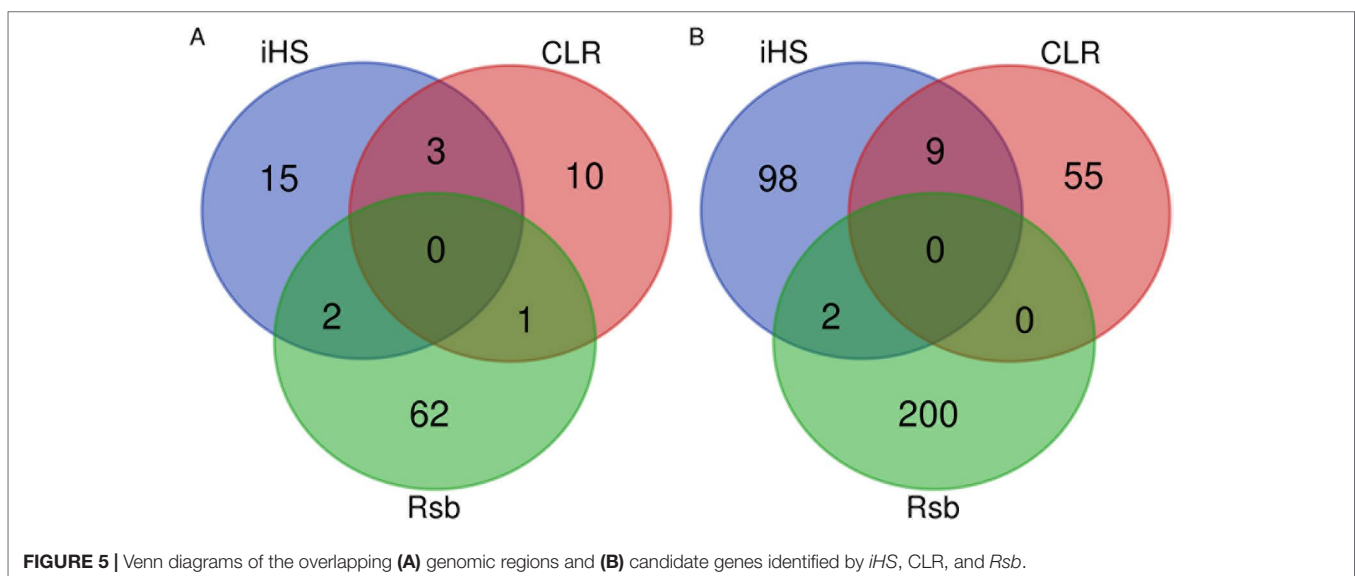
Our findings show strong selective sweeps (Figures 4A–C) in the genomic regions around the selected signature genes of Table 1 (Supplementary Table 11). This might indicate that the mutations in these genes have reached fixation or are near

fixation. Therefore, the identified candidate signature genes in Table 1 might play a major role in the natural tolerance attributes of Sheko against trypanosomiasis. Moreover, the comparison of candidate signature genes identified by the *iHS*, CLR, and *Rsb* methods show more overlaps between *iHS* and CLR than between *Rsb* and *iHS* or CLR analyses (Figures 5A, B), in agreement with *Rsb* being a powerful method to detect selection signature when the selected allele has reached fixation (Tang et al., 2007; Oleksyk et al., 2010; Bahbahani et al., 2018).

Among the 15 identified candidate signature genes (Table 1), the *MIGA1*, *RAETG*, and *PPP1R1AC* genes are not significantly functionally enriched ($\alpha = 0.05$). This might indicate that these candidate signature genes in Sheko could be specific to the environmental pressure in the region such as trypanosomiasis. Moreover, the identified signature regions of the three methods were compared with trypanotolerant QTL regions which were reported by Hanotte et al. (2003). Among the 55 trypanotolerant QTL, which were identified by crossing trypanotolerant N'Dama and susceptible Boran, 6 regions were overlapping with trypanotolerant QTL in N'Dama (Supplementary Table 12). Interestingly, among the identified candidate signature genes in Table 1, the *AP1M1* and *GNAS* genes are found in these overlapping regions. The overlapping regions and genes of Sheko and N'Dama might indicate occurrence of selection at the same genes in these two breeds against the same environmental pressures.

Functional Annotation of Candidate Signature Genes

In order to characterize the biological functions of functionally enriched candidate genes, a treemap was produced using the geneXplain platform (Krull et al., 2006). The treemap shows the clusters of 30 functional terms. Most of these terms are associated with cellular transport, metabolic processes and biological regulation (Figure 6). Interestingly, among the 30



enriched terms, two GO-terms are T-cell chemotaxis and cell-cell adhesion which play a critical role in the immune system (Springer, 1990; Gerard and Rollins, 2001; Bach et al., 2007). T-cell chemotaxis (chemoattractant cytokines) is a process that requires the movement of T-cells in response to a certain signal or external stimulus. The movement or circulation of immune cells in the blood and lymph as non-adherent cells and in tissues as adherent cells is critical for patrolling the body against infectious organisms effectively (Springer, 1990). For instance, β defensin is chemotactic for chemokine receptors of macrophages, natural killer cells, immature dendritic cells, and memory T-cells. Therefore, the recruitment of these cells to the site of a microbial invasion provides a link between innate and adaptive immunity (Yang et al., 1999). Likewise, T-cell mediated migration of thymocyte toward chemokines was observed following trypanosome infection in human (Mendes-da Cruz et al., 2006). In the presence of infectious organisms (foreign antigens), the immune cells aggregate at the site of the infection and through their adhesion receptors they adhere to cells bearing a foreign antigen (Springer, 1990).

Identification of Overrepresented Pathways in the Candidate Signature Gene Sets

Pathway analysis has become a powerful tool in order to refine the molecular mechanisms of disease tolerance. The rationale

of pathway analysis lies in the detection of overrepresentation of biologically defined pathways based upon the functionally enriched candidate selected genes. We performed pathway analysis using the TRANSPATH database on the geneXplain platform. The TRANSPATH pathway analysis identified 15 genes out of 260 functionally enriched genes that are involved in 13 overrepresented TRANSPATH pathways (Table 2). Among these genes, the immunoproteasome PSMD7 gene is involved in most of the overrepresented pathways. This gene is involved in the processes of presenting antigens by the major histocompatibility complex (MHC) class I proteins to CD8+ T-lymphocytes (Morrot and Zavala, 2004; Goldszmid and Sher, 2010; Jordan and Hunter, 2010). Sufficient induction of CD8+ during infection leads to pathogen elimination. It has been reported that immunoproteasome subunits are key determinants of the CD8+ T-cell level and quality involved in host resistance to trypanosomes infection (Ersching et al., 2016). This gene plays a critical role in the development of adaptive immunity or tolerance (Doolan and Hoffman, 1999).

However, adaptive immunity also plays a key role for the emergence of auto-immunity. Previous studies indicate that trypanosome infection could deplete thymocytes. As a result, immature T-lymphocytes are released from the thymic central tolerance and differentiate into mature T-helper cells in the lymph nodes (Flávia Nardy et al., 2015). This process would induce auto-immunity against self-antigens. Moreover, during trypanosome infection, the red blood cell membrane might be damaged by

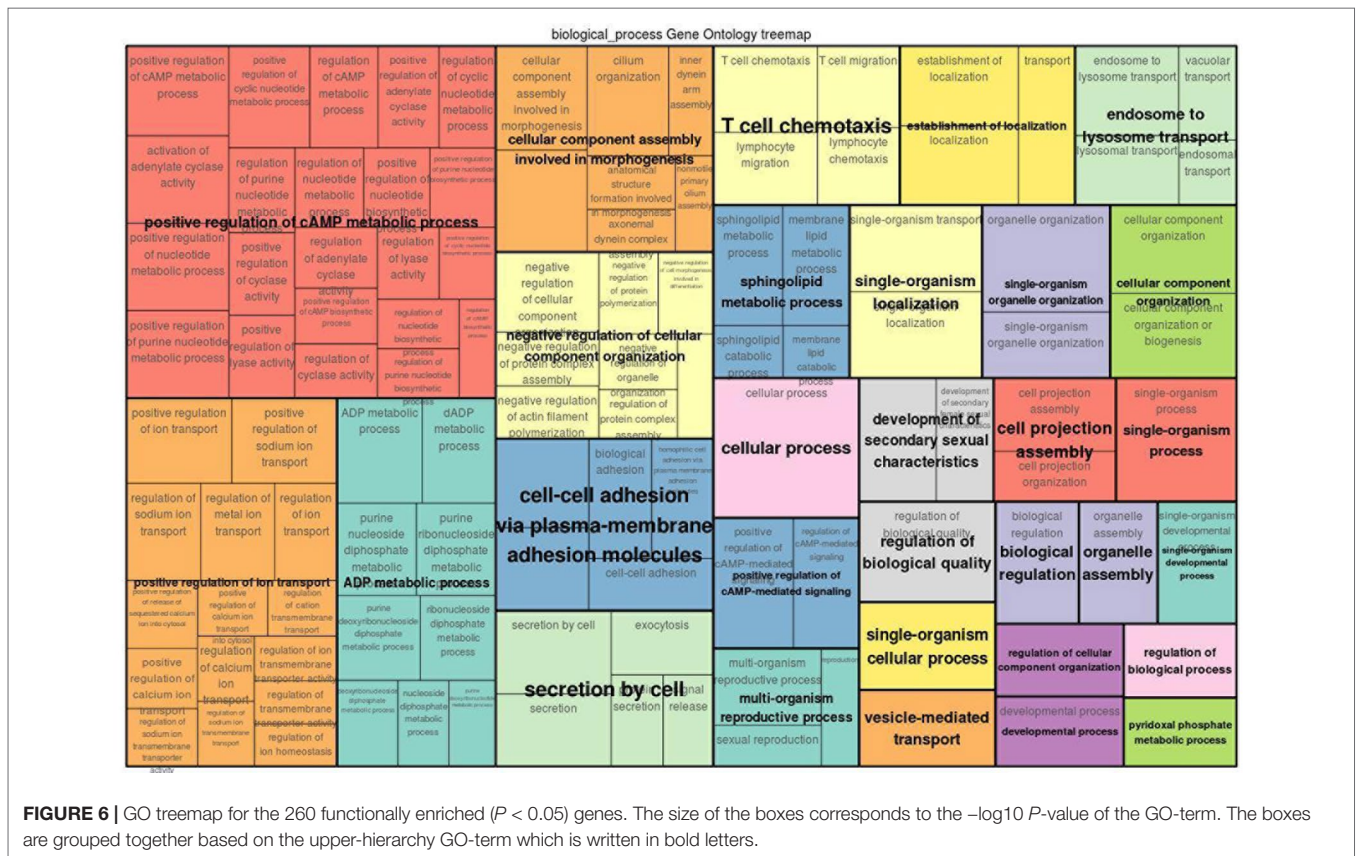


FIGURE 6 | GO treemap for the 260 functionally enriched ($P < 0.05$) genes. The size of the boxes corresponds to the $-\log_{10} P$ -value of the GO-term. The boxes are grouped together based on the upper-hierarchy GO-term which is written in bold letters.

TABLE 2 | Overrepresented pathways for the identified candidate signature genes.

Pathway	Raw P-value	Genes
PDGF B → STATs	0.003	STAT3, STAT5A
Stress-associated pathways	0.007	MBP, MEF2A, PSMD7, RAF1, RBX1, STAT3
E2F network	0.008	AKT3, CDC25C, PPP2R5A, PSMD7, RAF1, RBX1
G2/M phase (cyclin B:Cdk1)	0.015	AKT3, CDC25C, PSMD7, RBX1
IMP → ADP	0.025	AK5, AMPD3
ARIP1 → atrophin1	0.034	AKT3, APBA1
p38 pathway	0.039	MBP, MEF2A, STAT3
Plk1 cell cycle regulation	0.039	CDC25C, PSMD7, RBX1
IL-3 signaling	0.043	MBP, RAF1, STAT5A
Aurora-B cell cycle regulation	0.045	CENPE, PSMD7, RBX1
Oxygen independent HIF-1alpha degradation	0.045	PSMD7, RBX1, UBE2R2
Cul3 →/Nrf2	0.047	PSMD7, RBX1
S phase (Cdk2)	0.048	CDC25C, RAF1, RBX1

The names of the pathways are provided by the TRANSPATH database on the geneXplain platform.

parasite enzymes such as proteases or phospholipases. This could expose epitopes which are not recognized as self-antigens and would trigger immune-mediated hemolysis due to antibody response against these self-antigens (Taylor, 1998). This could be controlled by suppressing the development of auto-reactive immune cells through ubiquitination which is a degradative tag to be recognized by a proteasome complex such as PSMD7 (Lodish et al., 2004; Zinngrebe et al., 2014). Furthermore, some of the identified candidate signature genes are also associated with protein ubiquitination processes which might indicate that these genes are also involved in the functions described above (**Supplementary Tables 2–4**). To the best of our knowledge, our study is the first to show the potential of a molecular mechanism for controlling auto-reactive immune cells caused by trypanosomiasis in cattle. In agreement with our finding, Kierstein et al. (2006) reported that a trypanotolerant mouse strain showed overexpression of several genes encoding proteases.

In general, most of the overrepresented pathways (PDGFB → STATs, stress associated pathways, IMP → ADP, ARIP1 → atrophin 1, p38 pathway, IL-3 signaling, oxygen independent HIF-1alpha degradation and Cul3 →/Nrf2) pathways are activated by cellular stresses and antigens while others [E2F network, G2/M phase (cyclin B:Cdk1), S phase (Cdk2), Plk1 cell cycle regulation and Aurora-B cell cycle regulation] pathways are involved in cell cycle processes.

The first two pathways in **Table 2** (PDGFB → STATs and stress associated pathways) are related to the immune system and anemia. Especially, in stress associated pathways we find MBP, RAF1, MEF2A, and STAT3 genes that are involved in the immune and nervous systems. In the MBP gene, there are eight different mRNAs due to alternative splicing of exons (Zelenika et al., 1993). Three of the eight splice variants are expressed in the brain, macrophages and hemolymphopoietic tissues such as spleen, bone marrow, and thymus (Zelenika et al., 1993). This gene is also involved in the interleukin (IL)-3 signaling pathway.

IL-3 is a T-cell-derived hematopoiesis stimulating cytokine involved in the production, differentiation and function of granulocytes and macrophages (Ymer et al., 1985; Dorssers et al., 1987). This suggested that the expression of alternatively spliced MBP mRNAs is related with the immune system in response to trypanosome infection or the presence of a pathogen in the central nervous system. The serine/threonine kinase proto-oncogene RAF1 is also related with the stress associated pathway and is involved in inducing adaptive immunity by regulating the expression of cytokines that are important for the differentiation of T-helper cells (Gringhuis et al., 2009).

Moreover, STATs family members are also involved in the activation of various cytokines and in the promotion of cell survival by inducing the expression of antiapoptotic BCL2L1/BCL-X(L) genes (Benito et al., 1996; Packham et al., 1998; Yuan et al., 2004). For instance, STAT3 activation by trypanosomes was associated with the survival of cardiomyocytes during infection (Ponce et al., 2012; Stahl et al., 2013). The other gene involved in defense response is MEF2A which is associated with promoting antimicrobial peptide expression during infection (Clark et al., 2013). This gene is also involved in neuronal cell survival and loss of function (Gong et al., 2003; She et al., 2011). As reported by She et al. (2012), neurotoxins induce ubiquitination of MEF2A in response to toxic stress which leads to the loss of neuronal viability. Furthermore, He et al. (2015) reported that increased platelet-derived growth factor (PDGF)-B related signaling is associated with induced chemokine secretion which is a mediator of innate and adaptive immune responses (Kim and Broxmeyer, 1999). In addition, knock-out mice for PDGF-B develop anemia (Kaminski et al., 2001) which indicates that the PDGFB → STATs pathway is also involved in this disease.

The E2F network as well as the Cdk1 and Cdk2 related pathways are also associated with anemia which is the most prominent and consistent clinical sign of trypanosome infection (Kaminski et al., 2001; Dimova and Dyson, 2005; Noyes et al., 2011; Hu and Sun, 2016). The tumor suppressor retinoblastoma (Rb) is the inhibitor of E2Fs. When Rb binds to E2Fs, it prevents E2F mediated activation of transcriptional genes. In quiescent cells, E2F is required for the cell differentiation through a series of signal transduction cascades, including Cdks activation and phosphorylation. The Aurora-B and Plk1 pathways are involved in the activation and phosphorylation of Cdks, respectively. As a result of these and several other signaling cascades, E2Fs is activated while inactivating Rb. The activated E2F mediates quiescent cells for S phase entry and cell cycle progression (Dyson, 1998; Nevins, 1998; Trimarchi and Lees, 2002; Dimova and Dyson, 2005; Song et al., 2007). Hu et al. (2012) reported that mice deficient for both E2F8 (i.e., E2F gene family) and Rb show severe anemia.

Furthermore, the hypoxia inducible factor (HIF) and the nuclear factor-erythroid 2-related factor 2 (NRF2) pathways are related with anemia (Lee et al., 2004; Silva and Faustino, 2015). During hypoxia, HIF facilitates a high production of red blood cell (erythropoiesis) in order to overcome shortage of oxygen (Silva and Faustino, 2015). The other pathway, NRF2, regulates the expression of antioxidant responsive element-driven genes and plays a critical role in the antioxidant

responsive element-driven cellular protection (Cho et al., 2002). In addition, knockout mice for NRF2 show regenerative immune-mediated hemolytic anemia which indicates that this pathway is involved in erythrocyte maintenance during oxidative stress (Lee et al., 2004).

Intriguingly, serine/threonine kinase family isoforms of the AKT gene are involved in the E2F, Cdk1, IMP-ADP, and ARIP1-atrophia-1 pathways. This gene is activated in the host cells during trypanosome infection (Woolsey et al., 2003; Chuenkova and PereiraPerrin, 2009). The host kinase AKT promotes infected host cell survival and restricts the growth of intracellular parasites (Caradonna et al., 2013). AKT3 is also a key mediator of down stream signaling pathways of activated receptor tyrosine kinases which play a role in STAT3 activation (Yuan et al., 2004; Chuenkova and PereiraPerrin, 2009). The different isoforms of the kinase AKT regulate the development of immunity and autoimmunity. Zhang et al. (2013) reported that AKT is predominantly expressed in the innate immune cells. The isoforms of AKT are primarily involved in regulating inflammatory responses although it has been reported that AKT also modulates adaptive immune responses (Liu et al., 2002).

Moreover, the AKT related pathway Atrophia-1 plays a role in erythroid and lymphoid cell differentiation and in E3 ubiquitin ligase atrophia-1 interacting protein 4 (ITCH) signaling cascades. Atrophia-1 is involved in the regulation of immune responses through Notch-mediated signaling pathways (Qiu et al., 2000; You et al., 2009; Aki et al., 2015). It is also associated with spinocerebellar degeneration caused by extended CAG repeats encoding several glutamine units (polyglutamine tract) in the atrophia-1 protein (Kanazawa, 1999). The disease is characterized by neurological symptoms such as ataxia which is one of the clinical signs of trypanosome infection (Tuntasuvan et al., 1997; Suzuki and Yazawa, 2011; Giordani et al., 2016).

Further important pathways are p38, IMP → ADP, and the aurora B-cell cycle regulation pathways that are involved in the host defense mechanism. The p38 pathway is a MAPK-related pathway which is activated by various physical and chemical stresses, such as hypoxia and various cytokines. The activation of the p38 pathway is critical for normal immunity and inflammatory responses (Roux and Blenis, 2004). Moreover, the AK5 and AMPD3 genes are involved in the IMP → ADP pathway and play a central role in the regulation of inflammation and red blood cell homeostasis (Tavazzi et al., 2000; Mabley and Szabo, 2008). AK5 is associated with double positive thymocyte and auto-immunity regulation in the brain and pancreatic tissues (Stanojevic et al., 2008) while the AMPD3 gene is involved in the regulation of the energy state of red blood cells during oxidative stress (hypoxia) (Tavazzi et al., 2000). In addition to that, the aurora B-cell cycle regulation pathway is involved in the progression of T-lymphocytes which play a critical role for the development of innate and adaptive immunity (Song et al., 2007; Paul et al., 2011). To this end, the HIF and NRF2 related pathways are directly associated with the induction of host innate and adaptive immunity under oxidative stress (Singh et al., 1997; Cramer et al., 2003; Jantsch et al., 2011; McNamee et al., 2013; Battino et al., 2018).

In summary, our findings of the search for signature genes appear to be well substantiated by the results of the

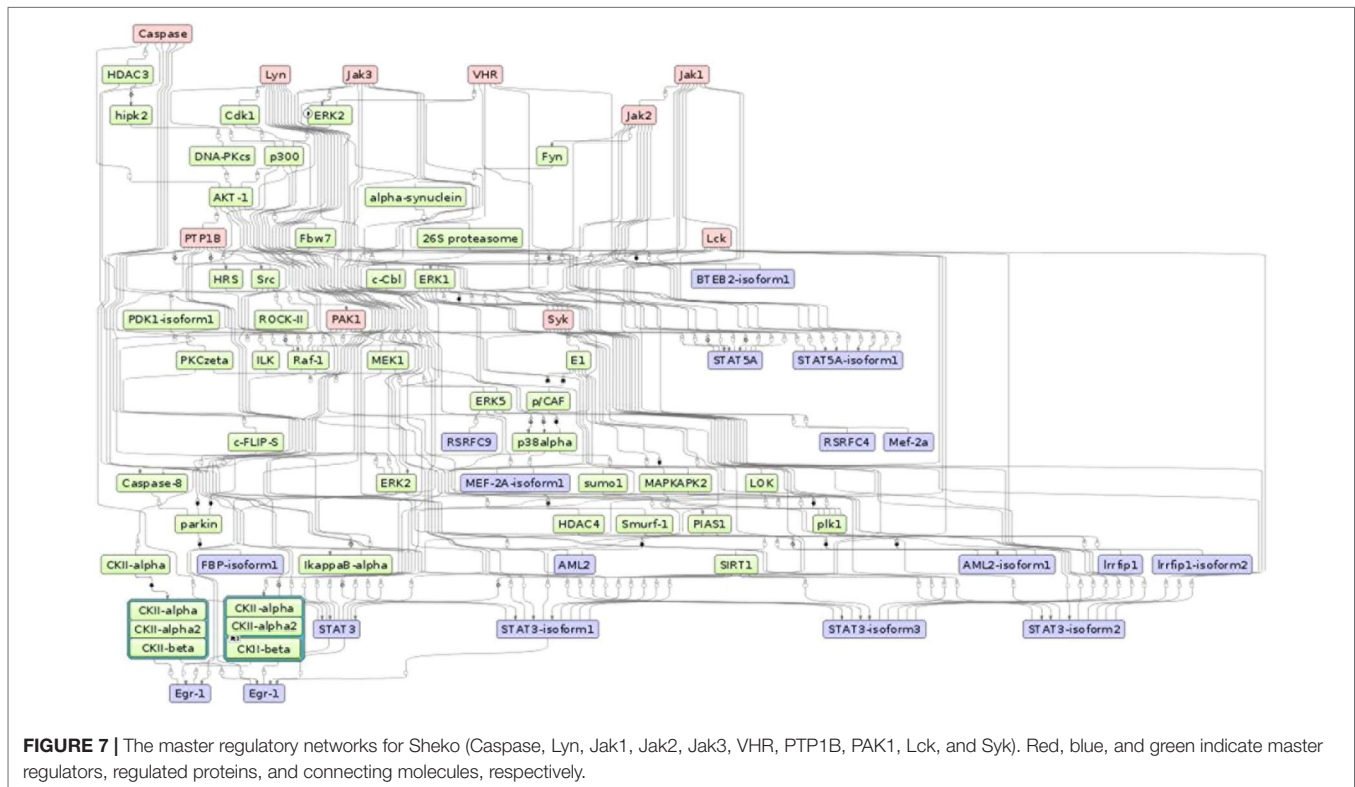
overrepresented pathways analysis. This implies that most of the overrepresented pathways are mainly associated with host defense mechanisms against pathogens and anemia. Particularly, stress-associated, HIF and NRF2 related pathways are involved in oxidative stress responses. Interestingly, trypanosome infection induces the production of superoxide, hydrogen peroxide, peroxy radicals, and hydroxyl radicals which are known to cause oxidative stress followed by tissue damage and hemolysis (Saleh et al., 2009). Under oxidative stress (hypoxia), erythrocytes are important mobile oxidative sinks (antioxidant) for themselves, other cells, and tissues. However, these properties of the red blood cells during oxidative stress contribute to its susceptibility toward hemolysis which leads to anemia (Chan et al., 2001; Sangokoya et al., 2010). In order to overcome the shortage of oxygen, stress-associated, HIF, and NRF2 related pathways play a critical role in the production of red blood cells in which hemoglobin acts as oxygen repository for red blood cells and other cells (Chan et al., 2001; Sangokoya et al., 2010; Silva and Faustino, 2015).

None of the most significant candidate signature genes (Table 1) was contained in the overrepresented pathway gene list (Table 2). This indicates that the candidate signature genes might be involved in the evolutionary gear particularly toward trypanotolerance in Sheko. For instance, candidate signature genes involved in the regulation of hypoxia (ERN1 and CAPG) are not identified in the overrepresented hypoxia related pathways. This might indicate that these candidate signature genes might be specific to oxidative stress tolerance attributes in Sheko. Hence, trypanotolerance of Sheko could be controlled by some major selected genes whose major effect close to fixation in the breed (become breed characteristic) and cohorts of genes with minor effects.

Identification of Master Regulators Based on Candidate Signature Genes

To gain more insight into the regulatory mechanisms of the identified candidate signature genes, we performed a master regulatory network analysis using the TRANSPATH database in the geneXplain platform. Applying the maximum radius of 10 steps upstream in the regulatory hierarchy, we identified ten master regulators (Figure 7). Remarkably, the master regulator Caspase, which is a family of protease enzymes, is associated mainly with regulating the reduction of the load of intracellular parasites, induction of nitric oxide production, increasing the level of CD4 and CD8+ T-cells, secretion of IFN γ , and control of trypanosome infection by macrophages (Gonçalves et al., 2013). This master regulator is involved in programmed cell death such as pyroptosis and necroptosis. These types of programmed cell deaths play a role for protecting an organism against oxidative stress (stress signals) and pathogenic attack (Shalini et al., 2015). In addition, Caspase also plays a role in the normal erythroid differentiation in the terminal stages (Zermati et al., 2001).

Most of the regulatory molecules (Syk, Lck, Lyn, Jak1, Jak2, and Jak3) are protein tyrosine kinases while others (VHR and PTP1B) are protein tyrosine phosphatases and activated kinase (PAK1). These master regulators are mainly associated with innate and adaptive immune responses and are critical for the



functioning of the nervous and immune systems. For instance, the activation of the regulatory molecule Syk requires the regulatory molecule Lck to phosphorylate immunoreceptor tyrosine-based activation motifs. Then, the phosphorylated immunoreceptor tyrosine-based activation motif modulates T-cell proliferation and differentiation by recruiting Syk protein tyrosine kinases (Acuto et al., 2008; Au-Yeung et al., 2009). In addition, coupling of the other master molecules JAK1 and JAK3 occurs on the cell surface receptor of IFN γ , followed by phosphorylation of the IFN γ receptor 1. This process leads to the activation of the STAT1 protein. The STAT1 protein binds to the target element of the IFN γ inducible gene in the nucleus and facilitates the transcription of the target regions during immunity responses (Rosenzweig and Holland, 2005; Casanova and Abel, 2007). Another reported regulator molecule VHR is also involved in the phosphorylation of STAT proteins and in the T-lymphocyte physiology (Alonso et al., 2001; Hoyt et al., 2007). Moreover, the master molecule JAK2 plays a critical role in the maintenance of hematopoiesis. It has been shown that selective deletion of JAK2 results in lethal anemia in adult mice (Grisouard et al., 2014).

Furthermore, a related master molecule, the protein tyrosine phosphatase 1B (PTP1B), is reported to modulate the activation of macrophages and plays a key role in mediating the central dendritic cell function of bridging innate and adaptive immunity (Heinonen et al., 2006; Martin-Granados et al., 2015). The kinase family regulator molecule Lyn is also involved in the regulation of innate and adaptive immune responses (Ingleby, 2012). Lyn is also known for mediating the production of type I interferone (IFN-I) which is involved in host defense mechanisms against invading

pathogens (Kawai and Akira, 2007; Blasius and Beutler, 2010; McNab et al., 2015). The related kinase regulatory molecule PAK1 is highly expressed in most leukocytes that are involved in immune responses. PAK1 also plays an important role in the activation of MAP-kinase pathways which are involved in all aspects of immune responses, from innate immunity to the activation of adaptive immune responses (Yi et al., 1991; Adachi et al., 1992; Zhang et al., 1995; Dong et al., 2002; Wang et al., 2002; Traves et al., 2014). In general, these proteins and master regulatory molecules are a large family of signaling enzymes that are expressed in various immune cells and regulate immune cell differentiation, cytokine production, and immune responses. Therefore, to maintain the tolerance against a pathogen, the regulation of these signaling pathways is critical (Manning et al., 2002; Salmond et al., 2009).

Strikingly, stress-induced protein kinases could also induce or aggravate auto-immunity by phosphorylating self-antigens to be recognized by auto-antibodies (Utz et al., 1997; Patterson et al., 2014). However, Caspase-mediated apoptosis plays an important role in arresting the development of auto-immunity by eliminating auto-reactive and pro-inflammatory cells (Eguchi, 2001). Moreover, the activation of Caspase and JAK2 is essential for the processes of erythroid differentiation and for the maintenance of hematopoiesis (Zermati et al., 2001). On the other hand, the inhibition of Caspase dependent mechanisms contributes to cell survival (Lamkanfi et al., 2007). We believe that the candidate signature genes involved in anemia, neurological dysfunction, listlessness, and immune tolerance might be governed by the top master regulator Caspase in harmony with other regulatory molecules. In general, our study provides a first report on the top

master regulators for trypanotolerance of Sheko and the overall analysis framework might be helpful to understand the underlying mechanisms of different cattle diseases in future works.

MATERIALS AND METHODS

SNP Genotyping and Quality Control

sDNA was extracted from 67 blood and tissue samples according to the QIAGEN DNA extraction protocol (Supplementary Table 13). 19 samples from Gindeberet, 12 from Sheko, 13 from Nuer, 12 from Benshangul and 11 from Fogera breeds were collected. All samples were taken randomly from unrelated animals based on the information given by livestock keepers at the time of sampling. All samples were genotyped for 777,962 SNPs using the Illumina BovineHD Genotyping Bead chip. In addition, the genotyping data of two west African breeds (24 N'Dama and 8 Muturu), and five east African breeds (92 EASZ, 25 Ankole, 16 Karamojong, 23 Nganda, and 12 Serere) were obtained from the International Livestock Research Institute (ILRI, Addis Ababa, Ethiopia; Bahbahani et al. (2017)). For quality control, Plink1.9 (Purcell et al., 2007) was used on 735,293 autosomal SNPs. SNPs with minor allele frequency of less than 1% were excluded (19,581 SNPs). Minimum genotyping call rate (<95%) and maximum identity-by-state (IBS) ($\geq 95\%$) were also used as filtering criteria. Two Benshangul samples failed the genotyping call rate criteria and were excluded from the analysis but no pair of samples was excluded due to the IBS filtering criterion. The total sample size for the down stream analysis consisted of 265 samples and 715,712 SNPs. BEAGLE 4 (Browning and Browning, 2007) was used for inferring haplotype phasing and imputing the missing alleles. The imputation was performed by fitting 83 sliding windows across the autosomes in which on average 8600 markers were included. With in each window 12 iterations were executed. Since our samples consist of indigenous African breeds, the total of 264 ($n - 1$) animals included in this study are used as a background to impute the missing alleles in the context of indigenous African cattle genome (i.e., without using the reference genome).

Genetic Background of the Cattle Population

In the eastern part of Africa, the mixture of African taurine and indicine cattle populations is common which reflects the wave of these two different ancestral aurochs in the region (Hanotte et al., 2000; Salim et al., 2014; Bahbahani et al., 2017). Regarding these two ancestral populations, the N'Dama and Muturu breeds are considered as African taurine whereas the Fogera, EASZ, Ankole, Karamojong, and Serere breeds are referred to as African zebu (Bahbahani et al., 2017). The Nuer and Ankole breeds are classified as African sanga (DAGRIS, 2007) while the Nganda breed is assigned to African zenga (Bahbahani et al., 2017). The sanga and zenga cattle are crossbreds between the indigenous humpless cattle and zebu. The latter have higher zebu genetic introgression than the former (Rege, 1999). Interestingly, the Sheko breed is

considered as the last oddments of the primordial *Bos taurus* cattle in eastern Africa. However, some animals in the present population of Sheko display small humps which indicates the genetic introgression of zebu cattle (DAGRIS, 2007). Yet, there is no research publication or documentation available on the genetic background of the Benshangul and Gindeberet breeds which are included in this study. The breed type and origin of the cattle samples included in this study are presented in Table 3.

Breed Differentiation, Genetic Relationship, and Structure

In order to understand the genomic structure of Sheko, we considered in total 12 indigenous African breeds genotyped with the Illumina BovineHD Genotyping BeadChip. To assess the within and between population genetic structure and admixture, PCA and admixture analyses were conducted. PCA was performed using Plink 1.9 to estimate the eigenvectors of the variance-standardized relationship matrix of all samples. In order to refine the genetic structure of the indigenous Ethiopian cattle breeds, separate PCA calculation were made for samples that were collected in Ethiopia (Sheko, Benshangul, Gindeberet, Fogera, and Nuer). Admixture analysis was performed using the ADMIXTURE 1.3 software with CV and 200 bootstraps for the hypothetical number of ancestries K ($2 \leq K \leq 7$). Both PCA and admixture analyses were used to determine the level of admixture and genetic differentiation of the populations. Furthermore, admixture analysis was used to determine the level of indicine and taurine ancestries of each breed at the genome-wide level. In particular, PCA and admixture analyses were performed to show the taurine background of Sheko.

Analysis of Signatures of Positive Selection

In general, methods for the detection of selection signatures are based on the spatial distribution of allele frequencies and the property of segregating haplotypes in the population (Hayes et al., 2010). As suggested by Ma et al. (2015) and

TABLE 3 | Cattle breeds included in the study.

Breed name	Breed category*	Breed origin
N'Dama	African taurine	Guinea
Muturu	African taurine	Nigeria
Ankole	Sanga	Uganda
Karamojong	African zebu	Uganda
Serere	African zebu	Uganda
Nganda	Zenga	Uganda
EASZ	African zebu	Kenya
Sheko	African taurine and zebu	Ethiopia
Nuer	Sanga	Ethiopia
Gindeberet	Not available	Ethiopia
Benshangul	Not available	Ethiopia
Fogera	African zebu	Ethiopia

*Breed category according to DAGRIS (2007).
EASZ, East African Shorthorn Zebu.

Vatsiou et al. (2016), combining these methods would help to reach a higher power than with single analysis. In this paper, we used EHH and spatial distribution of allele frequency-based methods to identify signatures of positive selection in the genome of the Sheko breed. This denotes that integrated haplotype score (*iHS*) and CLR analyses were performed on Sheko (12) while the ratio of site-specific EHH (EHHS) between populations (*Rsb*) analysis were performed between Sheko (12) and combined trypanosusceptible reference cattle populations (179) [EASZ (92) (Muhanguzi et al., 2014; Van Wyk et al., 2014), Ankole (25) (Magona et al., 2004), Karamojong (16) (Muhanguzi et al., 2017), Nganda (23) (FAO, 2004), Serere (12) (Ocaido et al., 2005) and Fogera (11) (Sinshaw et al., 2006)]. The results of these tests were combined into one gene set.

Extended Haplotype Homozygosity Based Methods

Rsb and *iHS* are LD based approaches which are implemented in R package *rehh*. Both *Rsb* and *iHS* are used to identify genome-wide signatures of selection (Gautier and Vitalis, 2012). These tests start with a core haplotype (i.e., a set of closely linked SNPs in which recombination does not take place) identification (Sabeti et al., 2002; Skipper, 2002). Then, the decay of LD as a function of the distance from the core haplotypes is analyzed (Sabeti et al., 2002). The *Rsb* analysis was performed between Sheko and the combined group of trypanosusceptible breeds. For each group, integrated site-specific EHH of each SNP (*iES*) was calculated. Standardized log-ratio between *iES* of the two groups was used to calculate *Rsb* values. The *iHS* values were calculated for Sheko as the natural log ratio of integrated EHH (*iHH*) between reference and alternative alleles for each SNP (Gautier and Vitalis, 2012; Bahbahani et al., 2018). The bovine reference genome (UMD3.1) is used as the reference allele while the study population (Sheko) is considered as the alternative allele. The *iHS* values were standardized based on the calculated mean and standard deviation values. This allows direct comparisons among different SNPs regardless of their allele frequencies (Gautier and Vitalis, 2012). For the standardization of *Rsb* values, median and standard deviation values were used. One-tailed Z-tests for *Rsb* and two-tailed Z-tests for *iHS* were applied on the standardized and normally distributed *Rsb* and *iHS* values (Supplementary Figures 2A, B) to identify statistically significant SNPs that are under positive selection. For one-tailed Z-tests, $P = 1 - \Phi(Rsb)$, whereas $P = 1 - 2|\Phi(iHS) - 0.5|$ was used for the two sided tests with Φ being the Gaussian cumulative density function. For both *Rsb* and *iHS* *P*-values, the significance threshold of $\alpha = 10^{-4}$ was applied following the study of Bahbahani et al. (2018) to identify candidate regions.

Spatial Distribution of Allele Frequency Based Method

The CLR test is an LD based selective sweep searching algorithms using the information from the spatial distribution of allele frequencies (Charlesworth, 2012). CLR is used to identify skewed patterns of the allele frequency spectrum

toward excess of rare alleles and high frequency alternative alleles due to the hitchhiking effect (Kim and Stephan, 2002; Nielsen et al., 2005; Qanbari et al., 2014). The *P*-values were calculated by the rank of the genome wide scan of CLR values. As suggested by Wilches et al. (2014), the 95th quantile of the distribution of the top CLR *P*-values was used to identify a significance threshold of $\alpha = 10^{-5}$ (Supplementary Figure 3). For CLR analysis, the Sweepfinder2 (DeGiorgio et al., 2016) software was used for each chromosome with a window size of 50kb including on average 226 SNPs per window. Sweepfinder2 estimates CLR in the context of background selection to identify sweeps (DeGiorgio et al., 2016; Huber et al., 2016).

Functional Annotation of Selected Candidate Regions

Genes found within 25 kb around the most significant SNP were considered as candidate genes (Bahbahani et al., 2018). Protein-coding and RNA genes found within the candidate regions were retrieved using the BioMart tool (Kinsella et al., 2011). The R package *Enrichr* (Kuleshov et al., 2016) was used to determine the candidate signature genes that are functionally enriched in GO terms with respect to the whole bovine reference genome background ($\alpha = 0.05$). These functionally enriched candidate signature genes were used to produce a treemap which shows clusters of functional terms based on the biological functions of the candidate signature genes.

To gain more insight into the functional properties and molecular mechanisms involved in trypanotolerance, overrepresented pathways were analyzed using the TRANSPATH database (Krull et al., 2006) of the geneXplain platform (<http://genexplain.com/>). Furthermore, to understand the regulatory mechanisms of the candidate signature genes and the signaling cascades in the regulatory hierarchy involved in trypanotolerance, the identification of master regulators was conducted using the TRANSPATH database.

CONCLUSION

For generations, African animal trypanosomiasis has been the major selection pressure in the region. We have identified the candidate causative genes, pathways, and master regulators associated with the adaptation of the Sheko breed to its natural environmental pressure. Most of the identified candidate signature genes, overrepresented pathways, and master regulator molecules were involved in immune tolerance, neurological dysfunction, and anemia. This entails that the genome of Sheko was targeted by these environmental pressures which are associated with trypanosomiasis. Therefore, this study helps as an input for designing and implementing genetic intervention strategies to improve the performance of susceptible as well as animals which are relatively tolerant toward higher trypanotolerance.

The improvement of the cattle health contributes to increase the production of milk and meat. The improvement of the cattle health enhances the draft power of the animal which is associated with increasing crop production. This implies that, increasing animal and crop production significantly contributes to eradicate poverty in the area. In general, this study contributes to the existing literature in two ways: 1) The genetic controls of Sheko against trypanosomiasis have not been well studied and this study examines the genomic signatures in response to trypanosomiasis in detail; 2) this study presents pathways and master regulators which could help to understand the upstream biological processes involved in trypanotolerance. Particularly, this study for the first time identifies the master regulators involved in the regulatory mechanisms of trypanotolerance in relation to signatures of selection not only for Sheko breed but also in the context of cattle genomics, which can be used for the development of effective new drugs. However, additional studies such as differential expressions of targeted genes and regulatory molecules may be required to further confirm the validity of the results reported in this paper.

DATA AVAILABILITY STATEMENT

The SNP data in this study can be found in the European Variation Archive (EVA): PRJEB34751.

ETHICS STATEMENT

Standard techniques were used to collect blood. The procedure was reviewed and approved by the University of Edinburgh Ethics Committee (reference number OS 03-06) and also by the Institute Animal Care and Use Committee of the International Livestock Research Institute, Nairobi.

AUTHOR CONTRIBUTIONS

YM, MG, and AS participated in the design of the study. YM conducted computational and statistical analyses as well as identified the signature genes. AS and MG supervised the computational and statistical analyses. YM interpreted the results with MG. YM carried out the literature survey and prepared the first draft of the manuscript. OH and KE were involved in the interpretation of the results. YM and KE were involved in collecting blood and tissue samples for this study. YM prepared the DNA samples. YM and AS were involved in the preparation of the genotyping data. YM and MG wrote the final version of the manuscript. YM, AS, and MG conceived and managed the project. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

We acknowledge support by the German Research Foundation and Open Access Publication Funds of the Göttingen University. We thank Henner Simianer for his helpful advise and insights at the early stages of this project and for comments on the methods. We would like to thank Tariku Abena for his support during the preparation of DNA samples. We would also like to thank Felix Heinrich and Faisal Ramzan for proofreading the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.01095/full#supplementary-material>

SUPPLEMENTARY FIGURE 1 | Cross validation error in relation to the number of hypothetical ancestral populations for the Admixture analyses.

SUPPLEMENTARY FIGURE 2 | Histogram of standardized *Rsb* and *iHS* values.

SUPPLEMENTARY FIGURE 3 | Box plot of CLR $-\log(P\text{-values})$.

SUPPLEMENTARY TABLE 1 | Proportion of admixture within each of the analyzed breeds.

SUPPLEMENTARY TABLE 2 | Functionally annotated gene list identified by *iHS* analysis.

SUPPLEMENTARY TABLE 3 | Functionally annotated gene list identified by CLR analysis.

SUPPLEMENTARY TABLE 4 | Functionally annotated gene list identified by *Rsb* analysis.

SUPPLEMENTARY TABLE 5 | Functionally enriched gene list identified by *iHS* analysis.

SUPPLEMENTARY TABLE 6 | Functionally enriched gene list identified by CLR analysis.

SUPPLEMENTARY TABLE 7 | Functionally enriched gene list identified by *Rsb* analysis.

SUPPLEMENTARY TABLE 8 | List of intergenic variants identified by *iHS* analysis.

SUPPLEMENTARY TABLE 9 | List of intergenic variants identified by CLR analysis.

SUPPLEMENTARY TABLE 10 | List of intergenic variants identified by *Rsb* analysis.

SUPPLEMENTARY TABLE 11 | Summary of the genomic regions identified by *iHS*, CLR and *Rsb*.

SUPPLEMENTARY TABLE 12 | QTL regions overlapping between Sheko and N'Dama.

SUPPLEMENTARY TABLE 13 | Summary of the blood and tissue samples collected from indigenous Ethiopian cattle breeds for DNA extraction.

REFERENCES

- Abebe, G. (2005). Current situation of Trypanosomiasis: In review article on: Trypanosomiasis in Ethiopia. *Ethiop. J. Biol. Sci.* 4, 75–121. doi: 10.4314/ejbs.v4i1.39017
- Acuto, O., Di Bartolo, V., and Michel, F. (2008). Tailoring T-cell receptor signals by proximal negative feedback mechanisms. *Nat. Rev. Immunol.* 8, 699–712. doi: 10.1038/nri2397
- Adachi, M., Sekiya, M., Arimura, Y., Takekawa, M., Itoh, F., and Hinoda, Y. (1992). Protein-tyrosine phosphatase expression in pre-B cell NALM-6. *Cancer Res.* 52, 737–740.
- Aki, D., Zhang, W., and Liu, Y. C. (2015). The E3 ligase Itch in immune regulation and beyond. *Immunol. Rev.* 266, 6–26. doi: 10.1111/imr.12301
- Alexander, D., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655–1664. doi: 10.1101/gr.094052.109
- Allam, L., Ogwu, D., Agbede, R., and Sackey, A. (2011). Hematological and serum biochemical changes in gilts experimentally infected with xperimentally infected with *Trypanosoma brucei*. *Vet. Arhiv.* 81, 597–609. doi: 10.1016/j.tvjl.2011.04.021
- Alonso, A., Saxena, M., Williams, S., and Mustelin, T. (2001). Inhibitory role for dual specificity phosphatase VHR in T cell antigen receptor and CD28-induced Erk and Jnk activation. *J. Biol. Chem.* 276, 4766–4771. doi: 10.1074/jbc.M006497200
- Andrew, A. H. (2004). *Bovine Medicine Diseases and Husbandry of Cattle*. 9600 Garsington Road, Oxford OX42DQ, UKs: Blackwell Science Ltd.
- Au-Yeung, B. B., Deindl, S., Hsu, L. Y., Palacios, E. H., Levin, S. E., and Kuriyan, J. (2009). The structure, regulation, and function of ZAP-70. *Immunol. Rev.* 228, 41–57. doi: 10.1111/j.1600-065X.2008.00753.x
- Bach, T. L., Chen, Q. M., Kerr, W. T., Wang, Y., Lian, L., and Choi, J. K. (2007). Phospholipase cbeta is critical for T cell chemotaxis. *J. Immunol.* 179, 2223–2227. doi: 10.4049/jimmunol.179.4.2223
- Bahbahani, H., Salim, B., Almuthen, F., Al Enezi, F., Mwacharo, J. M., and Hanotte, O. (2018). Signatures of positive selection in African Butana and Kenana dairy zebu cattle. *PLoS One* 13, e0190446. doi: 10.1371/journal.pone.0190446
- Bahbahani, H., Tijjani, A., Mukasa, C., Wragg, D., Almuthen, F., and Nash, O. (2017). Signatures of selection for environmental adaptation and zebu x taurine hybrid fitness in east african shorthorn zebu. *Front. Genet.* 8, 68. doi: 10.3389/fgene.2017.00068
- Barrett, A. D., and Stanberry, L. R. (2009). *Vaccines for biodefense and emerging and neglected diseases*. (London UK: Academic Press).
- Bastepe, M. (2008). The GNAS locus and pseudohypoparathyroidism. *Adv. Exp. Med. Biol.* 626, 27–40. doi: 10.1007/978-0-387-77576-0_3
- Batista, J. S., Riet-Correa, F., Teixeira, M. M., Madruga, C. R., Simoes, S. D., and Maia, T. F. (2007). Trypanosomiasis by *Trypanosoma vivax* in cattle in the Brazilian semi-arid: description of an outbreak and lesions in the nervous system. *Vet. Parasitol.* 143, 174–181. doi: 10.1016/j.vetpar.2006.08.017
- Batista, J. S., Rodrigues, C. M., Garcia, H. A., Bezerra, F. S., Olinda, R. G., and Teixeira, M. M. (2011). Association of *Trypanosoma vivax* in extracellular sites with central nervous system lesions and changes in cerebrospinal fluid in experimentally infected goats. *Vet. Res.* 42, 63. doi: 10.1186/1297-9716-42-63
- Battino, M., Giampieri, F., Pistollato, F., Sureda, A., de Oliveira, M. R., and Pittala, V. (2018). Nrf2 as regulator of innate immunity: a molecular Swiss army knife! *Biotechnol. Adv.* 36, 358–370. doi: 10.1016/j.biotechadv.2017.12.012
- Benito, A., Silva, M., Grillot, D., Nunez, G., and Fernandez-Luna, J. L. (1996). Apoptosis induced by erythroid differentiation of human leukemia cell lines is inhibited by Bcl-XL. *Blood* 87, 3837–3843. doi: 10.1182/blood.V87.9.3837.bloodjournal8793837
- Berthier, D., Brenière, S. F., Bras-Gonçalves, R., Lemesre, J.-L., Jamongne, V., and Solano, P. (2016). Tolerance to trypanosomatids: a threat, or a key for disease elimination?. *Trends In Parasitology* 32, 157–168. doi: 10.1016/j.pt.2015.11.001
- Blasius, A. L., and Beutler, B. (2010). Intracellular toll-like receptors. *Immun.* 32, 305–315. doi: 10.1016/j.immuni.2010.03.012
- Bomba, L., Nicolazzi, E. L., Milanese, M., Negrini, R., Mancini, G., and Biscarini, F. (2015). Relative extended haplotype homozygosity signals across breeds reveal dairy and beef specific signatures of selection. *Genet. Selection Evol.* 47, 25. doi: 10.1186/s12711-015-0113-9
- Bonfiglio, S., Ginja, C., De Gaetano, A., Achilli, A., Olivieri, A., and Colli, L. (2012). Origin and spread of *Bos taurus*: new clues from mitochondrial genomes belonging to haplogroup T1. *PLoS One* 7, e38601. doi: 10.1371/journal.pone.0038601
- Bosschaerts, T., Morias, Y., Stijlemans, B., Herin, M., Porta, C., and Sica, A. (2011). IL-10 limits production of pathogenic TNF by M1 myeloid cells through induction of nuclear NF- κ B p50 member in *Trypanosoma congolense* infection-resistant C57BL/6 mice. *Eur. J. Immunol.* 41, 3270–3280. doi: 10.1002/eji.201041307
- Bradley, D. G., MacHugh, D. E., Cunningham, P., and Loftus, R. T. (1996). Mitochondrial diversity and the origins of African and European cattle. *Proc. Natl. Acad. Sci. U.S.A.* 93, 5131–5135. doi: 10.1073/pnas.93.10.5131
- Browning, S. R., and Browning, B. L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* 81, 1084–1097. doi: 10.1086/521987
- Caradonna, K. L., Engel, J. C., Jacobi, D., Lee, C. H., and Burleigh, B. A. (2013). Host metabolism regulates intracellular growth of *Trypanosoma cruzi*. *Cell Host Microbe* 13, 108–117. doi: 10.1016/j.chom.2012.11.011
- Casanova, J. L., and Abel, L. (2007). Human genetics of infectious diseases: a unified theory. *EMBO J.* 26, 915–922. doi: 10.1038/sj.emboj.7601558
- Chan, K., Han, X. D., and Kan, Y. W. (2001). An important function of Nrf2 in combating oxidative stress: detoxification of acetaminophen. *Proc. Natl. Acad. Sci. U.S.A.* 98, 4611–4616. doi: 10.1073/pnas.081082098
- Chanie, M., Adula, D., and Bogale, B. (2013). Socio-economic assessment of the impacts of trypanosomiasis on cattle in girja district, southern oromia region, southern ethiopia. *Acta Parasitol. Globalis* 4, 80–85. doi: 10.5829/idosi.app.2013.4.3.7523
- Charlesworth, B. (2007). A hitch-hiking guide to the genome: a commentary on 'The hitch-hiking effect of a favourable gene' by John Maynard Smith and John Haigh. *Genet. Res.* 89, 389–390. doi: 10.1017/S0016672308009580
- Charlesworth, B. (2012). The effects of deleterious mutations on evolution at linked sites. *Genet.* 190, 5–22. doi: 10.1534/genetics.111.134288
- Chen, H. K., Fernandez-Funez, P., Acevedo, S. E., Lam, Y. C., Kaytor, M. D., and Fernandez, M. H. (2003). Interaction of Akt-phosphorylated ataxin-1 with 14-3-3 mediates neurodegeneration in spinocerebellar ataxia type 1. *Cell* 113, 457–468. doi: 10.1016/S0092-8674(03)00349-0
- Chen, T. H., Kambal, A., Krysiak, K., Walshauer, M. A., Raju, G., and Tibbitts, J. F. (2011). Knockdown of Hspa9, a del(5q31.2) gene, results in a decrease in hematopoietic progenitors in mice. *Blood* 117, 1530–1539. doi: 10.1182/blood-2010-06-293167
- Cho, H. Y., Jedlicka, A. E., Reddy, S. P., Kensler, T. W., Yamamoto, M., and Zhang, L. Y. (2002). Role of NRF2 in protection against hyperoxic lung injury in mice. *Am. J. Respir. Cell Mol. Biol.* 26, 175–182. doi: 10.1165/ajrcmb.26.2.4501
- Chuenkova, M. V., and PereiraPerrin, M. (2009). *Trypanosoma cruzi* targets Akt in host cells as an intracellular antiapoptotic strategy. *Sci. Signal* 2, ra74. doi: 10.1126/scisignal.2000374
- Clark, R. I., Tan, S. W. S., Péans, C. B., Roostalu, U., Vivancos, V., and Bronda, K. (2013). MEF2 Is an In Vivo Immune-Metabolic Switch. *Cell* 155, 435–447. doi: 10.1016/j.cell.2013.09.007
- Codjia, V., Mulatu, W., Majiwa, P. A., Leak, S. G., Rowlands, G. J., and Authie, E. (1993). Epidemiology of bovine trypanosomiasis in the Ghibe valley, southwest Ethiopia. 3. Occurrence of populations of *Trypanosoma congolense* resistant to diminazene, isometamidium and homidium. *Acta Trop.* 53, 151–163. doi: 10.1016/0001-706X(93)90026-8
- Cramer, T., Yamanishi, Y., Clausen, B. E., Forster, I., Pawlinski, R., and Mackman, N. (2003). HIF-1alpha is essential for myeloid cell-mediated inflammation. *Cell* 112, 645–657. doi: 10.1016/S0092-8674(03)00154-5
- DAGRIS. (2007). *Domestic Animal Genetic Resources Information System (DAGRIS)*. Rege, JHanotte, OMamo, YAsrat, BDessie, T. Addis Ababa, Ethiopia: International Livestock Research Institute (ILRI).
- Daoud, H., Tetreault, M., Gibson, W., Guerrero, K., Cohen, A., and Gburek-Augustat, J. (2013). Mutations in POLR3A and POLR3B are a major cause of hypomyelinating leukodystrophies with or without dental abnormalities and/or hypogonadotropic hypogonadism. *J. Med. Genet.* 50, 194–197. doi: 10.1136/jmedgenet-2012-101357
- DeGiorgio, M., Huber, C. D., Hubisz, M. J., Hellmann, I., and Nielsen, R. (2016). SweepFinder2: increased sensitivity, robustness and flexibility. *Bioinf.* 32, 1895–1897. doi: 10.1093/bioinformatics/btw051

- Dgany, O., Avidan, N., Delaunay, J., Krasnov, T., Shalmon, L., and Shalev, H. (2002). Congenital dyserythropoietic anemia type I is caused by mutations in codanin-1. *Am. J. Hum. Genet.* 71, 1467–1474. doi: 10.1086/344781
- Dimova, D. K., and Dyson, N. J. (2005). The E2F transcriptional network: old acquaintances with new faces. *Oncogene* 24, 2810–2826. doi: 10.1038/sj.onc.1208612
- Dong, C., Davis, R. J., and Flavell, R. A. (2002). MAP kinases in the immune response. *Annu. Rev. Immunol.* 20, 55–72. doi: 10.1146/annurev.immunol.20.091301.131133
- Doolan, D. L., and Hoffman, S. L. (1999). IL-12 and NK cells are required for antigen-specific adaptive immunity against malaria initiated by CD8+ T cells in the *Plasmodium yoelii* model. *J. Immunol.* 163, 884–892.
- Dorssers, L., Burger, H., Bot, F., Delwel, R., Geurts van Kessel, A. H., and Lowenberg, B. (1987). Characterization of a human multilineage-colony-stimulating factor cDNA clone identified by a conserved noncoding sequence in mouse interleukin-3. *Gene* 55, 115–124. doi: 10.1016/0378-1119(87)90254-X
- Dyson, N. (1998). The regulation of E2F by pRB-family proteins. *Genes Dev.* 12, 2245–2262. doi: 10.1101/gad.12.15.2245
- Eagle, R. A., and Trowsdale, J. (2007). Promiscuity and the single receptor: NKG2D. *Nat. Rev. Immunol.* 7, 737–744. doi: 10.1038/nri2144
- Eguchi, K. (2001). Apoptosis in autoimmune diseases. *Internal Med.* 40, 275–284. doi: 10.2169/internalmedicine.40.275
- Ersching, J., Vasconcelos, J. R., Ferreira, C. P., Caetano, B. C., Machado, A. V., Bruna-Romero, O., et al. (2016). The combined deficiency of immunoproteasome subunits affects both the magnitude and quality of Pathogen- and genetic vaccination- induced CD8+ T Cell responses to the human Protozoan parasite *Trypanosoma cruzi*. *PLoS Pathog.* 12 (4), e1005593
- FAO. (2004). *The uganda country report as part of the state of the world's animal genetic resources (sow-angr) report*. Food and Agriculture Organization of the United Nations, Rome, Italy.
- Fatih, M., Adamu, S., Ibrahim, N., Euvie, L., and Esievo, K. (2009). The effect of experimental *Trypanosoma vivax* infection on the thyroid gland in Zebu bulls. *Vet. Arhiv.* 79, 429–437. doi: 10.4103%2Fijem.IJEM_12_17
- Flávia Nardy, A., Freire-de-Lima, C. G., and Morrot, A. (2015). Immune evasion strategies of *trypanosoma cruzi*. *J. Immunol. Res.* 2015, 1–7. doi: 10.1155/2015/178947
- Ganz, T. (2003). Defensins: antimicrobial peptides of innate immunity. *Nat. Rev. Immunol.* 3, 710–720. doi: 10.1038/nri1180
- Gautier, M., and Vitalis, R. (2012). rehh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. *Bioinf.* 28, 1176–1177. doi: 10.1093/bioinformatics/bts115
- Gerard, C., and Rollins, B. J. (2001). Chemokines and disease. *Nat. Immunol.* 2, 108–115. doi: 10.1038/84209
- Gibbs, R. A., Taylor, J. F., Van Tassel, C. P., Barendse, W., Eversole, K. A., and Gill, C. A. (2009). Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Sci.* 324, 528–532. doi: 10.1126/science.1167936
- Giordani, F., Morrison, L. J., Rowan, T. G., DE Koning, H. P., and Barrett, M. P. (2016). The animal trypanosomiasis and their chemotherapy: a review. *Parasitology* 143, 1862–1889. doi: 10.1017/S0031182016001268
- Goldszmid, R. S., and Sher, A. (2010). Processing and presentation of antigens derived from intracellular protozoan parasites. *Curr. Opin. Immunol.* 22, 118–123. doi: 10.1016/j.coi.2010.01.017
- Gonçalves, V. M., Matteucci, K. C., Buzzo, C. L., Miozzo, B. H., Ferrante, D., and Torrecilhas, A. C. (2013). NLRP3 controls *Trypanosoma cruzi* infection through a caspase-1-dependent IL-1R-independent NO production. *PLoS Negl. Trop. Dis.* 7, e2469. doi: 10.1371/journal.pntd.0002469
- Gong, X., Tang, X., Wiedmann, M., Wang, X., Peng, J., and Zheng, D. (2003). Cdk5-mediated inhibition of the protective effects of transcription factor MEF2 in neurotoxicity-induced apoptosis. *Neuron* 38, 33–46. doi: 10.1016/S0896-6273(03)00191-0
- Gringhuis, S. I., den Dunnen, J., Litjens, M., van der Vlist, M., Wevers, B., and Bruijns, S. C. (2009). Dectin-1 directs T helper cell differentiation by controlling noncanonical NF- κ B activation through Raf-1 and Syk. *Nat. Immunol.* 10, 203–213. doi: 10.1038/ni.1692
- Grisouard, J., Hao-Shen, H., Dirnhofer, S., Wagner, K.-U., and Skoda, R. C. (2014). Selective deletion of jak2 in adult mouse hematopoietic cells leads to lethal anemia and thrombocytopenia. *Haematologica.* 99, e52–e54. doi: 10.3324/haematol.2013.100016
- Guillemot, J., and Seidah, N. G. (2015). PACE4 (PCSK6): another proprotein convertase link to iron homeostasis?. *Haematologica.* 100, e377. doi: 10.3324/haematol.2015.127175
- Hanke, J. H., Gardner, J. P., Dow, R. L., Changelian, P. S., Brissette, W. H., and Weringer, E. J. (1996). Discovery of a novel, potent, and Src family-selective tyrosine kinase inhibitor. Study of Lck- and FynT-dependent T cell activation. *J. Biol. Chem.* 271, 695–701. doi: 10.1074/jbc.271.2.695
- Hanotte, O., Ronin, Y., Agaba, M., Nilsson, P., Gelhaus, A., and Horstmann, R. (2003). Mapping of quantitative trait loci controlling trypanotolerance in a cross of tolerant West African N'Dama and susceptible East African Boran cattle. *Proc. Natl. Acad. Sci. U.S.A.* 100, 7443–7448. doi: 10.1073/pnas.1232392100
- Hanotte, O., Tawah, C. L., Bradley, D. G., Okomo, M., Verjee, Y., and Ochieng, J. (2000). Geographic distribution and frequency of a taurine *Bos taurus* and an indicine *Bos indicus* Y specific allele amongst sub-saharan African cattle breeds. *Mol. Ecol.* 9, 387–396. doi: 10.1046/j.1365-294x.2000.00858.x
- Hayes, B. J., Pryce, J., Chamberlain, A. J., Bowman, P. J., and Goddard, M. E. (2010). Genetic architecture of complex traits and accuracy of genomic prediction: coat colour, milk-fat percentage, and type in Holstein cattle as contrasting model traits. *PLoS Genet.* 6, e1001139. doi: 10.1371/journal.pgen.1001139
- Haynes, M. P., Li, L., Sinha, D., Russell, K. S., Hisamoto, K., and Baron, R. (2003). Src kinase mediates phosphatidylinositol 3-kinase/Akt-dependent rapid endothelial nitric-oxide synthase activation by estrogen. *J. Biol. Chem.* 278, 2118–2123. doi: 10.1074/jbc.M210828200
- He, C., Medley, S. C., Hu, T., Hinsdale, M. E., Lupu, F., and Virmani, R. (2015). PDGFR2 signalling regulates local inflammation and synergizes with hypercholesterolaemia to promote atherosclerosis. *Nat. Commun.* 6, 7770. doi: 10.1038/ncomms8770
- Heinonen, K. M., Dube, N., Bourdeau, A., Lapp, W. S., and Tremblay, M. L. (2006). Protein tyrosine phosphatase 1B negatively regulates macrophage development through CSF-1 signaling. *Proc. Natl. Acad. Sci. U.S.A.* 103, 2776–2781. doi: 10.1073/pnas.0508563103
- Hoare, C. A. (1972). “The trypanosomes of mammals,” in *A zoological monograph* (Oxford and Edinburgh: UKs: Publications BS).
- Hoyt, R., Zhu, W., Cerignoli, F., Alonzo, A., Mustelin, T., and David, M. (2007). Cutting edge: selective tyrosine dephosphorylation of interferon-activated nuclear STAT5 by the VHR phosphatase. *J. Immunol.* 179, 3402–3406. doi: 10.4049/jimmunol.179.6.3402
- Hu, H., and Sun, S. C. (2016). Ubiquitin signaling in immune responses. *Cell Res.* 26, 457–483. doi: 10.1038/cr.2016.40
- Hu, T., Ghazaryan, S., Sy, C., Wiedmeyer, C., Chang, V., and Wu, L. (2012). Concomitant inactivation of Rb and E2f8 in hematopoietic stem cells synergizes to induce severe anemia. *Blood* 119, 4532–4542. doi: 10.1182/blood-2011-10-388231
- Huber, C. D., DeGiorgio, M., Hellmann, I., and Nielsen, R. (2016). Detecting recent selective sweeps while controlling for mutation rate and background selection. *Mol. Ecol.* 25, 142–156. doi: 10.1111/mec.13351
- Ingley, E. (2012). Functions of the Lyn tyrosine kinase in health and disease. *Cell Commun. Signal* 10, 21. doi: 10.1186/1478-811X-10-21
- Jabara, H. H., Boyden, S. E., Chou, J., Ramesh, N., Massaad, M. J., and Benson, H. (2016). A missense mutation in TFRC, encoding transferrin receptor 1, causes combined immunodeficiency. *Nat. Genet.* 48, 74–78. doi: 10.1038/ng.3465
- Jackson, P. K. (2012). TTBK2 kinase: linking primary cilia and cerebellar ataxias. *Cell.* 151, 697–699. doi: 10.1016/j.cell.2012.10.027
- Jaggy, A., Oliver, J. E., Ferguson, D. C., Mahaffey, E. A., and Glaus, T. (1994). Neurological manifestations of hypothyroidism: a retrospective study of 29 dogs. *J. Vet. Intern. Med.* 8, 328–336. doi: 10.1111/j.1939-1676.1994.tb03245.x
- Jantsch, J., Wiese, M., Schodel, J., Castiglione, K., Glasner, J., and Kolbe, S. (2011). Toll-like receptor activation and hypoxia use distinct signaling pathways to stabilize hypoxia-inducible factor 1 α (HIF1A) and result in differential HIF1A-dependent gene expression. *J. Leukoc. Biol.* 90, 551–562. doi: 10.1189/jlb.1210683
- Johnson, K. R., Gagnon, L. H., and Chang, B. (2016). A hypomorphic mutation of the gamma-1 adaptin gene (*Aplg1*) causes inner ear, retina, thyroid, and testes abnormalities in mice. *Mamm. Genome* 27, 200–212. doi: 10.1007/s00335-016-9632-0
- Jordan, K. A., and Hunter, C. A. (2010). Regulation of CD8+ T cell responses to infection with parasitic protozoa. *Exp. Parasitol.* 126, 318–325. doi: 10.1016/j.exppara.2010.05.008
- Kaminski, W. E., Lindahl, P., Lin, N. L., Broudy, V. C., Crosby, J. R., and Hellstrom, M. (2001). Basis of hematopoietic defects in platelet-derived growth factor (PDGF)-B and PDGF beta-receptor null mice. *Blood* 97, 1990–1998. doi: 10.1182/blood.V97.7.1990

- Kanazawa, I. (1999). Molecular pathology of dentatorubral-pallidolusian atrophy. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 354, 1069–1074. doi: 10.1098/rstb.1999.0460
- Kawai, T., and Akira, S. (2007). Antiviral signaling through pattern recognition receptors. *J. Biochem.* 141, 137–145. doi: 10.1093/jb/mvm032
- Kierstein, S., Noyes, H., Naessens, J., Nakamura, Y., Pritchard, C., and Gibson, J. (2006). Gene expression profiling in a mouse model for African trypanosomiasis. *Genes Immun.* 7, 667–679. doi: 10.1038/sj.gene.6364345
- Kim, C. H., and Broxmeyer, H. E. (1999). Chemokines: signal lamps for trafficking of T and B cells for development and effector function. *J. Leukoc. Biol.* 65, 6–15. doi: 10.1002/jlb.65.1.6
- Kim, Y., and Stephan, W. (2002). Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genet.* 160, 765–777.
- Kinsella, R. J., Kahari, A., Haider, S., Zamora, J., Proctor, G., and Spudich, G. (2011). Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database (Oxford)* 2011, bar030. doi: 10.1093/database/bar030
- Kristjanson, P. M., Swallow, B. M., Rowlands, G. J., Kruska, R. L., and De Leeuw, P. N. (1999). Measuring the costs of African animal trypanosomiasis, the potential benefits of control and returns to research. *Agric. Syst.* 59 (7), 79–98. doi: 10.1016/S0308-521X(98)00086-9
- Krull, M., Pistor, S., Voss, N., Kel, A., Reuter, I., and Kronenberg, D. (2006). Transpath®: an information resource for storing and visualizing signaling pathways and their pathological aberrations. *Nucleic Acids Res.* 34, D546–D551. doi: 10.1093/nar/gkj107
- Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., and Wang, Z. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* 44, W90–W97. doi: 10.1093/nar/gkw377
- La Greca, F., Haynes, C., Stijlemans, B., De Trez, C., and Magez, S. (2014). Antibody-mediated control of *Trypanosoma vivax* infection fails in the absence of tumour necrosis factor. *Parasite Immunol.* 36, 271–276. doi: 10.1111/pim.12106
- Lamkanfi, M., Festjens, N., Declercq, W., Vanden Berghe, T., and Vandennebeele, P. (2007). Caspases in cell survival, proliferation and differentiation. *Cell Death Differ.* 14, 44–55. doi: 10.1038/sj.cdd.4402047
- Lanier, L. L. (2015). NKG2D Receptor and Its Ligands in Host Defense. *Cancer Immunol. Res.* 3, 575–582. doi: 10.1158/2326-6066.CIR-15-0098
- Leach, R. M., and Treacher, D. F. (1998). Oxygen transport-2. Tissue hypoxia. *BMJ* 317, 1370–1373. doi: 10.1136/bmj.317.7169.1370
- Leak, S. G., Peregrine, A. S., Mulatu, W., Rowlands, G. J., and D'Ieteren, G. (1996). Use of insecticide-impregnated targets for the control of tsetse flies (*Glossina* spp.) and trypanosomiasis occurring in cattle in an area of south-west Ethiopia with a high prevalence of drug-resistant trypanosomes. *Trop. Med. Int. Health* 1, 599–609. doi: 10.1111/j.1365-3156.1996.tb00085.x
- Lee, J. M., Chan, K., Kan, Y. W., and Johnson, J. A. (2004). Targeted disruption of Nrf2 causes regenerative immune-mediated hemolytic anemia. *Proc. Natl. Acad. Sci. U.S.A.* 101, 9751–9756. doi: 10.1073/pnas.0403620101
- Lemecha, H., Mulatu, W., Hussein, I., Rege, E., Tekle, T., and Abdicho, S. (2006). Response of four indigenous cattle breeds to natural tsetse and trypanosomiasis challenge in the Ghibe valley of Ethiopia. *Vet. Parasitol.* 141, 165–176. doi: 10.1016/j.vetpar.2006.04.035
- Liao, S. H., Zhao, X. Y., Han, Y. H., Zhang, J., Wang, L. S., and Xia, L. (2009). Proteomics-based identification of two novel direct targets of hypoxia-inducible factor-1 and their potential roles in migration/invasion of cancer cells. *Proteomics* 9, 3901–3912. doi: 10.1002/pmic.200800922
- Linnekin, D., DeBerry, C. S., and Mou, S. (1997). Lyn associates with the juxtamembrane region of c-Kit and is activated by stem cell factor in hematopoietic cell lines and normal progenitor cells. *J. Biol. Chem.* 272, 27450–27455. doi: 10.1074/jbc.272.43.27450
- Liu, G., Burns, S., Huang, G., Boyd, K., Proia, R. L., and Flavell, R. A. (2009). The receptor S1P1 overrides regulatory T cell-mediated immune suppression through Akt-mTOR. *Nat. Immunol.* 10, 769–777. doi: 10.1038/ni.1743
- Liu, Q. R., Zhang, P. W., Zhen, Q., Walther, D., Wang, X. B., and Uhl, G. R. (2002). KEPI, a PKC-dependent protein phosphatase 1 inhibitor regulated by morphine. *J. Biol. Chem.* 277, 13312–13320. doi: 10.1074/jbc.M107558200
- Lodish, H., Berk, A., Matsudaira, P. C., K., Krieger, M., and Scott, M. (2004). *Molecular Cell Biology*. 41 Madison Avenue, New York, USA: W. H. Freeman and Company.
- Lutje, V., Taylor, K. A., Kennedy, D., Authie, E., Boulange, A., and Gettinby, G. (1996). *Trypanosoma congolense*: a comparison of T-cell-mediated responses in lymph nodes of trypanotolerant and trypanosusceptible cattle during primary infection. *Exp. Parasitol.* 84, 320–329. doi: 10.1006/expr.1996.0120
- Ma, Y., Ding, X., Qanbari, S., Weigend, S., Zhang, Q., and Simianer, H. (2015). Properties of different selection signature statistics and a new strategy for combining them. *Heredity (Edinb)* 115, 426–436. doi: 10.1038/hdy.2015.42
- Mabley, J. G., and Szabo, C. (2008). Inflammatory disease and sunlight: the vitamin D-poly (ADP-ribose) polymerase connection. *Future Rheumatol.* 3, 169–181. doi: 10.2217/17460816.3.2.169
- Magona, J. W., Walubengo, J., and Odimim, J. J. (2004). Differences in susceptibility to trypanosome infection between Nkedi Zebu and Ankole cattle, under field conditions in Uganda. *Ann. Trop. Med. Parasitol.* 98, 785–792. doi: 10.1179/000349804225021532
- Mamoudou, A., Njanloga, A., Hayatou, A., Suh, P. F., and Achukwi, M. D. (2016). Animal trypanosomiasis in clinically healthy cattle of north Cameroon: epidemiological implications. *Parasit Vectors.* 9, 206. 5600, 1912–1934 doi: 10.1186/s13071-016-1498-1
- Manning, G., Whyte, D. B., Martinez, R., Hunter, T., and Sudarsanam, S. (2002). The protein kinase complement of the human genome. *Science* 298 (5600), 1912–1934. doi: 10.1126/science.1075762
- Martin-Granados, C., Prescott, A. R., Le Sommer, S., Klaska, I. P., Yu, T., and Muckersie, E. (2015). A key role for PTP1B in dendritic cell maturation, migration, and T cell activation. *J. Mol. Cell Biol.* 7, 517–528. doi: 10.1093/jmcb/mjv032
- Matilla-Duenas, A. (2012). The ever expanding spinocerebellar ataxias. *Editorial Cerebellum* 11, 821–827. doi: 10.1007/s12311-012-0376-4
- Mbole-Kariuki, M. N., Sonstegard, T., Orth, A., Thumbi, S. M., Bronsvort, B. M., and Kiara, H. (2014). Genome-wide analysis reveals the ancient and recent admixture history of East African Shorthorn Zebu from Western Kenya. *Heredity (Edinb)* 113, 297–305. doi: 10.1038/hdy.2014.31
- McGuire, S. E., and McGuire, A. L. (2008). Don't throw the baby out with the bathwater: enabling a bottom-up approach in genome-wide association studies. *Genome Res.* 18, 1683–1685. doi: 10.1101/gr.083584.108
- McNab, F., Mayer-Barber, K., Sher, A., Wack, A., and O'Garra, A. (2015). Type I interferons in infectious disease. *Nat. Rev. Immunol.* 15, 87–103. doi: 10.1038/nri3787
- McNamee, E. N., Korn Johnson, D., Homann, D., and Clambey, E. T. (2013). Hypoxia and hypoxia-inducible factors as regulators of T cell development, differentiation, and function. *Immunol. Res.* 55, 58–70. doi: 10.1007/s12026-012-8349-8
- Mendes-da Cruz, D. A., Silva, J. S., Cotta-de Almeida, V., and Savino, W. (2006). Altered thymocyte migration during experimental acute *Trypanosoma cruzi* infection: combined role of fibronectin and the chemokines CXCL12 and CCL4. *Eur. J. Immunol.* 36, 1486–1493. doi: 10.1002/eji.200535629
- Meyer, A., Holt, H. R., Oumarou, F., Chilongo, K., Gilbert, W., and Fauron, A. (2018). Integrated cost-benefit analysis of tsetse control and herd productivity to inform control programs for animal african trypanosomiasis. *Parasites Vectors* 11, 154. doi: 10.1186/s13071-018-2679-x
- Minchenko, O. H., Tsymbal, D. O., Minchenko, D. O., Kovalevska, O. V., Karbovskyi, L. L., and Bikfalvi, A. (2015). Inhibition of ERN1 signaling enzyme affects hypoxic regulation of the expression of E2F8, EPAS1, HOXC6, ATF3, TBX3 and FOXF1 genes in U87 glioma cells. *Ukr Biochem. J.* 87, 76–87. doi: 10.15407/ubj87.02.076
- MoARD. (2004). Ministry Of Agriculture and Rural Development of the Government of Ethiopia (MoARD). Tsetse and trypanosomiasis prevention and control strategies. Paper presented on Farming In Tsetse Controlled Areas (FITCA), Ethiopia final workshop. Adama, Ethiopia .
- Molenaar, P. C., Newsom-Davis, J., Polak, R. L., and Vincent, A. (1982). Eaton-Lambert syndrome: acetylcholine and choline acetyltransferase in skeletal muscle. *Neurol.* 32, 1061–1065. doi: 10.1212/WNL.32.9.1061
- Morrison, L. J., Vezza, L., Rowan, T., and Hope, J. C. (2016). Animal African trypanosomiasis: time to increase focus on clinically relevant parasite and host species. *Trends Parasitol.* 32, 599–607. doi: 10.1016/j.pt.2016.04.012
- Morrot, A., and Zavala, F. (2004). Effector and memory CD8+ T cells as seen in immunity to malaria. *Immunol. Rev.* 201, 291–303. doi: 10.1111/j.0105-2896.2004.00175.x
- Moura, I. C., Centelles, M. N., Arcos-Fajardo, M., Malheiros, D. M., Collawn, J. F., and Cooper, M. D. (2001). Identification of the transferrin receptor as a novel immunoglobulin (Ig)A1 receptor and its enhanced expression on mesangial cells in IgA nephropathy. *J. Exp. Med.* 194, 417–425. doi: 10.1084/jem.194.4.417

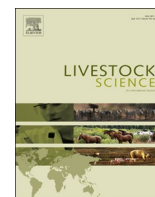
- Muhanguzi, D., Mugenyi, A., Bigirwa, G., Kamusiime, M., Kitibwa, A., and Akurut, G. G. (2017). African animal trypanosomiasis as a constraint to livestock health and production in Karamoja region: a detailed qualitative and quantitative assessment. *BMC Vet. Res.* 13, 355. doi: 10.1186/s12917-017-1285-z
- Muhanguzi, D., Picozzi, K., Hatendorf, J., Thrusfield, M., Welburn, S. C., and Kabasa, J. D. (2014). Improvements on restricted insecticide application protocol for control of Human and Animal African Trypanosomiasis in eastern Uganda. *PLoS Negl. Trop. Dis.* 8, e3284. doi: 10.1371/journal.pntd.0003284
- Mulugeta, W., Wilkes, J., Mulatu, W., Majiwa, P. A., Masake, R., and Peregrine, A. S. (1997). Long-term occurrence of Trypanosoma congolense resistant to diminazene, isometamidium and homidium in cattle at Ghibe, Ethiopia. *Acta Trop.* 64, 205–217. doi: 10.1016/S0001-706X(96)00645-6
- Murray, M., Trail, J. C., and D'Ieteren, G. D. (1990). Trypanotolerance in cattle and prospects for the control of trypanosomiasis by selective breeding. *Rev. - Off. Int. Epizoot.* 9, 369–386. doi: 10.20506/rst.9.2.506
- Mwai, O., Hanotte, O., Kwon, Y.-J., and Cho, S. (2015). African indigenous cattle: unique genetic resources in a rapidly changing world. *Asian-Australasian J. Anim. Sci.* 28, 911. doi: 10.5713/ajas.15.0002R
- Naessens, J. (2006). Bovine trypanotolerance: A natural ability to prevent severe anaemia and haemophagocytic syndrome?. *Int. J. Parasitol.* 36, 521–528. doi: 10.1016/j.ijpara.2006.02.012
- Nantulya, V. M. (1986). Immunological approaches to the control of animal trypanosomiasis. *Parasitol. Today (Regul. Ed.)* 2, 168–173. doi: 10.1016/0169-4758(86)90148-1
- Naula, C., and Burchmore, R. (2003). A plethora of targets, a paucity of drugs: progress towards the development of novel chemotherapies for human African trypanosomiasis. *Expert Rev. Anti Infect. Ther.* 1, 157–165. doi: 10.1586/14787210.1.1.157
- Nevins, J. R. (1998). Toward an understanding of the functional complexity of the E2F and retinoblastoma families. *Cell Growth Differ.* 9, 585–593.
- Nielsen, R., Williamson, S., Kim, Y., Hubisz, M. J., Clark, A. G., and Bustamante, C. (2005). Genomic scans for selective sweeps using SNP data. *Genome Res.* 15, 1566–1575. doi: 10.1101/gr.4252305
- Noyes, H., Brass, A., Obara, I., Anderson, S., Archibald, A. L., and Bradley, D. G. (2011). Genetic and expression analysis of cattle identifies candidate genes in pathways responding to Trypanosoma congolense infection. *Proc. Natl. Acad. Sci. U.S.A.* 108, 9304–9309. doi: 10.1073/pnas.1013486108
- NTTICC. (2004). National Tsetse and Trypanosomiasis Investigation and Control Center. Report for the period 7th June 2003 to 6th July 2004. Bedede, Ethiopia. 21–24.
- Ocaido, M., Otim, C. P., Okuna, N. M., Erume, J., Ssekitto, C., and Wafula, R. Z. O., et al (2005). Socio-economic and livestock disease survey of agro-pastoral communities in Serere County, Soroti District, Uganda. *Livestock Research for Rural Development* 17.
- Ohtsuka, R., Abe, Y., Fujii, T., Yamamoto, M., Nishimura, J., and Takayanagi, R. (2007). Mortalin is a novel mediator of erythropoietin signaling. *Eur. J. Haematol.* 79, 114–125. doi: 10.1111/j.1600-0609.2007.00870.x
- Oleksyk, T. K., Smith, M. W., and O'Brien, S. J. (2010). Genome-wide scans for footprints of natural selection. *Philos. Trans. R. Soc Lond. B. Biol. Sci.* 365, 185–205. doi: 10.1098/rstb.2009.0219
- Oosthuyse, B., Moons, L., Storkebaum, E., Beck, H., Nuyens, D., and Brusselmanns, K. (2001). Deletion of the hypoxia-response element in the vascular endothelial growth factor promoter causes motor neuron degeneration. *Nat. Genet.* 28, 131–138. doi: 10.1038/88842
- Packham, G., White, E. L., Eischen, C. M., Yang, H., Parganas, E., and Ihle, J. N. (1998). Selective regulation of Bcl-XL by a Jak kinase-dependent pathway is bypassed in murine hematopoietic malignancies. *Genes Dev.* 12, 2475–2487. doi: 10.1101/gad.12.16.2475
- Patterson, H., Nibbs, R., McInnes, I., and Siebert, S. (2014). Protein kinase inhibitors in the treatment of inflammatory and autoimmune diseases. *Clin. Exp. Immunol.* 176, 1–10. doi: 10.1111/cei.12248
- Paul, W. E., Steinman, R., Beutler, B., and Hoffmann, J. (2011). Bridging innate and adaptive immunity. *Cell* 147, 1212–1215. doi: 10.1016/j.cell.2011.11.036
- Ponce, N. E., Cano, R. C., Carrera-Silva, E. A., Lima, A. P., Gea, S., and Aoki, M. P. (2012). Toll-like receptor-2 and interleukin-6 mediate cardiomyocyte protection from apoptosis during Trypanosoma cruzi murine infection. *Med. Microbiol. Immunol.* 201, 145–155. doi: 10.1007/s00430-011-0216-z
- Pulst, S. M. (2016). Genetics of neurodegenerative diseases. *Neurol. Genet.* 2, e52. doi: 10.1212/NXG.0000000000000052
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., and Bender, D. (2007). PLINK: a toolset for whole-genome association and population-based linkage analysis. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795
- Qanbari, S., Pausch, H., Jansen, S., Somel, M., Strom, T. M., and Fries, R. (2014). Classic selective sweeps revealed by massive sequencing in cattle. *PLoS Genet.* 10, e1004148. doi: 10.1371/journal.pgen.1004148
- Qiu, L., Joazeiro, C., Fang, N., Wang, H. Y., Elly, C., and Altman, Y. (2000). Recognition and ubiquitination of Notch by Itch, a hec-type E3 ubiquitin ligase. *J. Biol. Chem.* 275, 35734–35737. doi: 10.1074/jbc.M007300200
- Ran, Q., Wadhwa, R., Kawai, R., Kaul, S. C., Sifers, R. N., and Bick, R. J. (2000). Extramitochondrial localization of mortalin/mthsp70/PBP74/GRP75. *Biochem. Biophys. Res. Commun.* 275, 174–179. doi: 10.1006/bbrc.2000.3237
- Rege, J. (1999). The state of african cattle genetic resources i. classification framework and identification of threatened and extinct breeds. *Anim. Genet. Resources/Resources génétiques Animales/Recursos genéticos animales* 25, 1–25. doi: 10.1017/S1014233900003448
- Renella, R., Roberts, N. A., Brown, J. M., De Gobbi, M., Bird, L. E., and Hassanali, T. (2011). Codanin-1 mutations in congenital dyserythropoietic anemia type 1 affect HP1 α localization in erythroblasts. *Blood* 117, 6928–6938. doi: 10.1182/blood-2010-09-308478
- Rius, J., Guma, M., Schachtrup, C., Akassoglou, K., Zinkernagel, A. S., and Nizet, V. (2008). NF-kappaB links innate immunity to the hypoxic response through transcriptional regulation of HIF-1 α . *Nat.* 453, 807–811. doi: 10.1038/nature06905
- Rosenzweig, S. D., and Holland, S. M. (2005). Defects in the interferon-gamma and interleukin-12 pathways. *Immunol. Rev.* 203, 38–47. doi: 10.1111/j.0105-2896.2005.00227.x
- Rouault, T. A. (2006). The role of iron regulatory proteins in mammalian iron homeostasis and disease. *Nat. Chem. Biol.* 2, 406–414. doi: 10.1038/nchembio807
- Roux, P. P., and Blenis, J. (2004). ERK and p38 MAPK-activated protein kinases: a family of protein kinases with diverse biological functions. *Microbiol. Mol. Biol. Rev.* 68, 320–344. doi: 10.1128/MMBR.68.2.320-344.2004
- Sabeti, P. C., Reich, D. E., Higgins, J. M., Levine, H. Z., Richter, D. J., and Schaffner, S. F. (2002). Detecting recent positive selection in the human genome from haplotype structure. *Nat.* 419, 832–837. doi: 10.1038/nature01140
- Saleh, M. A., Al-Salahy, M. B., and Sanousi, S. A. (2009). Oxidative stress in blood of camels (Camelus dromedaries) naturally infected with Trypanosoma evansi. *Vet. Parasitol.* 162, 192–199. doi: 10.1016/j.vetpar.2009.03.035
- Salim, B., Taha, K. M., Hanotte, O., and Mwacharo, J. M. (2014). Historical demographic profiles and genetic variation of the East African Butana and Kenana indigenous dairy zebu cattle. *Anim. Genet.* 45, 782–790. doi: 10.1111/age.12225
- Salmond, R. J., Filby, A., Qureshi, I., Caserta, S., and Zamoyska, R. (2009). T-cell receptor proximal signaling via the Src-family kinases, Lck and Fyn, influences T-cell activation, differentiation, and tolerance. *Immunol. Rev.* 228, 9–22. doi: 10.1111/j.1600-065X.2008.00745.x
- Sangokoya, C., Telen, M. J., and Chi, J. T. (2010). microRNA miR-144 modulates oxidative stress tolerance and associates with anemia severity in sickle cell disease. *Blood* 116, 4338–4348. doi: 10.1182/blood-2009-04-214817
- Schiffmann, R., and van der Knaap, M. S. (2009). Invited article: an MRI-based approach to the diagnosis of white matter disorders. *Neurol.* 72, 750–759. doi: 10.1212/01.wnl.0000343049.00540.c8
- Schmitz-Abe, K., Ciesielski, S. J., Schmidt, P. J., Campagna, D. R., Rahimov, F., and Schilke, B. A. (2015). Congenital sideroblastic anemia due to mutations in the mitochondrial HSP70 homologue HSPA9. *Blood* 126, 2734–2738. doi: 10.1182/blood-2015-09-659854
- Shalini, S., Dorstyn, L., Dawar, S., and Kumar, S. (2015). Old, new and emerging functions of caspases. *Cell Death Different.* 22, 526. doi: 10.1038/cdd.2014.216
- Shaw, A. P., Cecchi, G., Wint, G. R., Mattioli, R. C., and Robinson, T. P. (2014). Mapping the economic benefits to livestock keepers from intervening against bovine trypanosomiasis in Eastern Africa. *Prev. Vet. Med.* 113, 197–210. doi: 10.1016/j.prevetmed.2013.10.024
- She, H., Yang, Q., and Mao, Z. (2012). Neurotoxin-induced selective ubiquitination and regulation of MEF2A isoform in neuronal stress response. *J. Neurochem.* 122, 1203–1210. doi: 10.1111/j.1471-4159.2012.07860.x
- She, H., Yang, Q., Shepherd, K., Smith, Y., Miller, G., and Testa, C. (2011). Direct regulation of complex I by mitochondrial MEF2D is disrupted in a mouse model of Parkinson disease and in human patients. *J. Clin. Invest.* 121, 930–940. doi: 10.1172/JCI43871

- Silva, B., and Faustino, P. (2015). An overview of molecular basis of iron metabolism regulation and the associated pathologies. *Biochim. Biophys. Acta* 1852, 1347–1359. doi: 10.1016/j.bbadis.2015.03.011
- Singh, B., Soltys, B. J., Wu, Z. C., Patel, H. V., Freeman, K. B., and Gupta, R. S. (1997). Cloning and some novel characteristics of mitochondrial Hsp70 from Chinese hamster cells. *Exp. Cell Res.* 234, 205–216. doi: 10.1006/excr.1997.3609
- Singh, Y., Garden, O. A., Lang, F., and Cobb, B. S. (2016). MicroRNAs regulate T-cell production of interleukin-9 and identify hypoxia-inducible factor-2Ia as an important regulator of T helper 9 and regulatory T-cell differentiation. *Immunol.* 149, 74–86. doi: 10.1111/imm.12631
- Sinshaw, A., Abebe, G., Desquesnes, M., and Yoni, W. (2006). Biting flies and Trypanosoma vivax infection in three highland districts bordering lake Tana, Ethiopia. *Vet. Parasitol.* 142, 35–46. doi: 10.1016/j.vetpar.2006.06.032
- Skipper, M. (2002). Human genetics: Tracking positive selection. *Nat. Rev. Genet.* 3, 824. doi: 10.1038/nrg942
- Slingenbergh, J. (1992). Tsetse control and agricultural development in Ethiopia. *World Anim. Rev.* 70–71, 30–36.
- Song, J., Salek-Ardakani, S., So, T., and Croft, M. (2007). The kinases aurora B and mTOR regulate the G1-S cell cycle progression of T lymphocytes. *Nat. Immunol.* 8, 64–73. doi: 10.1038/ni1413
- Springer, T. A. (1990). Adhesion receptors of the immune system. *Nat.* 346, 425. doi: 10.1038/346425a0
- Stahl, P., Ruppert, V., Meyer, T., Schmidt, J., Campos, M. A., and Gazzinelli, R. T. (2013). Trypomastigotes and amastigotes of *Trypanosoma cruzi* induce apoptosis and STAT3 activation in cardiomyocytes in vitro. *Apoptosis* 18, 653–663. doi: 10.1007/s10495-013-0822-x
- Stanojevic, V., Habener, J. F., Holz, G. G., and Leech, C. A. (2008). Cytosolic adenylate kinases regulate K-ATP channel activity in human beta-cells. *Biochem. Biophys. Res. Commun.* 368, 614–619. doi: 10.1016/j.bbrc.2008.01.109
- Steverding, D. (2008). The history of African trypanosomiasis. *Parasit Vectors* 1, 3. doi: 10.1186/1756-3305-1-3
- Stijlemans, B., Vankrunkelsven, A., Caljon, G., Bockstal, V., Guilliams, M., and Bosschaerts, T. (2010). The central role of macrophages in trypanosomiasis-associated anemia: rationale for therapeutic approaches. *Endocr. Metab. Immune Disord. Drug Targets* 10, 71–82. doi: 10.2174/187153010790827966
- Suzuki, Y., and Yazawa, I. (2011). Pathological accumulation of atrophin-1 in dentatorubralpallidoluysian atrophy. *Int. J. Clin. Exp. Pathol.* 4, 378–384.
- Tang, K., Thornton, K. R., and Stoneking, M. (2007). A new approach for using genome scans to detect recent positive selection in the human genome. *PLoS Biol.* 5, e171. doi: 10.1371/journal.pbio.0050171
- Tavazzi, B., Di Pierro, D., Amorini, A. M., Fazzina, G., Tuttobene, M., and Giardina, B. (2000). Energy metabolism and lipid peroxidation of human erythrocytes as a function of increased oxidative stress. *Eur. J. Biochem.* 267, 684–689. doi: 10.1046/j.1432-1327.2000.01042.x
- Taylor, K. A. (1998). Immune responses of cattle to African trypanosomes: protective or pathogenic?. *Int. J. Parasitol.* 28, 219–240. doi: 10.1016/S0020-7519(97)00154-9
- Tomasec, P., Wang, E. C., Groh, V., Spies, T., McSharry, B. P., and Aichele, R. J. (2007). Adenovirus vector delivery stimulates natural killer cell recognition. *J. Gen. Virol.* 88, 1103–1108. doi: 10.1099/vir.0.82685-0
- Traves, P. G., Pardo, V., Pimentel-Santillana, M., Gonzalez-Rodriguez, A., Mojena, M., and Rico, D. (2014). Pivotal role of protein tyrosine phosphatase 1B (PTP1B) in the macrophage response to pro-inflammatory and anti-inflammatory challenge. *Cell Death Dis.* 5, e1125. doi: 10.1038/cddis.2014.90
- Trimarchi, J. M., and Lees, J. A. (2002). Sibling rivalry in the E2F family. *Nat. Rev. Mol. Cell Biol.* 3, 11–20. doi: 10.1038/nrm714
- Tuntasuvan, D., Sarataphan, N., and Nishikawa, H. (1997). Cerebral trypanosomiasis in native cattle. *Vet. Parasitol.* 73, 357–363. doi: 10.1016/S0304-4017(97)00128-3
- Utz, P. J., Hottel, M., Schur, P. H., and Anderson, P. (1997). Proteins phosphorylated during stress-induced apoptosis are common targets for autoantibody production in patients with systemic lupus erythematosus. *J. Exp. Med.* 185, 843–854. doi: 10.1084/jem.185.5.843
- Van Wyk, I. C., Goddard, A., de C Bronsvort, B. M., Coetzer, J. A., Handel, I. G., and Hanotte, O. (2014). The impact of co-infections on the haematological profile of East African Short-horn Zebu calves. *Parasitology* 141, 374–388. doi: 10.1017/S0031182013001625
- Vatsiou, A. I., Bazin, E., and Gaggiotti, O. E. (2016). Detection of selective sweeps in structured populations: a comparison of recent methods. *Mol. Ecol.* 25, 89–103. doi: 10.1111/mec.13360
- Vignali, D. A., Collison, L. W., and Workman, C. J. (2008). How regulatory T cells work. *Nat. Rev. Immunol.* 8, 523–532. doi: 10.1038/nri2343
- Wang, D., Sai, J., Carter, G., Sachpatzidis, A., Lolis, E., and Richmond, A. (2002). PAK1 kinase is required for CXCL1-induced chemotaxis. *Biochem. J.* 41, 7100–7107. doi: 10.1021/bi025902m
- Welburn, S. C., Molyneux, D. H., and Maudlin, I. (2016). Beyond tsetse-implications for research and control of human african trypanosomiasis epidemics. *Trends Parasitol.* 32, 230–241. doi: 10.1016/j.pt.2015.11.008
- Wilches, R., Voigt, S., Duchon, P., Laurent, S., and Stephan, W. (2014). Fine-mapping and selective sweep analysis of QTL for cold tolerance in *Drosophila melanogaster*. *G3: Genes, Genomes, Genetics.* 4, 1635–1645. doi: 10.1534/g3.114.012757
- Woolsey, A. M., Sunwoo, L., Petersen, C. A., Brachmann, S. M., Cantley, L. C., and Burleigh, B. A. (2003). Novel PI 3-kinase-dependent mechanisms of trypanosome invasion and vacuole maturation. *J. Cell. Sci.* 116, 3611–3622. doi: 10.1242/jcs.00666
- Wu, D., and Hersh, L. B. (1994). Choline acetyltransferase: celebrating its fiftieth year. *J. Neurochem.* 62, 1653–1663. doi: 10.1046/j.1471-4159.1994.62051653.x
- Yang, D., Chertov, O., Bykovskaia, S. N., Chen, Q., Buffo, M. J., and Shogan, J. (1999). Beta-defensins: linking innate and adaptive immunity through dendritic and T cell CCR6. *Sci.* 286, 525–528. doi: 10.1126/science.286.5439.525
- Yaro, M., Munyard, K. A., Stear, M. J., and Groth, D. M. (2016). Combatting African Animal Trypanosomiasis (AAT) in livestock: The potential role of trypanotolerance. *Vet. Parasitol.* 225, 43–52. doi: 10.1016/j.vetpar.2016.05.003
- Yi, T., Cleveland, J. L., and Ihle, J. N. (1991). Identification of novel protein tyrosine phosphatases of hematopoietic cells by polymerase chain reaction amplification. *Blood* 78, 2222–2228. doi: 10.1182/blood.V78.9.2222.2222
- Ymer, S., Tucker, W. Q., Sanderson, C. J., Hapel, A. J., Campbell, H. D., and Young, I. G. (1985). Constitutive synthesis of interleukin-3 by leukaemia cell line WEHI-3B is due to retroviral insertion near the gene. *Nat.* 317, 255–258. doi: 10.1038/317255a0
- You, F., Sun, H., Zhou, X., Sun, W., Liang, S., and Zhai, Z. (2009). PCB2 mediates degradation of the adaptor MAVS via the HECT ubiquitin ligase AIP4. *Nat. Immunol.* 10, 1300–1308. doi: 10.1038/ni.1815
- Yuan, Z. L., Guan, Y. J., Wang, L., Wei, W., Kane, A. B., and Chin, Y. E. (2004). Central role of the threonine residue within the p+1 loop of receptor tyrosine kinase in STAT3 constitutive phosphorylation in metastatic cancer cells. *Mol. Cell. Biol.* 24, 9390–9400. doi: 10.1128/MCB.24.21.9390-9400.2004
- Zelenika, D., Grima, B., and Pessac, B. (1993). A new family of transcripts of the myelin basic protein gene: expression in brain and in immune system. *J. Neurochem.* 60, 1574–1577. doi: 10.1111/j.1471-4159.1993.tb03325.x
- Zermati, Y., Garrido, C., Amsellem, S., Fishelson, S., Bouscary, D., and Valensi, F. (2001). Caspase activation is required for terminal erythroid differentiation. *J. Exp. Med.* 193, 247–254. doi: 10.1084/jem.193.2.247
- Zhang, R., Zhou, L., Li, Q., Liu, J., Yao, W., and Wan, H. (2009). Up-regulation of two actin-associated proteins prompts pulmonary artery smooth muscle cell migration under hypoxia. *Am. J. Respir. Cell Mol. Biol.* 41, 467–475. doi: 10.1165/rcmb.2008-0333OC
- Zhang, S., Han, J., Sells, M. A., Chernoff, J., Knaus, U. G., and Ulevitch, R. J. (1995). Rho family GTPases regulate p38 mitogen-activated protein kinase through the downstream mediator Pak1. *J. Biol. Chem.* 270, 23934–23936. doi: 10.1074/jbc.270.41.23934
- Zhang, Y., Wang, X., Yang, H., Liu, H., Lu, Y., and Han, L. (2013). Kinase AKT controls innate immune cell development and function. *Immunol.* 140, 143–152. doi: 10.1111/imm.12123
- Zinngrebe, J., Montinaro, A., Peltzer, N., and Walczak, H. (2014). Ubiquitin in the immune system. *EMBO Rep.* 15, 28–45. doi: 10.1002/embr.201338025

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Mekonnen, Gültas, Effa, Hanotte and Schmitt. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

A.2. Genetic Diversity and Population Structure of Six Ethiopian Cattle Breeds from Different Geographical Regions Using High Density Single Nucleotide Polymorphism Markers



Short communication

Genetic diversity and population structure of six Ethiopian cattle breeds from different geographical regions using high density single nucleotide polymorphisms



Selam Meseret^{a,e}, Yonatan A. Mekonnen^b, Bertram Brenig^{c,d}, Ekkehard Schütz^c, Olivier Hanotte^e, Mehmet Gültas^{b,d}, Armin O. Schmitt^{b,d,*}

^a Ethiopian Biotechnology Institute, P. O. Box 5954, Addis Ababa, Ethiopia

^b Universität Göttingen, Department of Animal Sciences, Breeding Informatics Group, Margarethe von Wrangell-Weg, 37083 Göttingen, Germany

^c Universität Göttingen, Institute of Veterinary Medicine, Burckhardtweg 2, 37077 Göttingen, Germany

^d Center for Integrated Breeding Research, Albrecht Thaer-Weg 3, 37075 Göttingen, Germany

^e LiveGene, International Livestock Research Institute, P. O. Box 5689, Addis Ababa, Ethiopia

ARTICLE INFO

Keywords:

Ethiopian indigenous cattle
Genetic diversity
Genetic structure
Genetic differentiation
Geographical distances

ABSTRACT

The aim of this study was to investigate the genetic diversity and population structure of six Ethiopian cattle breeds. Illumina's Bovine HD BeadChip was used to genotype 76 DNA samples from indigenous cattle breeds, which were named after their regions of origin, namely Benishangul ($n = 12$), Fogera ($n = 11$), Gindeberet ($n = 19$), Gofa ($n = 10$), Nuer ($n = 13$), and Sheko ($n = 11$). Additionally, two comparative European breeds, namely Holstein Friesian ($n = 15$) and Jersey ($n = 15$), were included in the analysis as outgroups. The relatively high within breed genetic diversity of the European and the Ethiopian indigenous cattle breeds was clearly evidenced by their allelic richness (1.747–1.879), effective number of alleles (1.437–1.522), observed heterozygosity (0.271–0.313), nucleotide diversity (0.264–0.315), and inbreeding coefficients (–0.025–0.051). Wright's fixation index F_{ST} and identity-by-state genetic distance revealed considerable genetic distinctiveness of the Ethiopian indigenous cattle compared to the European breeds. The F_{ST} values between Ethiopian cattle breeds were at the low side of the range (less than 0.03). The lowest F_{ST} values (0.002) were found between Benishangul, Fogera and Gindeberet, and the highest between Sheko and the other five Ethiopian breeds. Analysis of molecular variance (AMOVA) showed that the Ethiopian indigenous cattle breeds have maintained a high level of within-population genetic differentiation (explaining 97% of the molecular variability), the remainder being accounted for by differentiation among breeds (3%). The geographical distances and the F_{ST} values had a moderate positive correlation coefficient (0.283). The inter-breed comparisons and structure analysis clearly revealed that Sheko and Nuer were quite separate from the other indigenous breeds and that Benishangul, Fogera, Gindeberet and Gofa are genetically very closely related breeds. Their genetic closeness could be due to high within breed genetic variability, due to a common ancestor, or because of lack of selection intensity. Our results, particularly for the Benishangul, Gindeberet, Gofa and Nuer breeds, are to our knowledge the first of this kind and could be used as a foundation for further investigations of breed characterization or for the design of conservation strategies.

1. Introduction

Ethiopia exhibits a great diversity of ecosystems and abundant cattle resources comprising around 60 million animals (CSA, 2017). Most of these animals belong to indigenous breeds or populations which are

distributed across diverse topographic and climatic conditions in the country. The number of registered cattle breeds in Ethiopia amounts to 25 (IBC, 2004). The efforts to identify and characterize the indigenous cattle resources of Ethiopia were up to date limited. The registered breeds were mostly named after their original geographical region or

* Corresponding author.

E-mail addresses: s.meseret@cgiar.org (S. Meseret), yonatan-ayalew.mekonnen@agr.uni-goettingen.de (Y.A. Mekonnen), bbrenig@gwdg.de (B. Brenig), ekkehard.schuetz@agr.uni-goettingen.de (E. Schütz), olivier.hanotte@nottingham.ac.uk (O. Hanotte), gueltas@cs.uni-goettingen.de (M. Gültas), armin.schmitt@uni-goettingen.de (A.O. Schmitt).

<https://doi.org/10.1016/j.livsci.2020.103979>

Received 2 April 2019; Received in revised form 14 February 2020; Accepted 25 February 2020

1871-1413/ © 2020 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

after the herder's ethnic group. Due to migration, a breed originating from one region could be found in another region at a later time point, potentially resulting in the assignment of a new name in another study. This could give rise to multiple names for one and the same breed or to the assignment of one and the same name to several distinct breeds.

Knowledge about the genetic diversity and the population structure of indigenous cattle is crucial to precisely define breeds, design sustainable genetic improvement programs, and to develop utilization and conservation strategies (Groeneveld et al., 2010). Recently, high density genotyping arrays became available which allow to study the genetic variability and structure between individuals and breeds (Pérez-Pardal et al., 2016; Mekonnen et al., 2019). Ethiopia's livestock is characterized by an enormous genetic variability whose analysis is still in its bare infancy. Thus, to harness this genetic diversity for the improvement of livestock genetics, genotyping of the cattle populations and breeds represented in Ethiopia is an indispensable prerequisite.

In previous studies, microsatellite DNA markers have been used for assessing the genetic diversity and relatedness between individuals and breeds in different livestock species (Dadi et al., 2008; Zerabruk et al., 2011), but very few studies have been conducted using high density SNP markers (Edea et al., 2013; Edea et al., 2015). Here, we investigated the genetic diversity and population structure of six cattle breeds from Ethiopia, namely Benishangul, Fogera, Gindeberet, Gofa, Nuer, and Sheko, using the genome wide single nucleotide polymorphisms (SNP) represented on Illumina's BovineHD BeadChip (~780,000 SNPs). The Benishangul and Gindeberet populations have not yet been registered as breeds in the Domestic Animal Genetic Resources Information System (DAGRIS) database (<http://dagris.info/countries/192/breeds>). Nevertheless, we refer to them as 'breeds' in this text for the sake of simplicity. To the best of our knowledge, there is no research publication on the genotypic characterization of the indigenous cattle breeds under investigation in this study, except for Fogera and Sheko (Edea et al., 2013; Edea et al., 2015).

2. Materials and methods

2.1. Sample collection, preparation of genomic DNA and genotyping

The samples were randomly collected from 76 unrelated animals, based on the information obtained from the livestock keepers, of six indigenous cattle breeds in Ethiopia. Blood samples were taken from 12 Benishangul, 11 Fogera, 19 Gindeberet, 10 Gofa, 13 Nuer, and 11 Sheko animals. The samples were collected using EDTA VACUETTE® tubes (Bio-One Kremsmünster, Austria) and DNA was extracted using DNeasy Blood and Tissue Kits (Qiagen, Hilden, Germany), following the manufacturer's recommendations. The DNA samples were genotyped with the BovineHD BeadChip (800 K) according to Illumina's standard protocol. The International Livestock Research Institute (ILRI), Addis Ababa, Ethiopia, and the Göttingen University undertook the sample collection, extraction and genotyping. As a reference, 15 Holstein Friesian and 15 Jersey (Bahbahani et al., 2017) genotyped with the same Illumina BovineHD BeadChip were included.

2.2. Quality control

PLINK v1.9 (Purcell et al., 2007) was used to control the quality of the genotyping data. SNPs without known genomic position were excluded from the analysis. Only autosomal chromosomes were included, and the total number of such variants was 735,293. SNPs with an individual call rate of 100% were retained for the subsequent analysis. Furthermore, we removed SNPs with a minor allele frequency (MAF) of less than 1% and with a significant deviation from Hardy-Weinberg

equilibrium (HWE) ($P < 0.001$). By accepting only SNPs that could be called in all animals 348,149 SNPs were removed. By filtering for MAF and HWE, 10,456 and 79 SNPs were removed, respectively. The number of SNPs remaining after the quality control was 337,614 out of initially 777,960 SNPs (43%).

2.3. Estimates of within population genetic diversity

Five measures of genetic variability within populations were used to compare the levels of heterogeneity within the cattle breeds (allelic richness (A_R), effective number of alleles (N_e), observed heterozygosity (H_o), inbreeding coefficient (F), and nucleotide diversity (π)). A_R , N_e and H_o were estimated for each breed using the GenoDive 3.0 package (Meirmans and van Tienderen, 2004), while F was calculated with PLINK v1.9 (Purcell et al., 2007) under the default setting. Moreover, π was determined using VCFtools (Danecek et al., 2011).

2.4. Genetic distance and population differentiation

The pairwise identity-by-state (IBS) distances between all breeds were calculated using PLINK v1.9. Additionally, genetic distances between individuals and breeds were evaluated based on Nei's (1987) unbiased genetic distance using the R-package StAMPP (Pembleton et al., 2013). A neighbor-joining relationship tree was constructed using Figtree V1.4.4 (<http://tree.bio.ed.ac.uk/software/figtree/>) on the basis of Nei's genetic distance matrix. To assess pairwise genetic differentiation the fixation index F_{ST} (Weir and Cockerham, 1984) was calculated for all pairs of cattle breeds. Furthermore, analysis of molecular variance (AMOVA; Excoffier et al., 1992) was performed using the GenoDive 3.0 package (Meirmans and van Tienderen, 2004). AMOVA was performed in four ways: (i) six Ethiopian cattle breeds, (ii) four Ethiopian cattle breeds that are highly related according to their F_{ST} values (Benishangul, Gindeberet, Gofa and Fogera), (iii) all eight cattle breeds and (iv) Holstein-Friesian and Jersey as a group of European breeds and the six indigenous breeds as an Ethiopian group. For the first three analyses, the breeds were not assigned to groups.

2.5. Geographical distance and isolation by distance

For this study, the geographical coordinates, longitude and latitude, of the regions' centers were taken as points of origin of the respective breeds. Thus, the geographical distances between the breeds' geographical origins were calculated as the beeline distance with the GenoDive 3.0b program. The geographical distances and the pairwise F_{ST} values were compared using Mantel's test to understand the role of geographical distance as driving force for genetic differentiation.

2.6. Population structure

Population structure was analyzed for $K = 2$ to 8 hypothetical ancestral populations using ADMIXTURE V1.3.0 (Alexander et al., 2009) and the cross-validation (CV) error was used to estimate the most likely number (K) of underlying ancestral populations of the present populations.

3. Results and discussion

3.1. Within breed diversity

Table 1 presents measures of within breed genetic diversity for the eight cattle breeds. The Gindeberet breed had the highest allelic

Table 1
Sample size and genetic diversity measures within the eight cattle breeds under investigation.

Breed	N ^a	A _R ^b	N _e ^c	H _o ^d	π ^e	F ^f
Holstein Friesian	15	1.823	1.516	0.313	0.308	-0.017
Jersey	15	1.747	1.437	0.271	0.264	-0.025
Gindeberet	19	1.879	1.493	0.291	0.298	0.051
Sheko	11	1.845	1.515	0.309	0.313	-0.008
Nuer	13	1.869	1.522	0.313	0.315	-0.019
Benishangul	12	1.855	1.496	0.302	0.304	0.017
Gofa	10	1.837	1.494	0.302	0.305	0.015
Fogera	11	1.839	1.486	0.298	0.298	0.03

^a Total number of individuals sampled in each breed.

^b Allelic richness.

^c Effective number of alleles.

^d Observed heterozygosity.

^e Nucleotide diversity.

^f Inbreeding coefficient.

richness (A_R = 1.879) while Jersey had the lowest allelic richness (A_R = 1.747). The effective number of alleles was highest in Nuer followed by Holstein Friesian and Sheko. Looking across all six Ethiopian breeds, Nuer and Sheko cattle had the highest level of gene diversity (H_o and π), which was lowest in Jersey. The level of inbreeding was close to zero ranging from -0.025 (Jersey) to 0.051 (Gindeberet) which implies that there is still considerable genetic variability in the breeds under investigation. In general, our results showed that the within breed diversity values are comparable for European and Ethiopian cattle breeds.

3.2. Genetic differentiation between breeds and the role of geographical distance

In general, the fixation indices between the Ethiopian breeds were lower than those between the European and the Ethiopian breeds (Table 2). *F*_{ST} was at most 0.030 among the Ethiopian breeds (Gindeberet and Sheko). The *F*_{ST} values were statistically significantly greater than 0 at the level α = 0.05, except for the pairs Benishangul - Gindeberet, Fogera - Gindeberet, Fogera - Benishangul, and Gofa - Benishangul. The p-values were determined empirically (10,000 permutations for a random subset of 10,000 markers) and were corrected for

Table 2

Wright's fixation indices (*F*_{ST}) (below the diagonal) and Nei's genetic distances (above the diagonal) for all pairs of populations under investigation.

	HF ^a	JE ^b	GB ^c	SH ^d	NU ^e	BE ^f	GO ^g	FO ^h
HF ^a		0.099	0.211	0.175	0.182	0.208	0.211	0.213
JE ^b	0.172		0.239	0.202	0.208	0.235	0.238	0.240
GB ^c	0.293	0.340		0.030	0.022	0.015	0.020	0.016
SH ^d	0.245	0.299	0.030		0.028	0.031	0.034	0.034
NU ^e	0.253	0.304	0.018	0.019		0.025	0.028	0.027
BE ^f	0.284	0.335	0.002	0.025	0.015		0.022	0.020
GO ^g	0.284	0.338	0.007	0.028	0.018	0.005		0.025
FO ^h	0.290	0.342	0.002	0.031	0.018	0.002	0.009	

^a Holstein Friesian.

^b Jersey.

^c Gindeberet.

^d Sheko.

^e Nuer.

^f Benishangul.

^g Gofa.

^h Fogera.

Table 3

Analysis of molecular variance (AMOVA) for eight Ethiopian and European cattle breeds.

	Variance component (%)		
	Among groups (p-value)	Among populations (p-value)	Within populations
(i) Six Ethiopian breeds	-	2.9 (0.001)	97.1
(ii) Four closely related Ethiopian breeds	-	0.8 (0.001)	99.2
(iii) All eight breeds under investigation	-	30 (0.001)	70
(iv) Two groups, Ethiopian and European breeds	35.3 (0.038)	7.9 (0.001)	56.8

multiple testing according to Bonferroni. The lowest *F*_{ST} value (0.002) was found among the breeds Benishangul, Fogera, and Gindeberet. As expected, the greatest divergence (0.342) was found between an Ethiopian indigenous and a European breed, namely between Fogera and Jersey. Specifically, Sheko had the highest *F*_{ST} values among all indigenous breeds followed by Nuer. Slightly lower *F*_{ST} values (between 0.002 and 0.014) were reported for the five Ethiopian indigenous cattle breeds and populations, respectively, Ambo, Borana, Arsi, Horro, and Dankil (Edea et al., 2013). Zerabruk et al. (2011) reported similar values (0.001–0.027) between seven north Ethiopian cattle breeds on the basis of 20 microsatellites. Edea et al. (2015) reported nearly the same *F*_{ST} value (0.030) between Fogera and Sheko. In general, the low level of differentiation between the Ethiopian cattle breeds could be attributed to common recent ancestry, admixture from other populations, and lack of selection. The Nei's genetic distances between the breeds followed a similar pattern as the *F*_{ST} values (Table 2). The average genetic distance between individuals drawn from different breeds ranged from 0.015 (Benishangul versus Gindeberet) to 0.239 (Gindeberet versus Jersey). The correlation between the geographic distance (Table S1) and the *F*_{ST} values was positive (0.283), albeit statistically only marginally significant (*p* = 0.095, 1000 bootstrap permutations). A slightly smaller Pearson's correlation coefficient between genetic and geographic distances (*r* = 0.22), which was statistically not significant at the level of α = 0.05, was reported for five Ethiopian cattle breeds using random amplified polymorphic DNA markers (Hassen et al., 2007).

3.3. Analysis of molecular variance (AMOVA)

AMOVA revealed the percentage of genetic variability between and within breeds (Table 3). The AMOVA results illustrated that within breed genetic variance accounted for about 97% of the total variance among six Ethiopia breeds. Edea et al. (2013) reported an even higher within population genetic variance (99%) for the Ethiopian breeds Ambo, Borana, Arsi, Horro, and Danakil. The genetically closely related breeds Benishangul, Fogera, Gindeberet, and Gofa showed very high within population variance (more than 99%) and only 0.8% among populations variance. AMOVA of the eight Ethiopian and European breeds under investigation showed that there was considerable (30%) variance among breeds, and the variance increased to ~35% when the breeds were grouped into an Ethiopian (indigenous) and a European group. In this case, only 8% of the total variance was explained by variance among populations.

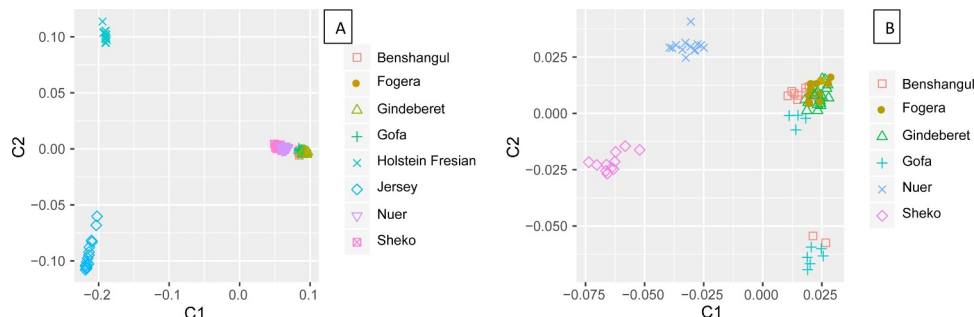


Fig. 1. The identity by state (IBS) distance based multidimensional scaling plot of 96 animals from a total of eight European and Ethiopian cattle breeds is shown in A and 66 samples from six Ethiopian indigenous breeds is shown in B.

3.4. Multidimensional scaling and population structure

We used multidimensional scaling (MDS), based on pairwise identity-by-state distance to analyze the relationship between eight cattle breeds (Fig. 1). The identity-by-state distance between breeds clearly separated the European breeds from the indigenous breeds. This result agrees with reports from Zerabruk et al. (2011) and Edea et al. (2013). MDS performed for only indigenous breeds showed clear clusters of indigenous breeds. Above all, Sheko and Nuer were clearly separated from the other Ethiopian breeds whilst the plot shows one cluster containing the Benishangul, Fogera, Gindeberet, and Gofa breeds. Five Gofa animals and two Benishangul animals, however, drop out of this cluster and seem to form a separate cluster. In general, there is clear evidence that Benishangul, Fogera, Gindeberet, and Gofa share the same genetic background. The cross validation results generated from ADMIXTURE runs (Fig. 2) gave an optimal number of $K = 4$ hypothetical ancestor populations according to the lowest cross-validation error (0.530). ADMIXTURE allows the assignment of individuals to groups based on their genetic similarities, thereby providing information about the most likely number of ancestral populations and the proportion of their contribution. The proportions of individuals in each breed inferred by ADMIXTURE are presented in Table 4 for $K = 2, 3,$ and 4. The result obtain for $K = 4$ revealed that practically the whole Holstein-Friesian genome was assigned to cluster 1, 97% of Jersey was assigned to cluster 2, with 3% of its genome being assigned to cluster 1. About 99% of the genome of Fogera and Gindeberet were assigned to cluster 3. The genome of Benishangul was assigned to clusters 3 (91.7%) and 4 (8.3%). A similarly high genome proportion of Gofa (92%) was assigned to cluster 3 and 7.6% to cluster 4. Nearly 100% of Sheko was assigned to cluster 4 whilst 62% of Nuer was assigned to

Table 4

Proportion of membership of the analyzed eight breeds in each of the $K = 4$ ancestral groups as inferred by the ADMIXTURE program. The highest proportion is highlighted in bold face for each breed and population, respectively.

Breed	Numbers of inferred cluster			
	1	2	3	4
Holstein Friesian	0.997	0.003	0	0
Jersey	0.029	0.971	0	0
Benishangul	0	0	0.917	0.083
Fogera	0	0	0.989	0.011
Gindeberet	0	0	0.989	0.011
Gofa	0.001	0.001	0.922	0.076
Nuer	0.001	0.001	0.377	0.621
Sheko	0	0	0.003	0.997

cluster 4, with 37% being assigned to cluster 3. With $K = 2$ putative origins, the estimated main ancestry of Ethiopian cattle breeds was *Bos indicus* with a marginal contribution from the putatively European *Bos taurus* ancestor, which was found especially in the Benishangul, Fogera, Gindeberet, and Gofa breeds. Fig. 2 reveals that Nuer and Sheko share common ancestry across more than half of their genomes ($K = 2$ and 3). The other indigenous cattle breeds had some admixture from cluster 4. Interestingly, despite the long period of time that has passed since the importation of European breeds to Ethiopia the admixture from European cattle to the indigenous cattle was still negligible. Additionally, we constructed a neighbor-joining tree based on Nei's individual genetic distances and the result is in line with MDS and ADMIXTURE analysis (Fig. S1).

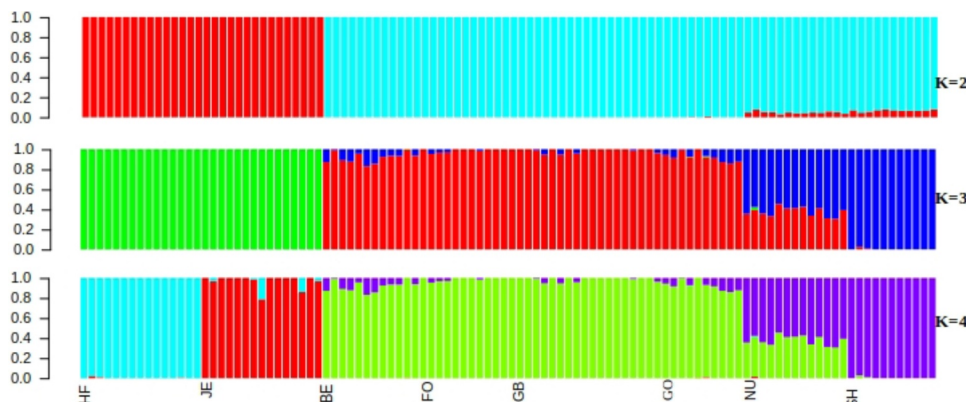


Fig. 2. Population structure analysis for $K = 2, 3,$ and 4 clusters. $K = 4$ turned out to be the most likely number according to the cross-validation error. The genetic composition is shown for each animal of the Holstein Friesian (HF), Jersey (JE), Benishangul (BE), Fogera (FO), Gindeberet (GB), Gofa (GO), Nuer (NU) and Sheko (SH) breeds. The proportion of the ancestral contribution is shown on the y-axis.

4. Conclusion

This study revealed that the Ethiopian breeds under investigation are genetically clearly differentiated from European breeds. The indigenous Ethiopian cattle breeds have maintained high levels of within breed genetic diversity. Particularly Sheko and Nuer are clearly distinct Ethiopian breeds, whilst Benishangul, Fogera, Gindeberet and Gofa are very closely related. It needs to be discussed further if it is justified to consider them as separate breeds. The small number of animals in the range between 10 and 19 per breed used in this study is obviously a limitation. However, we believe that the major results will not be heavily compromised by this since they are supported by distinct lines of analysis. To corroborate our results, it would be desirable to consider also morphological and phenotypic characteristics of the indigenous cattle breed.

Author contributions

SM performed the analyses and wrote the manuscript, YAM collected a part of the samples and provided general advice, BB and ES contributed to the genotyping, OH provided samples and genotyping data, MG provided support to the analysis, AOS conceived the study.

Data availability statement

The genotypic data for the Ethiopian breeds used in this work are available under study number PRJEB34751 of the European Variant Archive (EVA). The genotypic data of the Holstein Friesian and Jersey breeds can be obtained from the Dryad Digital Repository under <https://doi.org/10.5061/dryad.38jp6>.

Declaration of Competing Interest

The authors declare no competing interests.

Acknowledgement

Selam Meseret was supported as a visiting scientist to the Breeding Informatics group by the German Research Foundation (DFG) in the framework of the World Academy of Sciences (TWAS) program.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.livsci.2020.103979](https://doi.org/10.1016/j.livsci.2020.103979).

References

- Alexander, D.H., Novembre, J., Lange, K., 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 19, 1655–1664. <https://doi.org/10.1101/gr.094052.109>.
- Bahbahani, H., Tijjani, A., Mukasa, C., Wragg, D., Almathen, F., Nash, O., Akpa, G.N., Mbole-Kariuki, M., Malla, S., Woolhouse, M., Sonstegard, T., Van Tassell, C., Blythe, M., Huson, H., Hanotte, O., 2017. Signatures of selection for environmental adaptation and Zebu x taurine hybrid fitness in East African Shorthorn Zebu. *Front. Genet.* 8, 68. <https://doi.org/10.3389/fgene.2017.00068>.
- CSA, 2017. Federal democratic republic of Ethiopia (Central statistical agency), Agricultural sample survey report on livestock and livestock characteristics.
- Dadi, H., Tibbo, M., Takahashi, Y., Nomura, K., Hanada, H., Amano, A., 2008. Microsatellite analysis reveals high genetic diversity but low genetic structure in Ethiopian indigenous cattle populations. *Anim. Genet.* 39, 425–443. <https://doi.org/10.1111/j.1365-2052.2008.01748.x>.
- Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., Durbin, R., 1000 Genomes Project Analysis Group, 2011. The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>.
- Edea, Z., Bhuiyan, M.S.A., Dessie, T., Rothschild, M.F., Dadi, H., Kim, K.S., 2015. Genome-wide genetic diversity, population structure and admixture analysis in Africa and Asian cattle breeds. *Animal* 9, 218–226. <https://doi.org/10.1017/S1751731114002560>.
- Edea, Z., Dadi, H., Kim, S., Dessie, T., Lee, T., Kim, H., Kim, J.J., Kim, K.S., 2013. Genetic diversity, population structure and relationships in indigenous cattle populations of Ethiopia and Korean Hanwoo breeds using SNP markers. *Front. Genet.* 4, 35. <https://doi.org/10.3389/fgene.2013.00035>.
- Excoffier, L., Smouse, P.E., Quattro, J.M., 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131, 479–491.
- Groeneveld, L.F., Lenstra, J.A., Eding, H., Toro, M.A., Scherf, B., Pilling, D., Negrini, R., Finlay, E.K., Jianlin, H., Groeneveld, E., Weigend, S., GLOBALDIV Consortium, 2010. Genetic diversity in farm animals – a review. *Anim. Genet.* 41, 6–31. <https://doi.org/10.1111/j.1365-2052.2010.02038.x>.
- Hassen, F., Bekele, E., Ayalew, W., Dessie, T., 2007. Genetic variability of five indigenous Ethiopian cattle breeds using RAPD markers. *Afr. J. Biotechnol.* 6, 2274–2279. <https://doi.org/10.5897/AJB2007.000-2356>.
- Institute of Biodiversity Conservation, 2004. The State of Ethiopia's Farm Animal Genetic Resources: Country Report. A Contribution to the First Report on the State of the World's Animal Genetic Resources. IBC, Addis Ababa, Ethiopia May 2004.
- Meirmans, P.G., Van Tienderen, P.H., 2004. GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. *Mol. Ecol. Notes* 4, 792–794. <https://doi.org/10.1111/j.1471-8286.2004.00770.x>.
- Mekonnen, Y.A., Gültas, M., Effa, K., Hanotte, O., Schmitt, A.O., 2019. Identification of candidate signature genes and key regulators associated with trypanotolerance in the Sheko Breed. *Front. Genet.* 10, 1095. <https://doi.org/10.3389/fgene.2019.01095>.
- Pembleton, L.W., Cogan, N.O.L., Forster, J.W., 2013. StAMPP: an R package for calculation of Genetic differentiation and structure of mixed-ploidy level populations. *Mol. Ecol. Resour.* 13, 946–952. <https://doi.org/10.1111/1755-0998>.
- Pérez-Pardal, L., Saglam, I.K., Costa, V., Chen, S., Miller, M.R., Beja-Pereira, A., 2016. Comparison of high-density SNP chip versus RAD sequencing in cattle and related species. *J. Anim. Sci.* 94 (Issue suppl 4), 23. <https://doi.org/10.2527/jas2016.94supplement423x>.
- Weir, B.S., Cockerham, C.C., 1984. Estimating F-statistics for the analysis of population structure. *Evolution* 38, 1358–1370. <https://doi.org/10.2307/2408641>.
- Zerabruk, M., Li, M.H., Kantanen, J., Olsaker, I., Ibeagha-Awemu, E.M.G., Erhardt, G., Vangen, O., 2011. Genetic diversity and admixture of indigenous cattle from North Ethiopia: implications of historical introgressions in the gateway region to Africa. *Anim. Genet.* 43, 257–266. <https://doi.org/10.1111/j.1365-2052.2011.02245.x>.

A.3. Evolutionary Responses Associated with Trypanotolerance in Three Ethiopian Breeds – A Comparative Genomics Approach

Evolutionary Responses Associated with Trypanotolerance in Three Ethiopian Cattle Breeds – A Comparative Genomics Approach

Yonatan Ayalew Mekonnen¹, Mehmet Gültas^{1,2}, Olivier Hanotte^{3,4} and Armin O. Schmitt^{1,2}

¹ *Breeding Informatics Group, Department of Animal Sciences, University of Göttingen, Göttingen, Germany*

² *Center for Integrated Breeding Research (CiBreed), University of Göttingen, Göttingen, Germany*

³ *Cells, Organisms and Molecular Genetics, School of Life Sciences, University of Nottingham, Nottingham, United Kingdom*

⁴ *LiveGene, International Livestock Research Institute (ILRI), Addis Ababa, Ethiopia*

Correspondence*:

Armin O. Schmitt

armin.schmitt@uni-goettingen.de

ABSTRACT

Trypanosomiasis in Ethiopia is a cattle disease that is posing a major challenge in rearing livestock. However, for generations, cattle breeds such as Benshangul, Gindeberet, and Nuer are kept in the infected regions without proper prevention and medication. These breeds might display some degree of trypanotolerance due to adaptive responses against this particular environmental pressure. However, the tolerance levels of Gindeberet, Nuer, and Benshangul are not well investigated. We perform genome wide signature gene identification using within-population (*iHS* and CLR) and between-population (*Rsb*) methods for these breeds. The identified genes were used to determine functionally enriched GO terms. Based on the enriched gene lists, we perform (i) gene network search using GeneMANIA in the Cytoscape platform, (ii) overrepresented pathways and master regulator search using TRANSPATH in the geneXplain platform to reveal the genetic controls of trypanotolerance. The trypanotolerance tendencies of the Nuer, Benshangul, and Gindeberet breeds are assessed by comparing the candidate genomic regions, genes, hub genes, overrepresented pathways, and master regulators identified in each breed with those of the trypanotolerant Sheko breed. The Sheko breed is used in this paper to comparatively identify the trypanotolerance tendencies of the Nuer, Benshangul, and Gindeberet breeds. We find that the polymorphisms in the FTMT, RAET1G, and PPP1R14C genes which are shared by Benshangul, Gindeberet, Sheko, and Nuer breeds, are related to oxidative stress, anemia, and immune responses based on a literature survey. Moreover, we identify the hub gene KCNIP4 common to Sheko, Benshangul, and Gindeberet which is involved in neurological dysfunctions. Our results also reveal the common overrepresented pathways in Gindeberet, Sheko, and Benshangul that are involved in immune responses and oxidative stresses. Furthermore, the hub genes, overrepresented pathways, and master regulators in the Nuer, Benshangul, and Gindeberet breeds which are in common with Sheko are identified. These results indicate that the Nuer, Benshangul, and Gindeberet breeds have undergone similar evolutionary responses against trypanosomiasis. Strikingly, the top master regulator Caspase is identified commonly in Sheko, Gindeberet, and Benshangul which might be involved in governing the evolutionary responses against trypanosomiasis. Our findings might be used as an input for genetic interventions to enhance the performance of trypanosusceptible cattle. Moreover, the identification of master regulator Caspase suggests potential candidate therapeutic targets for the development of new drugs for trypanosomiasis treatment.

Keywords: comparative genomics, gene networks, master regulators, overrepresented pathways, signature genes, trypanosomiasis, trypanotolerance

1 INTRODUCTION

Trypanosomiasis is a disease caused by a uni-cellular protozoan parasite. This disease has become a major challenge for livestock keepers in sub-Saharan African countries (Yaro et al., 2016). Ethiopia is one of the sub-Saharan African countries where trypanosomiasis is a major challenge for keeping livestock. The infection zone in the country ranges from south western to north western regions covering 22,000 km² (NTTICC, 2004; Andrew, 2004). The disease is transmitted from infected animals to healthy animals by the vector tsetse fly (Welburn et al., 2016). The infected animal might show symptoms such as severe anemia, weight loss and reduced productivity. If it is not treated, it can cause serious problems that could eventually lead to death. In Ethiopia, cattle are used in farming processes such as pulling plows which is the initial cultivation process of sowing seed. If the animal is infected with trypanosome, it loses the power to pull plows and the whole cultivation processes could be interrupted which entails decreasing crop production (Nantulya, 1986; Noyes et al., 2011).

There have been numerous measures to control the spread of this disease such as chemotherapy, insecticides, insect traps, and insecticide-treated cattle (Slingenbergh, 1992; Barrett and Stanberry, 2009; Leak et al., 1996). But none of these controlling measures have provided the optimum solution for preventing the spread of this disease. In addition, wild animals function as a reservoir of the parasite that leads to the deteriorating conditions towards controlling the spread of the disease (NTTICC, 2004; Codjia et al., 1993; Mulugeta et al., 1997; Nantulya, 1986). However, there are naturally trypanotolerant breeds which are kept in the infected regions without major health and production trait performance issues. Moreover, cattle that are kept in the tsetse infested regions for several generations might display some degree of trypanotolerance due to the environmental pressure (Hanotte et al., 2003; Bahbahani et al., 2018) (Hanotte, 2003; Bahbahani, 2018). A study on the tolerance level of the indigenous Ethiopian breed Sheko showed that it has better trypanotolerance attributes than other breeds such as Horro and Gurage (Lemecha et al., 2006). The Nuer, Benshangul and Gindeberet breeds are also kept in tsetse infested regions with a trypanosome challenge (Tasew and Duguma, 2012; Leta et al., 2016). Therefore, these breeds might also display a certain level of trypanotolerance. Therefore, designing a cross between breeds with a certain level of trypanotolerance together with vector control could provide an effective strategy for controlling the spread of this disease (Lutje et al., 1996). Hence, designing a marker assisted selection strategy will significantly contribute to improve the health of the cattle. However, this requires the identification of genes and genomic regions controlling trypanotolerance. Despite the fact that there have been some efforts to identify genes and genomic regions controlling trypanotolerance in N'Dama (Hanotte et al., 2003; Kim et al., 2017), reports on the genetic controls and mechanisms of trypanotolerance in cattle which are necessary for the implementation of marker assisted selection strategies are limited. Therefore, the aim of

this study is to reveal the genetic controls and regulatory mechanisms that are involved in trypanotolerance tendencies in the Nuer, Gindeberet, and Benshangul cattle breeds.

In order to study their tolerance levels and mechanisms of these breeds, we consider trypanosomiasis as biotic stress for these breeds that occurs in their natural habitats. This selection pressure plays a critical role in the evolutionary adaptive responses which leads to trypanotolerant breed formation. The molecular mechanism for such breed formation could be a genetic drift that results in an increase in the frequency of beneficial alleles in a population. This means that under positive selection, strong and long range linkage disequilibrium (LD) and unexpectedly high local haplotype homozygosity might occur in the genome (Gautier and Vitalis, 2012; Bomba et al, 2015). Therefore, unexpectedly high local homozygosity in the genome of a given breed is identified as a signature of selection. In this paper, we use selection signature detecting methods, extended haplotype homozygosity (EHH), and composite likelihood ratio (CLR) methods, to comparatively identify signatures of positive selection in the genomes of the Nuer, Gindeberet, and Benshangul breeds. Combining the spatial distribution of allele frequency and EHH based methods increase the power of the analyses to identify the signature of positive selection in the genome of the Nuer, Benshangul, and Gindeberet breeds than a single analysis (Ma et al., 2015; Vatsiou et al., 2016; Mekonnen et al., 2019).

Moreover, to understand the interactions between genes (gene networks) and to gain more insights into the regulatory mechanisms of trypanotolerance, the combined gene sets identified by *iHS*, CLR, *Rsb* methods for each breed are used in the analysis pipeline to construct: i) Gene networks to identify hub genes using the GeneMANIA prediction server in the Cytoscape platform (Warde-Farley et al., 2010); ii) overrepresented pathways and master regulators using the TRANSPATH database in the geneXplain platform (Wlochowitz et al., 2016). To identify the trypanotolerance tendencies of the Nuer, Benshangul, and Gindeberet breeds, the published signature genes, overrepresented pathways, and master regulators identified in the trypanotolerant Sheko (Mekonnen et al., 2019) are compared.

2 MATERIAL AND METHODS

2.1 Cattle Population, SNP Genotyping and Quality Control

We used published genotyping data of 12 Sheko, 13 Nuer, 19 Gindeberet, and 12 Benshangul (Mekonnen et al., 2019). In addition, we obtained genotyping data from the International Livestock Research Institute (ILRI) consisting of 25 Ankole, 12 Serere, 23 Nganda, 16 Karamojong, and 92 East African Shorthorn Zebu (EASZ) (Bahbahani et al., 2017). Plink 1.9 (Purcell et al., 2007) was used for quality controls. Due to the minor allele frequency criterion ($MAF < 1\%$), 19,581 SNPs were excluded. Two Benshangul samples were excluded from the analysis due to genotyping call rate ($<95\%$) filtering criterion. But no pair of

samples were excluded due to the identity by state (IBS) (>95%) criterion. The total sample size for downstream analysis consists of 222 samples and 715,712 SNPs. BEAGLE 4 (Browning and Browning, 2007) was used for inferring haplotype phasing and imputing the missing alleles. The breed origin and type of cattle samples included in this study are presented in Table 1.

Table 1 Cattle breeds included in the study. Modified from Mekonnen et al. (2019)).

Breed name	*Breed category	Breed origin	No. of animals	Trypanosusceptible
Ankole	Sanga	Uganda	25	Yes (Magona et al., 2004)
Karamojong	African zebu	Uganda	16	Yes (Muhanguzi et al., 2017)
Serere	African zebu	Uganda	12	Yes (Ocaido et al., 2005)
Nganda	Zenga	Uganda	23	Yes (FAO, 2004)
EASZ	African zebu	Kenya	92	Yes (Muhanguzi et al., 2014; Van Wyk et al., 2014)
Sheko	African taurine and zebu	Ethiopia	12	No (Lemecha et al., 2006)
Nuer	Sanga	Ethiopia	13	Not available
Gindeberet	Not available	Ethiopia	19	Not available
Benshangul	Not available	Ethiopia	12	Not available
Fogera	African zebu	Ethiopia	11	Yes (Sinshaw et al., 2006)

*Breed category according to DAGRIS (2009).

2.2 Analysis of Signatures of Positive Selection and Selective Sweep

For the detection of signature genes, we used extended haplotype homozygosity (EHH) and composite likelihood ratio (CLR) methods which implement LD based statistics. The EHH estimation is computed from the decay of LD as a function of the increasing distance from the core bi-allelic SNPs (Sabeti et al., 2002). The *iHS* and *Rsb* values were calculated using the R package *rehh* (Gautier and Vitalis, 2012). We used the bovine reference genome (UMD3.1) as the ancestral allele and the study population as a derived allele (Bahbahani et al., 2018; Tijjani et al., 2019; Mekonnen et al., 2019). Therefore, the *iHS* values were calculated from the natural log ratio of integrated EHH (iHH) between reference and alternative alleles for each SNP (Bahbahani et al., 2018; Tijjani et al., 2019). Therefore, the *iHS* is given as:

$$iHS(\text{unstandardized}) = \ln \left(\frac{iHH_A}{iHH_D} \right), \quad (1)$$

To allow the direct comparisons among different SNPs regardless of their allele frequencies, the *iHS* values are standardized using their mean and standard deviation (Voight et al., 2006; Gautier and Vitalis, 2012).

$$iHS(\text{standardized}) = \frac{\ln\left(\frac{iHHA}{iHHD}\right) - E_p\left[\ln\left(\frac{iHHA}{iHHD}\right)\right]}{SD_p\left[\ln\left(\frac{iHHA}{iHHD}\right)\right]}, \quad (2)$$

the empirical distribution of the SNPs with a frequency p of the core SNP is used to estimate the expectation $E_p\left[\ln\left(\frac{iHHA}{iHHD}\right)\right]$ and standard deviation $SD_p\left[\ln\left(\frac{iHHA}{iHHD}\right)\right]$ of the unstandardized iHS (Voight et al., 2006).

The Rsb values were computed as the log ratio of integrated site-specific EHH (iES) between each of Benshangul, Gindeberet, and Nuer breeds against the trypanosusceptible group (EASZ, Ankole, Karamojong, Nganda, Serere, and Fogera breeds). Therefore, integrated iES is defined as the area under the EHHS curve against the distance from the core haplotype to the last haplotype carrying the core SNP. The iES is used to summarize the decay of EHHS for a single SNP site in a population. Hence, the relative integrated site-specific EHH between populations (Rsb) can be given as:

$$Rsb(\text{unstandardized}) = \ln\left(\frac{iES_{pop1}}{iES_{pop2}}\right), \quad (3)$$

where $pop1$ refers to the study population and $pop2$ refers to the reference population. Due to a slower decay of EHH in one population compared to the other, there might be extreme values of Rsb . Therefore, unlike iHS , the standardization of Rsb values uses the median instead of the mean, since the median is less sensitive to extreme values (Tang et al., 2007). Therefore, the standardized Rsb is given as:

$$Rsb(\text{standardized}) = \frac{\ln\left(\frac{iES_{pop1}}{iES_{pop2}}\right) - med\left[\ln\left(\frac{iES_{pop1}}{iES_{pop2}}\right)\right]}{SD\left[\ln\left(\frac{iES_{pop1}}{iES_{pop2}}\right)\right]}, \quad (4)$$

where $med\left[\ln\left(\frac{iES_{pop1}}{iES_{pop2}}\right)\right]$ is the median and $SD\left[\ln\left(\frac{iES_{pop1}}{iES_{pop2}}\right)\right]$ is the standard deviation of the unstandardized Rsb . The median and standard deviation of the unstandardized Rsb are estimated from the empirical distribution of the total set of SNPs (Tang et al., 2007). The significance threshold of $\alpha = 10^{-4}$ was used for both iHS and Rsb P -values following the study of Bahbahani et al. (2018) and Tijjani et al. (2019).

To identify the skewed pattern of the allele frequency spectrum based upon the information from the spatial distribution of the allele frequency, we applied the CLR test using Sweepfinder2 (DeGiorgio et al., 2016). Therefore, the composite likelihood ratio test statistic T is given as:

$$T = 2\{\log CL(\hat{p}_{v \leftrightarrow b}; v \leftrightarrow b) - \log CL(\hat{p}; v \leftrightarrow b)\}, \quad (5)$$

where $\log CL(\hat{p}_{v \leftrightarrow b}; v \leftrightarrow b)$ is the log likelihood of the data under the alternative model, $\log CL(\hat{p}; v \leftrightarrow b)$ is the log likelihood of the data under the null model, and \hat{p} is the estimate of the maximum composite likelihood of p from SNP v to SNP b (Nielsen et al., 2005; Pavlidis et al., 2008).

The Sweepfinder2 software was applied to each chromosome with a window of size 50k in which on average 226 SNPs are analyzed per window. The significant threshold of $\alpha = 10^{-5}$ was identified using the 95th quantile distribution of the top CLR P -values (Wilches et al., 2014).

2.3 Functional Characterization and Gene Network Analysis of Candidate Selected Regions

If a gene was found within 25 kb from the significant SNP, then the gene was considered as a candidate gene (Bahbahani et al., 2018). Protein-coding and RNA genes found within the candidate regions were retrieved using the BioMart tool (Kinsella et al., 2011) (Supplementary Table 1 - 4). All the retrieved gene lists were used to determine the functionally enriched genes. Significantly enriched functional terms ($P < 0.05$) in comparison to the whole bovine reference genome background were identified using the enrichment analysis tool “Enrichr” (Kuleshov et al., 2016). Enriched gene ontology (GO) terms were used to produce a Treemap in the geneXplain platform. Since trypanosomiasis is considered as environmental pressure, the enriched signature genes might work together in a network to govern the genes that are involved in the evolving breed against this disease. In order to identify the gene networks, we applied the GeneMANIA prediction server in the Cytoscape platform (Warde-Farley et al., 2010; Shannon et al., 2003). We considered co-expression, physical interaction, predicted functional relationships, genetic interactions, and co-localization between genes for the network construction.

For the better insights of the functional properties and the regulatory mechanisms involved in trypanotolerance, we used TRANSPATH in the geneXplain platform for the identification of overrepresented pathways and master regulators. The identified hub genes, overrepresented pathways, and master regulators are compared among breeds.

3 RESULT

3.1 Comparative Identification of Signature Genes Associated with Trypanotolerance

The combined gene sets of *iHS*, CLR, and *Rsb* analyses contain 220, 449, 1022, and 439 candidate genes with a signature of positive selection in 22, 27, 29, and 29 autosomes in Sheko (Mekonnen et al.,

2019), Benshangul, Gindeberet, and Nuer, respectively (Figure 1). The comparison of the genomic regions and signature genes among these breeds using the combined gene sets of *iHS*, *CLR*, and *Rsb* analyses reveal that Benshangul and Gindeberet share the biggest overlaps in both genomic regions and genes while Sheko and Benshangul share few genes (– shared genes – referring to genes found in shared haplotypes) (Figure 1). To explore the trypanotolerance tendencies of Benshangul, Nuer, and Gindeberet breeds, the identified signature genes of these breeds are compared with those of Sheko (Mekonnen et al., 2019). Sheko shares 15 genes with Benshangul, 38 genes with Gindeberet and 34 genes with Nuer. On these overlapping gene sets, we perform a literature survey and identify 6 genes in Benshangul (*STXBP3*, *GNAS*, *KCNIP4*, *DHX33*, *C1QBP*, and *AK5*), 3 genes in Gindeberet (*JMJD1C*, *EGR1*, and *SLC24A3*), and 2 genes (*CD86* and *L3MBTL3*) in Nuer which are associated with trypanotolerance attributes. The *STXBP3*, *GNAS*, and *KCNIP4* genes are involved in neuronal dysfunctions. Especially, the *KCNIP4* gene encodes a potassium ion transporter protein which is directly involved in neuronal functions (Castillo-Rodríguez et al., 2018). The polymorphisms in the *KCNIP4* gene are associated with neuro-developmental disorders such as Attention-deficit/hyperactivity disorder (ADHD) (Weißflog et al., 2013). Moreover, the *STXBP3* gene is involved in neurological dysfunction such as hearing loss (Kelsen et al., 2018), whereas the *GNAS* gene is associated with pseudohypoparathyroidism (low level of calcium and a high phosphate level in the blood) (Bastepe, 2008).

Another gene identified in common in both Benshangul and Sheko is *DHX33*. This gene is involved in the regulation of RNA induced NLRP3 inflammasome activation (Mitoma et al., 2013). Inflammasomes facilitate the activation of pro-inflammatory Caspase-1 which catalyzes the maturation of pro-IL-1 β and pro-IL-18. The activated IL-1 β is involved in antigen presenting processes in host defence and in the production of CD4⁺ and CD8⁺ T-cells. These T-cells produce IFN- γ which activates the phagocytic cells to kill invading parasites (Paroli et al., 2018). Moreover, the *C1QBP* is reported as a direct target gene of the *ZNF32* gene which plays a critical role in the cellular resistance to oxidative stress (Li et al., 2015). In addition, the *AK5* gene is involved in the regulation of thymocyte maturation and auto-immunity (Sheridan, 2007; Stanojevic et al., 2008).

Furthermore, the *JMJD1C* and *EGR1* genes are identified in common in Gindeberet and Sheko show upregulation in hypoxic conditions that are induced by the hypoxia-inducible factor (HIF)-1 (Sperandio et al., 2009; Hancock et al., 2015). The *SLC24A3* gene is involved in cellular calcium ion homeostasis and its expression is induced by chronic hypoxia (Gaudet et al., 2011). Moreover, the *ARG1* gene is involved in the defence response against protozoans (GO:0042832). This gene is also involved in parasite-induced tissue damage repairing processes (Gordon, 2003; El Kasmi et al., 2008).

Furthermore, the CD86 gene in Nuer and Sheko plays an important role in T-cell activation and co-stimulation of naive T-cells (Short et al., 2006) while the L3MBTL3 gene is involved in erythropoiesis. The knockout experiment on the L3MBTL3 gene results severe anemia in mice (Arai and Miyazaki, 2005).

3.2 Treemap Comparisons Between Breeds

The biological functions for the combined signature gene sets from each Benshangul, Gindeberet, and Nuer breeds are clustered into a treemap of 30 functional terms (Supplementary Figure 1). Most of these terms are involved in cellular transport, metabolic process, and regulation. Among the 30 enriched functional terms, one of the GO-terms in Sheko and Nuer (Figure 2) is T cell chemotaxis which plays a critical role in immune responses (Gerard and Rollins, 2001; Bach et al., 2007).

3.3 Comparison with the Trypanotolerant N'Dama and Muturu for Signature of Positive Selection against Trypanosomiasis

Among the 55 trypanotolerant QTL identified in N'Dama (Hanotte et al., 2003), 23 regions in Benshangul, 21 regions in Nuer, and 26 regions in Gindeberet overlap (Supplementary Table 5 - 7). Moreover, among the genes that are associated with anemia in N'Dama (Noyes et al., 2011; Kim et al., 2017), RPS26 gene in Benshangul and Gindeberet, GNA14 gene in Gindeberet and Sheko (Mekonnen et al., 2019), LCT and MCM6 genes in Gindeberet, FGF20 gene in Sheko (Mekonnen et al., 2019), and MADCAM1 gene in Benshangul are commonly identified. Moreover, the ENSBTAG00000047764 and ST6GALNAC5 genes in Benshangul, Gindeberet, Nuer, and Sheko (Mekonnen et al., 2019) are commonly identified in the Muturu breed which is reported in the literature (Tijjani et al., 2019). Furthermore, Benshangul and Nuer share the JAZF1, PPP2R2A, and BNIP3L genes while only Benshangul and Sheko (Mekonnen et al., 2019) share the AK5 gene with Muturu (Tijjani et al., 2019).

The ENSBTAG00000047764 gene is orthologous to the human gene ABCC4 which is involved in cellular detoxification (Bai et al., 2004) while the ST6GALNAC5 and JAZF1 genes are involved in the glycolipid biosynthesis (Ikehara et al., 1999; Jang et al., 2014). The PPP2R2A gene is expressed as the sub-unit of the regulatory protein phosphatases 2 (PP2A) which is entangled in a variety of cellular processes involved in the (negative) control of cell growth and division (Janssens et al., 2005). Moreover, the BNIP3L gene is involved in apoptosis and cellular responses to hypoxia (Bellot et al., 2009; Kitamura et al., 2011).

3.4 Comparative Identification of Hub Genes Involved in Trypanotolerance

We performed a gene network search analysis using GeneMANIA and identified the top three hub genes that have the highest connectivity with the nodes as shown in Figure 3 - 6. Among the top three hub genes, the KCNIP4 gene is shared by Gindeberet, Sheko, and Benshangul breeds. The Foxp1 hub gene is found in

both Sheko and Benshangul, while the FYN hub gene is shared by Benshangul and Nuer. Moreover, the PSMD7 hub gene in Sheko, CTNNA2 and RDX hub genes in Gindeberet, AAK1 and SYT1 hub genes in Nuer are unique for each breed. The KCNIP4 gene encodes a potassium ion transporter protein which is directly involved in neuronal functions (Castillo-Rodríguez et al., 2018). The polymorphisms in KCNIP4 gene are associated with neuro-developmental disorders such as Attention-deficit/hyperactivity disorder (ADHD) (Weißflog et al., 2013).

Moreover, the hub gene FOXP1 is involved in the transcriptional regulatory networks of B lymphopoiesis, cardiomyocyte proliferation, motor neuron, and regulatory T-cell functions (Hu et al., 2006; Wang and Morrisey, 2010; Adams et al., 2015; Konopacki et al., 2019). Further, the hub gene FYN is activated by hypoxia (Hu et al., 2014) and interacts with Nrf2 by phosphorylating tyrosine 568 of Nrf2 which is involved in the activation of defence genes during stressful conditions (Kaspar and Jaiswal, 2011). In addition, the hub gene FYN plays a critical role in the initiation of T-cell receptor signaling pathways (Salmond et al., 2009).

Furthermore, mutations in the hub genes RDX and CTNNA2 are associated with neurological dysfunctions such as hearing loss (Flex et al., 2003; Khan et al., 2007). The RDX hub gene is also involved in leukocyte extravasation signaling processes (Kanaan et al., 2010). The other hub gene, AAK1, which is identified in Nuer is involved in hypoxia induced cellular processes (Chen et al., 2006), whereas the hub gene SYT1 functions as calcium sensor in the cell that triggers cellular response to a calcium ion (Yoo et al., 2013). The function of the hub gene PSMD7 identified in Sheko is related with the regulation of auto-reactive immune cells (Lodish et al., 2004; Zinngrebe et al., 2014).

3.5 Comparative Identification of Overrepresented Pathway Involved in Trypanotolerance

The TRANSPATH pathway analysis reveals 15, 29, 18, and 32 genes involved in 13, 20, 12, and 22 overrepresented TRANSPATH pathways in Sheko (Mekonnen et al., 2019), Benshangul, Nuer, and Gindeberet, respectively (Table 2). Sheko and Gindeberet share two overrepresented pathways involved in PDGF B → STATs and ARIP1 → atrophin1 signaling cascades. Gindeberet and Benshangul share two overrepresented pathways involved in E1 –/ ERBB3 and p53 related signaling cascades. The PDGF B → STATs pathway is the topmost overrepresented pathway in Sheko which is involved in the induction of chemokine secretion and plays an important role in innate and adaptive immunity responses (Kim and Broxmeyer, 1999). The other overrepresented pathway shared by Sheko and Gindeberet is ARIP1 → atrophin1 which is involved in erythroid and lymphoid differentiation. Moreover, atrophin1 related pathways play a role in immune responses (Qiu et al., 2000; You et al., 2009; Aki et al., 2015) and are also associated with spinocerebellar degeneration.

The E1 → ERBB3 pathway, shared by Gindeberet and Benshangul, is the second most overrepresented pathway in Benshangul which is involved in the ubiquitin-proteasome system and plays a key role in the degradation of cellular proteins (Nobuyuki et al., 1992). The topmost overrepresented pathway in Benshangul is a platelet-activating factor (PAF) which is related to host defence system (Zimmerman et al., 2002). The other overrepresented pathway commonly shared by Gindeberet and Benshangul is p53 → cytochrome C which is involved in the induction of apoptosis by Caspase activation (Schuler et al., 2000). The two topmost overrepresented pathways in Gindeberet are N-cadherin and parkin related pathways which are involved in immunity and neurological dysfunctions (Springer, 1990; Kitada et al., 1998; Derycke and Bracke, 2004). The N-cadherin related pathway is involved in cell migration and cell-cell contact (Derycke and Bracke, 2004) which play an important role to patrol the body against pathogens (Springer, 1990). The parkin related pathway is related to parkin protein. This protein is associated with a neurological disorder known as autosomal recessive juvenile parkinsonism (AR-JP) (Kitada et al., 1998).

There were no overlapping overrepresented pathways between Nuer and the other three breeds. The first overrepresented pathway in Nuer is Syk → RhoA which is involved in the engulfment of the pathogen by macrophages (Tohyama and Yamamura, 2006), whereas the second overrepresented pathway is acetyl-CoA → palmitic acid which is involved in lipid metabolism (Li et al., 2010).

Table 2 Overrepresented pathway analyses of the four populations.

Pathways	P-value	Genes
Gindeberet		
N-cadherin:plakoglobin, homophilic ligation	0.003	CTNNA1, CTNNA2, JUP
NOS →/ parkin	0.003	NOS2, PRKN
E1 →/ ErbB3	0.008	ERBB3, RNF41, UBE2D2
H2O2 → Pyk2	0.009	PLD2, PTK2B
9-cis-retinol → 9-cis-retinoic acid	0.009	HSD17B6, RDH5
Nrdp1 →/ apollon	0.013	BIRC6, PSMA7, PSMD3, RNF41, UBE2D2
Cdk2 → p53	0.018	CDK2, TP53
ARIP1 → atrophin1	0.024	AKT1, AKT3, APBA1, CTNNA1
N-cadherin network	0.027	CTNNA1, CTNNA2, JUP
citrate cycle	0.028	ACLY, CS, DLST, MDH1
CH000000331	0.028	CTNNA1, JUP
OSM → STAT3	0.028	OSMR, STAT3
EGF → STAT3	0.028	PTK2B, STAT3
AKT-1 →/ JNK1	0.028	AKT1, MAP3K5
p53 → cytochrome C	0.028	AKT1, AKT3
uridine → beta-alanine	0.028	DPYD, DPYS
2'-deoxythymidine → deoxythymidine 5'-monophosphate	0.028	DPYD, DPYS

Table 2 – continued from previous page

Pathways	P-value	Genes
PDGF B → STATs	0.041	STAT3, STAT5A
Bad → 14-3-3	0.041	AKT1, PRKACB
N-cadherin:beta-catenin, homophilic ligation	0.041	CTNNA1, CTNNA2
VE-cadherin, ligation	0.041	CTNNA1, CTNNA2
Nuer		
Syk → RhoA	0.002	RHOA, SYK
acetyl-CoA → palmitic acid	0.002	FASN, PPT1
fatty acid synthesis	0.007	FASN, PPT1
PMCA4 —/ nNOS	0.007	DAG1, SNTA1
fatty acid chain elongation	0.016	FASN, HADHA
acyl-CoA, malonyl-CoA → fatty acyl-CoA	0.016	FASN, HADHA
glucose-1-p → UDP-D-galactose	0.020	B4GALT1, B4GALT2
lactose metabolism	0.020	B4GALT1, B4GALT2
cytidine 5'-phosphate → RNA-P-C	0.025	POLR1B,POLR1E, POLR2B, POLR3D
histidine metabolism	0.026	AOC1, AOX1
biosynthesis of saturated and n - 9 series of MUFA and PUFA	0.026	FASN, SCD
L-methionine → dimethylglycine	0.032	BHMT2, DNMT3B
Benshangul		
1-alkyl-glycerol 3-phosphate → platelet-activating factor	0.001	LPCAT1, LPCAT2
E1 —/ ErbB3	0.002	ERBB3, RNF41, UBE2D2
Cdk2 → p53	0.006	CDK2, TP53
p53 → cytochrome C	0.011	BID, TP53
kennedy pathway	0.014	CDS2, LPCAT1, LPCAT2, PLPP2
CH000000249	0.016	MBD3, TP53
CH000000250	0.016	MBD3, TP53
sn-glycerol 3-phosphate → lysophosphatidylserine	0.019	LPCAT1, LPCAT2, PLPP2
Fer → beta-cateninTyr142	0.022	CTNNA1, FYN
dsRNA → p50:RelA	0.026	CDC34, MAP3K1, UBE2D2
TNF-alpha → c-Jun	0.029	MAP3K5, TRADD
alpha-D-Ribose 5-phosphate → inosine 5'-phosphate	0.029	PAICS, PPAT
synthesis of purine ribonucleotides	0.029	PAICS, PPAT
platelet-activating factor → fatty aldehyde	0.029	PAFAH1B, PLA2G7
TNF —MEKK1 → c-Jun	0.029	MAP3K1, TRADD
CH000000332	0.036	CYBA, MAP3K1
LKB1 —AMPKalpha-2, AMPKbeta-2 → PPARalpha	0.036	PRKCZ, STK11
metabolism of androgens	0.043	HSD17B1, HSD17B6,SRD5A3
fMLP → NADPH oxidase	0.044	CYBA, PRKCZ
EP2 → VEGFA	0.048	CYBA, GNAS, GNB1

Table 2 – continued from previous page

Pathways	P-value	Genes
Sheko (Adapted from Mekonnen et al. (2019))		
PDGF B → STATs	0.003	STAT3, STAT5A
stress-associated pathways	0.006	MBP, MEF2A, PSMD7, RAF1, RBX1, STAT3
E2F network	0.008	AKT3, CDC25C, PPP2R5A, PSMD7, RAF1, RBX1
G2/M phase (cyclin B:Cdk1)	0.015	AKT3, CDC25C, PSMD7, RBX1
IMP → ADP	0.025	AK5, AMPD3
ARIP1 → atrophin1	0.034	AKT3, APBA1
p38 pathway	0.039	MBP, MEF2A, STAT3
Plk1 cell cycle regulation	0.039	CDC25C, PSMD7, RBX1
IL-3 signaling	0.043	MBP, RAF1, STAT5A
Aurora-B cell cycle regulation	0.046	CENPE, PSMD7, RBX1
oxygen independent HIF-1alpha degradation	0.046	PSMD7, RBX1, UBE2R2
Cul3 —/ Nrf2	0.047	PSMD7, RBX1
S phase (Cdk2)	0.048	CDC25C, RAF1, RBX1

3.6 Comparative Identification of Master Regulators

The trypanotolerance attributes are further elucidated by the identification of 10 master regulators in Nuer, Gindeberet, and Benshangul breeds using the maximum radius of 10 steps upstream in the regulatory hierarchy using the TRANSPATH database. The master regulator Caspase family proteins are shared mostly by Sheko (Mekonnen et al., 2019), Gindeberet, and Benshangul as shown in Figure 7 - 10. Caspase is a protease enzyme which is involved in nitric oxide production, induction of CD8+ and CD4 T-cells, production of IFN γ , programmed cell death, erythroid differentiation, and regulation of macrophages (Gonçalves et al., 2013; Shalini et al., 2015; Zermati et al., 2001).

The top master regulator ITCH is only found in Nuer which is involved in tolerance and immune responses including T-cell activation and T-helper cell differentiation (Xiao et al., 2014). In addition, only Sheko and Nuer share the master regulator PAK1 which is involved in the activation of MAP-kinase pathway and plays an important role in innate and adaptive immunity (Yi et al., 1991; Adachi et al., 1992; Zhang et al., 2008; Traves et al., 2014). Moreover, all the master regulators of Benshangul and Gindeberet are identical.

4 DISCUSSION

In this study, the signature genes, hub genes, pathways, and master regulators of Gindeberet, Nuer, and Benshangul breeds are compared with the trypanotolerant Sheko breed (Mekonnen et al., 2019) to detect

the tendency of trypanotolerance in those breeds. Trypanosomiasis is considered as a major environmental pressure in the region where Gindeberet, Nuer, and Benshangul breeds are kept. Comparing the signature genes of these breeds with Sheko and among themselves shows a similar or distinct pattern of adaptation towards trypanotolerance. The common signature genes shared by Nuer, Gindeberet, and Benshangul with those of Sheko are related to trypanotolerance attributes that are involved in neurological dysfunction, anemia, and immunity. Since these trypanotolerant attributes are the most common symptoms of the disease (Murray et al., 1990; Tuntasuvan et al., 1997; Naessens, 2006; Giordani et al., 2016), polymorphisms in STXBP3, GNAS, KCNIP4, DHX33, C1QBP, AK5, JMJD1C, EGR1, SLC24A3, CD86, and L3MBTL3 genes might be related with an evolutionary response against the effect of the parasite.

Specifically, the STXBP3, C1QBP, DHX33, GNAS, KCNIP4, and AK5 genes which are commonly identified in Benshangul and Sheko are associated with trypanotolerance attributes related to immune tolerance and neurological dysfunctions. Moreover, the JMJD1C, EGR1, SLC24A3 and ARG1 genes in Gindeberet which are shared by Sheko are involved in trypanotolerance attributes related to immune responses. The CD86 and L3MBTL3 genes shared by Nuer and Sheko are involved in trypanotolerance attributes related to immune responses and anemia, respectively. Hence, the signature genes involved in immune responses identified in Benshangul, Nuer, and Gindeberet in common with Sheko might be engaged in a similar pattern of adaptation (similar selection responses) against the same environmental pressure related to trypanosomiasis. Consistent with our results, trypanotolerant breeds such as N'Dama show a shift from innate to adaptive responses during trypanosome infection (Stijlemans et al., 2010; Bosschaerts et al., 2011).

Intriguingly, Sheko and Nuer shared the GO term known as T-cell chemotaxis. This indicates that there is a positive selection response to diseases in these breeds. In agreement with our result, mice infected with trypanosomes show the differentially expressed genes that are involved in immune responses such as T-chemotaxis, T-cell maturation, and presenting antigens by MHC class II (Amin et al., 2010).

Our results are further substantiated by the identification of the hub genes involved in the gene networks. In response to an environmental pressure or stress, the expression of genes is regulated in a coordinated manner (Martin and Sung, 2018). Likewise, the identified signature genes could also work in such a coordinated manner and there might be a few genes (hub genes) that are involved in the regulation of the expression of other genes in response to a given environmental pressure. Interestingly, the gene PSMD7 which is involved in most of the overrepresented pathways, is also identified as a hub gene. Hence, the PSMD7 gene might be obliged in harmonizing proteins (proteases) to degrade auto-reactive immune cells caused by trypanosomiasis (Lodish et al., 2004; Zinngrebe et al., 2014). In addition, most of the hub genes are involved in the immune system, neurological dysfunctions, and hypoxia. Remarkably, the hub genes

CTNNA2 (in Gindeberet) and KCNIP4 (in Gindeberet, Benshangul, and Nuer) are also reported in the known trypanotolerant N'Dama (Kim et al., 2017) which are associated with neurological dysfunctions (Flex et al., 2003; Khan et al., 2007). The CTNNA2 gene is also reported in N'Dama as a major hub gene that plays a role in orchestrating the immune cells in response to the presence of a pathogen (Kim et al., 2017).

Moreover, trypanosomes show the ability to manipulate the immune response of the host by reducing the HIF-1 α (Oosthuysen et al., 2001). The reduction of HIF-1 α negatively affects the cellular responses such as high production of red blood cells (Silva and Faustino, 2015; Lee et al., 2004) to overcome the shortage of oxygen. In addition, progressive degradation of the motor neurons in the spinal cord is associated with the reduction of hypoxic response elements in mice (Oosthuysen et al., 2001). Therefore, the hub genes RDX, AAK1, SYT1, and FOXP1 which are related to neurological dysfunctions and hypoxia induced cellular processes might be involved in the regulation of adaptive evolutionary responses against the host immune response escaping mechanisms of trypanosomes.

Furthermore, the identification of overrepresented pathways reveals the underlying molecular mechanisms of tolerance against trypanosomiasis. Notably, the common pathways identified in Gindeberet, Sheko and Benshangul are involved in immune responses and oxidative stresses. Not only the common pathways between these breeds but also the first two overrepresented pathways of each breed are related to immunity responses and oxidative stresses. Consistent with our results, trypanosome infection is known to induce oxidative stress. Therefore, these results strengthen the hypothesis that Nuer, Gindeberet, and Benshangul are evolving or are in the processes of rapid evolutionary change in response to the common environmental stress (e.g. trypanosomiasis).

It has been reported in previous studies that trypanosomes utilize lipid nutrients of the host for growth and differentiation (Katunga-Rwakishaya et al., 1991). Trypanotolerant breeds such as N'Dama show a low level of plasma cholesterol and triglycerides compared to trypanosusceptible zebu breeds such as White Fulani (Ogunsanmi et al., 2000). This entails that trypanotolerant cattle has a superior ability of lipid clearance and hormonal control of lipid synthesis in order to limit the lipid nutrients which is required by trypanosomes (Ogunsanmi et al., 2000). Interestingly, the second most overrepresented pathway in Nuer, the acetyl-CoA \rightarrow palmitic acid pathway, which is involved in lipid metabolism, might play a significant role similar to N'Dama control of lipid synthesis. Remarkably, polymorphisms in the ST6GALNAC5 gene in Benshangul, Gindeberet, Nuer, Sheko, and Muturu, as well as the JAZF1 gene in Benshangul and Nuer are involved in lipid metabolism (Ikehara et al., 1999; Jang et al., 2014). Polymorphisms in these genes might be related to lipid clearance trypanotolerance attributes mentioned above. In agreement with this

result, the overexpression of the JAZF1 gene is associated with suppression of lipid accumulation (lipid clearance) in mice (Jang *et al.*, 2014).

For the further understanding of the regulatory mechanisms of trypanotolerance, we identify key master regulators that are potentially involved in the trypanotolerance attributes in Nuer, Benshangul, and Gindeberet breeds. Strikingly, the top master regulator Caspase family proteins are commonly identified in the Sheko (Mekonnen *et al.*, 2019), Gindeberet, and Benshangul breeds. Caspase might be involved in governing the evolutionary responses against trypanosomiasis (Mekonnen *et al.*, 2019). Consistent with our findings Paroli *et al.* (Paroli *et al.*, 2018) reported that Caspase-1 deficient mice show an increment on the level of parasitemia compared to the wild type during trypanosome infection. The knockout mice for Caspase-1 show IFN- γ , CD4+, and CD8+ cell reduction which plays a major role in the induction of adaptive immunity (Paroli *et al.*, 2018). Therefore, Caspase protein families might be involved in orchestrating the expression of the signature genes which are involved in trypanotolerance. Complementary to Caspase, the master regulator ITCH might play an important role in adaptive immunity in the Nuer breed (Fang *et al.*, 2002; Xiao *et al.*, 2014). Moreover, most of the master regulators of Sheko and Nuer are kinase families which might have similar biological functions and show ubiquitous expressions in various tissues. Interestingly, the Benshangul and Gindeberet master regulators are all identical and most of these regulators are phosphatase family proteins which might indicate that these two breeds have undergone a similar environmental pressure for generations.

In addition, the trypanotolerance tendencies of the Nuer, Benshangul, and Gindeberet breeds are supported by the overlapping QTL regions of trypanotolerant N'Dama. Gindeberet shares the largest trypanotolerant QTL regions with N'Dama which provides additional evidence that this breed has a certain level of trypanotolerance. However, the small number of overlapping trypanotolerant QTL regions reported by Mekonnen *et al.* (2019) between Sheko and N'Dama (Supplementary Table 8) compared to the Nuer, Benshangul, and Gindeberet (Supplementary Table 5 - 7) breeds indicate that the Sheko and N'Dama breeds could share regions with similar evolutionary changes while other regions are unique for each individual breed's trypanotolerance attributes. In agreement with this, Noyes *et al.* (2011) reported unique genomic regions under positive selection in the N'Dama that are associated with trypanotolerance in comparison to the Sheko breed.

5 CONCLUSION

Trypanosomiasis has been one of the major selection pressures for cattle that are kept in infested regions in Ethiopia. The Nuer, Gindeberet, and Benshangul breeds are kept in the tsetse challenged regions for generations. The genome of these breeds might undergo an evolutionary change against this environmental

pressure. Comparatively, we have identified traces of these changes as candidate signature genes which are associated with adaptation of these breeds to their natural environmental pressure. This result provides the first evidence on comparative genome wide signature gene detection associated with trypanosomiasis. We found that these breeds have common as well as unique tolerance mechanisms against trypanosome infection. This indicates that designing and implementing cross breeding with different levels of tolerance and mechanisms would produce a synthetic breed that would combine the tolerance attributes of the two parental populations. Hence, this study serves as an input for designing and implementing genetic intervention strategies to improve the performance of these breeds towards higher trypanotolerance. On top of that, our findings reveal the importance of pathway and master regulator analyses which can be used as targets for the development of effective drugs for trypanosome infection. To validate the results reported in this paper, differential expression or gene knockout experiments on selected genes and regulatory proteins may be required.

CONFLICT OF INTEREST STATEMENT

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

AUTHOR CONTRIBUTIONS

YAM and AOS participated in the design of the study. YAM conducted computational and statistical analyses as well as identified the signature genes. AOS and MG supervised the computational and statistical analyses. YAM interpreted the results. YAM carried out the literature survey and prepared the first draft of the manuscript. OH and MG were involved in the interpretation of the results. YAM wrote the final version of the manuscript. YAM and AOS conceived and managed the project. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

We acknowledge support by the German Research Foundation and Open Access Publication Funds of the Göttingen University. We thank Henner Simianer for his helpful advise and insights at the early stages of this project and for comments on the methods.

SUPPLEMENTAL DATA

Supplementary Figure 1 Gene Ontology treemap for the functionally enriched ($P < 0.05$) genes for Sheko (adapted from Mekonnen et al. (2019)), Gindeberet, Nuer and Benshangul.

Supplementary Table 1 Functionally annotated gene list identified by *iHS*, CLR, and *Rsb* analysis in Sheko (adapted from Mekonnen et al. (2019)).

Supplementary Table 2 Functionally annotated gene list identified by *iHS*, CLR, and *Rsb* analysis in Benshangul.

Supplementary Table 3 Functionally annotated gene list identified by *iHS*, CLR, and *Rsb* analysis in Gindeberet.

Supplementary Table 4 Functionally annotated gene list identified by *iHS*, CLR, and *Rsb* analysis in Nuer.

Supplementary Table 5 QTL regions overlapping between Nuer and N'Dama.

Supplementary Table 6 QTL regions overlapping between Benshangul and N'Dama.

Supplementary Table 7 QTL regions overlapping between Gindeberet and N'Dama.

Supplementary Table 8 QTL regions overlapping between Sheko and N'Dama (adapted from Mekonnen et al. (2019)).

REFERENCES

- Adachi, M., Sekiya, M., Arimura, Y., Takekawa, M., Itoh, F., Hinoda, Y., et al. (1992). Protein-tyrosine phosphatase expression in pre-B cell NALM-6. *Cancer Res.* 52, 737–740
- Adams, K. L., Rousso, D. L., Umbach, J. A., and Novitch, B. G. (2015). Foxp1-mediated programming of limb-innervating motor neurons from mouse and human embryonic stem cells. *Nat Commun* 6, 6778
- Aki, D., Zhang, W., and Liu, Y. C. (2015). The E3 ligase Itch in immune regulation and beyond. *Immunol. Rev.* 266, 6–26
- Amin, D. N., Ngoyi, D. M., Nkwachi, G.-M., Palomba, M., Rottenberg, M., Büscher, P., et al. (2010). Identification of stage biomarkers for human african trypanosomiasis. *The American journal of tropical medicine and hygiene* 82, 983–990
- Andrew, A. H. (2004). *Bovine Medicine Diseases and Husbandry of Cattle* (9600 Garsington Road, Oxford OX42DQ, UKs: Blackwell Science Ltd)
- Arai, S. and Miyazaki, T. (2005). Impaired maturation of myeloid progenitors in mice lacking novel polycomb group protein mbt-1. *The EMBO journal* 24, 1863–1873
- Bach, T. L., Chen, Q. M., Kerr, W. T., Wang, Y., Lian, L., Choi, J. K., et al. (2007). Phospholipase cbeta is critical for T cell chemotaxis. *J. Immunol.* 179, 2223–2227
- Bahbahani, H., Salim, B., Almathen, F., Al Enezi, F., Mwacharo, J. M., and Hanotte, O. (2018). Signatures of positive selection in African Butana and Kenana dairy zebu cattle. *PLoS ONE* 13, e0190446

- Bahbahani, H., Tijjani, A., Mukasa, C., Wragg, D., Almathen, F., Nash, O., et al. (2017). Signatures of Selection for Environmental Adaptation and Zebu x Taurine Hybrid Fitness in East African Shorthorn Zebu. *Front Genet* 8, 68
- Bai, J., Lai, L., Yeo, H. C., Goh, B. C., and Tan, T. M. (2004). Multidrug resistance protein 4 (mrp4/abcc4) mediates efflux of bimanane-glutathione. *The international journal of biochemistry & cell biology* 36, 247–257
- Barrett, A. D. and Stanberry, L. R. (2009). *Vaccines for biodefense and emerging and neglected diseases* (Academic Press)
- Bastepe, M. (2008). The GNAS locus and pseudohypoparathyroidism. *Adv. Exp. Med. Biol.* 626, 27–40
- Bellot, G., Garcia-Medina, R., Gounon, P., Chiche, J., Roux, D., Pouysségur, J., et al. (2009). Hypoxia-induced autophagy is mediated through hypoxia-inducible factor induction of bnip3 and bnip3l via their bh3 domains. *Molecular and cellular biology* 29, 2570–2581
- Bosschaerts, T., Morias, Y., Stijlemans, B., Herin, M., Porta, C., Sica, A., et al. (2011). IL-10 limits production of pathogenic TNF by M1 myeloid cells through induction of nuclear NF- κ B p50 member in *Trypanosoma congolense* infection-resistant C57BL/6 mice. *Eur. J. Immunol.* 41, 3270–3280
- Browning, S. R. and Browning, B. L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* 81, 1084–1097
- Castillo-Rodríguez, R. A., Dávila-Borja, V. M., and Juárez-Méndez, S. (2018). Data mining of pediatric medulloblastoma microarray expression reveals a novel potential subdivision of the group 4 molecular subgroup. *Oncology letters* 15, 6241–6250
- Chen, Z., Krmar, R. T., Dada, L., Efendiev, R., Leibiger, I. B., Pedemonte, C. H., et al. (2006). Phosphorylation of adaptor protein-2 μ 2 is essential for Na^+ , K^+ -ATPase endocytosis in response to either G protein-coupled receptor or reactive oxygen species. *American journal of respiratory cell and molecular biology* 35, 127–132
- Codjia, V., Mulatu, W., Majiwa, P. A., Leak, S. G., Rowlands, G. J., Authie, E., et al. (1993). Epidemiology of bovine trypanosomiasis in the Ghibe valley, southwest Ethiopia. 3. Occurrence of populations of *Trypanosoma congolense* resistant to diminazene, isometamidium and homidium. *Acta Trop.* 53, 151–163
- DeGiorgio, M., Huber, C. D., Hubisz, M. J., Hellmann, I., and Nielsen, R. (2016). SweepFinder2: increased sensitivity, robustness and flexibility. *Bioinformatics* 32, 1895–1897
- Derycke, L. D. and Bracke, M. E. (2004). N-cadherin in the spotlight of cell-cell adhesion, differentiation, embryogenesis, invasion and signalling. *International Journal of Developmental Biology* 48, 463–476

- El Kasmi, K. C., Qualls, J. E., Pesce, J. T., Smith, A. M., Thompson, R. W., Henao-Tamayo, M., et al. (2008). Toll-like receptor–induced arginase 1 in macrophages thwarts effective immunity against intracellular pathogens. *Nature immunology* 9, 1399
- Fang, D., Elly, C., Gao, B., Fang, N., Altman, Y., Joazeiro, C., et al. (2002). Dysregulation of t lymphocyte function in itchy mice: a role for itch in t h 2 differentiation. *Nature immunology* 3, 281
- FAO (2004). THE UGANDA COUNTRY REPORT AS PART OF THE STATE OF THE WORLD'S ANIMAL GENETIC RESOURCES (SoW-AnGR) REPORT
- Flex, E., Mangino, M., Mazzoli, M., Martini, A., Migliosi, V., Colosimo, A., et al. (2003). Mapping of a new autosomal dominant non-syndromic hearing loss locus (DFNA43) to chromosome 2p12. *J. Med. Genet.* 40, 278–281
- Gaudet, P., Livstone, M. S., Lewis, S. E., and Thomas, P. D. (2011). Phylogenetic-based propagation of functional annotations within the Gene Ontology consortium. *Brief. Bioinformatics* 12, 449–462
- Gautier, M. and Vitalis, R. (2012). rehh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. *Bioinformatics* 28, 1176–1177
- Gerard, C. and Rollins, B. J. (2001). Chemokines and disease. *Nat. Immunol.* 2, 108–115
- Giordani, F., Morrison, L. J., Rowan, T. G., DE Koning, H. P., and Barrett, M. P. (2016). The animal trypanosomiasis and their chemotherapy: a review. *Parasitology* 143, 1862–1889
- Gonçalves, V. M., Matteucci, K. C., Buzzo, C. L., Miollo, B. H., Ferrante, D., Torrecilhas, A. C., et al. (2013). NLRP3 controls Trypanosoma cruzi infection through a caspase-1-dependent IL-1R-independent NO production. *PLoS Negl Trop Dis* 7, e2469
- Gordon, S. (2003). Alternative activation of macrophages. *Nature reviews immunology* 3, 23
- Hancock, R. L., Dunne, K., Walport, L. J., Flashman, E., and Kawamura, A. (2015). Epigenetic regulation by histone demethylases in hypoxia. *Epigenomics* 7, 791–811
- Hanotte, O., Ronin, Y., Agaba, M., Nilsson, P., Gelhaus, A., Horstmann, R., et al. (2003). Mapping of quantitative trait loci controlling trypanotolerance in a cross of tolerant West African N'Dama and susceptible East African Boran cattle. *Proc. Natl. Acad. Sci. U.S.A.* 100, 7443–7448
- Hu, H., Takano, N., Xiang, L., Gilkes, D. M., Luo, W., and Semenza, G. L. (2014). Hypoxia-inducible factors enhance glutamate signaling in cancer cells. *Oncotarget* 5, 8853–8868
- Hu, H., Wang, B., Borde, M., Nardone, J., Maika, S., Allred, L., et al. (2006). Foxp1 is an essential transcriptional regulator of b cell development. *Nature immunology* 7, 819
- Ikehara, Y., Shimizu, N., Kono, M., Nishihara, S., Nakanishi, H., Kitamura, T., et al. (1999). A novel glycosyltransferase with a polyglutamine repeat; a new candidate for gd1 α synthase (st6galnac v) 1. *FEBS letters* 463, 92–96

- Jang, W. Y., Bae, K. B., Kim, S. H., Yu, D. H., Kim, H. J., Ji, Y. R., et al. (2014). Overexpression of jazf1 reduces body weight gain and regulates lipid metabolism in high fat diet. *Biochemical and biophysical research communications* 444, 296–301
- Janssens, V., Goris, J., and Van Hoof, C. (2005). Pp2a: the expected tumor suppressor. *Current opinion in genetics & development* 15, 34–41
- Kanaan, Z., Qadan, M., Eichenberger, M. R., and Galandiuk, S. (2010). The actin-cytoskeleton pathway and its potential role in inflammatory bowel disease-associated human colorectal cancer. *Genet Test Mol Biomarkers* 14, 347–353
- Kaspar, J. W. and Jaiswal, A. K. (2011). Tyrosine phosphorylation controls nuclear export of Fyn, allowing Nrf2 activation of cytoprotective gene expression. *FASEB J.* 25, 1076–1087
- Katunga-Rwakishaya, E., MURRAY, M., and HOLMES, P. H. (1991). Hematological erythrokinetic and blood lipid changes in sheep infected with trypanosoma congolense. international scientific council for trypanosomiasis research and control (oau/isctrc) 21st meeting. *IUBMB life* , 257
- Kelsen, J. R., Ouahed, J., Spessott, W. A., Kooshesh, K., Sanmillan, M. L., Dawany, N., et al. (2018). 25 mutations in stxbp3 contribute to very early onset of ibd, immunodeficiency and hearing loss. *Gastroenterology* 154, S40–S41
- Khan, S. Y., Ahmed, Z. M., Shabbir, M. I., Kitajiri, S., Kalsoom, S., Tasneem, S., et al. (2007). Mutations of the RDX gene cause nonsyndromic hearing loss at the DFNB24 locus. *Hum. Mutat.* 28, 417–423
- Kim, C. H. and Broxmeyer, H. E. (1999). Chemokines: signal lamps for trafficking of T and B cells for development and effector function. *J Leukoc Biol* 65, 6–15
- Kim, S.-J., Ka, S., Ha, J.-W., Kim, J., Yoo, D., Kim, K., et al. (2017). Cattle genome-wide analysis reveals genetic signatures in trypanotolerant n'dama. *BMC genomics* 18, 371
- Kinsella, R. J., Kahari, A., Haider, S., Zamora, J., Proctor, G., Spudich, G., et al. (2011). Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database (Oxford)* 2011, bar030
- Kitada, T., Asakawa, S., Hattori, N., Matsumine, H., Yamamura, Y., Minoshima, S., et al. (1998). Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. *nature* 392, 605
- Kitamura, N., Nakamura, Y., Miyamoto, Y., Miyamoto, T., Kabu, K., Yoshida, M., et al. (2011). Mieap, a p53-inducible protein, controls mitochondrial quality by repairing or eliminating unhealthy mitochondria. *PloS one* 6, e16060
- Konopacki, C., Pritykin, Y., Rubtsov, Y., Leslie, C. S., and Rudensky, A. Y. (2019). Transcription factor foxp1 regulates foxp3 chromatin binding and coordinates regulatory t cell function. *Nature immunology* 20, 232

- Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* 44, W90–97
- Leak, S. G., Peregrine, A. S., Mulatu, W., Rowlands, G. J., and D'Ieteren, G. (1996). Use of insecticide-impregnated targets for the control of tsetse flies (*Glossina* spp.) and trypanosomiasis occurring in cattle in an area of south-west Ethiopia with a high prevalence of drug-resistant trypanosomes. *Trop. Med. Int. Health* 1, 599–609
- Lee, J. M., Chan, K., Kan, Y. W., and Johnson, J. A. (2004). Targeted disruption of Nrf2 causes regenerative immune-mediated hemolytic anemia. *Proc. Natl. Acad. Sci. U.S.A.* 101, 9751–9756
- Lemecha, H., Mulatu, W., Hussein, I., Rege, E., Tekle, T., Abdicho, S., et al. (2006). Response of four indigenous cattle breeds to natural tsetse and trypanosomiasis challenge in the Ghibe valley of Ethiopia. *Vet. Parasitol.* 141, 165–176
- Leta, S., Alemayehu, G., Seyoum, Z., and Bezie, M. (2016). Prevalence of bovine trypanosomiasis in Ethiopia: a meta-analysis. *Parasit Vectors* 9, 139
- Li, K., Gao, B., Li, J., Chen, H., Li, Y., Wei, Y., et al. (2015). Znf32 protects against oxidative stress-induced apoptosis by modulating c1qbp transcription. *Oncotarget* 6, 38107
- Li, L. O., Klett, E. L., and Coleman, R. A. (2010). Acyl-coa synthesis, lipid metabolism and lipotoxicity. *Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids* 1801, 246–251
- Lodish, H., Berk, A., Matsudaira, P., C., K., Krieger, M., Scott, M., et al. (2004). *Molecular Cell Biology* (41 Madison Avenue, New York, USA: W. H. Freeman and Company), 5 edn.
- Lutje, V., Taylor, K. A., Kennedy, D., Authie, E., Boulange, A., and Gettinby, G. (1996). Trypanosoma congolense: a comparison of T-cell-mediated responses in lymph nodes of trypanotolerant and trypanosusceptible cattle during primary infection. *Exp. Parasitol.* 84, 320–329
- Ma, Y., Ding, X., Qanbari, S., Weigend, S., Zhang, Q., and Simianer, H. (2015). Properties of different selection signature statistics and a new strategy for combining them. *Heredity (Edinb)* 115, 426–436
- Magona, J. W., Walubengo, J., and Odimim, J. J. (2004). Differences in susceptibility to trypanosome infection between Nkedi Zebu and Ankole cattle, under field conditions in Uganda. *Ann Trop Med Parasitol* 98, 785–792
- Martin, E. and Sung, M.-H. (2018). Challenges of decoding transcription factor dynamics in terms of gene regulation. *Cells* 7, 132
- Mekonnen, Y. A., Gültas, M., Effa, K., Hanotte, O., and Schmitt, A. O. (2019). Identification of candidate signature genes and key regulators associated with trypanotolerance in the sheko breed. *Frontiers in Genetics* 10

- Mitoma, H., Hanabuchi, S., Kim, T., Bao, M., Zhang, Z., Sugimoto, N., et al. (2013). The DHX33 RNA helicase senses cytosolic RNA and activates the NLRP3 inflammasome. *Immunity* 39, 123–135
- Muhanguzi, D., Mugenyi, A., Bigirwa, G., Kamusiime, M., Kitibwa, A., Akurut, G. G., et al. (2017). African animal trypanosomiasis as a constraint to livestock health and production in Karamoja region: a detailed qualitative and quantitative assessment. *BMC Vet. Res.* 13, 355
- Muhanguzi, D., Picozzi, K., Hatendorf, J., Thrusfield, M., Welburn, S. C., Kabasa, J. D., et al. (2014). Improvements on restricted insecticide application protocol for control of Human and Animal African Trypanosomiasis in eastern Uganda. *PLoS Negl Trop Dis* 8, e3284
- Mulugeta, W., Wilkes, J., Mulatu, W., Majiwa, P. A., Masake, R., and Peregrine, A. S. (1997). Long-term occurrence of *Trypanosoma congolense* resistant to diminazene, isometamidium and homidium in cattle at Ghibe, Ethiopia. *Acta Trop.* 64, 205–217
- Murray, M., Trail, J. C., and D'Ieteren, G. D. (1990). Trypanotolerance in cattle and prospects for the control of trypanosomiasis by selective breeding. *Rev. - Off. Int. Epizoot.* 9, 369–386
- Naessens, J. (2006). Bovine trypanotolerance: A natural ability to prevent severe anaemia and haemophagocytic syndrome? *Int. J. Parasitol.* 36, 521–528
- Nantulya, V. M. (1986). Immunological approaches to the control of animal trypanosomiasis. *Parasitol. Today (Regul. Ed.)* 2, 168–173
- Nielsen, R., Williamson, S., Kim, Y., Hubisz, M. J., Clark, A. G., and Bustamante, C. (2005). Genomic scans for selective sweeps using SNP data. *Genome Res.* 15, 1566–1575
- Nobuyuki, I., Sumiko, K., Yukiko, N., Takeshi, S., Dai, A., Fumio, H., et al. (1992). Cloning and sequence of a functionally active cDNA encoding the mouse ubiquitin-activating enzyme e1. *Gene* 118, 279–282
- Noyes, H., Brass, A., Obara, I., Anderson, S., Archibald, A. L., Bradley, D. G., et al. (2011). Genetic and expression analysis of cattle identifies candidate genes in pathways responding to *Trypanosoma congolense* infection. *Proc. Natl. Acad. Sci. U.S.A.* 108, 9304–9309
- NTTICC (2004). National Tsetse and Trypanosomiasis Investigation and Control Center. Report for the period 7th June 2003 to 6th July 2004. Bedele, Ethiopia. 21-24
- Ocaido, M., Otim, C. P., Okuna, N. M., Erume, J., Ssekitto, C., Wafula, R. Z. O., et al. (2005). Socio-economic and livestock disease survey of agro-pastoral communities in Serere County, Soroti District, Uganda. *Livestock Research for Rural Development* 17
- Ogunsanmi, A., Taiwo, V., Onawumi, B., Mbagwu, H., Okoronkwo, C., et al. (2000). Correlation of physiological plasma lipid levels with resistance of cattle to trypanosomiasis. *Veterinarski arhiv* 70, 251–258

- Oosthuysen, B., Moons, L., Storkebaum, E., Beck, H., Nuyens, D., Brusselmans, K., et al. (2001). Deletion of the hypoxia-response element in the vascular endothelial growth factor promoter causes motor neuron degeneration. *Nat. Genet.* 28, 131–138
- Paroli, A. F., Gonzalez, P. V., Díaz-Luján, C., Onofrio, L. I., Arocena, A., Cano, R. C., et al. (2018). Nlrp3 inflammasome and caspase-1/11 pathway orchestrate different outcomes in the host protection against trypanosoma cruzi acute infection. *Frontiers in immunology* 9
- Pavlidis, P., Hutter, S., and Stephan, W. (2008). A population genomic approach to map recent positive selection in model species. *Molecular Ecology* 17, 3585–3598
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). PLINK: a toolset for whole-genome association and population-based linkage analysis. *American Journal of Human Genetics* 81, 559–575
- Qiu, L., Joazeiro, C., Fang, N., Wang, H. Y., Elly, C., Altman, Y., et al. (2000). Recognition and ubiquitination of Notch by Itch, a hec-type E3 ubiquitin ligase. *J. Biol. Chem.* 275, 35734–35737
- Sabeti, P. C., Reich, D. E., Higgins, J. M., Levine, H. Z., Richter, D. J., Schaffner, S. F., et al. (2002). Detecting recent positive selection in the human genome from haplotype structure. *Nature* 419, 832–837
- Salmond, R. J., Filby, A., Qureshi, I., Caserta, S., and Zamoyska, R. (2009). T-cell receptor proximal signaling via the src-family kinases, lck and fyn, influences t-cell activation, differentiation, and tolerance. *Immunological reviews* 228, 9–22
- Schuler, M., Bossy-Wetzell, E., Goldstein, J. C., Fitzgerald, P., and Green, D. R. (2000). p53 induces apoptosis by caspase activation through mitochondrial cytochrome c release. *Journal of Biological Chemistry* 275, 7337–7342
- Shalini, S., Dorstyn, L., Dawar, S., and Kumar, S. (2015). Old, new and emerging functions of caspases. *Cell death and differentiation* 22, 526
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research* 13, 2498–2504
- Sheridan, J. M. (2007). Generation of a compartmentalised thymus organoid in vitro using fetal thymic epithelial progenitor cells
- Short, J. J., Vasu, C., Holterman, M. J., Curiel, D. T., and Pereboev, A. (2006). Members of adenovirus species b utilize cd80 and cd86 as cellular attachment receptors. *Virus research* 122, 144–153
- Silva, B. and Faustino, P. (2015). An overview of molecular basis of iron metabolism regulation and the associated pathologies. *Biochim. Biophys. Acta* 1852, 1347–1359
- Sinshaw, A., Abebe, G., Desquesnes, M., and Yoni, W. (2006). Biting flies and Trypanosoma vivax infection in three highland districts bordering lake Tana, Ethiopia. *Vet. Parasitol.* 142, 35–46

- Slingenbergh, J. (1992). Tsetse control and agricultural development in Ethiopia. *World Anim Rev.* 70-71, 30–36
- Sperandio, S., Fortin, J., Sasik, R., Robitaille, L., Corbeil, J., and de Belle, I. (2009). The transcription factor *egr1* regulates the *hif-1 α* gene during hypoxia. *Molecular Carcinogenesis: Published in cooperation with the University of Texas MD Anderson Cancer Center* 48, 38–44
- Springer, T. A. (1990). Adhesion receptors of the immune system. *Nature* 346, 425
- Stanojevic, V., Habener, J. F., Holz, G. G., and Leech, C. A. (2008). Cytosolic adenylate kinases regulate K-ATP channel activity in human beta-cells. *Biochem. Biophys. Res. Commun.* 368, 614–619
- Stijlemans, B., Vankrunkelsven, A., Caljon, G., Bockstal, V., Guilliams, M., Bosschaerts, T., et al. (2010). The central role of macrophages in trypanosomiasis-associated anemia: rationale for therapeutical approaches. *Endocr Metab Immune Disord Drug Targets* 10, 71–82
- Tang, K., Thornton, K. R., and Stoneking, M. (2007). A new approach for using genome scans to detect recent positive selection in the human genome. *PLoS biology* 5, e171
- Tasew, S. and Duguma, R. (2012). Cattle anaemia and trypanosomiasis in western Oromia State, Ethiopia. *evue Méd. Vét.* 163, 581–588
- Tijjani, A., Utsunomiya, Y. T., Ezekwe, A., Nash, O., and Hanotte, O. H. (2019). Genome sequence analysis reveals selection signatures in endangered trypano-tolerant west african muturu cattle. *Frontiers in Genetics* 10, 442
- Tohyama, Y. and Yamamura, H. (2006). Complement-mediated phagocytosis-the role of syk. *IUBMB life* 58, 304–308
- Traves, P. G., Pardo, V., Pimentel-Santillana, M., Gonzalez-Rodriguez, A., Mojena, M., Rico, D., et al. (2014). Pivotal role of protein tyrosine phosphatase 1B (PTP1B) in the macrophage response to pro-inflammatory and anti-inflammatory challenge. *Cell Death Dis* 5, e1125
- Tuntasuvan, D., Sarataphan, N., and Nishikawa, H. (1997). Cerebral trypanosomiasis in native cattle. *Vet. Parasitol.* 73, 357–363
- Van Wyk, I. C., Goddard, A., de C Bronsvooort, B. M., Coetzer, J. A., Handel, I. G., Hanotte, O., et al. (2014). The impact of co-infections on the haematological profile of East African Short-horn Zebu calves. *Parasitology* 141, 374–388
- Vatsiou, A. I., Bazin, E., and Gaggiotti, O. E. (2016). Detection of selective sweeps in structured populations: a comparison of recent methods. *Molecular ecology* 25, 89–103
- Voight, B. F., Kudaravalli, S., Wen, X., and Pritchard, J. K. (2006). A map of recent positive selection in the human genome. *PLoS biology* 4, e72
- Wang, Y. and Morrissey, E. (2010). Regulation of cardiomyocyte proliferation by Foxp1. *Cell Cycle* 9, 4251–4252

- Warde-Farley, D., Donaldson, S. L., Comes, O., Zuberi, K., Badrawi, R., Chao, P., et al. (2010). The genemania prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic acids research* 38, W214–W220
- Weißflog, L., Scholz, C.-J., Jacob, C. P., Nguyen, T. T., Zamzow, K., Groß-Lesch, S., et al. (2013). Kcnp4 as a candidate gene for personality disorders and adult adhd. *European Neuropsychopharmacology* 23, 436–447
- Welburn, S. C., Molyneux, D. H., and Maudlin, I. (2016). Beyond Tsetse—Implications for Research and Control of Human African Trypanosomiasis Epidemics. *Trends Parasitol.* 32, 230–241
- Wilches, R., Voigt, S., Duchon, P., Laurent, S., and Stephan, W. (2014). Fine-mapping and selective sweep analysis of QTL for cold tolerance in *Drosophila melanogaster*. *G3 (Bethesda)* 4, 1635–1645
- Wlochowitz, D., Haubrock, M., Arackal, J., Bleckmann, A., Wolff, A., Beissbarth, T., et al. (2016). Computational Identification of Key Regulators in Two Different Colorectal Cancer Cell Lines. *Front Genet* 7, 42
- Xiao, N., Eto, D., Elly, C., Peng, G., Crotty, S., and Liu, Y.-C. (2014). The e3 ubiquitin ligase itch is required for the differentiation of follicular helper t cells. *Nature immunology* 15, 657
- Yaro, M., Munyard, K. A., Stear, M. J., and Groth, D. M. (2016). Combatting African Animal Trypanosomiasis (AAT) in livestock: The potential role of trypanotolerance. *Vet. Parasitol.* 225, 43–52
- Yi, T., Cleveland, J. L., and Ihle, J. N. (1991). Identification of novel protein tyrosine phosphatases of hematopoietic cells by polymerase chain reaction amplification. *Blood* 78, 2222–2228
- Yoo, J. C., yeon Lim, T., Park, J. S., Hah, Y.-S., Park, N., Hong, S.-G., et al. (2013). Syt14l, especially its c2 domain, is involved in regulating melanocyte differentiation. *Journal of dermatological science* 72, 246–251
- You, F., Sun, H., Zhou, X., Sun, W., Liang, S., Zhai, Z., et al. (2009). PCBP2 mediates degradation of the adaptor MAVS via the HECT ubiquitin ligase AIP4. *Nat. Immunol.* 10, 1300–1308
- Zermati, Y., Garrido, C., Amsellem, S., Fishelson, S., Bouscary, D., Valensi, F., et al. (2001). Caspase activation is required for terminal erythroid differentiation. *J. Exp. Med.* 193, 247–254
- Zhang, C., Zhang, J., Niu, J., Zhou, Z., Zhang, J., and Tian, Z. (2008). Interleukin-12 improves cytotoxicity of natural killer cells via upregulated expression of NKG2D. *Hum. Immunol.* 69, 490–500
- Zimmerman, G. A., McIntyre, T. M., Prescott, S. M., and Stafforini, D. M. (2002). The platelet-activating factor signaling system and its regulators in syndromes of inflammation and thrombosis. *Critical care medicine* 30, S294–S301
- Zinngrebe, J., Montinaro, A., Peltzer, N., and Walczak, H. (2014). Ubiquitin in the immune system. *EMBO Rep.* 15, 28–45

FIGURE CAPTIONS

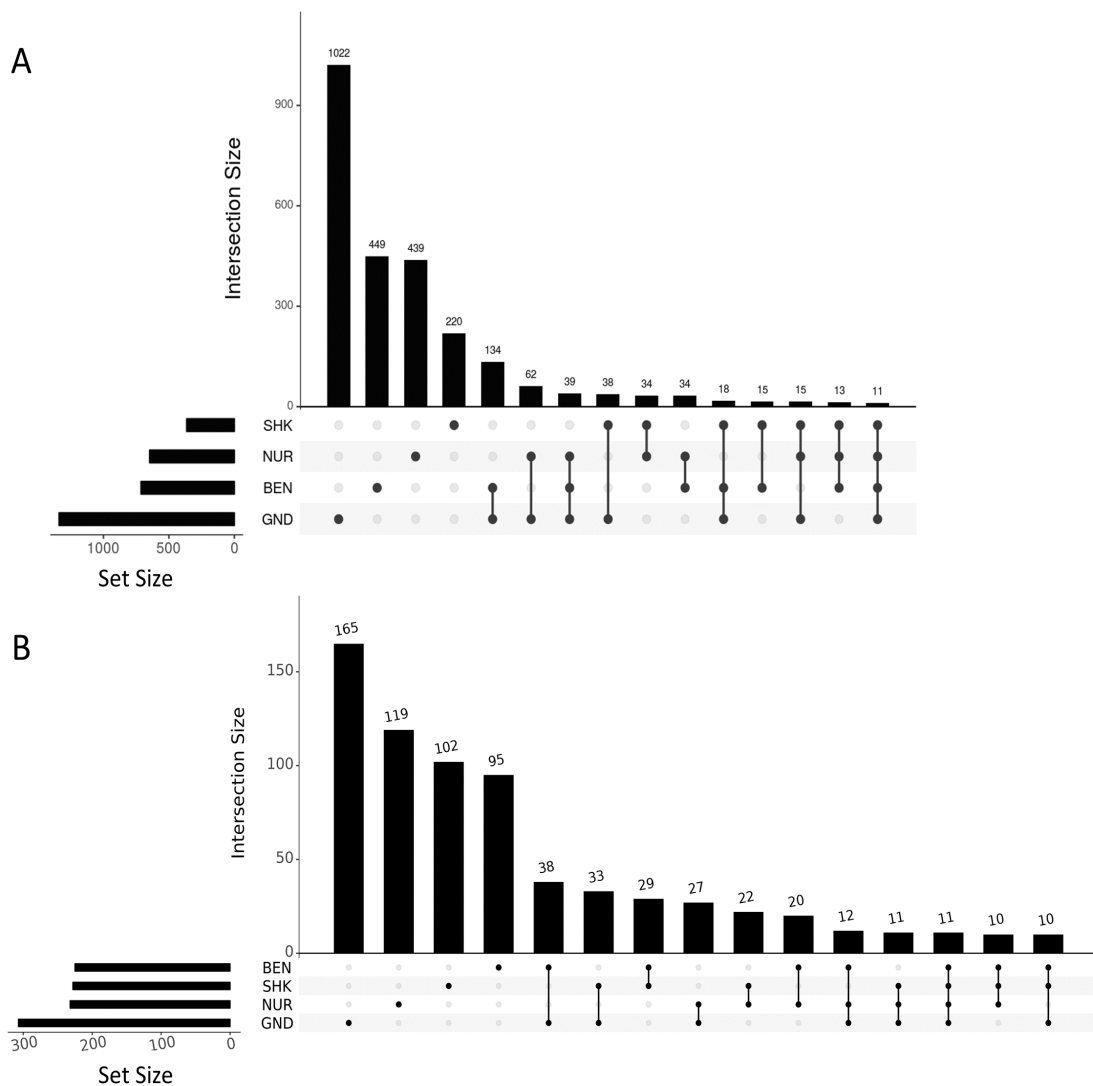


Figure 1. UpSet plots of the overlapping candidate genes (A) and genomic regions (B) identified by *iHS*, CLR, and *Rsb*. The Set Size refers to the total number of unique and overlapping genes (A) and genomic regions (B). Below the X-axis, the first four dots corresponds to the four breeds and dots connected by lines corresponds to the common genes (A) and genomic regions (B) shared by the corresponding breeds. In the Y-axis, the Intersection Size refers to the number of unique (the first four columns) and shared genes (A) and genomic regions (B) for the corresponding breeds. BEN, Benshangul; SHK, Sheko; NUR, Nuer; GND, Gindeberet.



Figure 2. Gene Ontology treemap for the 260 functionally enriched ($p < 0.05$) genes identified in (A) Nuer and (B) Sheko. The size of the boxes corresponds to the $-\log_{10} P$ -value of the GO-term. The boxes are grouped together based on the upper-hierarchy GO-term which is written in bold letters. Figure B is adapted from Mekonnen et al. (2019).

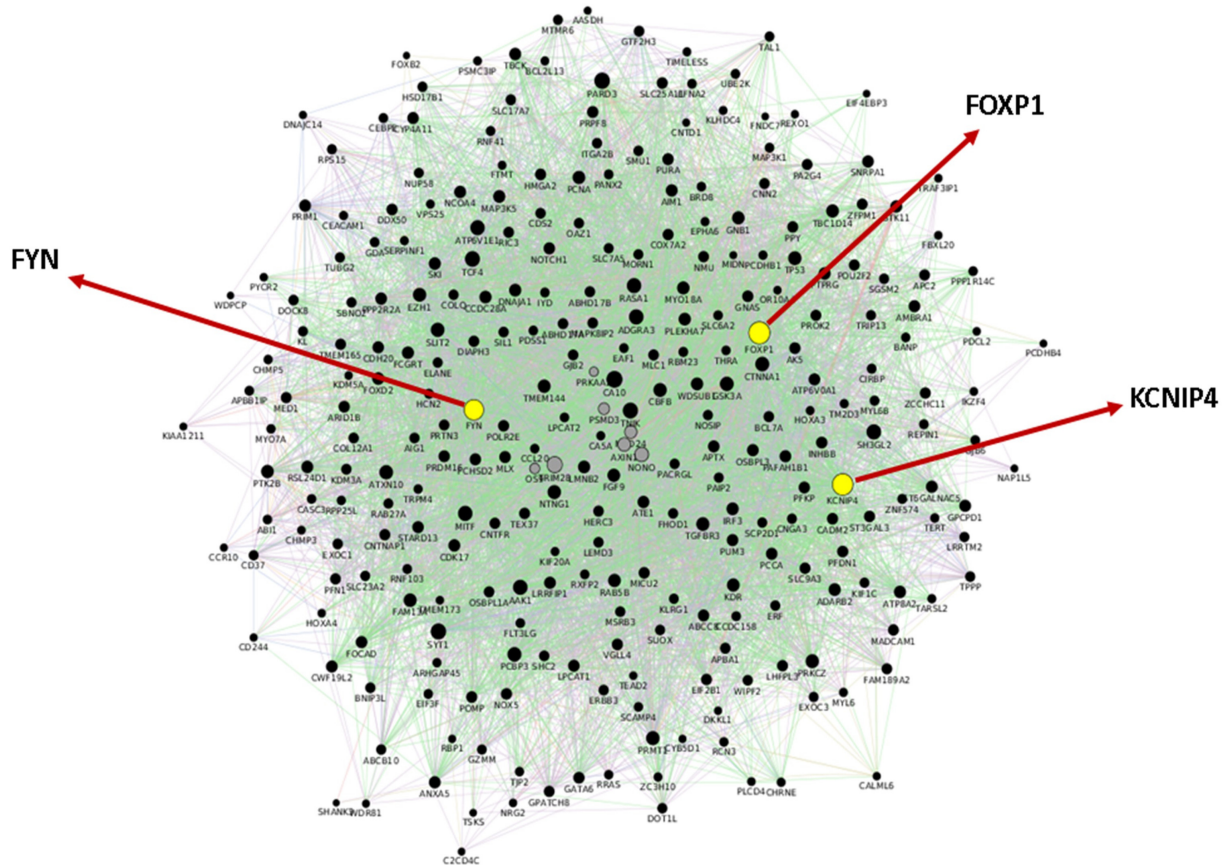


Figure 3. Gene networks identified in Benshangul. The highlighted circles indicate hub genes. The black and gray circle indicate the query genes and connecting genes from the GeneMANIA databases, respectively. The connecting line colours between genes indicate: green, genetic interaction; purple, co-expression; orange, predicted functional relationships; pink, physical interaction, yellow; shared protein domains.

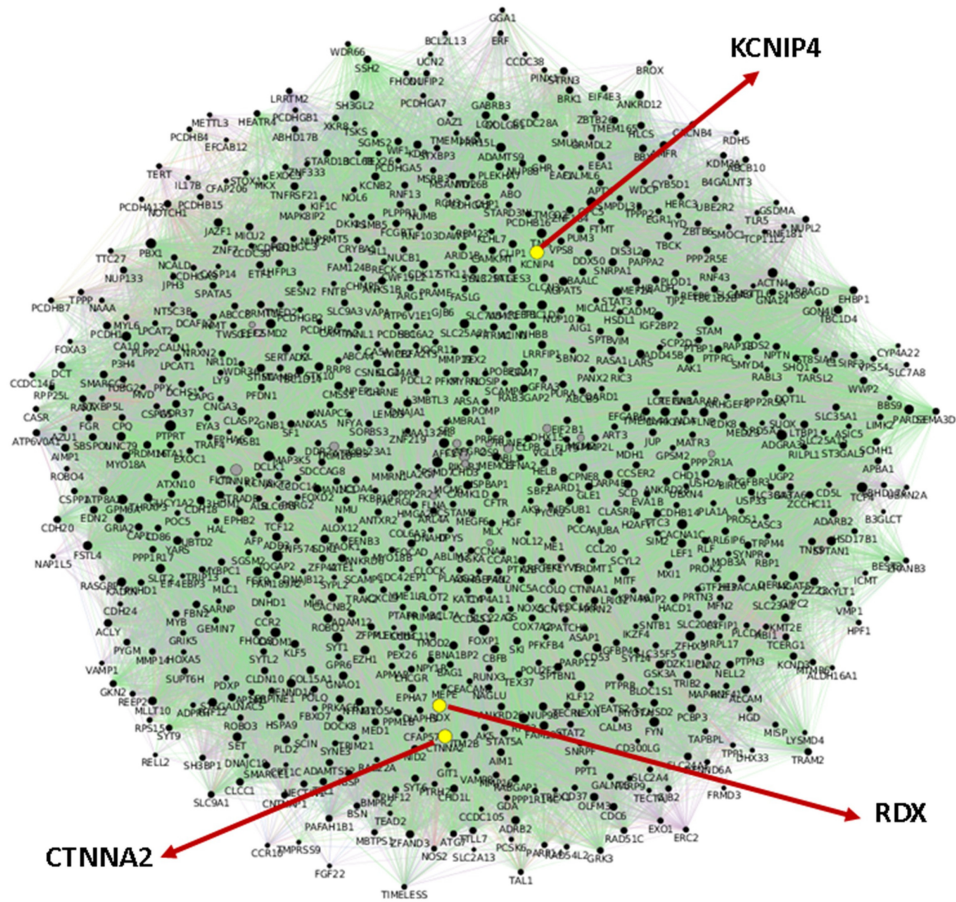


Figure 4. Gene networks identified in Gindeberet. The highlighted circles indicate hub genes. The black and gray circle indicate the query genes and connecting genes from the GeneMANIA databases, respectively. The connecting line colours between genes indicate: green, genetic interaction; purple, co-expression; orange, predicted functional relationships; pink, physical interaction, yellow; shared protein domains.

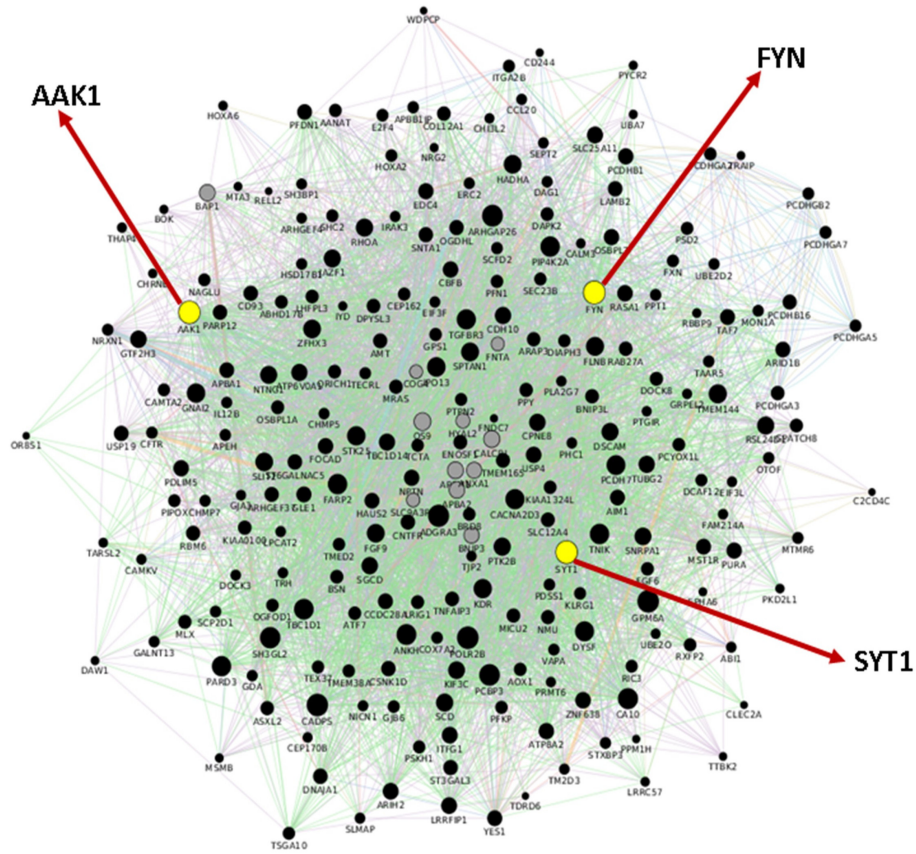


Figure 5. Gene networks identified in Nuer. The highlighted circles indicate hub genes. The black and gray circle indicate the query genes and connecting genes from the GeneMANIA databases, respectively. The connecting line colours between genes indicate: green, genetic interaction; purple, co-expression; orange, predicted functional relationships; pink, physical interaction, yellow; shared proteins domains.

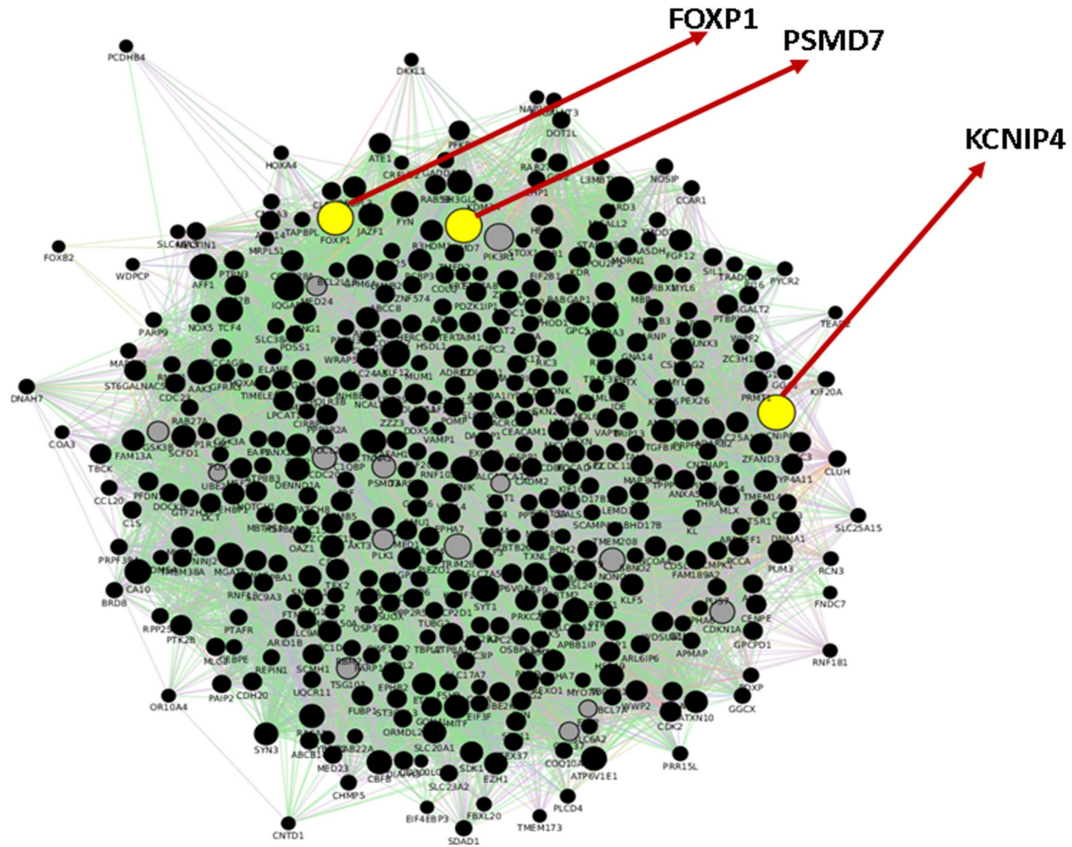


Figure 6. Gene networks identified in Sheko. The highlighted circles indicate hub genes. The black and gray circle indicate the query genes and connecting genes from the GeneMANIA databases, respectively. The connecting line colours between genes indicate: green, genetic interaction; purple, co-expression; orange, predicted functional relationships; pink, physical interaction, yellow; shared proteins domains.

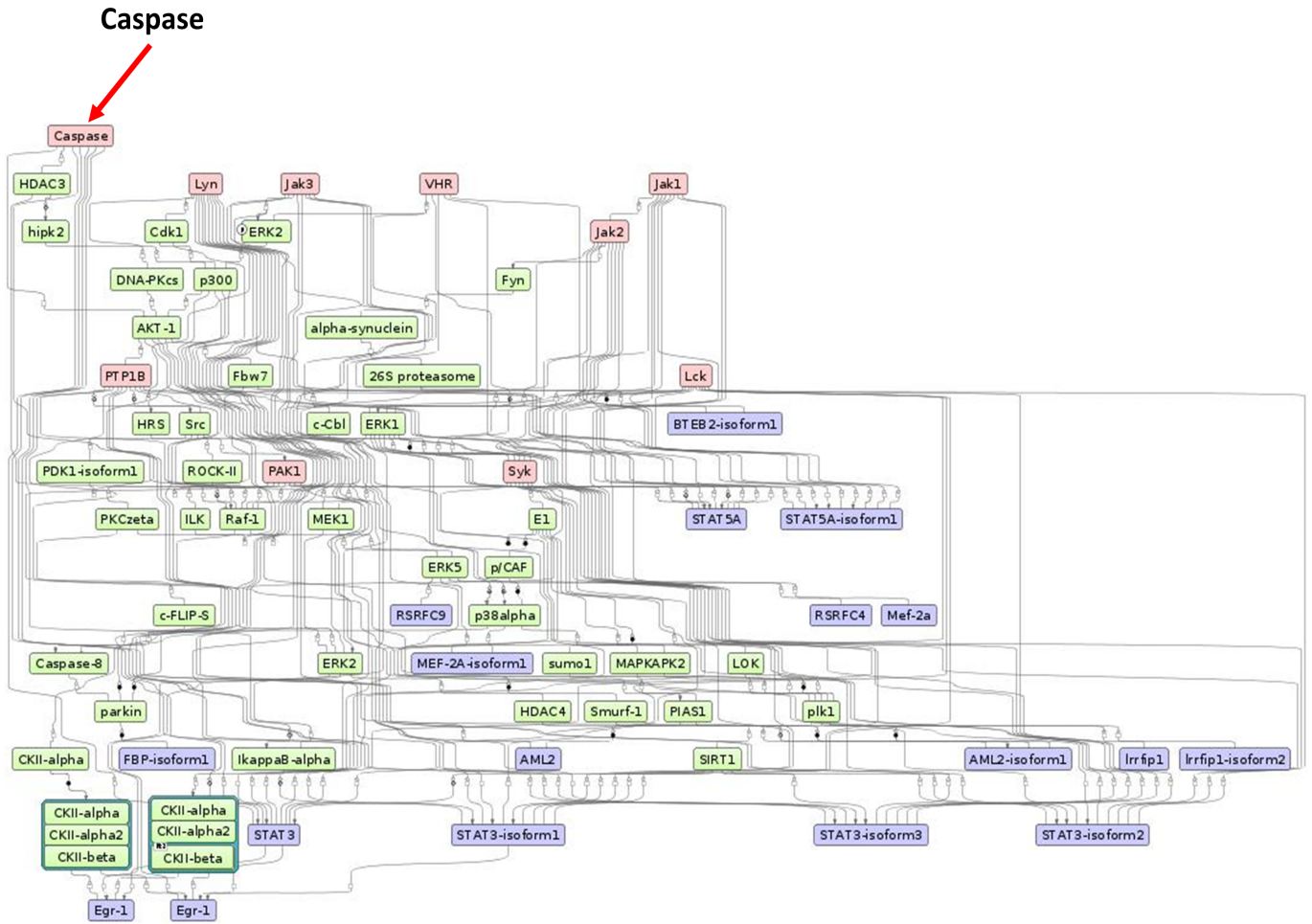


Figure 7. The master regulatory networks identified in Sheko. The colors red, blue, and green indicate master regulators, regulated proteins, and connecting molecules, respectively. Adapted from Mekonnen et al. (2019).

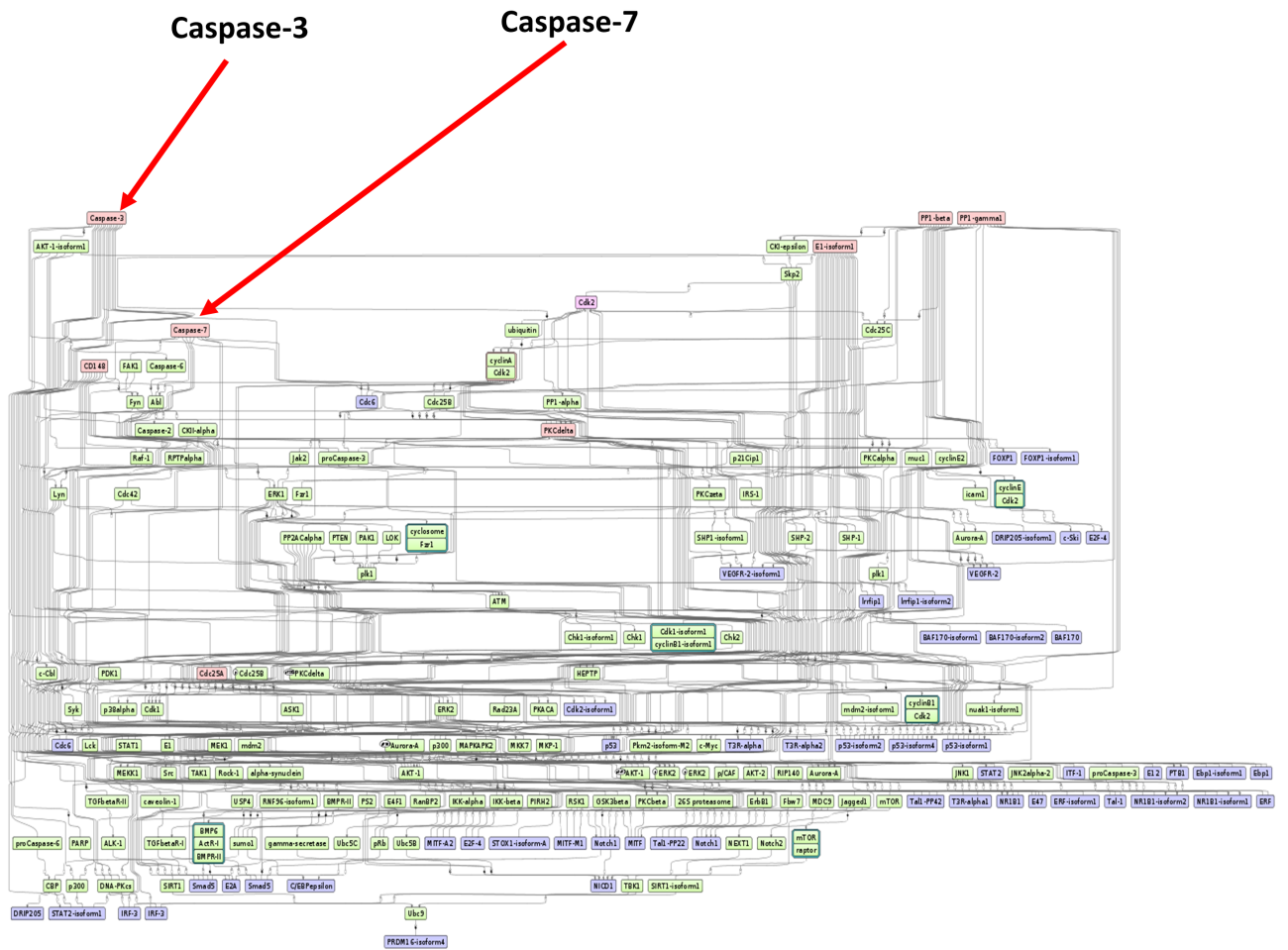


Figure 8. The master regulatory networks identified in Gindeberet. The colors red, blue, and green indicate master regulators, regulated proteins, and connecting molecules, respectively. The red/pink coloration of the master regulators represents the intensity of the degree of overrepresentation for the respective master regulator.

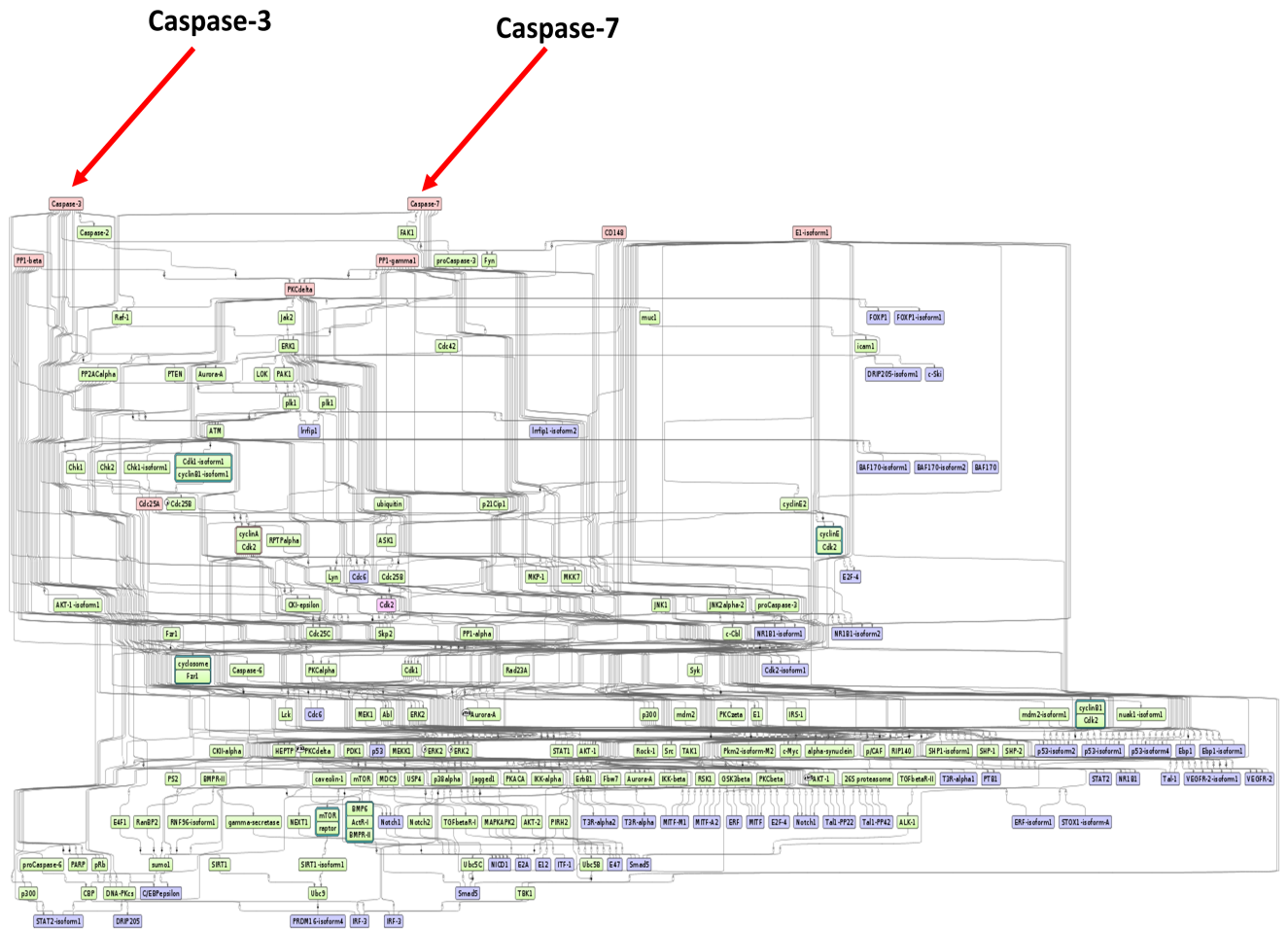


Figure 9. The master regulatory networks identified in Benshangul. The colors red, blue, and green indicate master regulators, regulated proteins, and connecting molecules, respectively. The red/pink coloration of the master regulators represents the intensity of the degree of overrepresentation for the respective master regulator.

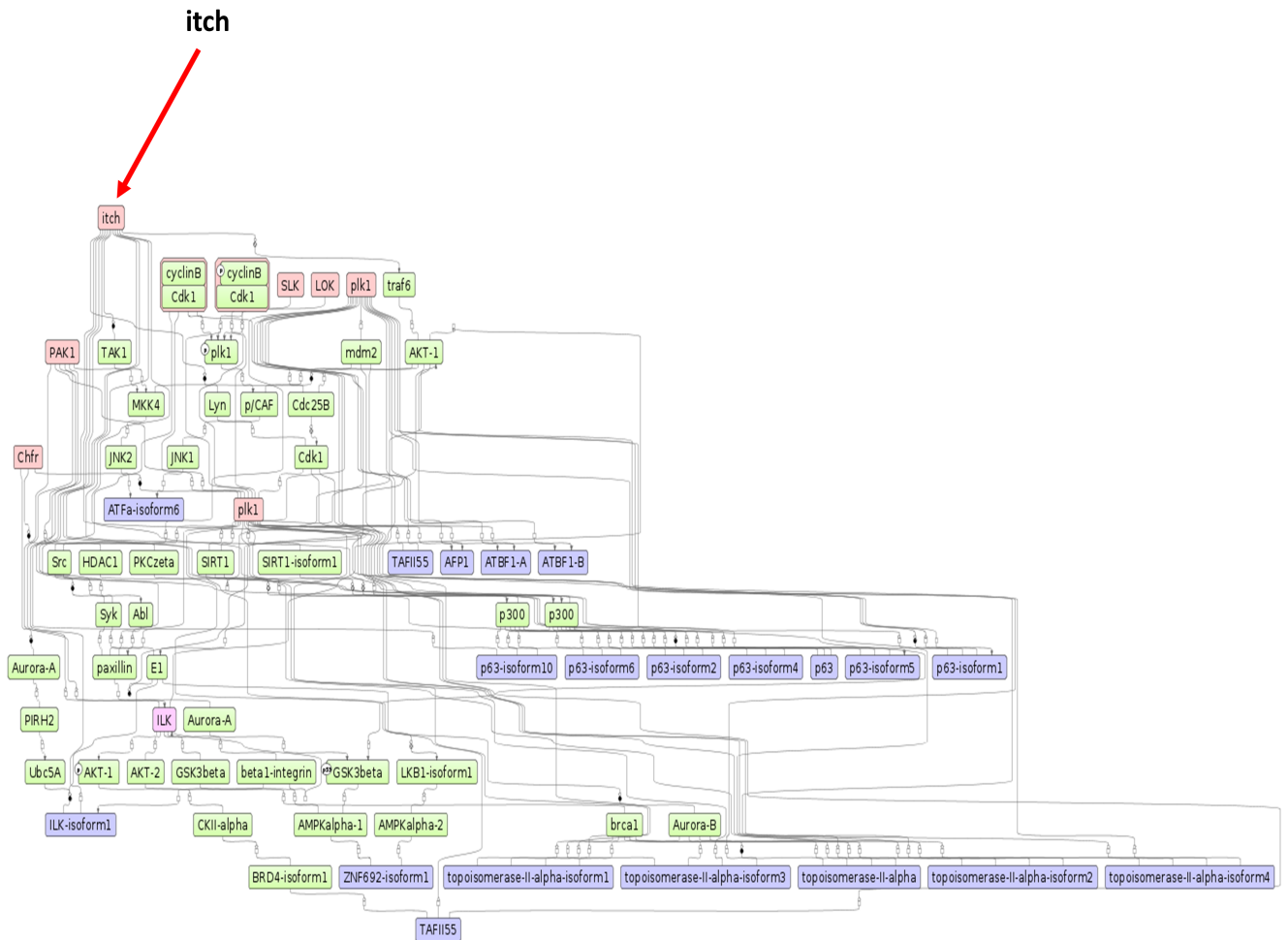


Figure 10. The master regulatory networks identified in Nuer. The colors red, blue, and green indicate master regulators, regulated proteins, and connecting molecules, respectively. The red/pink coloration of the master regulators represents the intensity of the degree of overrepresentation for the respective master regulator.