Deciphering the molecular mechanisms underlying complex traits using bioinformatics and computational biology approaches

Dissertation

zur Erlangung des naturwissenschaftlichen Doktorgrades

"Doctor rerum naturalium"

der Georg-August-Universität Göttingen

im Internationalen Promotionsprogramm für Agrarwissenschaften in Göttingen (IPAG)

der Graduiertenschule Forst- und Agrarwissenschaften (GFA)

vorgelegt von

Abirami Rajavel
aus Trichy, Indien

Göttingen
2022

Supervisory committee

Prof. Dr. Armin O. Schmitt,
Department of Animal Sciences, Georg August University of Göttingen.

Prof. Dr. Klaus Jung,
Institute for Animal Breeding and Genetics, University of Veterinary Medicine, Hannover.

Prof. Dr. Mehmet Gültas,
Faculty of Agriculture, South Westphalia University of Applied Sciences, Soest.

Date of thesis defense: 25.10.2022

Summary

With the advent of high-throughput sequencing technologies, omics studies (genomics, transcriptomics, proteomics, and metabolomics) are becoming increasingly popular to decipher the molecular patterns in association with a disease or a biological process. Access to such data has revolutionized the field of agriculture and provides novel perspectives for several systems biology studies. Furthermore, the usage of multi-omics data has proven to be powerful and accurate to study the biologial processes in association with the growth, development, adaptation and disease progression in higher organisms. Moreover, systems biology approaches enable the integration of multi-omics data to create a holistic understanding of the molecular mechanisms underpinning complex traits.

This thesis addresses the regulatory mechanisms underlying two complex traits in agricultural species namely (i) African animal trypanosomiasis disease (AAT) resistance in cattle breeds and (ii) seed oil content of the oilcrop rapeseed. Chapters 2, 3 and 4 detail projects corresponding to the AAT disease progression in cattle and Chapter 5 deals with the investigation of the seed oil content of *Brassica napus*. The thesis is organised into the chapters as follows:

In **Chapter 1**, I present a brief introductory overview of the main concepts in this thesis, especially transcription factor (TF) co-operations and monotonically expressed genes (MEGs). Additionally, this chapter reviews the current knowledge in (i) African animal trypanosomiasis disease progression in cattle breeds and (ii) the seed oil content of the oilcrop *Brassica napus*.

I investigate in **Chapter 2** a continuous transcription profiling time-series microarray dataset obtained from three tissues (liver-, spleen-, and lymph node tissues) of the two cattle breeds: trypanosusceptible Boran and trypanotolerant N'Dama, after being infected with *Trypanosoma congolense*. In this study, I attempt to capture the transcriptional events while considering the multistage progression process of AAT disease through the identification of MEGs. As a result, I identify several tissue-specific TF cooperations for

the tissues of both cattle breeds, from which I observe and explain the role of preferential partner choices of TFs in association with the trypanosusceptibility and trypanotolerance mechanisms.

In **Chapter 3**, I focus on the upstream regulatory processes underlying the multi-stage progression process of the AAT disease. Therefore, I consider the MEGs with regular ascending or descending expression pattern during the disease to computationally identify their upstream master regulators and over-represented pathways governing the regulatory mechanisms of trypanotolerance level of both cattle breeds. Consequently, this study unravels unique, cattle breed-specific master regulators and over-represented signaling pathways for three different tissues (liver, spleen and lymph node) of these two cattle breeds, respectively. Furthermore, pathway analysis also bolsters the crucial roles of these master regulators.

To complement the previous studies focusing on upstream regulatory processes, I study the influence of downstream regulatory events involving the effector molecules and their complex interplay with the regulatory SNPs and gene expression in **Chapter 4**. For this purpose, I combine transcriptomic and genomic datasets of Boran and N'Dama cattle breeds. Further, I focus on the MEGs found for three tissues of both cattle breeds and variants (SNPs) that are located in the promoter regions of the MEGs. Moreover, I manually analyse and annotate the gene expression profiles of MEGs for each tissue in order to find highly interesting gene expression profiles, differentiating between the cattle breeds.

Knowledge regarding transcriptional regulation is important to gain insights into the developmental switches between the cultivars of *Brassica napus* L., namely Zhongshuang11 (ZS11), a double-low accession with high-oil- content, and Zhongyou821 (ZY821), a double-high accession with low-oil-content. In **Chapter 5**, I analyse a time series RNA-seq dataset of seed tissue from both cultivars by mainly focusing on the MEGs. The consideration of the MEGs enables the capturing of the multi-stage progression process that is orchestrated by the cooperative TFs and, thus, facilitates the understanding of the molecular mechanisms determining seed oil content. In this study, I report that TF families, such as NAC, MYB, DOF, GATA, and HD-ZIP, are highly involved in the seed developmental processes. Particularly, their preferential partner choices as well as changes in their gene expression profiles seem to be strongly associated with the differentiation of the oil content between the two cultivars.

In **Chapter 6**, I discuss the main findings of the thesis and conclude in **Chapter 7** with a summary and an outlook on future potential directions.

Zusammenfassung

Mit dem Aufkommen von Hochdurchsatz-Sequenzierungstechnologien werden Omics Studien (Genomik, Transkriptomik, Proteomik und Metabolomik) immer beliebter, um die mit einer Krankheit oder einem biologischen Prozess verbundenen molekularen Funktionen zu entschlüsseln. Der Zugang zu solchen Daten hat die Agrarwissenschaften revolutioniert und bietet neue Perspektiven für verschiedene systembiologische Studien. Darüber hinaus hat sich die Verwendung von Multi-Omics-Daten als leistungsfähig und genau erwiesen, um biologische Prozesse im Zusammenhang mit Wachstum, Entwicklung, Anpassung und Krankheitsverlauf in einem Organismus zu untersuchen. Außerdem haben systembiologische Ansätze die Integration von Multi-Omics-Daten ermöglicht, um ein ganzheitliches Verständnis der molekularen Mechanismen zu erlangen, die komplexen Merkmalen zugrunde liegen.

Diese Arbeit befasst sich mit den Regulationsmechanismen, die zwei komplexen Merkmalen bei landwirtschaftlich relevanten Arten zugrunde liegen (i) der Resistenz gegen die Afrikanische Trypanosomiasis-Krankheit (AAT) beim Rind und (ii) dem Samenölgehalt von Raps (*Brassica napus*). In den Kapiteln 2, 3 und 4 werden Projekte zum Verlauf der AAT-Krankheit bei Rindern beschrieben, und Kapitel 5 befasst sich mit der Untersuchung des Samenölgehalts von *Brassica napus*. Die Arbeit ist wie folgt gegliedert:

In **Kapitel 1** gebe ich einen kurzen Überblick über die wichtigsten Konzepte dieser Arbeit, insbesondere über die Zusammenarbeit von Transkriptionsfaktoren (TFs) und monoton exprimierten Genen (MEGs). Darüber hinaus gibt dieses Kapitel einen Überblick über den aktuellen Wissensstand bezüglich (i) des Krankheitsverlaufs der Afrikanischen Trypanosomiasis beim Rind und (ii) des Samenölgehalts von *Brassica napus*.

In **Kapitel 2** untersuche ich einen kontinuierlichen Transkriptions-Profiling-Zeitreihen-Microarray-Datensatz, der aus drei Geweben (Leber-, Milz- und Lymphknotengewebe) von zwei Rinderrassen gewonnen wurde: den trypanosuszeptiblen Boran und den trypanotoleranten N'Dama, nachdem sie

mit *Trypanosoma congolense* infiziert wurden. In dieser Studie versuche ich, die transkriptionellen Ereignisse unter Berücksichtigung des mehrstufigen Verlaufs der AAT-Krankheit durch die Identifizierung von MEGs zu erfassen. Als Ergebnis finde ich mehrere gewebespezifische TF-Kooperationen für die Gewebe beider Rinderrassen, anhand derer ich die Rolle der bevorzugten Partnerwahl von TFs in Verbindung mit den Mechanismen der Trypanosanfälligkeit und Trypanotoleranz beobachte und erkläre.

In **Kapitel 3** konzentriere ich mich auf die upstream gelegenen regulatorischen Prozesse, die dem mehrstufigen Fortschreiten der AAT-Krankheit zugrunde liegen. Dazu betrachte ich MEGs mit monotonen auf- oder absteigenden Expressionsmustern im Krankheitsverlauf, um ihre upstream gelegenen Masterregulatoren und überrepräsentierten Signalwege zu identifizieren, die die Regulationsmechanismen der Trypanotoleranz in beiden Rinderrassen bestimmen. Folglich deckt diese Studie einzigartige, rinderrassenspezifische Masterregulatoren und überrepräsentierte Signalwege für drei verschiedene Gewebe (Leber, Milz und Lymphknoten) dieser beiden Rinderrassen auf. Darüber hinaus untermauert die Signalweganalyse die entscheidende Rolle dieser Hauptregulatoren.

Zur Ergänzung der bisherigen Studien, die sich auf upstream gelegene regulatorische Prozesse konzentrieren, untersuche ich in **Kapitel 4** den Einfluss downstream regulatorischer Ereignisse, an denen Effektormoleküle beteiligt sind, und deren komplexes Zusammenspiel mit regulatorischen SNPs und der Genexpression. Zu diesem Zweck kombiniere ich Transkriptom- und Genom-Datensätze von den Rinderrassen Boran und N'Dama. Ich konzentriere mich auch auf die MEGs, die in drei Geweben beider Rinderrassen gefunden wurden, und auf Einzelnukleotid-Polymorphismen (SNPs), die sich in den Promotorregionen der MEGs befinden. Darüber hinaus analysiere und annotiere ich die Genexpressionsprofile der MEGs für jedes Gewebe manuell, um hochinteressante Genexpressionsprofile zu finden, die sich zwischen den Rinderrassen unterscheiden.

Kenntnisse über die Transkriptionsregulierung sind wichtig, um Einblicke in die Entwicklungsschalter zwischen den Sorten von *Brassica napus* L. zu gewinnen, nämlich Zhongshuang11 (ZS11), einer Sorte mit niedrigen Erucasäure- und Glucosinolat-Anteil (Doppelnull-Sorte), und Zhongyou821 (ZY821), einer Sorte mit hohen Erucasäure- und Glucosinolat-Anteil (Doppelplus-Sorte). In **Kapitel 5** analysiere ich einen Zeitreihen-RNA-seq-Datensatz vom Samengewebe beider Sorten, wobei ich mich hauptsächlich auf die MEGs konzentriere. Die Berücksichtigung der MEGs ermöglicht die Erfassung eines mehrstufigen Entwicklungsprozesses, der von den kooperativen TFs gesteuert wird, und erleichtert somit das Verständnis der moleku-

laren Mechanismen, die den Ölgehalt der Samen bestimmen. In dieser Studie berichte ich, dass TF-Familien wie NAC, MYB, DOF, GATA und HD-ZIP stark in den Entwicklungsprozess der Samen involviert sind. Insbesondere ihre bevorzugte Partnerwahl sowie Veränderungen in ihren Genexpressionsprofilen scheinen stark mit der Differenzierung des Ölgehalts zwischen den beiden Kultivaren verbunden zu sein.

In **Kapitel 6** diskutiere ich die wichtigsten Ergebnisse der Arbeit und schließe diese Thesis in **Kapitel 7** mit einer Zusammenfassung und einem Ausblick auf zukünftige Forschungsmöglichkeiten ab.

Acknowledgement

A scientific endeavour is a quest that cannot be accomplished alone. This journey would not have been possible without the encouragement, teaching, and assistance of many people. I am sincerely thankful for all the love and support.

First and foremost, I would like to thank Prof. Dr. Armin Schmitt for providing me the opportunity to work in the Breeding Informatics group and for being my first supervisor. I am forever thankful to him for sharing his wealth of knowledge, continuous support and for providing a wonderful scientific environment which is nurturing new ideas.

I am eternally grateful to Prof. Dr. Mehmet Gültas, for providing me consistent input, encouragement and fantastic guidance all throughout my Ph.D. I cannot thank you enough for the invaluable suggestions, mentoring and for always finding time to offer constant assistance.

I sincerely thank Prof. Dr. Klaus Jung for being the member of my thesis committee and for providing insights to help refine my work. I would like to repeat my thanks to Prof. Dr. Kun Lu for his collaboration in the *Brassica napus* project.

I would like to convey my gratitude to Mrs. Monika Siebert for her friendly support and her constant help in the office.

I am completely lucky to be a member of this incredible team of Breeding Informatics group. Having this opportunity was without a doubt one of the greatest delights of my Ph.D. I cannot have been luckier with my colleagues.

I am deeply thankful for Felix Heinrich for taking care of computational problems and for helping me with data analysis. Special thanks extends to Selina for providing me outstanding support in analysing regulatory SNPs, in proofreading and completing manuscripts. For several of my analysis, Martin provided support to write R scripts for which I am greatly thankful. His inclination to help has comforted me. Thank you Martin. I specially thank Dr. Faisal Ramzan, who is a former colleague of the Breeding Informatics

group. I am thankful to Faisal for teaching GWAS analysis and his invaluable discussions on everything inside and outside of the projects. His presence has always brought spark in the office in the two years. I would like to express my gratitude to my other colleagues Haleem, Thomas and Hendrik for creating a wonderful atmosphere and for their incredible support. I wish to thank Magdalena and Johanna for being my tutorial buddies in teaching. I have learnt a lot from both of them. I extend my thanks to Johanna for helping me with the enrichment analysis and generation of treemaps. Thank you all for being my wonderful colleagues and for bringing camaraderie and joy all throughout these years. Moreover, it has been a pleasure to work with Yuehan Hui and Muhammad Jawad (bachelor and master students), who have performed their scientific projects with me.

I am deeply thankful to the Göttingen Nirmithi group for their support as a community to deal with everyday life in Göttingen. My sincerest gratitude to my friend Madhumalar whose constant support both personally and professionally inspires me for which I cannot do justice with nothing, but by the reaffirmation of friendship. I extend my thanks to Srini who is always there for me as my brother abroad. I am forever thankful to both.

I thank my parents for believing in me and my decisions, and for showing me the meaning of hard work and purpose of life. The unconditional support from my brother Radha has filled me with courage and joy to always move forward not only in my Ph.D. journey, but also in life. I cannot thank you enough.

At last, I thank the universe for making me realise this dream and to make it come true.

Thank you!

Abirami Rajavel

Contents

List of tables

List of figures

Abbreviations

AAT	African animal trypanosomiasis
DNA	Deoxyribonucleic acid
DEG	Differentially expressed gene
FDR	False discovery rate
MEG	Monotonically expressed gene
PMI	Pointwise mutual information
PWM	Position weight matrix
QTL	Quantitative trait loci
RNA	Ribonucleic acid
SNP	Single Nucleotide Polymorphism
rSNP	Regulatory Single Nucleotide Polymorphism
TAG	Triacylglycerol
TF	Transcription factor
TFBS	Transcription factor binding site
TPM	Transcript per million
TSS	Transcription start site
ZS11	Zhongshuang11
ZY821	Zhongyou821

Chapter 1

General Introduction

Recent advancements in high-throughput sequencing technologies during the last decades and decreasing costs of sequencing revolutionised the field of agriculture. Large amounts of data generated from high-throughput sequencing is indispensable for answering important biological questions as they provide information for investigating multiple molecular layers including genomics, transcriptomics, proteomics, metabolomics, epigenomics, commonly regarded as "omics". In particular, molecular biology is becoming rapidly data-driven based on the improvement of multi-omics data integration and its applications. In this thesis, I will present my research work on the computational investigation, interpretation, and application of omics data to understand the molecular mechanisms underlying the complex traits in agricultural species. Specifically, in this chapter, I will give a brief introductory overview of the concepts used in this thesis.

1.1 Gene regulation

Gene expression is regulated by the collaborative action of several transcription factors (TFs). TFs are a special class of regulatory proteins that specifically bind to short DNA sequences, which allows them to activate or repress the expression of specific genes [1]. These proteins perform their regulatory functions by directly binding to the specific regulatory DNA sequences called transcription factor binding sites (TFBSs) [2]. As a result, TFs are in direct interaction with the transcriptional machinery and they could change the structure of chromatin through DNA modifications. Taking their molecular details into account, TFs contain the following domains i) a DNA binding domain, ii) an oligomerization domain, iii) a regulatory domain and iv) a trans-activation domain [3]. Importantly, the DNA binding domain is in-

volved in the recognition of the specific DNA sequences and in facilitating the binding of the TF protein to the DNA. Furthermore, the regulatory domain governs the activity of a TF [4].

Today it is well known that the complex interplay of TFs is essential for gene regulation. Several of the TF proteins physically co-operate with other TFs, forming functional homodimers or heterodimers [5]. TF-TF cooperations could take place between TFs binding the same DNA region or different DNA regions. The preferential partner choice of the TFs influences the regulation of a biological process by stimulating a chain of molecular events which thereby changes the fate of the cell. Therefore, TF dimerization could contribute to the complexity of the gene regulation process, while considering the availability of an ample number of TFs [6]. However, despite the computational challenges using bioinformatics approaches, TF co-operations could be deciphered by the prediction of TFBSs.

The orchestration of cell differentiation changes its direction in terms of developmental switches and specifying cellular fate in eukaryotes, depending on the specific gene regulatory events that are governed by TFs and their preferential partner choices [7]. Therefore, the knowledge regarding TFs and their cooperations is important to gain new insights into the genetic programs regulating the biological processes.

1.2 Gene expression

Gene expression occurs when DNA is transcribed into RNA which is then translated into proteins. The rate of production of functional proteins in the cell is regulated at many stages of gene expression, primarily at the level of transcription but also at post-transcriptional levels. It is convenient to describe gene expression as a series of sequential steps, from transcription to post-translational protein modifications [8].

In several biological studies, researchers investigate how gene or protein expression changes over time, for example, while analysing the stages of a disease progression [9]. A classical approach to address the changes in gene expression is to identify the differentially expressed genes (DEGs), especially between two timepoints. However, in time-series microarray data analysis consisting of measurements from several timepoints, finding genes with increasing or decreasing monotonic expression patterns over time is important. Monotonically expressed genes (MEGs) are genes whose expression pattern follows a monotonic pattern either in ascending or descending order as a disease progresses or over a time period (see Figure 1.1 and 1.2). Recently, a

study by Tian [10] investigated the MEGs using the Monotonic Feature Selector (MFSelector) method [11] for a dataset consisting of different stages of non-small cell lung cancer. Similarly, considering African animal trypanosomiasis (AAT) disease as a multi stage progression process, identifying the MEGs involved in this disease could provide insights into the underlying molecular mechanisms. To find monotonic genes, we apply the MFSelector technique [11], which is based on the idea of total discriminating error (DE_{total}). MFSelector computes DE_{total} for each gene by considering several stages in stage order and is capable of detecting genes with monotonic features.

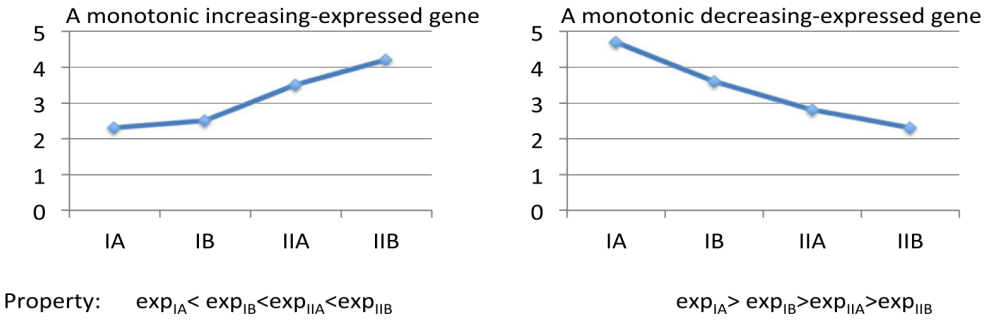

Figure 1.1: Definition for a MEG which is monotonically expressed either in the increasing or decreasing order (adapted from [10]).

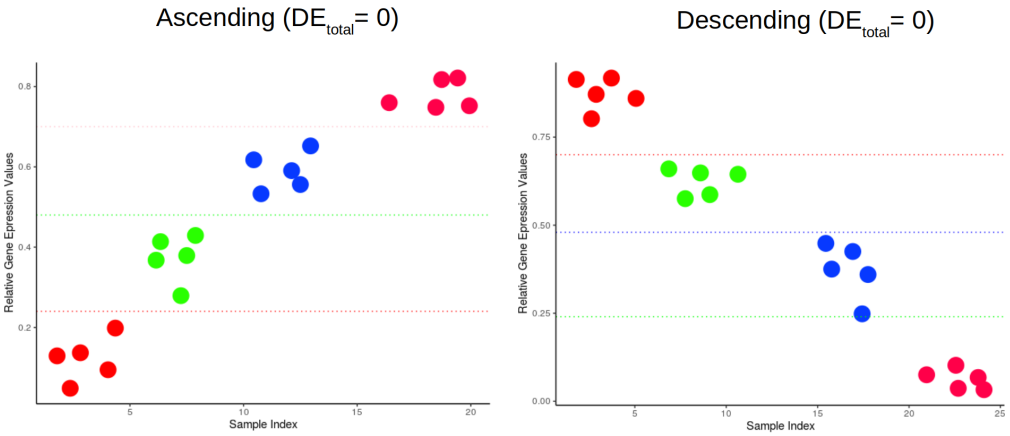

Figure 1.2: Graphical representation of monotonically expressed genes following an increasing or a decreasing pattern over time.

1.3 Multi-omics technologies

Multi-omics is an integrative approach which combines multiple omics data including genomics, transcriptomics and proteomics for measuring the expression of a gene or the amounts of protein. Moreover, multi-omics analysis facilitates a holistic understanding of molecular events involved in biological processes such as development, response to stimuli and disease progression. Furthermore, integrated multi-omics data analysis establishes genotype to phenotype associations and assists in the discovery of novel therapeutic targets and biomarkers [12]. I present a brief description in the following paragraph about different omics studies (see Figure 1.3), techniques and technologies available for profiling.

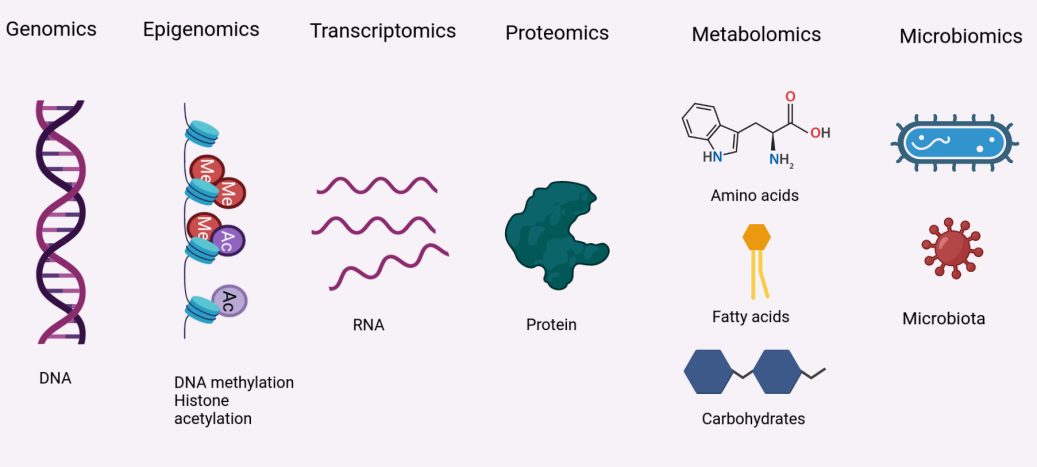

Figure 1.3: A schematic representation of multi-omics data (Created with www.biorender.com).

Genome refers to the complete genetic information of an individual and genomics refers to the study of structure, function, mapping, and editing of information within the genome [13]. Profiling of genomic sequences and their associated information is performed by using techniques such as whole-genome sequencing [14, 15], exome sequencing [16], targeted sequencing and microarrays. Secondly, the transcriptome is the complete set of RNA transcripts including messenger RNA (mRNA), ribosomal RNA (rRNA), transfer RNA (tRNA), and other non-coding RNAs within a cell or a population of cells [17]. Transcriptomics is the study of the transcriptome and how it changes while responding to the disease mechanisms or other regulatory processes [13]. Transcriptomic data is obtained through mRNA-seq, whole transcriptome, and targeted RNA sequencing techniques [18]. Thirdly, pro-

teome is the entire set of proteins, expressed by the genome of an organism at a particular time in a specific tissue type/cell. Proteomics is the characterization of a large set of proteins, particularly their structure, function and their interaction [12] which determines the expression patterns of proteins in response to certain stimuli. The laboratory techniques in use to study the proteome include two-dimensional gel electrophoresis and mass spectrometry [19].

1.4 Application of integrated multi-omics data analysis in agricultural species

With the advent of high-throughput sequencing technologies, it is made possible to investigate the molecular mechanisms underlying important traits in agricultural species. Having analysed both animal and plant datasets in this thesis, the following sections will provide background information in association with the two datasets analysed. The first section 1.4.1 comprises of information regarding the AAT disease, on the other hand, the second section 1.4.2 addresses the seed oil content of the crop *Brassica napus*.

1.4.1 African trypanosomiasis

African trypanosomiasis is caused by the unicellular protozoan parasites from the genus *Trypanosoma* and it is mainly spread by the tsetse fly, which carries the parasite in its saliva [20, 21, 22]. This disease infects a broad range of mammalian hosts, which include humans, cattle, sheep, goats, pigs, horses and donkeys [23, 24, 25, 26]. Further, it is predominantly found in the sub-Saharan countries, leading to huge economic losses of up to 5 billion US dollars every year [27]. When trypanosomes infect humans, the disease is called sleeping sickness as it interferes with the sleep-wake cycle [28, 29]. Human African trypanosomiasis is mainly caused by the subspecies *Trypanosoma brucei* [30, 31]. On the other hand, when the trypanosomes infect animals, it is regarded as African animal trypanosomiasis (AAT). The most pathogenic species affecting animals include *Trypanosoma brucei*, *Trypanosoma vivax*, and *Trypanosoma congolense* [32].

1.4.1.1 Trypanosomes

Trypanosomes are unicellular protozoan parasites, which live in the bloodstream and tissue fluids of the mammalian host. They are flagellated organisms that multiply in the body of the hosts. There are different species of

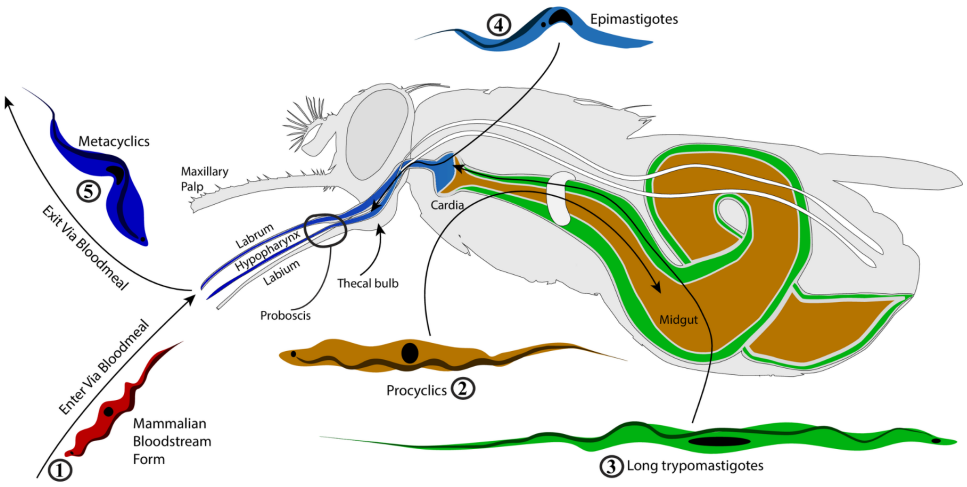

Figure 1.4: The life cycle of *Trypanosoma congolense* (adapted from [33]).

Trypanosoma having a variety of morphological characteristics which enable the distinct species identification [34]. The genus *Trypanosoma* is classified into two sections: stercoraria and salivaria [35]. Stercorarian trypanosomes grow in the hindgut of a vector such as the tsetse fly. They are transmitted to the host via faeces [35, 36]. On the other hand, salivarian trypanosomes grow in the front portion of the digestive tract of the fly, either in the salivary glands (*T. brucei*) or in the proboscis (*T. congolense* and *T. vivax*). These species are transmitted through the saliva [35]. In Figure 1.4, the life cycle of *Trypanasoma congolense* is depicted which demonstrates the entry and exit of *T. congolense* from the vector tsetse fly.

1.4.1.2 Economic importance of AAT

AAT poses a serious concern in several countries despite the continuous research for many years, constraining economic development in sub-Saharan Africa [38]. AAT affects domestic animals, thereby creating a negative impact with serious health issues as well as a hindrance for crop-livestock production systems [39]. Due to this disease, the major losses to livestock farming are the mortality and morbidity of cattle, treatment costs and reduction in meat and milk production [40]. This disease results in an increased mortality rate of the animal in acute stages and decreased productivity at chronic stages. In trypanosomiasis prevalent areas, animals were reported with a 6-20% increase in calf mortality per year, 6-19% reduced calving rate, 20% reduction

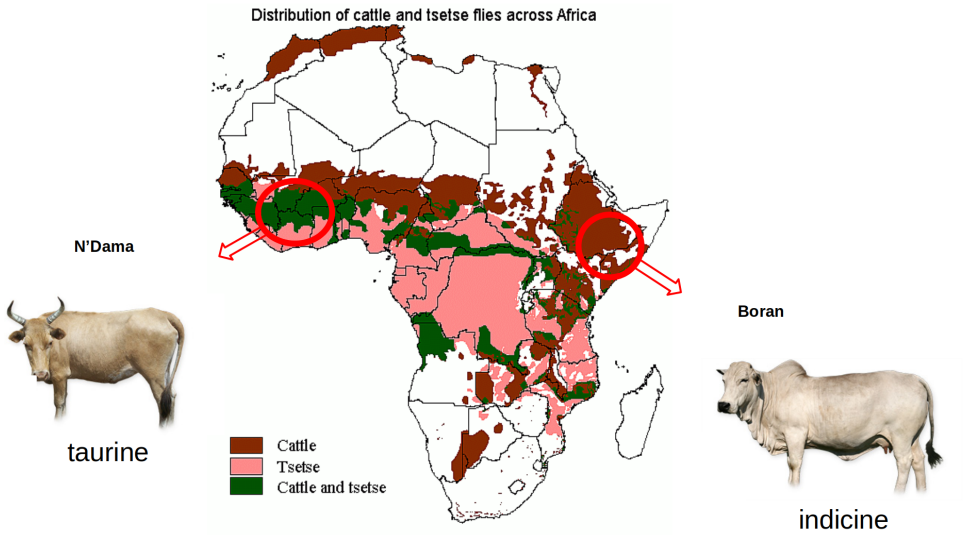

Figure 1.5: Distribution of cattle and tsetse flies across Africa, indicating the survival of trypanotolerant *Bos taurus* in West Africa and trypanosusceptible *Bos indicus* breeds in East Africa (The figure is adapted from [37] and www.krankykids.com accessed on 22/02/2022).

in milk production and up to 38% loss of body weight [41]. Annually, farmers face a huge loss of 3 million cattle [42]. Furthermore, the diseased animals influence the crop farming due to decreased productivity [39].

1.4.1.3 Clinical features of AAT

In the initial acute stages of the AAT disease, the animal shows symptoms like intermittent fever, weight loss, eye discharge and anaemia [43]. Furthermore, the disease leads to reduction in milk production as well as of their growth and draught power. In the chronic stages of the disease, the animals suffer from neuronal dysfunctions, abortion, and congestive heart failure which results in the death of the animal. Anaemia has been observed in both acute and chronic stages of the infection. In the acute infection stage, anaemia has been related with the host immune response. On the other hand, it is associated with splenomegaly and haemodilution at the chronic stages of the disease [2, 45, 46, 47].

1.4.1.4 Trypanotolerant cattle breeds

The ability of cattle to control anaemia and parasitaemia despite the infection with trypanosomes and their capacity to remain productive is termed as

the trait of trypanotolerance [23, 24, 25, 26]. Few west African cattle breeds are resistant to trypanosomiasis, in contrast to other cattle breeds which are susceptible. Cattle breeds such as N'Dama, Baoulé, and Muturu are the most widely known trypanotolerant breeds [48, 49]. The trait of trypanotolerance is prevalent in 6% of the total African cattle population and in 17% of the cattle in West Africa, making it the largest cattle population in this tsetse-challenged region [50]. Important constraints in using the trypanotolerant breeds for farming purposes are their small size, decreased agricultural productivity and reduced draught power in comparison with *Zebu* breeds [48, 49, 51, 52]. The distribution of the trypanotolerant N'Dama and the trypanosusceptible Boran cattle breed across Africa is represented in Figure 1.5.

Concerning the genetics of the trypanotolerant breeds, it is reported that cytokine production differentiates the immune responses between susceptible and tolerant hosts [53]. Studies also demonstrated the higher levels of antibody production in trypanotolerant hosts compared to susceptible hosts, thereby eliciting strong host immunity against the parasite proliferation within the host [54, 55]. Even though this trait is not absolute, resistance to the AAT disease could be improved by cross breeding with these breeds [56].

1.4.2 Oil crop *Brassica napus*

Brassica napus L. is the one of the largest oilseed crops which is grown all over the world for vegetable oil production [57]. It belongs to the Brassicaceae family (oilseed rape, 2n = 38). This allotetraploid crop resulted from the hybridisation of two diploid species, namely *Brassica rapa* L. (turnip rape, 2n = 20) and *Brassica oleracea* L. (cabbage, 2n = 20) [58, 59]. Currently, oilseed rape is a major oil crop with a wide range of applications as nutritious food oil, biodiesel, lubricants and animal fodder [60]. Attempts have been made continuously in the past two decades to increase the seed oil content ranging from 40 to 50 % for most of the cultivars [61]. Improving the seed quality with increased oil content has become an important trait for rapeseed breeding programs due to the growing global demand for oil production, for their use as bio-fuel, and vegetable oil [57].

1.4.2.1 The seed oil content

The most advantageous component of an oilseed crop is the seed oil content. Seed oil is present in the form of lipid droplets called oleosomes [62, 63]. Fatty acids are the major constituents of seed oil which account for the growing

economic importance of rapeseed. The oilseed rape comprises different fatty acids namely palmitic, palmitoleic, stearic, oleic, linoleic and linolenic acid. The fatty acid composition is affected by the environment [64, 65], for example, cold climate and high latitude lead to increasing amounts of acyl groups. In addition, genetics control the fatty acid composition to a greater extent. Conventional breeding has exploited these genetic mechanisms and increased the fatty acid content of the seed oil which makes it suitable for food and industrial applications [65].

Naturally, rapeseed oil contains a large amount of the anti-nutritive fatty acid called erucic acid which is classified as a natural toxicant due to health risks [66], that makes the oil unsuitable for human consumption. Therefore, breeding efforts led to lower levels of anti-nutritional components such as erucic acid and glucosinolates that are not useful for human consumption or as fodder for animal consumption. The variety of oilseed rape with reduced erucic acid and glucosinolates was introduced as double low quality oilseed rape or canola [67]. Furthermore, all modern varieties contain a mutation in both copies of the fatty acid elongase gene *FAE1* [68, 69], that produces seed oil with less C22:1 and more oleic acid (C18:1). Therefore, this variety is regarded as the nutritious plant-based oil [70]. Moreover, it is appropriate for biodiesel production and animal feed.

1.4.2.2 Storage lipids

Lipids play an important role as a biological molecule in the living cell. They perform biological functions such as energy storage and signaling [71, 72]. They are found in various forms including fats, sterols, phospholipids, fat-soluble vitamins, mono-, di-, triglycerides and glycolipids. The seeds function as storage organ for lipids which are called storage lipids (94%) comprised mostly of triacylglycerols (TAGs) [73, 74, 75].

At the maturation stage of the seed, oil is synthesized and serves as the crucial energy reserve which is necessary for the growth and development. Seed oil comprises of TAGs produced from fatty acids and glycerol-3-phosphate. They are formed in a sequence of steps which involve the synthesis of fatty acids and the export of acyl-coA thioesters [78]. Fatty acids are synthesized from acetyl-CoA which is catalysed in the plastid compartment by several enzymes including major enzymes like acetyl-CoA carboxylase (ACCase) and the FA synthase (FAS). ACCase plays the most important role as a rate-limiting enzyme catalysing the conversion of acetyl-CoA to malonyl-CoA. On the other hand, FAS is involved in the catalysis of malonyl-CoA to acyl-carrier protein by transferring the malonyl moiety. This step adds two car-

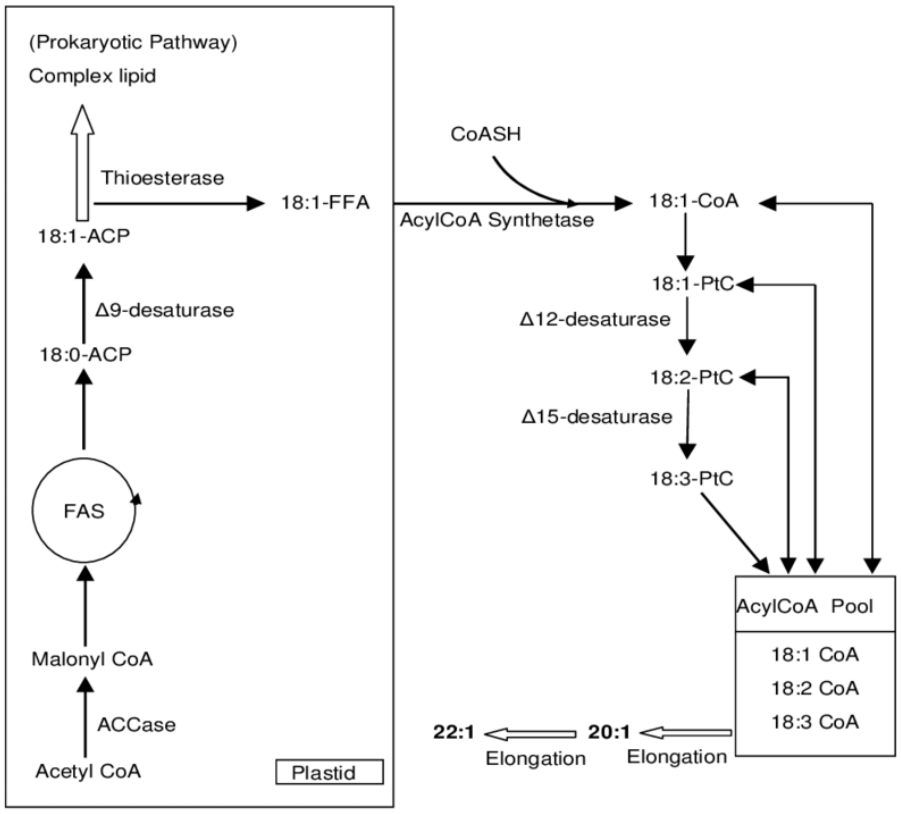

Figure 1.6: Fatty acid synthesis in *Brassica napus*. ACCase: Acetyl-CoA carboxylase, ACP: Acyl Carrier Protein, FAS: Fatty acid synthetase, FFA: Free fatty acid, PtC: Phosphatidyl choline, and CoASH: Activated CoA (adapted from [76, 77]).

bons, thereby forming acyl-ACP which is later transferred to the cytoplasm. FA dehydrogenase is involved in the catalysis, resulting in the formation of unsaturated fatty acids. Afterwards, FA elongase catalyse the addition of two carbons to form long chain FA. Consequently, TAG synthesis is initialized in the endoplasmic reticulum. A schematic representation of fatty acid synthesis in *Brassica napus* is shown in Figure 1.6. Synthesis of fatty acids and TAG are in close association with carbohydrate metabolism and photosynthesis [79]. Analysis of DNA microarray data showed that transcriptional control is important in regulating the fatty acid synthesis [80]. In addition, optimisation of the enzyme activity also regulates the process of fatty acid synthesis [81].

1.5 Aim of the thesis

The aim of the present dissertation is to mainly decipher the regulatory mechanisms using bioinformatics and systems biology approaches in association with (i) African animal trypanosomiasis disease progression in cattle breeds, and (ii) seed oil content of two cultivars of *Brassica napus*. The following chapters describe the projects that resulted in publications and the development of this thesis. Further, the context of the chapters is mainly based on my publications [82, 83, 84, 85].

- **Chapter 2:** I investigate the gene expression data from the cattle breeds Boran and N'Dama in which the former breed is known to be susceptible to trypanosomiasis and the latter to be resistant. Focussing on the monotonically expressed genes which could capture the progression process of a disease/a biological process, I identify the tissue-specific transcription factor co-operations for both cattle breeds and I report the importance of preferential partner choices of the TFs in governing the AAT susceptibility/resistance mechanisms.

- **Chapter 3:** With regard to AAT, I report the breed-specific master regulators which orchestrate the regulatory cascades influencing the level of trypanotolerance in cattle breeds. Further, I identify over-represented pathways governing the upstream regulatory mechanisms of trypanotolerance/trypanosusceptibility of the two cattle breeds.

- **Chapter 4:** I perform the combined analysis of transcriptomic and genomic datasets of Boran and N'Dama cattle breeds, to unravel the complex interplay of regulatory SNPs (rSNPs), MEGs and downstream effectors in orchestrating the molecular mechanism of the AAT disease progression.

- **Chapter 5:** I analyse a time series RNA-seq dataset of seed tissue from two cultivars of *Brassica napus* namely Zhongshuang11 (ZS11), a double-low accession with high-oil- content, and Zhongyou821 (ZY821), a double-high accession with low-oil-content. I inspect the TF co-operations for both cultivars and report significant TF families that are highly involved in the seed developmental process of *Brassica napus*. Moreover, I describe the association between the changes in their gene expression profiles and the differentiation of the oil content between the two cultivars.

Bibliography

[1] J. D. Watson, T. A. Baker, S. P. Bell, A. Gann, M. Levine, and R. Losick, "Molecular biology of the gene - international edition," 2008.

[2] E. Wingender, T. Schoeps, and J. Donitz, "TFClass: an expandable hierarchical classification of human transcription factors," *Nucleic Acids Res.*, vol. 41, pp. D165–170, Jan 2013.

[3] E. Wingender, "Classification Scheme of Eukaryotic Transcription Factors," *Molecular Biology*, vol. 31, no. 4, pp. 483–497, 1997.

[4] E. Wingender, "Classification Scheme of Eukaryotic Transcription Factors," *Molecular Biology*, vol. 31, no. 4, pp. 483–497, 1997.

[5] A. P. Funnell and M. Crossley, "Homo-and heterodimerization in transcriptional regulation," *Protein dimerization and oligomerization in biology*, pp. 105–121, 2012.

[6] G. D. Amoutzias, D. L. Robertson, Y. Van de Peer, and S. G. Oliver, "Choose your partners: dimerization in eukaryotic transcription factors," *Trends in biochemical sciences*, vol. 33, no. 5, pp. 220–229, 2008.

[7] K. Kaufmann, A. Pajoro, and G. C. Angenent, "Regulation of transcription in plants: mechanisms controlling developmental switches," *Nature Reviews Genetics*, vol. 11, no. 12, pp. 830–842, 2010.

[8] H. Lodish, A. Berk, C. A. Kaiser, C. Kaiser, M. Krieger, M. P. Scott, A. Bretscher, H. Ploegh, P. Matsudaira, *et al.*, *Molecular cell biology*. Macmillan, 2008.

[9] H.-W. Wang, H.-J. Sun, T.-Y. Chang, H.-H. Lo, W.-C. Cheng, G. C. Tseng, C.-T. Lin, S.-J. Chang, N. R. Pal, and I.-F. Chung, "Discovering monotonic stemness marker genes from time-series stem cell microarray data," in *BMC genomics*, vol. 16, p. S2, BioMed Central, 2015.

[10] S. Tian, "Identification of monotonically differentially expressed genes for non-small cell lung cancer," *BMC bioinformatics*, vol. 20, no. 1, p. 177, 2019.

[11] H.-W. Wang, H.-J. Sun, T.-Y. Chang, H.-H. Lo, W.-C. Cheng, G. C. Tseng, C.-T. Lin, S.-J. Chang, N. R. Pal, and I.-F. Chung, "Discovering monotonic stemness marker genes from time-series stem cell microarray data," in *BMC genomics*, vol. 16, p. S2, BioMed Central, 2015.

[12] Y. Hasin, M. Seldin, and A. Lusis, "Multi-omics approaches to disease," *Genome biology*, vol. 18, no. 1, pp. 1–15, 2017.

[13] Y. Hasin, M. Seldin, and A. Lusis, "Multi-omics approaches to disease," *Genome biology*, vol. 18, no. 1, pp. 1–15, 2017.

[14] D. C. Koboldt, K. M. Steinberg, D. E. Larson, R. K. Wilson, and E. R. Mardis, "The next-generation sequencing revolution and its impact on genomics," *Cell*, vol. 155, no. 1, pp. 27–38, 2013.

[15] E. T. Cirulli and D. B. Goldstein, "Uncovering the roles of rare variants in common disease through whole-genome sequencing," *Nature Reviews Genetics*, vol. 11, no. 6, pp. 415–425, 2010.

[16] S. B. Ng, E. H. Turner, P. D. Robertson, S. D. Flygare, A. W. Bigham, C. Lee, T. Shaffer, M. Wong, A. Bhattacharjee, E. E. Eichler, *et al.*, "Targeted capture and massively parallel sequencing of 12 human exomes," *Nature*, vol. 461, no. 7261, pp. 272–276, 2009.

[17] K. R. Kukurba and S. B. Montgomery, "Rna sequencing and analysis," *Cold Spring Harbor Protocols*, vol. 2015, no. 11, pp. pdb–top084970, 2015.

[18] Z. Wang, M. Gerstein, and M. Snyder, "Rna-seq: a revolutionary tool for transcriptomics," *Nature reviews genetics*, vol. 10, no. 1, pp. 57–63, 2009.

[19] B. Domon and R. Aebersold, "Mass spectrometry and protein analysis," *science*, vol. 312, no. 5771, pp. 212–217, 2006.

[20] M. P. Barrett, D. W. Boykin, R. Brun, and R. R. Tidwell, "Human african trypanosomiasis: pharmacological re-engagement with a neglected disease," *British journal of pharmacology*, vol. 152, no. 8, pp. 1155–1171, 2007.

[21] K. Kappmeier, E. Nevill, and R. Bagnall, "Review of tsetse flies and trypanosomosis in south africa," 1998.

[22] P. Van den Bossche, "Some general aspects of the distribution and epidemiology of bovine trypanosomosis in southern africa," *International journal for parasitology*, vol. 31, no. 5-6, pp. 592–598, 2001.

[23] G. d'Ieteren, E. Authié, N. Wissocq, and M. Murray, "Trypanotolerance, an option for sustainable livestock production in areas at risk from trypanosomosis.," *Revue scientifique et technique (International Office of Epizootics)*, vol. 17, no. 1, pp. 154–175, 1998.

[24] G. Roelants, "Natural resistance to african trypanosomiasis," *Parasite Immunology*, vol. 8, no. 1, pp. 1–10, 1986.

[25] É. Authie, "Contribution à l'étude des mécanismes immunologiques impliqués dans la trypanotolérance des taurins d'afrique," 1993.

[26] J. Ellis, J. Scott, N. D. MacHugh, G. Gettinby, and W. Davis, "Peripheral blood leucocytes subpopulation dynamics during trypanosoma congolense infection in boran and n'dama cattle: an analysis using monoclonal antibodies and flow cytometry," *Parasite immunology*, vol. 9, no. 3, pp. 363–378, 1987.

[27] M. Murray and A. Gray, "The current situation on animal trypanosomiasis in africa," *Preventive Veterinary Medicine*, vol. 2, no. 1-4, pp. 23–30, 1984.

[28] A. Buguet, J. Bert, P. Tapie, F. Tabaraud, F. Doua, J. Lonsdorfer, P. Bogui, and M. Dumas, "Sleep-wake cycle in human african trypanosomiasis.," *Journal of clinical neurophysiology: official publication of the American Electroencephalographic Society*, vol. 10, no. 2, pp. 190–196, 1993.

[29] F. Rijo-Ferreira, T. Carvalho, C. Afonso, M. Sanches-Vaz, R. M. Costa, L. M. Figueiredo, and J. S. Takahashi, "Sleeping sickness is a circadian disorder," *Nature communications*, vol. 9, no. 1, pp. 1–13, 2018.

[30] R. Brun, J. Blum, F. Chappuis, and C. Burri, "Human african trypanosomiasis," *The Lancet*, vol. 375, no. 9709, pp. 148–159, 2010.

[31] C. Burri, "Chemotherapy against human african trypanosomiasis: is there a road to success?," *Parasitology*, vol. 137, no. 14, pp. 1987–1994, 2010.

[32] L. J. Morrison, L. Vezza, T. Rowan, and J. C. Hope, "Animal african trypanosomiasis: time to increase focus on clinically relevant parasite and host species," *Trends in parasitology*, vol. 32, no. 8, pp. 599–607, 2016.

[33] E. O. Awuoche, B. L. Weiss, A. Vigneron, P. O. Mireji, E. Aksoy, B. Nyambega, G. M. Attardo, Y. Wu, M. O'Neill, G. Murilla, *et al.*, "Molecular characterization of tsetse's proboscis and its response to trypanosoma congolense infection," *PLoS neglected tropical diseases*, vol. 11, no. 11, p. e0006057, 2017.

[34] M. Wéry, *Protozoologie médicale*. De Boeck Université, 1995.

[35] C. A. Hoare, "The classification of mammalian trypanosomes," *Ergebnisse der Mikrobiologie Immunitätsforschung und Experimentellen Therapie*, pp. 43–57, 1966.

[36] J. Haag, C. O'hUigin, and P. Overath, "The molecular phylogeny of trypanosomes: evidence for an early divergence of the salivaria," *Molecular and biochemical parasitology*, vol. 91, no. 1, pp. 37–49, 1998.

[37] E. M. Kalundi, *Molecular analysis of the congopain gene family.* PhD thesis, 2008.

[38] M. Murray and A. Gray, "The current situation on animal trypanosomiasis in africa," *Preventive Veterinary Medicine*, vol. 2, no. 1-4, pp. 23–30, 1984.

[39] A. U. Kalu, "Current status of tsetse fly and animal trypanosomosis on the jos plateau, nigeria," *Preventive Veterinary Medicine*, vol. 27, no. 3-4, pp. 107–113, 1996.

[40] R. Oluwafemi, A. Ilemobade, and E. Laseinde, "The impact of african animal trypanosomosis and tsetse on the livelihood and wellbeing of cattle and their owners in the bicot study area of nigeria," *Sc Res Essays*, vol. 2, pp. 380–3, 2007.

[41] A. Shaw, "20 economics of african trypanosomiasis," *The trypanosomiases*, p. 369, 2004.

[42] M. J. Vreysen, "Prospects for area-wide integrated control of tsetse flies (diptera: Glossinidae) and trypanosomosis in sub-saharan africa," *Revista de la Sociedad Entomológica Argentina*, vol. 65, no. 1-2, pp. 1–21, 2006.

[43] B. Stijlemans, P. De Baetselier, S. Magez, J. A. Van Ginderachter, and C. De Trez, "African trypanosomiasis-associated anemia: the contribution of the interplay between parasites and the mononuclear phagocyte system," *Frontiers in immunology*, vol. 9, p. 218, 2018.

[44] A. Firesbhat and C. Desalegn, "Epidemiology and impacts of trypanosomiasis in cattle," *European Journal of Applied Sciences*, vol. 7, no. 5, pp. 220–225, 2015.

[45] O. M. Radostits, C. C. Gay, K. W. Hinchcliff, and P. D. Constable, *Veterinary Medicine E-Book: A textbook of the diseases of cattle, horses, sheep, pigs and goats*. Elsevier Health Sciences, 2006.

[46] G. J. Losos and B. Ikede, "Review of pathology of diseases in domestic and laboratory animals caused by trypanosoma congolense, t. vivax, t. brucei, t. rhodesiense and t. gambiense," *Veterinary Pathology*, vol. 9, no. 1_suppl, pp. 1–79, 1972.

[47] M. Yaro, K. Munyard, M. Stear, and D. Groth, "Combatting african animal trypanosomiasis (aat) in livestock: the potential role of trypanotolerance," *Veterinary parasitology*, vol. 225, pp. 43–52, 2016.

[48] J. Naessens, "Bovine trypanotolerance: A natural ability to prevent severe anaemia and haemophagocytic syndrome?," *International journal for parasitology*, vol. 36, no. 5, pp. 521–528, 2006.

[49] P. Starkey, "N'dama cattle-a productive trypanotolerant breed," *World Animal Review*, 1984.

[50] K. Agyemang, *Trypanotolerant livestock in the context of trypanosomiasis intervention strategies*, vol. 7. Food & Agriculture Org., 2005.

[51] M. Murray, G. D'Ieteren, E. Authie, and N. Wissocq, "Trypanotolerance, an option for sustainable livestock production in areas at risk from trypanosomosis," *OIE Revue Scientifique et Technique*, vol. 17, no. 1, pp. 154–175, 1998.

[52] J. Grootenhuis, R. Dwinger, R. Dolan, S. Moloo, and M. Murray, "Susceptibility of african buffalo and boran cattle to trypanosoma congolense transmitted by glossina morsitans centralis," *Veterinary parasitology*, vol. 35, no. 3, pp. 219–231, 1990.

[53] B. Namangala, P. De Baetselier, and A. Beschin, "Quantitative differences in immune responses in mouse strains that differ in their susceptibility to trypanosoma brucei brucei infection," *Journal of Veterinary Medical Science*, vol. 71, no. 7, pp. 951–956, 2009.

[54] S. Magez and M. Radwanska, "African trypanosomiasis and antibodies: implications for vaccination, therapy and diagnosis," *Future microbiology*, vol. 4, no. 8, pp. 1075–1087, 2009.

[55] S. Magez, M. Radwanska, M. Drennan, L. Fick, T. N. Baral, F. Brombacher, and P. D. Baetselier, "Interferon-γ and nitric oxide in combination with antibodies are key protective host immune factors during Trypanosoma congolense Tc13 infections," *The Journal of infectious diseases*, vol. 193, no. 11, pp. 1575–1583, 2006.

[56] S. Welburn, E. Fèvre, P. Coleman, I. Maudlin, *et al.*, "Epidemiology of human african trypanosomiasis.," *The trypanosomiases*, pp. 219–232, 2004.

[57] A. Abbadi and G. Leckband, "Rapeseed breeding for oil content, quality, and sustainability," *European Journal of Lipid Science and Technology*, vol. 113, no. 10, pp. 1198–1206, 2011.

[58] I. A. Parkin, A. Sharpe, and D. Lydiate, "Patterns of genome duplication within the brassica napus genome," *Genome*, vol. 46, no. 2, pp. 291–303, 2003.

[59] I. Parkin, A. Sharpe, D. Keith, and D. Lydiate, "Identification of the a and c genomes of amphidiploid brassica napus (oilseed rape)," *Genome*, vol. 38, no. 6, pp. 1122–1131, 1995.

[60] R. Snowdon, W. Lühs, and W. Friedt, "Oilseed rape," in *Oilseeds*, pp. 55–114, Springer, 2007.

[61] C. Jiang, J. Shi, R. Li, Y. Long, H. Wang, D. Li, J. Zhao, and J. Meng, "Quantitative trait loci that control the oil content variation of rapeseed (brassica napus l.)," *Theoretical and applied genetics*, vol. 127, no. 4, pp. 957–968, 2014.

[62] D. Fowler and R. Downey, "Lipid and morphological changes in developing rapeseed, brassica napus," *Canadian Journal of Plant Science*, vol. 50, no. 3, pp. 233–247, 1970.

[63] G. Stringam, D. McGregor, and P. SH, "Chemical and morphological characteristics associated with seedcoat color in rapeseed," in *Fette Seifen Anstrichmittel*, vol. 76, pp. 302–303, 1974.

[64] J. K. Kramer, *High and low erucic acid in Rapeseed Oils*. Academic press, 2012.

[65] D. Kimber and D. McGregor, *Brassica oilseeds: production and utilization*. Cab international, 1995.

[66] W. Vetter, V. Darwisch, and K. Lehnert, "Erucic acid in brassicaceae and salmon–an evaluation of the new proposed limits of erucic acid in food," *NFS journal*, vol. 19, pp. 9–15, 2020.

[67] J. Daun, "Composition and use of canola seed, oil, and meal," *Cereal Foods World*, vol. 29, no. 5, pp. 291–+, 1984.

[68] W. Ecke, M. Uzunova, and K. Weissleder, "Mapping the genome of rapeseed (brassica napus l.). ii. localization of genes controlling erucic acid synthesis and seed oil content," *Theoretical and Applied Genetics*, vol. 91, no. 6, pp. 972–977, 1995.

[69] S. Das, T. Roscoe, M. Delseny, P. Srivastava, and M. Lakshmikumaran, "Cloning and molecular characterization of the fatty acid elongase 1 (fae 1) gene from high and low erucic acid lines of brassica campestris and brassica oleracea," *Plant Science*, vol. 162, no. 2, pp. 245–250, 2002.

[70] M. de Lorgeril, P. Salen, F. Laporte, T. Foulon, N. Payen, and J. de Leiris, "Rapeseed oil and rapeseed oil-based margarine for the prevention and treatment of coronary heart disease," *European journal of lipid science and technology*, vol. 103, no. 7, pp. 490–495, 2001.

[71] E. Fahy, S. Subramaniam, R. C. Murphy, M. Nishijima, C. R. Raetz, T. Shimizu, F. Spener, G. van Meer, M. J. Wakelam, and E. A. Dennis, "Update of the lipid maps comprehensive classification system for lipids1," *Journal of lipid research*, vol. 50, pp. S9–S14, 2009.

[72] S. Subramaniam, E. Fahy, S. Gupta, M. Sud, R. W. Byrnes, D. Cotter, A. R. Dinasarapu, and M. R. Maurya, "Bioinformatics and systems biology of the lipidome," *Chemical reviews*, vol. 111, no. 10, pp. 6452–6490, 2011.

[73] J. L. Harwood, "Recent advances in the biosynthesis of plant fatty acids," *Biochimica et Biophysica Acta (BBA)-Lipids and Lipid Metabolism*, vol. 1301, no. 1-2, pp. 7–56, 1996.

[74] Y. Li-Beisson, B. Shorrosh, F. Beisson, M. X. Andersson, V. Arondel, P. D. Bates, S. Baud, D. Bird, A. DeBono, T. P. Durrett, *et al.*, "Acyl-lipid metabolism," *The Arabidopsis book/American Society of Plant Biologists*, vol. 11, 2013.

[75] G. Chen, J. Geng, M. Rahman, X. Liu, J. Tu, T. Fu, G. Li, P. B. McVetty, and M. Tahir, "Identification of qtl for oil content, seed yield,

and flowering time in oilseed rape (brassica napus)," *Euphytica*, vol. 175, no. 2, pp. 161–174, 2010.

[76] U. K. Nath, *Increasing erucic acid content in the seed oil of rapeseed (Brassica napus L.) by combining selection for natural variation and transgenic approaches*. PhD thesis, Niedersächsische Staats-und Universitätsbibliothek Göttingen, 2008.

[77] P. O'Hara, A. R. Slabas, and T. Fawcett, "Fatty acid synthesis in developing leaves of brassica napus in relation to leaf growth and changes in activity of 3-oxoacyl-acp reductase," *FEBS letters*, vol. 488, no. 1-2, pp. 18–22, 2001.

[78] T. J. Roscoe, "Identification of acyltransferases controlling triacylglycerol biosynthesis in oilseeds using a genomics-based approach," *European journal of lipid science and technology*, vol. 107, no. 4, pp. 256–262, 2005.

[79] H. Tan, X. Yang, F. Zhang, X. Zheng, C. Qu, J. Mu, F. Fu, J. Li, R. Guan, H. Zhang, *et al.*, "Enhanced seed oil production in canola by conditional expression of brassica napus leafy cotyledon1 and lec1-like in developing seeds," *Plant physiology*, vol. 156, no. 3, pp. 1577–1588, 2011.

[80] W. I. Mentzen, J. Peng, N. Ransom, B. J. Nikolau, and E. S. Wurtele, "Articulation of three core metabolic processes in arabidopsis: fatty acid biosynthesis, leucine catabolism and starch metabolism," *BMC Plant Biology*, vol. 8, no. 1, pp. 1–15, 2008.

[81] T. J. Buckhout and O. Thimm, "Insights into metabolism obtained from microarray analysis," *Current opinion in plant biology*, vol. 6, no. 3, pp. 288–296, 2003.

[82] A. Rajavel, F. Heinrich, A. O. Schmitt, and M. Gültas, "Identifying cattle breed-specific partner choice of transcription factors during the african trypanosomiasis disease progression using bioinformatics analysis," *Vaccines*, vol. 8, no. 2, p. 246, 2020.

[83] A. Rajavel, S. Klees, Y. Hui, A. O. Schmitt, and M. Gültas, "Deciphering the molecular mechanism underlying african animal trypanosomiasis by means of the 1000 bull genomes project genomic dataset," *Biology*, vol. 11, no. 5, 2022.

[84] A. Rajavel, S. Klees, J.-S. Schlüter, H. Bertram, K. Lu, A. O. Schmitt, and M. Gültas, "Unravelling the complex interplay of transcription factors orchestrating seed oil content in brassica napus l.," *International journal of molecular sciences*, vol. 22, no. 3, p. 1033, 2021.

[85] A. Rajavel, A. O. Schmitt, and M. Gültas, "Computational identification of master regulators influencing trypanotolerance in cattle," *International Journal of Molecular Sciences*, vol. 22, no. 2, p. 562, 2021.

Chapter 2

Identifying Cattle Breed-Specific Partner Choice of Transcription Factors during the African Trypanosomiasis Disease Progression Using Bioinformatics Analysis

Abirami Rajavel[1], Felix Heinrich[1], Armin Otto Schmitt[1,2] and Mehmet Gültas[1,2,*]

[1] Breeding Informatics Group, Department of Animal Sciences, Georg-August University, Margarethe von Wrangell-Weg 7, 37075 Göttingen, Germany
[2] Center for Integrated Breeding Research (CiBreed), Albrecht-Thaer-Weg 3, Georg-August University, 37075 Göttingen, Germany
* Author to whom correspondence should be addressed.

Bibliographic Information

Rajavel, A.; Heinrich, F.; Schmitt, A.O.; Gültas, M. Identifying Cattle Breed-Specific Partner Choice of Transcription Factors during the African Trypanosomiasis Disease Progression Using Bioinformatics Analysis. Vaccines 2020, 8, 246. https://doi.org/10.3390/vaccines8020246.

Author contribution of Abirami Rajavel

The author contributed to the design of the study, and conducted computational analyses. Furthermore, the author interpreted the results and wrote a significant part of the final version of the manuscript.

2.1 Abstract

African Animal Trypanosomiasis (AAT) is a disease caused by pathogenic trypanosomes which affects millions of livestock every year causing huge economic losses in agricultural production especially in sub-Saharan Africa. The disease is spread by the tsetse fly which carries the parasite in its saliva. During the disease progression, the cattle are prominently subjected to anaemia, weight loss, intermittent fever, chills, neuronal degeneration, congestive heart failure, and finally death. According to their different genetic programs governing the level of tolerance to AAT, cattle breeds are classified as either resistant or susceptible. In this study, we focus on the cattle breeds N'Dama and Boran which are known to be resistant and susceptible to trypanosomiasis, respectively. Despite the rich literature on both breeds, the gene regulatory mechanisms of the underlying biological processes for their resistance and susceptibility have not been extensively studied. To address the limited knowledge about the tissue-specific transcription factor (TF) cooperations associated with trypanosomiasis, we investigated gene expression data from these cattle breeds computationally. Consequently, we identified significant cooperative TF pairs (especially $DBP - PPARA$ and $DBP - THAP1$ in N'Dama and $DBP - PAX8$ in Boran liver tissue) which could help understand the underlying AAT tolerance/susceptibility mechanism in both cattle breeds.

2.2 Introduction

Climate change is likely to increase the risk of several vector-borne diseases including human and animal Trypanosomiasis [1]. African Animal Trypanosomiasis (AAT), also known as nagana disease, is a chronic parasitic infection which affects livestock in large numbers, prevalently found in the cattle of sub-Saharan Africa [2, 3]. Trypanosomes are unicellular protozoans, which are transmitted through the saliva of the vector tsetse fly. They survive in the bloodstream of the animal by escaping and manipulating the host's immune response, thereby causing serious health problems in cattle [4]. Due to this debilitating infection, the animal becomes diseased which results in lower economic productivity such as reduced meat and milk production and reduced draught power for agricultural production, thus imposing huge financial losses to farmers in sub-Saharan Africa [5]. Especially the main causative species *Trypanosoma congolense* and *Trypanosoma vivax* severely impair the health of the cattle population [6, 7].

The animals display numerous clinical signs in the early and later stages of

the disease with anaemia being the most prominent pathological feature of AAT [8]. Other clinical signs include fever, intermittent chills, weight loss, lethargy, emaciation, neurological disorders, infertility, abortion, difficulty in breathing, loss of appetite, and congestive heart failure leading to death if left untreated [2, 3, 5, 9].

Some cattle breeds are capable of resisting the disease [after trypanosome infection] despite the parasite's infection in the body. Those cattle breeds are known to be resistant to trypanosomiasis and this trait is termed as trypanotolerance. Trypanotolerance is a distinctive trait of few West African taurine breeds, which has been gained through natural selection by continuous interaction of host and parasite [10, 11]. One such trypanotolerant *Bos taurus* cattle breed is N'Dama [12]. Even though N'Dama cattle are trypanotolerant, they are not particularly advantageous for agricultural purposes because of their low productivity and their small size. Whereas the other cattle breed Boran is beneficial for their productivity, heat tolerance and performance, it is highly susceptible to trypanosomiasis [13]. Therefore, understanding the gene regulatory mechanisms underlying the biological processes for their susceptibility/resistance is useful for the selective breeding of this trait.

Trypanosomes are able to escape the host's natural and adaptive immunity due to their antigenic variation [14, 15, 16, 17]. This speculative adaptation of the parasite is of great interest in recent times because of the existing and increasing risk of drug resistance to the currently available trypanocides [18, 19]. Henceforth, there is an urgent need for the understanding of the molecular mechanism underlying this infectious tropical disease.

For this purpose, Noyes *et al.* [20] performed gene expression analysis of a transcription profiling time series microarray dataset for liver, spleen and lymph node tissues in the cattle breeds Boran and N'Dama. These tissues are vital lymphoid organs which mount host immune responses to pathogenic invasion by generating high numbers of macrophages involved in phagocytosis and further production of pro-inflammatory cytokines [21, 22, 23]. By analysing this dataset, Noyes *et al.* identified the differentially expressed genes (DEGs) based on which the candidate genes within previously reported QTLs in the regulation of immune responses were obtained. However, according to its clinical signs, AAT could be considered as a multistage progression process. While mainly focusing on the DEGs, Noyes *et al.* studied the differences in expression levels at certain timepoints, but their analysis strategy could not capture the importance of the genes with regular monotonic expression patterns as the disease advances. Recently, Wang *et al.* [24] and Suyan Tian [25] have pointed out that the genes with monotonic expression patterns are quintessential for gaining complete insight into the multistage

progression of the disease.
Despite the rich literature on trypanotolerance and molecular studies addressing the identification of candidate markers in AAT, its underlying molecular mechanism in cattle has not been well studied. As of now, there has been no study performed in AAT especially in cattle, with the aspect of examining the role of the regulatory elements involved in gene regulation, for example, TFs and their cooperations. Today it is well known that TFs and their complex interplay have critical roles in the progression of disease [26, 27]. In order to address the importance of cooperative TFs in the AAT disease, we analysed in this study the dataset published by Noyes *et al.* [20]. Unlike this study, we focus on the identification as well as the analysis of monotonically expressed genes (MEGs) to completely capture the multistage progression process of the AAT disease. Further, we applied our previously published computational PC-TraFF approach [26] to the promoter regions of the MEGs in order to identify specific cooperative TFs in different tissues of Boran and N'Dama cattle breeds. Our results suggest that the preferential partner choice of TFs could be related to the gene regulatory mechanisms determining the level of AAT-tolerance of the cattle breeds. Particularly with regard to AAT-disease, the partner choice of the transcription factor *DBP* is likely to orchestrate the genetic programs governing the molecular mechanism of the level of trypanotolerance of both cattle breeds. Especially, focusing on DBP's function in association with the circadian rhythm, we attempted to highlight the significant role of the circadian transcriptional program in regulation of immune responses to the pathogen infection at the tissue level (see the review [28] for details regarding the circadian regulation of immunity).

Conserved Functions of Transcription Factors across Mammals

Transcription factors (TFs) are proteins which bind to short DNA sequences known as Transcription Factor Binding Sites (TFBSs), involved in regulating the transcription of genes [26, 29, 30]. These two interacting regulatory elements (TFs and TFBSs) are two of the significant functional elements which are involved in the regulation of various cellular processes [29, 30]. According to the widely accepted phylogenetics footprinting approach, the basic assumption is that the functional elements are likely to be more conserved than the non-functional elements in response to selective pressure [31, 32]. Therefore, the functions of transcription factors and the binding sites are expected to be well conserved across multi-species, particularly across evolutionarily closely related mammalian species. Several studies also confirm

the evolutionary conservation of binding specificities of TFs in a wide range of species [33, 34, 35, 36]. In our study, the predictions are performed for the tissues of two different cattle breeds. Regarding the conserved functions of TFs across mammals, we interpreted the results for cattle breeds based on experimental studies which have been designed and performed for other mammalian species (such as model organisms including human and mouse).

Transcription Factors, Potential Targets for Vaccine Development

Transcription factors play an important role in mounting immune responses especially during pathogenic invasion [37, 38]. Both innate and adaptive immunity of the immune system are controlled at the transcriptional level, which thereby provides valuable drug targets for regulating the gene expression of several immune cells [39]. Almost 10% of anti-cancer drugs approved by Food and Drug Administration (FDA) targets the transcription process [40]. Recently, transcription therapy has been proposed as a state-of-the-art approach targeting transcription factors for therapeutic interventions [41]. Therefore, transcription factors, being converging point for many signalling pathways, can serve as promising drug candidates for vaccine development [42]. In our study, we focused on transcription factors and their complex interplay which orchestrate the genetic programs underlying the level of trypanotolerance of both cattle breeds. Our findings could provide novel drug targets for the development of effective vaccine-mediated control of the AAT disease.

2.3 Materials and Methods

In this section we describe the microarray gene expression dataset that we analyzed and the methods applied in this study. Our analysis follows the structure as represented in Figure 2.1.

2.3.1 Microarray Dataset

A microarray experiment was undertaken by Noyes *et al.* (http://www.ebi.ac.uk/arrayexpress/, accession no. E-MEXP-1778) [20] to survey the genome of two cattle breeds for differentially expressed genes and their related pathways in different tissues. The dataset contains the gene expression obtained from a trypanosomiasis-susceptible (Boran) breed and a trypanosomiasis-resistant (N'Dama) breed at different time points (days: 0, 12, 15, 18, 21,

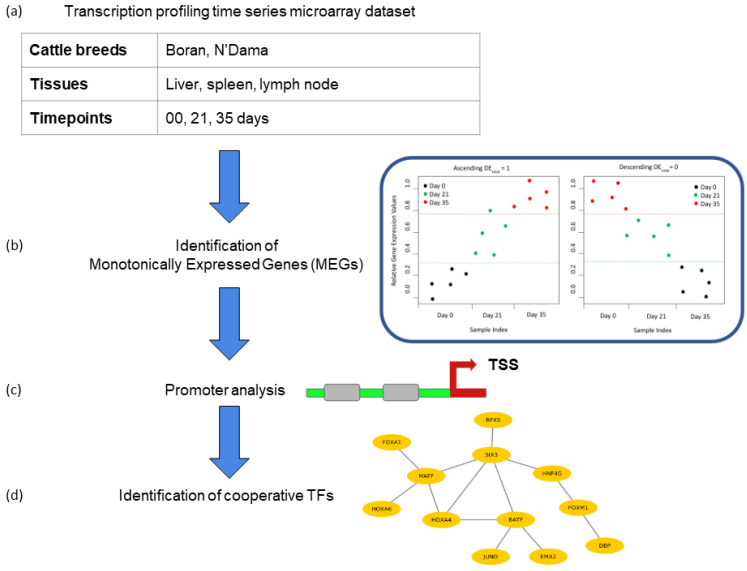

Figure 2.1: Flowchart of analysis. (**a**) Processing of the microarray dataset; (**b**) Identification of Monotonically Expressed Genes using the MFSelector approach; (**c**) Promoter analysis (TSS: transcription start site); (**d**) identification of cooperative TFs using PC-TraFF approach.

26, 29, 32 and 35) for three tissues, namely liver, spleen and lymph nodes. In their study, Noyes *et al.* [20] included 25 trypanosomiasis-free (healthy) animals of each breed which were chosen from herds at a tsetse fly free- and trypanosomiais-free zone of the ILRI Kapiti Plains ranch. Furthermore, before being transferred to the ILRI research facility at Kabete, they were tested for tick-borne parasites and confirmed negative. Afterwards, all animals were infected with *T. congolense*, which is one of the infectious species causing trypanosomiasis. N'Dama was selected for its trypanotolerance and Boran for its trypanosusceptibility. At different time points after infection, the tissues were collected from the animals. For control experiments, the tissue collection was undertaken similarly for the two breeds before infection and noted as day 0. Tissues from liver, spleen and lymph nodes were harvested on days 0, 21 and 35 after infection. Furthermore, liver tissues at additional time points on days: 12, 15, 18, 21, 26, 29 and 32, were collected by biopsy. RNA from each tissue was extracted and hybridised on arrays. Consequently, the dataset altogether consisted of 160 samples from three cattle tissues, containing expression values for 13,934 cattle genes. The expression strength is given in transcript per million (TPM) values.

2.3.2 Identification of Monotonically Expressed Genes

We applied the monotonic feature selector (MFSelector) approach [24] in order to identify the genes with a strong monotonic pattern (either in ascending or descending order) in their expression profiles over time points during disease progression.
The MFSelector approach requires gene expression datasets measured over several time points as input and assesses the confidence of the monotonicity of each gene by calculating its total discriminating error ($\mathrm{DE}_{\mathrm{total}}$).
Let g_{ijs} be an expression value of gene g_i measured at time point $j = 1, \cdots, T$ from sample $s = 1, \cdots, S$. The discriminating error (DE) of g_i is calculated by comparing its expression values g_{ijs} observed for all samples S at time point j against the remaining time points as:

$$\mathrm{DE}(g_{ijs}) = \begin{cases} 1 & \text{if} \quad g_{ijs} > \tau_j \\ 0 & \text{otherwise} \end{cases} \tag{2.1}$$

In Equation (2.1), τ_j refers to the threshold for discriminating line which distinguishes in ascending order the g_{ijs} values of all samples at time point j from other time points and simultaneously ensures the minimum total discriminating error for j. Consequently, the $\mathrm{DE}_{\mathrm{total}}$ score for g_i is calculated as:

$$\mathrm{DE}_{\mathrm{total}}(g_i) = \sum_{j=1}^{T} \sum_{s=1}^{S} \mathrm{DE}(g_{ijs}) \tag{2.2}$$

Similar as in Equation (2.1), the discriminating error of genes is calculated to assess their monotonicity in descending order as:

$$\mathrm{DE}(g_{ijs}) = \begin{cases} 1 & \text{if} \quad g_{ijs} < \eta_j \\ 0 & \text{otherwise} \ , \end{cases} \tag{2.3}$$

where η_j is the threshold for the discriminating line at j and lastly the $\mathrm{DE}_{\mathrm{total}}$ score is calculated again using Equation (2.2).
In the next step, MFSelector performs a permutation test for the assessment of the statistical significance of the $\mathrm{DE}_{\mathrm{total}}$ scores by calculating their

unadjusted p-values and q-values which were adjusted p-values for multiple testing. Finally, based on the Sample Variance for Discriminating Error (*svde*), the level of confidence of a monotonic gene is determined. A small *svde* value indicates clear monotonicity of the corresponding gene.
For the application of the MFSelector package, we have to define parameters such as *permut*, *svdetimes* and *svdenoise* in order to set the level of stringency for the monotonicity. *permut* controls the statistical significance of the DE_{total} index, *svdetimes* represents the number of times the SVDE procedure along with the random noise will be repeated and *svdenoise* indicates the strength of the noise in the experiment. The greater the values of the aforementioned parameters, the higher the stringency of the selection. For our analysis, to obtain the MEGs which are expressed with strong monotonic pattern, we applied the MFSelector approach by setting its parameters: *permut* 100, *svdetimes* 100 and *svdenoise* 0.1.

2.3.3 Identification of Transcription Factor Cooperation

In order to identify cooperative transcription factors, we applied our previously developed PC-TraFF method [26] which is a well established information theory based approach. The PC-TraFF algorithm uses the concept of pointwise mutual information (PMI) for the identification of TF cooperation by mainly considering the co-occurrence of their transcription factor binding sites (TFBSs) in the promoters of genes [26].
As input parameters the algorithm needs a set of regulatory sequences, a library of position weight matrices (PWMs) and pre-defined TFBS distance constraints.

- **Promoter sequences:** The promoter sequence (covering the -500 to 100 bp regions relative to a transcription start site) of each significant monotonically expressed gene (MEG) is extracted from the UCSC genome browser [43].

- **Creation of the PWM library and TFBS detection:** For the detection of TFBSs in the promoters of MEGs, we obtained PWMs from the TRANSFAC database (release 2018.1) [44].

  Until now, based on the functional analysis and comprehensive performance evaluation strategies, different studies have shown that the computational TFBS detection methods using PWMs are well established and highly applied. However, their prediction performance is

prone to high rates of false positive predictions [30, 45, 46]. In order to eliminate the false predictions to some extent in our analysis, we manually created a specific PWM library following our previous study [47]. For this purpose, we first obtained all available cattle TFs from AnimalTFDB 2.0 [48] and identified the expression values (TPM values) of their corresponding TF genes in the gene expression dataset, under study. Second, the TFs were excluded from further analysis if the TPM values of their TF genes were zero. After that, the corresponding PWMs of the remaining TFs were obtained from the TRANSFAC database [44]. Finally, based on the Pearson correlation between these PWMs, we applied hierarchical clustering and used only the PWMs with the highest information content from each cluster as representative to create our final non-redundant vertebrate PWM library (see Supplementary Table S1).

In addition, we applied the Match^{TM} program [51] using these PWMs and their TRANSFAC specific profile parameter *minSum* to minimize the sum of false positive and false negative rate for the detection of putative TFBSs in promoter sequences.

- **Pre-defined distances:** For the identification of cooperative TFs based on the co-occurrence of their TFBSs, the PC-TraFF algorithm requires pre-defined minimum and maximum distance thresholds. In this study, the recommended distance values of ≥ 5 and ≤ 20 were used for the minimum and maximum distance, respectively.

The PC-TraFF algorithm provides a PMI(T_a; T_b)-score for each cooperative TF-pair (T_a and T_b), which is transformed in the next step into the *z-score*. A cooperation between any T_a and T_b is considered to be statistically significant if they have a *z-score* ≥ 3.

2.4 Results and Discussion

2.4.1 Data Processing

We analysed a time series dataset which consisted of gene expression values for three tissues (liver, spleen and lymph node) from two cattle breeds, trypanosusceptible and trypanotolerant, during the disease progression after *T. congolense* infection. Although the dataset for liver tissue consisted of gene expression values from several time points (days 0, 12, 15, 18, 21, 26, 29, 32 and 35), we considered for this tissue the data only for 3 time points (days:

0, 21 and 35) similar to spleen and lymph node tissues, to ensure the purpose of maintaining uniformity throughout the analyses.

2.4.2 Identification of MEGs

To begin with the analysis, we organized the dataset for each breed separately and arranged the dataset unique for the tissue type of each breed in ascending order of time points. Afterwards, we applied the MFSelector package to the gene sets of each tissue, using its parameters as mentioned in the Materials and Methods section. Subsequently, we obtained for each tissue in both cattle breeds two lists of MEGs in ascending and descending order based on their monotonicity of expression during the disease progression. Finally, we defined for our further analysis a gene to be a statistically significant MEG if its corresponding q-value ≤ 0.05 and its *svde* value ≤ 1. The numbers of statistically significant MEGs are given in Table 2.1 and all lists of MEGs are provided in Supplementary Table S2.

Table 2.1: Numbers of statistically significant Monotonically Expressed Genes in ascending and descending order for liver-, spleen- and lymph node tissues for the cattle breeds Boran and N'Dama.

	Boran		**N'Dama**	
	Ascending	Descending	Ascending	Descending
Liver	741	308	757	124
Spleen	669	126	13	139
Lymph node	87	5	119	114

The analysis of significant ascending and descending MEGs for spleen and lymph node tissues revealed that the vast majority of them are unique in both breeds and only a small minority of these MEGs is common to both breeds. On the other hand, there are 429 overlapping genes in the significant ascending MEGs and 39 overlapping genes in the descending MEGs in the liver tissue of Boran and N'Dama, respectively (see Figure 2.2). In our further analysis, we omitted the MEGs which are found to be significant and overlapping for a certain tissue in both breeds.

2.4.3 Identification of Cooperative TFs

Applying the PC-TraFF approach [26] to the promoters of MEGs for each tissue in both cattle breeds individually, we have identified significant cooperative TF pairs for liver, spleen and lymph node tissues which are listed

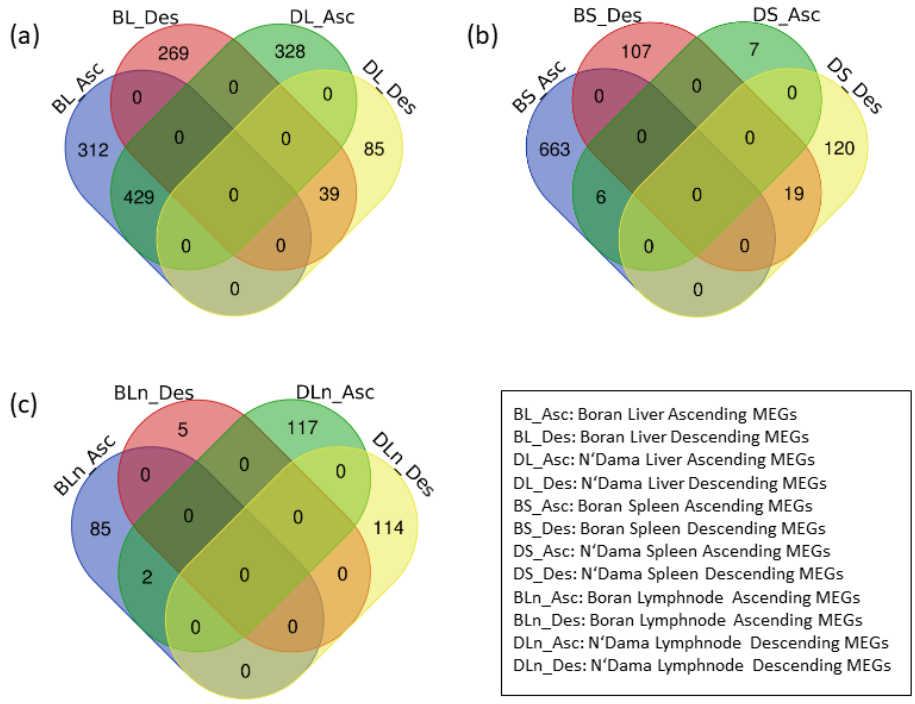

Figure 2.2: Venn-diagram of the MEGs in the ascending and descending orders of (a) liver-, (b) spleen- and (c) lymph node tissues for the progressive three time points (day 0, day 21, and day 35) after trypanosomiasis infection of the two cattle breeds.

in Table 2.2. The TFs *E2F1*, *PPARA*, *THAP1*, *HAND1E47* and *TFAP2A* are frequently observed in all three tissues of both cattle breeds. The factor *E2F1* belongs to the E2F family of transcription factors [50] and is involved in promoting the process of adipogenesis and in regulating lipolysis [51, 52]. Denechaud *et al.* pointed out the association of *E2F1* with several processes in pancreas, liver, heart, muscle, and adipocytes including lipogenesis, cholesterol transport, bile acid synthesis, glucose oxidation, and oxidative metabolism [52]. Furthermore, it was also reported that *E2F1* mRNA levels in the adipose tissue correlated with circulating free fatty acid levels in obese human subjects [52, 53]. Interestingly, E2F has been reported as one of the promising candidates in the circadian transcriptional regulators [54].
Kersten *et al.* explained in their review [55] *PPARA* is activated by ligands and is abundantly found in liver. In mouse, it is reported as the master regulator of lipid metabolism in liver during fasting. In human liver, it is reported

Table 2.2: Cooperative TF pairs specific for liver-, spleen-, and lymph node tissues of Boran and N'Dama obtained from PC-TraFF approach.

	Cooperative Transcription Factor pairs		
Breed	**Liver**	**Spleen**	**Lymph node**
Boran	*PPARG – RFX5* *HAND1E47 – THAP1* *HAND1E47 – TFAP2A* *THAP1 – E2F1* *HOXA6 – BATF* *TTF1 – RFX5* *DBP – PAX8* *E2F1 – PPARA* *E2F1 – E2F1*	*EMX2 – BATF* *FOXM1 – JUND* *HOXA4 – HOXB7* *PPARA – E2F1* *E2F1 – E2F1* *E2F1 – TFAP2A*	*JUND – BATF* *HOXA6 – MAFF* *SIX3 – MAFF* *HAND1E47 – E2F1* *HOXA4 – BATF* *FOXM1 – HNF4G* *DBP – FOXM1* *SIX3 – RFX5* *HAND1E47 – FOSL1* *FOXA1 – MAFF* *HOXA4 – MAFF* *SIX3 – BATF* *SIX3 – HNF4G* *HOXA4 – SIX3* *THAP1 – PPARA* *E2F1 – PPARA* *EMX2 – BATF* *THAP1 – THAP1*
N'Dama	*PPARG – SIX5* *FOXM1 – DLX3* *HAND1E47 – E2F1* *HAND1E47 – USF2* *SIX5 – PPARA* *SIX3 – THAP1* *THAP1 – THAP1* *PPARA – DBP* *PPARA – TFAP2A* *E2F1 – E2F1*	*HMBOX – BATF* *HOXB7 – BATF* *SIX5 – THAP1* *HOXA4 – HMBOX1* *HAND1E47 – DBP* *HOXA6 – BATF* *SIX5 – E2F1* *HMBOX1 – RFX5* *E2F1 – TFAP2A*	*SMAD4 – E2F1* *HOXA6 – BATF* *HOXA4 – HOXB7* *TCF4 – HNF4G* *E2F1 – TFAP2A* *THAP1 – HNF4G* *DBP – TFAP2A* *HAND1E47 – RFX5* *THAP1 – PPARA* *E2F1 – PPARA* *EMX2 – BATF* *HOXA4 – SIX3* *THAP1 – THAP1*

that *PPARA* induces several genes involved in numerous metabolic pathways including bile acid synthesis, lipoprotein metabolism, synthesis and breakdown of triglycerides and lipid droplets. Moreover, they play suppressive roles in inflammation and acute phase response.
The factor THAP domain containing apoptosis associated protein 1 abbreviated to *THAP1* belongs to the THAP protein family [50, 56, 57]. They are found in the interaction and co-localization with the promyelocytic leukemia nuclear bodies [56]. Another important factor is *TFAP2A* which is a member of the *TFAP2* (AP-2) family of basic helix-span-helix transcription factors. A crucial role of these transcription factors has been shown in [58] as the master regulator of lipid droplet biogenesis, in which lipid droplets are known for various other functions including inflammatory responses, host-pathogen interaction, and other metabolic processes [58, 59, 60]. Importantly, it was found to be significantly over-represented in the promoter regions of clock controlled genes [54].

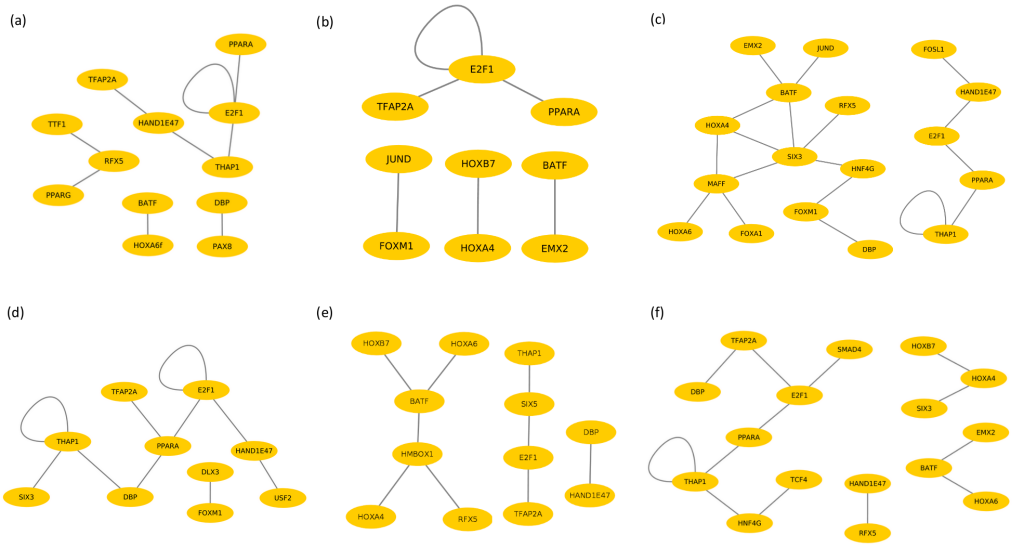

Figure 2.3: Cooperation networks for the TF pairs of (a) liver tissue of Boran, (b) spleen tissue of Boran, (c) lymph node tissue of Boran, (d) liver tissue of N'Dama, (e) spleen tissue of N'Dama and (f) lymph node tissue of N'Dama.

On the other hand, the factor *HAND1E47* from the basic helix-loop-helix (bHLH) transcription factor family is reported in the development of heart [61, 62, 63, 64] and to be associated with cardiac defects [65]. Its association with the host pathogen interplay is currently biologically unconfirmed. In order to gain a better understanding of the underlying molecular mechanism of AAT in different tissues and compare the results of both breeds, we created cooperation networks for each tissue based on its specific TF pairs as suggested in our previous studies [26, 47, 66, 67]. The nodes represent the TFs and the edges represent their co-operation in these networks which are presented in Figure 2.3. The cooperation networks of liver-, spleen- and lymph node tissues consist of 9, 6 and 18 cooperative TF pairs in Boran and 10, 9 and 13 pairs in N'Dama, respectively.

2.4.4 Cooperative TFs in Liver Tissue

The analysis of the cooperation networks for liver tissue (Figure 2.3a,d) reveals that, although several (single) TFs are overlapping in both networks, they change their partners in both breeds. Among others, cooperation of the factor *DBP* in these two cattle breeds highlights the difference in its remarkable preferential partner choice, namely *DBP–PAX8* in Boran, whereas

DBP–THAP1 and *DBP–PPARA* in N'Dama.
The albumin Site D-binding Protein (*DBP*) is regarded as a clock target gene which regulates primarily the sleep-wake cycle in mammalian species [54, 68, 69, 70, 71]. Further, *DBP* is associated with the circadian rhythm [68], which is normally disrupted in humans during the disease progression [72]. In Human African Trypanosomiasis (HAT), an improper disorganised sleeping pattern is common wherein the infected persons sleep more during day time and stay awake during night time [72]. The relationship between clock genes and the circadian rhythm are well established in several studies [73, 74, 75]. Further, *DBP* is a liver specific transcriptional activator, expressed in a circadian manner [76]. Circadian rhythm in peripheral tissues like liver is crucial for the normal hepatic metabolism [77, 78], especially of lipids [79]. Several binding sites for *DBP* are reported in the promoter regions of the gene *CYP7* which is involved in the rate limiting step of the pathway converting cholesterol to bile acids in mammalian species [80].
In the liver tissue of the cattle breed Boran, *DBP* forms dimers with the factor *PAX8* which is one of the thyroid-specific transcription factors essential for the development of the thyroid gland [81, 82]. Mutations in *PAX8* have been reported to cause hypothyroidism [83]. *PAX8* is a highly sensitive marker for thyroid and renal tumors [84]. The levels of the Thyroid Stimulating Hormone (TSH) are influenced by the circadian rhythm [85]. In relation to AAT, one of the endocrine organs affected by trypanosome infection is thyroid. Interestingly, decrease in the levels of T4 has been observed in goats after *T. congolense* infection, consequently indicating the impairment of the thyroid function [86]. Studies on Boran cattle infected with *Trypanosoma congolense* have confirmed the impairment of the pituitary gland [87].
Using neonatal *Pax8* $-/-$ mice, about a 10-fold increase of accumulation of hepatic triglycerides has been observed [88]. After the administration of thyroid hormone (TH), hepatic triglycerides were mobilised and processed [88]. Thyroid hormones and the functioning of the liver, particularly in the lipid metabolism, are interconnected with each other. Additionally, it is proven that thyroid hormone regulates a variety of metabolic processes by interacting with several important signalling pathways, thereby influencing energy metabolism and energy homeostasis [89]. They critically regulate the cholesterol metabolism in rat [90]. Specifically in liver, T3 and T4 hormones regulate the lipoprotein metabolism [88, 91, 92]. These findings support the hypothesis that the cooperation between the TFs *DBP* and *PAX8* could be strongly associated with the circadian rhythm, thyroid hormones and the lipid metabolism of the AAT susceptible breed Boran.
On the other hand, the TF cooperations *DBP–PPARA* and *DBP–THAP1* in the liver tissue of N'Dama might be noticeably changing the host and

parasite interaction in a direction that is opposite to that of Boran. The factor *PPARA* from the Peroxisome proliferator-activated receptor family is reported to be an important regulator of lipid metabolism, predominantly expressed in liver. It belongs to the nuclear receptor hormone superfamily which are ligand-induced [93]. PPARs are increasingly studied in inflammation as they are involved directly in the negative regulation of inflammation. Remarkably, *PPARA* plays a significant anti-inflammatory role in the regulation of the immune system [93]. The second partner of *DBP* is the zinc fnger transcription factor *THAP1* [56, 94, 95]. Mutations in *THAP1* result in neuronal dysfunction leading to dystonia, a brain disorder which is characterized by involuntary muscle contractions and abnormal postures [96].
After *T. congolense* infection, cerebral lesions and enlargement of several organs in particular liver, spleen, lungs, heart, and lymph nodes, are observed during the pathogenesis of the disease [97, 98]. This could be an indication that the infected cattle are attempting to remove the parasites from the body via chronic inflammation. Preferential partner choice of the factor *DBP* in the liver of both breeds could play an influential role in their AAT-tolerance mechanisms. In a study performed by Kierstein *et al.* [99] in mouse models infected with *T. congolense*, *DBP* has been identified as one of the Differentially Expressed Genes (DEGs) between susceptible and tolerant mice [99]. In a similar study, genes related to lipid transport and metabolism are frequently reported during the progressive stages of the disease [99]. Contrarily in N'Dama, the cooperation between *DBP–PPARA* could be leading to the regulation of lipid metabolism and inflammation. The parasitic trypanosomes might have altered the aforementioned regulatory mechanism of the host by changing the TF cooperation, especially that of *DBP*. This might be the hidden link between metabolic and immune system related pathways. In Boran liver, the *DBP-PAX8* cooperation might be favouring the survival of the parasite in manipulating the pathways for lipid metabolism, which are essential for the parasite. On the other hand in N'Dama, strong transcriptional regulation of metabolism and inflammation might be serving as a critical switch in AAT-tolerance. Regarding the function of *DBP* in controlling the circadian rhythm of liver tissue in mammals and its relation to the AAT disease, our findings suggest that the specific partners of *DBP* in both breeds could be associated with different genetic programs governing their susceptibility or tolerance.
Another interesting TF found in the liver tissue of Boran is *RFX5*. The factor *RFX5* belongs to the family of *RFX* (Regulatory Factor X) gene transcription factors [50] and its over-expression has been observed in hepatocellular carcinoma [100]. Previous studies have shown its critical importance in the regulation of the MHC (Major Histocompatibility Complex) class II gene for

which *RFX5* activates the expression of those genes essential for the initiation and propagation of the antigen-specific immune T cells [101, 102, 103]. In this regard, MHC II genes are shown to play an important role in the adaptive immunity [101, 102, 103]. In the thymus, they are important for the positive and negative selection of T-cells [101, 102, 103].

In the liver tissue of Boran, the *RFX5* cooperates with the factors *TTF1* and *PPARG* (see Figure 2.3a). The factor *TTF1* (Thyroid Transcription Factor 1) is a nuclear protein expressed in the thyroid and the pulmonary epithelium [104, 105]. It serves as a specific marker for lung and rectal adenocarcinoma [106, 107, 108]. Together with the factor *PAX8*, the factor *TTF1* is a particularly important player in the organogenesis of the thyroid gland; both are reported in several studies of thyroid carcinomas [109, 110, 111]. The second cooperation partner of *RFX5* is *PPARG*, which is a nuclear receptor with anti-inflammatory role and it contributes to cardiovascular diseases [112]. Furthermore, it regulates the expression of *CD36* upon induction, which is involved in processes such as angiogenesis and inflammation. *PPARG* acts as modulator in adipogenesis, insulin sensitivity and the whole-body lipid metabolism [113, 114, 115, 116].

On the other hand, the cooperation network of liver of N'Dama contains the transcription factors *USF2* and *FOXM1* which are strongly associated with immune responses (Figure 2.3d). Upstream Stimulatory Factor (*USF2*) is a member of the basic helix-loop-helix family that has been identified as one of the controllers of insulin synthesis [117]. Further, *USF2* participates in the regulation of important cellular processes like metabolism, embryonic development, brain function, fertility, iron homeostasis and immune responses [118, 119]. In a breast cancer study, levels of *USF2* were reported abnormal and were suggested to play a role in cancer progression [120].

The factor *FOXM1* belongs to the Forkhead box (FOX) transcription factor family [50] and is involved in a variety of biological processes including DNA damage response, drug resistance, cell death, and cell proliferation [121]. Furthermore, *FOXM1* is regarded as the master regulator for DNA damage response and genotoxic agent resistance [122, 123]. Based on its regulatory role, it is studied as a potential target for prognosis and treatment of cancers [124].

Taken together, the cooperative TF pairs in the liver tissue of both breeds could provide promising information to elucidate their regulatory genetic programs governing their susceptibility and tolerance traits.

2.4.5 Cooperative TFs in Spleen Tissue

Examining the cooperation networks for the spleen tissue of Boran and for N'Dama (Figure 2.3b,e) illustrates that all single TFs have different partners in the two breeds except for the TF pair *E2F1–TFAP2A*. Interestingly, the factor *DBP* forms only one dimer in the spleen tissue of N'Dama and it is absent in the cooperation network of Boran spleen tissue. The absence of *DBP* might be indirectly a significant implication of the disruption of the circadian rhythm and the related rhythmic processes in the spleen tissue as a result of trypanosome infection in Boran.

In contrast to Boran, the factor *DBP* forms a cooperative pair with the factor *HAND1E47* in the TF network of the spleen tissue of N'Dama. Taking into account the significance of the interaction between *DBP* and *HAND1E47*, the factor *HAND1E47* belongs to the bHLH transcription factor family, which is mainly involved in the cardiogenesis and hematopoiesis processes, as per the studies in Drosophila model [125]. Furthermore, loss-of-function mutation in *HAND1* has been reported in dilated cardiomyopathy, which is the continuous enlargement and loss of contraction of the ventricular chamber in the heart [126]. The *DBP-HAND1E47* cooperation in the spleen tissue seems to have great importance in the AAT disease resistance of N'Dama, because the AAT-infected cattle die in the final stage of the AAT disease, from cardiovascular defects wherein the TF cooperation *DBP-HAND1E47* could play a defensive role in N'Dama.

A closer look at the cooperation networks of the spleen tissue further reveals that there are several homeobox transcription factors in both networks. Particularly, the factors *HOXA4* and *HOXB7* are found in the networks of both breeds, however with different TF partners. The TFs *HOXA*, *HOXB*, and *HMBOX1* are involved in the regulation of differentiation of haematopoietic cells [127, 128, 129, 130] and development of the embryo [131]. *HOX* genes have been shown to be master regulators of haematopoiesis and are related to haematopoietic disorders [132]. Consequently, their TF cooperations might also play important roles during the AAT disease as the cattle suffers from anaemia which is the destruction of blood cells. Moreover, *HOXA4* has been reported in relation to Chronic Myeloid Leukaemia (CML) [133] and increased expression of *HOXA6* in Acute Myeloid Leukaemia [134]. Furthermore, the factor *HMBOX1* in the network of N'Dama, functions as a transcriptional repressor of the cell cytolytic activity of NK cells [135]. The factor *HMBOX1* cooperates with *BATF*, *HOXA4*, and *RFX5* only in the network of N'Dama which could lead to a significant difference in the regulatory events of the spleen tissue between Boran and N'Dama. *HMBOX1* regulates the process of cell cytolysis, which could thereby be controlling the destruction

of blood cells and maintaining the normal blood count of blood cells. According to its known molecular functions, the factor *HMBOX1* might have a protective role in the spleen tissue against splenomegaly and anaemia which are prominent features of AAT.
Although the factor *BATF* has been identified as significant in both networks, it switches its partner (see Figure 2.3b,e). *BATF* belongs to the bZIP family of transcription factors and is predominantly expressed in lymphocytes. Its preferential partner choice could be strongly associated with the production of immune responses since this TF is specialized in controlling the differentiation of Th17 cells [136].
Another interesting factor is *SIX5* which is found only in the network of N'Dama. *SIX5* belongs to the Sine Oculis homeobox homolog family and is mainly involved in the process of differentiation, migration, and organogenesis [137]. In mouse, *SIX5*-deficient animals displayed characteristics of myotonic dystrophy [137, 138, 139], which is characterized by muscle weakness, cataracts, heart conduction complications and impaired cognitive functions [139]. The absence of *SIX5* in Boran could also lead to the different AAT-disease signs.
Collectively, our findings in the spleen tissue of both breeds suggest that the specific partner choice of TFs could potentially contribute to splenomegaly, anaemia and immune responses in the susceptible breed Boran.

2.4.6 Cooperative TFs in Lymph Node Tissue

Analysis of the cooperation networks of the lymph node tissue (Figure 2.3c,f) demonstrates that few single TFs are common in the networks of both breeds, however, with different partners. In particular, the transcription factor *DBP* forms different dimers in these cooperation networks. Interestingly, in Boran lymph node *DBP* cooperates with *FOXM1*, which is a crucial mediator of inflammatory responses. Further, knocking out *FOXM1* has resulted in the reduction of inflammatory response in osteoarthritis [140], suggesting that *FOXM1* could play a crucial role in chronic inflammation in Boran during the disease. On the other hand, in N'Dama, the cooperation partner of *DBP* is *TFAP2A* which belongs to the *AP2* transcription factors. Surprisingly similar to TFs found in liver tissue, also the TF *TFAP2A* is strongly related to lipid droplet biogenesis, which plays an important role in host-pathogen interaction [58, 141]. According to the regulatory functions of *TFAP2A*, the cooperation of *DBP-TFAP2A* could be remarkably in strong favour of the rhythmic regulation of lipid droplet biogenesis process in the trypanotolerant breed N'Dama.
The cooperation *DBP-FOXM1* in Boran could be in relation with the reg-

ulation of inflammation processes. In contrast to Boran, the *DBP-TFAP2A* cooperation might be involved in the regulation of the circadian rhythm in lymph node tissues of N'Dama.
Another transcription factor found in the cooperation network of Boran is *MAFF*, a leucine zipper (bZIP)-type transcription factor that cooperates with the following factors: *FOXA1*, *SIX3*, *HOXA4* and *HOXA6*. Remarkably, SNPs in *MAFF* are experimentally reported to be in association with Chronic Myeloid Leukaemia [142]. The symptoms of Chronic Myeloid Leukaemia closely resemble the condition of AAT-affected cattle which suffer from similar weight loss, lymphadenopathy, splenomegaly, hepatomegaly, and cardiac failure during the disease [2, 143].
The factor *SMAD4* which is found in the network of N'Dama, is crucial for the regulation of differentiation of Th17 cells. As previously mentioned, Th17 cells are important in the investigation of inflammatory and autoimmune diseases. Furthermore, mutations on *SMAD4* had resulted in the loss of suppression of Th17 cell differentiation and therefore they also serve as therapeutic target for autoimmune disorders [144]. As shown in the network of Boran, *BATF*, and *JUN* complexes have been studied in cartilage destruction through gene regulation in chondrocytes and therefore were identified as targets for osteoarthritis, a degenerative arthritis which affects joint tissues [145].
The factor *TCF4* present in the network of N'Dama, belongs to a basic helix-loop-helix family which plays an integral role in Wnt signalling and neuronal differentiation especially in the brain development [146, 147, 148, 149, 150]. Furthermore, *TCF4* is also involved in the immune responses through the production of plasmacytoid Dendritic Cells (pDCs), which respond to viral nucleic acids and autoimmune diseases, by the secretion of cytokines such as type I interferons [151, 152, 153, 154]. Genetic alterations in *TCF4* are easy targets and therefore mutations in *TCF4* have been reported in the most common form of lymphoma which is the diffuse large B-cell lymphoma and Angelman syndrome [155, 156]. The factor *TCF4* has also been identified as the master regulator of schizophrenia, a severe complicated mental disorder [157]. It is reported that disruption in *TCF4* regulatory networks is associated with neuropsychiatric diseases namely schizoprenia, autism, the Pitt-Hopkins syndrome, and depression [158, 159]. In connection with AAT, the cattle suffers from fever, listlessness, oedema, depression, and paralysis during the progressive stages of the disease [2, 160].
Similar to the networks of liver and spleen tissues, the cooperation network of lymph node tissue reveals the significance of the preferential partner choice of the factor *DBP* and, additionally, it provides a hint that the circadian rhythm in lymph node tissue could be associated with the generation of immune

responses, which also includes inflammatory cytokines and the regulation of lipid droplets during the AAT disease.

2.5 Conclusions

Knowledge about TFs and their complex interplay is pivotal to understand the regulation of genetic programs which maintain adaptation of the animal to different pathophysiological stresses like parasitic infections. Our findings indicate that given the AAT disease progression, the preferential partner choice of TFs is strongly related to the tissue type and the susceptibility/resistance of the cattle breeds. Especially the results emphasized the higher relevance of the factor *DBP* along with its partners in circadian rhythm and lipid metabolism, which could be associated with the pathogenesis of AAT in trypanotolerant N'Dama and trypanosusceptible Boran. Importantly, the recent study of Solis *et al.* [28] on the crucial role of circadian regulation for the coordination of the immune functions lend support to our findings that the circadian control of the immune system influenced by host-pathogen interaction might have resulted in the transcriptional reprogramming of regulation determining the level of AAT-tolerance of the cattle. To the best of our knowledge, this is the first study in this field which mainly focuses on the importance of TFs and their cooperation to reveal the genetic programs underlying the AAT disease. Our results could be used in future works for deciphering the master regulators which could support experimental studies in generating novel hypotheses for potential drug targets.

Supplementary Information

The supplementary files can be accessed via the original publication (https://www.mdpi.com/2076-393X/8/2/246/s1). Table S1: The library of non-redundant position weight matrices (PWMs) used in our study, Table S2: Lists of Monotonically Expressed Genes obtained from the MFSelector approach.

Author contributions

M.G. designed and supervised the research. A.R. participated in the design of the study and conducted computational analyses together with M.G. and performed the literature survey. F.H. and A.O.S. interpreted the results with A.R. and M.G. A.R. and M.G. wrote the final version of the manuscript.

M.G. conceived of and managed the project. All authors read and approved the final manuscript.

Data availability

Not applicable.

Acknowledgments

We acknowledge support by the German Research Foundation and the Open Access Publication Funds of the Göttingen University.

Conflicts of Interest

The authors declare no conflict of interest.

Bibliography

[1] D. Biello, "Deadly by the dozen: 12 diseases climate change may worsen," *Sci Am*, vol. 8, pp. 12–8, 2008.

[2] A. Firesbhat and C. Desalegn, "Epidemiology and impacts of trypanosomiasis in cattle," *European Journal of Applied Sciences*, vol. 7, no. 5, pp. 220–225, 2015.

[3] O. M. Radostits, C. C. Gay, K. W. Hinchcliff, and P. D. Constable, *Veterinary Medicine E-Book: A textbook of the diseases of cattle, horses, sheep, pigs and goats*. Elsevier Health Sciences, 2006.

[4] J. E. Donelson, K. L. Hill, and N. M. El-Sayed, "Multiple mechanisms of immune evasion by african trypanosomes," *Molecular and biochemical parasitology*, vol. 91, no. 1, pp. 51–66, 1998.

[5] G. J. Losos and B. Ikede, "Review of pathology of diseases in domestic and laboratory animals caused by trypanosoma congolense, t. vivax, t. brucei, t. rhodesiense and t. gambiense," *Veterinary Pathology*, vol. 9, no. 1_suppl, pp. 1–79, 1972.

[6] L. J. Morrison, L. Vezza, T. Rowan, and J. C. Hope, "Animal african trypanosomiasis: time to increase focus on clinically relevant parasite and host species," *Trends in parasitology*, vol. 32, no. 8, pp. 599–607, 2016.

[7] P. M. Kristjanson, B. M. Swallow, G. Rowlands, R. Kruska, and P. De Leeuw, "Measuring the costs of african animal trypanosomosis, the potential benefits of control and returns to research," *Agricultural systems*, vol. 59, no. 1, pp. 79–98, 1999.

[8] B. Stijlemans, P. De Baetselier, S. Magez, J. A. Van Ginderachter, and C. De Trez, "African trypanosomiasis-associated anemia: the contribution of the interplay between parasites and the mononuclear phagocyte system," *Frontiers in immunology*, vol. 9, p. 218, 2018.

[9] M. Yaro, K. Munyard, M. Stear, and D. Groth, "Combatting african animal trypanosomiasis (aat) in livestock: the potential role of trypanotolerance," *Veterinary parasitology*, vol. 225, pp. 43–52, 2016.

[10] J. Naessens, "Bovine trypanotolerance: a natural ability to prevent severe anaemia and haemophagocytic syndrome?," *International Journal for Parasitology*, vol. 36, no. 5, pp. 521–528, 2006.

[11] M. Murray, G. D'Ieteren, E. Authie, and N. Wissocq, "Trypanotolerance, an option for sustainable livestock production in areas at risk from trypanosomosis," *OIE Revue Scientifique et Technique*, vol. 17, no. 1, pp. 154–175, 1998.

[12] P. Starkey, "N'dama cattle-a productive trypanotolerant breed," 1984.

[13] J. Grootenhuis, R. Dwinger, R. Dolan, S. Moloo, and M. Murray, "Susceptibility of african buffalo and boran cattle to trypanosoma congolense transmitted by glossina morsitans centralis," *Veterinary parasitology*, vol. 35, no. 3, pp. 219–231, 1990.

[14] P. Borst, "Antigenic variation and allelic exclusion," *Cell*, vol. 109, no. 1, pp. 5–8, 2002.

[15] G. A. Cross, L. E. Wirtz, and M. Navarro, "Regulation of vsg expression site transcription and switching in trypanosoma brucei," *Molecular and biochemical parasitology*, vol. 91, no. 1, pp. 77–91, 1998.

[16] J. D. Barry and R. McCulloch, "Antigenic variation in trypanosomes: enhanced phenotypic variation in a eukaryotic parasite.," *Advances in parasitology*, vol. 49, pp. 1–70, 2001.

[17] J. Cnops, S. Magez, and C. De Trez, "Escape mechanisms of african trypanosomes: why trypanosomosis is keeping us awake," *Parasitology*, vol. 142, no. 3, pp. 417–427, 2015.

[18] B. Anene, D. Onah, and Y. Nawa, "Drug resistance in pathogenic african trypanosomes: what hopes for the future?," *Veterinary Parasitology*, vol. 96, no. 2, pp. 83–100, 2001.

[19] L. Michel-Todó, P. Bigey, P. A. Reche, M.-J. Pinazo, J. Gascón, and J. Alonso-Padilla, "Design of an epitope-based vaccine ensemble for animal trypanosomiasis by computational methods," *Vaccines*, vol. 8, no. 1, p. 130, 2020.

[20] H. Noyes, A. Brass, I. Obara, S. Anderson, A. L. Archibald, D. G. Bradley, P. Fisher, A. Freeman, J. Gibson, M. Gicheru, *et al.*, "Genetic and expression analysis of cattle identifies candidate genes in pathways responding to trypanosoma congolense infection," *Proceedings of the National Academy of Sciences*, vol. 108, no. 22, pp. 9304–9309, 2011.

[21] S. M. Kuriakose, R. Singh, and J. E. Uzonna, "Host intracellular signaling events and pro-inflammatory cytokine production in african trypanosomiasis," *Frontiers in immunology*, vol. 7, p. 181, 2016.

[22] N. H. Ruddle and E. M. Akirav, "Secondary lymphoid organs: responding to genetic and environmental cues in ontogeny and the immune response," *The Journal of Immunology*, vol. 183, no. 4, pp. 2205–2212, 2009.

[23] C. E. Clayton, M. E. Selkirk, C. A. Corsini, B. M. Ogilvie, and B. A. Askonas, "Murine trypanosomiasis: cellular proliferation and functional depletion in the blood, peritoneum, and spleen related to changes in bone marrow stem cells," *Infection and immunity*, vol. 28, no. 3, pp. 824–831, 1980.

[24] H.-W. Wang, H.-J. Sun, T.-Y. Chang, H.-H. Lo, W.-C. Cheng, G. C. Tseng, C.-T. Lin, S.-J. Chang, N. R. Pal, and I.-F. Chung, "Discovering monotonic stemness marker genes from time-series stem cell microarray data," in *BMC genomics*, vol. 16, p. S2, BioMed Central, 2015.

[25] S. Tian, "Identification of monotonically differentially expressed genes for non-small cell lung cancer," *BMC bioinformatics*, vol. 20, no. 1, p. 177, 2019.

[26] C. Meckbach, R. Tacke, X. Hua, S. Waack, E. Wingender, and M. Gültas, "Pc-traff: identification of potentially collaborating transcription factors using pointwise mutual information," *BMC bioinformatics*, vol. 16, no. 1, p. 400, 2015.

[27] D. Wlochowitz, M. Haubrock, J. Arackal, A. Bleckmann, A. Wolff, T. Beißbarth, E. Wingender, and M. Gültas, "Computational identification of key regulators in two different colorectal cancer cell lines," *Frontiers in genetics*, vol. 7, p. 42, 2016.

[28] R. Orozco-Solis and L. Aguilar-Arnal, "Circadian regulation of immunity through epigenetic mechanisms," *Frontiers in Cellular and Infection Microbiology*, vol. 10, p. 96, 2020.

[29] D. S. Latchman, "Transcription factors: an overview," *The international journal of biochemistry & cell biology*, vol. 29, no. 12, pp. 1305–1312, 1997.

[30] T. W. Whitfield, J. Wang, P. J. Collins, E. C. Partridge, S. F. Aldred, N. D. Trinklein, R. M. Myers, and Z. Weng, "Functional analysis of transcription factor binding sites in human promoters," *Genome biology*, vol. 13, no. 9, p. R50, 2012.

[31] M. Blanchette and M. Tompa, "Discovery of regulatory elements by a computational method for phylogenetic footprinting," *Genome research*, vol. 12, no. 5, pp. 739–748, 2002.

[32] M. Blanchette, B. Schwikowski, and M. Tompa, "Algorithms for phylogenetic footprinting," *Journal of computational biology*, vol. 9, no. 2, pp. 211–223, 2002.

[33] A. Jolma, J. Yan, T. Whitington, J. Toivonen, K. R. Nitta, P. Rastas, E. Morgunova, M. Enge, M. Taipale, G. Wei, *et al.*, "Dna-binding specificities of human transcription factors," *Cell*, vol. 152, no. 1-2, pp. 327–339, 2013.

[34] K. R. Nitta, A. Jolma, Y. Yin, E. Morgunova, T. Kivioja, J. Akhtar, K. Hens, J. Toivonen, B. Deplancke, E. E. Furlong, *et al.*, "Conservation of transcription factor binding specificities across 600 million years of bilateria evolution," *Elife*, vol. 4, p. e04837, 2015.

[35] M. T. Weirauch, A. Yang, M. Albu, A. G. Cote, A. Montenegro-Montero, P. Drewe, H. S. Najafabadi, S. A. Lambert, I. Mann, K. Cook, *et al.*, "Determination and inference of eukaryotic transcription factor sequence specificity," *Cell*, vol. 158, no. 6, pp. 1431–1443, 2014.

[36] J. M. Franco-Zorrilla, I. López-Vidriero, J. L. Carrasco, M. Godoy, P. Vera, and R. Solano, "Dna-binding specificities of plant transcription factors and their potential to define target genes," *Proceedings of the National Academy of Sciences*, vol. 111, no. 6, pp. 2367–2372, 2014.

[37] Y. Xie, M. Moussaif, S. Choi, L. Xu, and J. Y. Sze, "Rfx transcription factor daf-19 regulates 5-ht and innate immune responses to pathogenic bacteria in caenorhabditis elegans," *PLoS genetics*, vol. 9, no. 3, 2013.

[38] N. Powell, J. Canavan, T. MacDonald, and G. Lord, "Transcriptional regulation of the mucosal immune system mediated by t-bet," *Mucosal immunology*, vol. 3, no. 6, pp. 567–577, 2010.

[39] E. V. Rothenberg, "Transcriptional control of early t and b cell developmental choices," *Annual review of immunology*, vol. 32, pp. 283–321, 2014.

[40] D. Ghosh and A. G. Papavassiliou, "Transcription factor therapeutics: long-shot or lodestone," *Current medicinal chemistry*, vol. 12, no. 6, pp. 691–701, 2005.

[41] M. Moustaqil, Y. Gambin, and E. Sierecki, "Biophysical techniques for target validation and drug discovery in transcription-targeted therapy," *International Journal of Molecular Sciences*, vol. 21, no. 7, p. 2301, 2020.

[42] K. A. Papavassiliou and A. G. Papavassiliou, "Transcription factor drug targets," *Journal of cellular biochemistry*, vol. 117, no. 12, pp. 2693–2696, 2016.

[43] D. Karolchik, A. S. Hinrichs, T. S. Furey, K. M. Roskin, C. W. Sugnet, D. Haussler, and W. J. Kent, "The ucsc table browser data retrieval tool," *Nucleic acids research*, vol. 32, no. suppl_1, pp. D493–D496, 2004.

[44] E. Wingender, "The TRANSFAC project as an example of framework technology that supports the analysis of genomic regulation," *Briefings in Bioinformatics*, vol. 9, no. 4, pp. 326–332, 2008.

[45] G. D. Stormo, "DNA binding sites: representation and discovery," *Bioinformatics*, vol. 16, pp. 16–23, jan 2000.

[46] M. Tompa, N. Li, T. L. Bailey, G. M. Church, B. De Moor, E. Eskin, A. V. Favorov, M. C. Frith, Y. Fu, W. J. Kent, *et al.*, "Assessing computational tools for the discovery of transcription factor binding sites," *Nature biotechnology*, vol. 23, no. 1, p. 137, 2005.

[47] L. Steuernagel, C. Meckbach, F. Heinrich, S. Zeidler, A. O. Schmitt, and M. Gültas, "Computational identification of tissue-specific transcription factor cooperation in ten cattle tissues," *PLOS ONE*, vol. 14, pp. 1–21, 05 2019.

[48] H.-M. Zhang, T. Liu, C.-J. Liu, S. Song, X. Zhang, W. Liu, H. Jia, Y. Xue, and A.-Y. Guo, "AnimalTFDB 2.0: a resource for expression, prediction and functional study of animal transcription factors," *Nucleic Acids Research*, vol. 43, pp. D76–D81, sep 2014.

[49] E. Wingender, X. Chen, E. Fricke, R. Geffers, R. Hehl, I. Liebich, M. Krull, V. Matys, H. Michael, R. Ohnhaeuser, *et al.*, "Match-a tool for searching transcription factor binding sites in dna sequences," *Nucl. Acids Res*, vol. 29, pp. 281–283, 2001.

[50] E. Wingender, T. Schoeps, and J. Dönitz, "Tfclass: an expandable hierarchical classification of human transcription factors," *Nucleic acids research*, vol. 41, no. D1, pp. D165–D170, 2012.

[51] L. Fajas, R. L. Landsberg, Y. Huss-Garcia, C. Sardet, J. A. Lees, and J. Auwerx, "E2fs regulate adipocyte differentiation," *Developmental cell*, vol. 3, no. 1, pp. 39–49, 2002.

[52] P.-D. Denechaud, L. Fajas, and A. Giralt, "E2f1, a novel regulator of metabolism," *Frontiers in endocrinology*, vol. 8, p. 311, 2017.

[53] Y. Haim, M. Blüher, N. Slutsky, N. Goldstein, N. Klöting, I. Harman-Boehm, B. Kirshtein, D. Ginsberg, M. Gericke, E. Guiu Jurado, *et al.*, "Elevated autophagy gene expression in adipose tissue of obese humans: A potential non-cell-cycle-dependent function of e2f1," *Autophagy*, vol. 11, no. 11, pp. 2074–2088, 2015.

[54] K. Bozek, A. Relógio, S. M. Kielbasa, M. Heine, C. Dame, A. Kramer, and H. Herzel, "Regulation of clock-controlled genes in mammals," *PloS one*, vol. 4, no. 3, 2009.

[55] S. Kersten and R. Stienstra, "The role and regulation of the peroxisome proliferator activated receptor alpha in human liver," *Biochimie*, vol. 136, pp. 75–84, 2017.

[56] M. Roussigne, C. Cayrol, T. Clouaire, F. Amalric, and J.-P. Girard, "Thap1 is a nuclear proapoptotic factor that links prostate-apoptosis-response-4 (par-4) to pml nuclear bodies," *Oncogene*, vol. 22, no. 16, p. 2432, 2003.

[57] V. Gervais, S. Campagne, J. Durand, I. Muller, and A. Milon, "Nmr studies of a new family of dna binding proteins: the thap proteins," *Journal of biomolecular NMR*, vol. 56, no. 1, pp. 3–15, 2013.

[58] C. C. Scott, S. Vossio, J. Rougemont, and J. Gruenberg, "Tfap2 transcription factors are regulators of lipid droplet biogenesis," *Elife*, vol. 7, p. e36330, 2018.

[59] R. C. Melo and P. F. Weller, "Lipid droplets in leukocytes: organelles linked to inflammatory responses," *Experimental cell research*, vol. 340, no. 2, pp. 193–197, 2016.

[60] M. Konige, H. Wang, and C. Sztalryd, "Role of adipose specific lipid droplet proteins in maintaining whole body energy homeostasis," *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, vol. 1842, no. 3, pp. 393–401, 2014.

[61] J. C. Cross, M. L. Flannery, M. A. Blanar, E. Steingrimsson, N. A. Jenkins, N. Copeland, W. Rutter, and Z. Werb, "Hxt encodes a basic helix-loop-helix transcription factor that regulates trophoblast cell development," *Development*, vol. 121, no. 8, pp. 2513–2523, 1995.

[62] S. M. Hollenberg, R. Sternglanz, P. F. Cheng, and H. Weintraub, "Identification of a new family of tissue-specific basic helix-loop-helix proteins with a two-hybrid system.," *Molecular and cellular biology*, vol. 15, no. 7, pp. 3813–3822, 1995.

[63] P. Cserjesi, D. Brown, G. E. Lyons, and E. N. Olson, "Expression of the novel basic helix-loop-helix gene ehand in neural crest derivatives and extraembryonic membranes during mouse development," *Developmental biology*, vol. 170, no. 2, pp. 664–678, 1995.

[64] D. Srivastava, P. Cserjesi, and E. N. Olson, "A subclass of bhlh proteins required for cardiac morphogenesis," *Science*, vol. 270, no. 5244, 1995.

[65] D. G. McFadden, A. C. Barbosa, J. A. Richardson, M. D. Schneider, D. Srivastava, and E. N. Olson, "The hand1 and hand2 transcription factors regulate expansion of the embryonic cardiac ventricles in a gene dosage-dependent manner," *Development*, vol. 132, no. 1, pp. 189–201, 2005.

[66] C. Meckbach, E. Wingender, and M. Gültas, "Removing background co-occurrences of transcription factor binding sites greatly improves the prediction of specific transcription factor cooperations," *Frontiers in genetics*, vol. 9, 2018.

[67] S. Zeidler, C. Meckbach, R. Tacke, F. S. Raad, A. Roa, S. Uchida, W.-H. Zimmermann, E. Wingender, and M. Gültas, "Computational detection of stage-specific transcription factor clusters during heart development," *Frontiers in genetics*, vol. 7, p. 33, 2016.

[68] L. Lopez-Molina, F. Conquet, M. Dubois-Dauphin, and U. Schibler, "The dbp gene is expressed according to a circadian rhythm in the suprachiasmatic nucleus and influences circadian behavior," *The EMBO Journal*, vol. 16, 1997.

[69] W. Olds, *Sleep, circadian rhythms, and metabolism: The rhythm of life*. Apple Academic Press, 2014.

[70] W. Huang, K. M. Ramsey, B. Marcheva, and J. Bass, "Circadian rhythms, sleep, and metabolism," *The Journal of clinical investigation*, vol. 121, no. 6, pp. 2133–2141, 2011.

[71] A. D. Laposky, J. Bass, A. Kohsaka, and F. W. Turek, "Sleep and circadian rhythms: key components in the regulation of energy metabolism," *FEBS letters*, vol. 582, no. 1, pp. 142–151, 2008.

[72] G. B. Lundkvist, K. Kristensson, and M. Bentivoglio, "Why trypanosomes cause sleeping sickness," *Physiology*, vol. 19, no. 4, pp. 198–206, 2004.

[73] S. M. Reppert and D. R. Weaver, "Molecular analysis of mammalian circadian rhythms," *Annual review of physiology*, vol. 63, no. 1, pp. 647–676, 2001.

[74] L. Borgs, P. Beukelaers, R. Vandenbosch, S. Belachew, L. Nguyen, and B. Malgrange, "Cell "circadian" cycle: new role for mammalian core clock genes," *Cell Cycle*, vol. 8, no. 6, pp. 832–837, 2009.

[75] A. Balsalobre, "Clock genes in mammalian peripheral tissues," *Cell and tissue research*, vol. 309, no. 1, pp. 193–199, 2002.

[76] J. Wuarin and U. Schibler, "Expression of the liver-enriched transcriptional activator protein dbp follows a stringent circadian rhythm," *Cell*, vol. 63, no. 6, pp. 1257–1266, 1990.

[77] H. Reinke and G. Asher, "Circadian clock control of liver metabolic functions," *Gastroenterology*, vol. 150, no. 3, pp. 574–580, 2016.

[78] G. Asher and U. Schibler, "Crosstalk between components of circadian and metabolic cycles in mammals," *Cell metabolism*, vol. 13, no. 2, pp. 125–137, 2011.

[79] G. Cretenet, M. Le Clech, and F. Gachon, "Circadian clock-coordinated 12 hr period rhythmic activation of the ire1α pathway controls lipid metabolism in mouse liver," *Cell metabolism*, vol. 11, no. 1, pp. 47–57, 2010.

[80] Y.-H. Lee, J. A. Alberta, F. J. Gonzalez, and D. J. Waxman, "Multiple, functional dbp sites on the promoter of the cholesterol 7 alpha-hydroxylase p450 gene, cyp7. proposed role in diurnal regulation of liver gene expression.," *Journal of Biological Chemistry*, vol. 269, no. 20, pp. 14681–14689, 1994.

[81] M. Zannini, H. Francis-Lang, D. Plachov, and R. Di Lauro, "Pax-8, a paired domain-containing protein, binds to a sequence overlapping the recognition site of a homeodomain and activates transcription from two thyroid-specific promoters.," *Molecular and Cellular Biology*, vol. 12, no. 9, pp. 4230–4241, 1992.

[82] D. Plachov, K. Chowdhury, C. Walther, D. Simon, J.-L. Guenet, and P. Gruss, "Pax8, a murine paired box gene expressed in the developing excretory system and thyroid gland," *Development*, vol. 110, no. 2, pp. 643–651, 1990.

[83] P. E. Macchia, P. Lapi, H. Krude, M. T. Pirro, C. Missero, L. Chiovato, A. Souabni, M. Baserga, V. Tassi, A. Pinchera, *et al.*, "Pax8 mutations associated with congenital hypothyroidism caused by thyroid dysgenesis," *Nature genetics*, vol. 19, no. 1, p. 83, 1998.

[84] A. R. Laury, R. Perets, H. Piao, J. F. Krane, J. A. Barletta, C. French, L. R. Chirieac, R. Lis, M. Loda, J. L. Hornick, *et al.*, "A comprehensive analysis of pax8 expression in human epithelial tumors," *The American journal of surgical pathology*, vol. 35, no. 6, pp. 816–826, 2011.

[85] Y. Patel, F. Alford, and H. Burger, "The 24-hour plasma thyrotrophin profile," *Clinical Science*, vol. 43, no. 1, pp. 71–77, 1972.

[86] B. Mutayoba and S. Gombe, "Effect of african trypanosomiasis on plasma cortisol and thyroxine concentration in goats," *Research in veterinary science*, vol. 47, no. 3, pp. 315–318, 1989.

[87] G. Abebe, *The integrity of hypothalamic-pituitary-adrenal axis in boran (Bos indicus) cattle infected with Trypanosoma congolense*. PhD thesis, 1991.

[88] D. Mukhopadhyay, M. Plateroti, S. Anant, F. Nassir, J. Samarut, and N. O. Davidson, "Thyroid hormone regulates hepatic triglyceride mobilization and apolipoprotein b messenger ribonucleic acid editing in a murine model of congenital hypothyroidism," *Endocrinology*, vol. 144, no. 2, pp. 711–719, 2003.

[89] E. A. McAninch and A. C. Bianco, "Thyroid hormone signaling in energy homeostasis and energy metabolism," *Annals of the New York Academy of Sciences*, vol. 1311, p. 77, 2014.

[90] D. Mathe and F. Chevallier, "Effects of the thyroid state on cholesterol metabolism in the rat," *Biochimica et Biophysica Acta (BBA)-Lipids and Lipid Metabolism*, vol. 441, no. 1, pp. 155–164, 1976.

[91] G. C. Ness, R. E. Dugan, M. Lakshmanan, C. M. Nepokroeff, and J. W. Porter, "Stimulation of hepatic β-hydroxy-β-methylglutaryl coenzyme a reductase activity in hypophysectomized rats by l-triiodothyronine," *Proceedings of the National Academy of Sciences*, vol. 70, no. 12, pp. 3839–3842, 1973.

[92] G. C. Ness, L. C. Pendleton, Y. C. Li, and J. Y. Chiang, "Effect of thyroid hormone on hepatic cholesterol 7α hydroxylase, ldl receptor, hmg-coa reductase, farnesyl pyrophosphate synthetase and apolipoprotein ai mrna levels in hypophysectomized rats," *Biochemical and biophysical research communications*, vol. 172, no. 3, pp. 1150–1156, 1990.

[93] T. Varga, Z. Czimmerer, and L. Nagy, "Ppars are a unique set of fatty acid regulated transcription factors controlling both lipid metabolism and inflammation," *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, vol. 1812, no. 8, pp. 1007–1022, 2011.

[94] T. Clouaire, M. Roussigne, V. Ecochard, C. Mathe, F. Amalric, and J.-P. Girard, "The thap domain of thap1 is a large c2ch module with zinc-dependent sequence-specific dna-binding activity," *Proceedings of the National Academy of Sciences*, vol. 102, no. 19, pp. 6907–6912, 2005.

[95] M. Roussigne, S. Kossida, A.-C. Lavigne, T. Clouaire, V. Ecochard, A. Glories, F. Amalric, and J.-P. Girard, "The thap domain: a novel protein motif with similarity to the dna-binding domain of p element transposase," *Trends in biochemical sciences*, vol. 28, no. 2, pp. 66–69, 2003.

[96] Z. Zakirova, T. Fanutza, J. Bonet, B. Readhead, W. Zhang, Z. Yi, G. Beauvais, T. P. Zwaka, L. J. Ozelius, R. D. Blitzer, *et al.*, "Mutations in thap1/dyt6 reveal that diverse dystonia genes disrupt similar neuronal pathways and functions," *PLoS genetics*, vol. 14, no. 1, p. e1007169, 2018.

[97] V. Valli, C. Forsberg, and G. Robinson, "The pathogenesis of trypanosoma congolense infection in calves. i. clinical observations and gross pathological changes," *Veterinary pathology*, vol. 15, no. 5, pp. 608–620, 1978.

[98] S. Silva Pereira, S. Trindade, M. De Niz, and L. M. Figueiredo, "Tissue tropism in parasitic diseases," *Open biology*, vol. 9, no. 5, p. 190036, 2019.

[99] S. Kierstein, H. Noyes, J. Naessens, Y. Nakamura, C. Pritchard, J. Gibson, S. Kemp, and A. Brass, "Gene expression profiling in a mouse model for african trypanosomiasis," *Genes & Immunity*, vol. 7, no. 8, pp. 667–679, 2006.

[100] Y. Zhao, X. Xie, W. Liao, H. Zhang, H. Cao, R. Fei, X. Wang, L. Wei, Q. Shao, and H. Chen, "The transcription factor rfx5 is a transcriptional activator of the tpp1 gene in hepatocellular carcinoma," *Oncology reports*, vol. 37, no. 1, pp. 289–296, 2017.

[101] S. Aftab, L. Semenec, J. S.-C. Chu, and N. Chen, "Identification and characterization of novel human tissue-specific rfx transcription factors," *BMC evolutionary biology*, vol. 8, no. 1, p. 226, 2008.

[102] J. Villard, M. Peretti, K. Masternak, E. Barras, G. Caretti, R. Mantovani, and W. Reith, "A functionally essential domain of rfx5 mediates activation of major histocompatibility complex class ii promoters by promoting cooperative binding between rfx and nf-y," *Molecular and cellular biology*, vol. 20, no. 10, pp. 3364–3376, 2000.

[103] C. A. Janeway Jr, "Thymic selection: two pathways to life and two to death," *Immunity*, vol. 1, no. 1, pp. 3–6, 1994.

[104] D. Lazzaro, M. Price, M. de Felice, and R. Di Lauro, "The transcription factor ttf-1 is expressed at the onset of thyroid and lung morphogenesis and in restricted regions of the foetal brain," *Development*, vol. 113, no. 4, pp. 1093–1104, 1991.

[105] K. Ikeda, J. C. Clark, J. R. Shaw-White, M. T. Stahlman, C. J. Boutell, and J. A. Whitsett, "Gene structure and expression of human thyroid transcription factor-1 in respiratory epithelial cells," *Journal of Biological Chemistry*, vol. 270, no. 14, pp. 8108–8114, 1995.

[106] Y.-C. Su, Y.-C. Hsu, and C.-Y. Chai, "Role of ttf-1, ck20, and ck7 immunohistochemistry for diagnosis of primary and secondary lung adenocarcinoma," *The Kaohsiung journal of medical sciences*, vol. 22, no. 1, pp. 14–19, 2006.

[107] A. Belalcazar, U. Krishnamurti, and W. Shaib, "Ttf-1 expression in rectal adenocarcinoma: a case report and review of the literature," *Case reports in oncology*, vol. 9, no. 3, pp. 747–751, 2016.

[108] S. Aversa and C. Bellan, "Ttf1 expression in pulmonary metastatic rectal adenocarcinoma," *Case reports in gastrointestinal medicine*, vol. 2018, 2018.

[109] D. Fabbro, C. Di Loreto, C. A. Beltrami, A. Belfiore, R. Di Lauro, and G. Damante, "Expression of thyroid-specific transcription factors ttf-1 and pax-8 in human thyroid neoplasms," *Cancer research*, vol. 54, no. 17, pp. 4744–4749, 1994.

[110] P. Ros, D. L. Rossi, A. Acebrón, and P. Santisteban, "Thyroid-specific gene expression in the multi-step process of thyroid carcinogenesis," *Biochimie*, vol. 81, no. 4, pp. 389–396, 1999.

[111] P. Zhang, H. Zuo, Y. Nakamura, M. Nakamura, T. Wakasa, and K. Kakudo, "Immunohistochemical analysis of thyroid-specific transcription factors in thyroid tumors," *Pathology international*, vol. 56, no. 5, pp. 240–245, 2006.

[112] W. Wahli and L. Michalik, "Ppars at the crossroads of lipid signaling and inflammation," *Trends in Endocrinology & Metabolism*, vol. 23, no. 7, pp. 351–363, 2012.

[113] R. A. Daynes and D. C. Jones, "Emerging roles of ppars in inflammation and immunity," *Nature Reviews Immunology*, vol. 2, no. 10, p. 748, 2002.

[114] M. Febbraio, D. P. Hajjar, and R. L. Silverstein, "Cd36: a class b scavenger receptor involved in angiogenesis, atherosclerosis, inflammation, and lipid metabolism," *The Journal of clinical investigation*, vol. 108, no. 6, pp. 785–791, 2001.

[115] E. D. Rosen, P. Sarraf, A. E. Troy, G. Bradwin, K. Moore, D. S. Milstone, B. M. Spiegelman, and R. M. Mortensen, "Pparγ is required for the differentiation of adipose tissue in vivo and in vitro," *Molecular cell*, vol. 4, no. 4, pp. 611–617, 1999.

[116] T. Yamauchi, J. Kamon, H. Waki, K. Murakami, K. Motojima, K. Komeda, T. Ide, N. Kubota, Y. Terauchi, K. Tobe, *et al.*,

"The mechanisms by which both heterozygous peroxisome proliferator-activated receptor γ (pparγ) deficiency and pparγ agonist improve insulin resistance," *Journal of Biological Chemistry*, vol. 276, no. 44, pp. 41245–41254, 2001.

[117] S. Spohrer, R. Groß, L. Nalbach, L. Schwind, H. Stumpf, M. D. Menger, E. Ampofo, M. Montenarh, and C. Götz, "Functional interplay between the transcription factors usf1 and pdx-1 and protein kinase ck2 in pancreatic β-cells," *Scientific reports*, vol. 7, no. 1, p. 16367, 2017.

[118] M. Sirito, Q. Lin, J. M. Deng, R. R. Behringer, and M. Sawadogo, "Overlapping roles and asymmetrical cross-regulation of the usf proteins in mice," *Proceedings of the National Academy of Sciences*, vol. 95, no. 7, pp. 3758–3763, 1998.

[119] S. Corre and M.-D. Galibert, "Upstream stimulating factors: highly versatile stress-responsive transcription factors," *Pigment cell research*, vol. 18, no. 5, pp. 337–348, 2005.

[120] Y. Tan, Y. Chen, M. Du, Z. Peng, and P. Xie, "Usf2 inhibits the transcriptional activity of smurf1 and smurf2 to promote breast cancer tumorigenesis," *Cellular signalling*, vol. 53, pp. 49–58, 2019.

[121] E. W.-F. Lam and A. R. Gomes, "Forkhead box transcription factors in cancer initiation, progression and chemotherapeutic drug response," *Frontiers in oncology*, vol. 4, p. 305, 2014.

[122] G.-B. Liao, X.-Z. Li, S. Zeng, C. Liu, S.-M. Yang, L. Yang, C.-J. Hu, and J.-Y. Bai, "Regulation of the master regulator foxm1 in cancer," *Cell Communication and Signaling*, vol. 16, no. 1, pp. 1–15, 2018.

[123] S. Zona, L. Bella, M. J. Burton, G. N. de Moraes, and E. W.-F. Lam, "Foxm1: an emerging master regulator of dna damage response and genotoxic agent resistance," *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, vol. 1839, no. 11, pp. 1316–1322, 2014.

[124] D. Nandi, P. S. Cheema, N. Jaiswal, and A. Nag, "Foxm1: repurposing an oncogene as a biomarker," in *Seminars in cancer biology*, vol. 52, pp. 74–84, Elsevier, 2018.

[125] Z. Han, P. Yi, X. Li, and E. N. Olson, "Hand, an evolutionarily conserved bhlh transcription factor required for drosophila cardiogenesis and hematopoiesis," *Development*, vol. 133, no. 6, pp. 1175–1182, 2006.

[126] Y.-M. Zhou, X.-Y. Dai, X.-B. Qiu, F. Yuan, R.-G. Li, Y.-J. Xu, X.-K. Qu, R.-T. Huang, S. Xue, and Y.-Q. Yang, "Hand1 loss-of-function mutation associated with familial dilated cardiomyopathy," *Clinical Chemistry and Laboratory Medicine (CCLM)*, vol. 54, no. 7, pp. 1161–1167, 2016.

[127] C. Abramovich and R. K. Humphries, "Hox regulation of normal and leukemic hematopoietic stem cells," *Current opinion in hematology*, vol. 12, no. 3, pp. 210–216, 2005.

[128] S. Chiba, "Homeobox genes in normal hematopoiesis and leukemogenesis.," *International journal of hematology*, vol. 68, no. 4, pp. 343–353, 1998.

[129] C. Cillo, M. Cantile, A. Faiella, and E. Boncinelli, "Homeobox genes in normal and malignant cells," *Journal of cellular physiology*, vol. 188, no. 2, pp. 161–169, 2001.

[130] D. Grier, A. Thompson, A. Kwasniewska, G. McGonigle, H. Halliday, and T. Lappin, "The pathophysiology of hox genes and their role in cancer," *The Journal of Pathology: A Journal of the Pathological Society of Great Britain and Ireland*, vol. 205, no. 2, pp. 154–171, 2005.

[131] J. Deschamps and D. Duboule, "Embryonic timing, axial stem cells, chromatin dynamics, and the hox clock," *Genes & Development*, vol. 31, no. 14, pp. 1406–1416, 2017.

[132] E. M. Collins and A. Thompson, "Hox genes in normal, engineered and malignant hematopoiesis," *The International Journal of Developmental Biology*, vol. 62, no. 11-12, 2020.

[133] M. H. Elias, A. A. Baba, A. Husin, S. Sulong, R. Hassan, G. A. Sim, A. Wahid, S. Fadilah, and R. Ankathil, "Hoxa4 gene promoter hypermethylation as an epigenetic mechanism mediating resistance to imatinib mesylate in chronic myeloid leukemia patients," *BioMed research international*, vol. 2013, 2013.

[134] G. J. Dickson, A. Kwasniewska, K. I. Mills, T. R. Lappin, and A. Thompson, "Hoxa6 potentiates short-term hemopoietic cell proliferation and extended self-renewal," *Experimental hematology*, vol. 37, no. 3, pp. 322–333, 2009.

[135] L. Wu, C. Zhang, and J. Zhang, "Hmbox1 negatively regulates nk cell functions by suppressing the nkg2d/dap10 signaling pathway," *Cellular & molecular immunology*, vol. 8, no. 5, p. 433, 2011.

[136] B. U. Schraml, K. Hildner, W. Ise, W.-L. Lee, W. A.-E. Smith, B. Solomon, G. Sahota, J. Sim, R. Mukasa, S. Cemerski, *et al.*, "The ap-1 transcription factor batf controls t h 17 differentiation," *Nature*, vol. 460, no. 7253, p. 405, 2009.

[137] J. Kumar, "The sine oculis homeobox (six) family of transcription factors as regulators of development and disease," *Cellular and Molecular Life Sciences*, vol. 66, no. 4, p. 565, 2009.

[138] T. R. Klesert, D. H. Cho, J. I. Clark, J. Maylie, J. Adelman, L. Snider, E. C. Yuen, P. Soriano, and S. J. Tapscott, "Mice deficient in six5 develop cataracts: implications for myotonic dystrophy," *Nature genetics*, vol. 25, no. 1, p. 105, 2000.

[139] G. Meola and R. Cardani, "Myotonic dystrophies: an update on clinical aspects, genetic, pathology, and molecular pathomechanisms," *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, vol. 1852, no. 4, pp. 594–606, 2015.

[140] R.-m. Zeng, X.-h. Lu, J. Lin, J. Hu, Z.-J. Rong, W.-C. Xu, Z.-W. Liu, and W.-t. Zeng, "Knockdown of foxm1 attenuates inflammatory response in human osteoarthritis chondrocytes," *International immunopharmacology*, vol. 68, pp. 74–80, 2019.

[141] A. L. Vallochi, L. Teixeira, K. d. S. Oliveira, C. M. Maya-Monteiro, and P. T. Bozza, "Lipid droplet, a key player in host-parasite interactions," *Frontiers in immunology*, vol. 9, p. 1022, 2018.

[142] A. Martínez-Hernández, H. Gutierrez-Malacatt, K. Carrillo-Sánchez, Y. Saldaña-Alvarez, A. Rojas-Ochoa, E. Crespo-Solis, A. Aguayo-González, A. Rosas-López, J. M. Ayala-Sanchez, X. Aquino-Ortega, *et al.*, "Small maf genes variants and chronic myeloid leukemia," *European journal of haematology*, vol. 92, no. 1, pp. 35–41, 2014.

[143] J. Mukiihi, C. Nyirenda, B. Paul, J. Aduwuyi, E. Mzula, and H. Malata, "Chronic myeloid leukaemia in central africans," *East African medical journal*, vol. 80, no. 9, pp. 470–475, 2003.

[144] S. Zhang, G. Zhang, and Y. Y. Wan, "Ski and smad4 are essential for il-21-induced th17 differentiation," *Molecular immunology*, vol. 114, pp. 260–268, 2019.

[145] J. Rhee, S.-H. Park, S.-K. Kim, J.-H. Kim, C.-W. Ha, C.-H. Chun, and J.-S. Chun, "Inhibition of batf/jun transcriptional activity protects against osteoarthritic cartilage destruction," *Annals of the rheumatic diseases*, vol. 76, no. 2, pp. 427–434, 2017.

[146] S. F. Grant, G. Thorleifsson, I. Reynisdottir, R. Benediktsson, A. Manolescu, J. Sainz, A. Helgason, H. Stefansson, V. Emilsson, A. Helgadottir, *et al.*, "Variant of transcription factor 7-like 2 (tcf7l2) gene confers risk of type 2 diabetes," *Nature genetics*, vol. 38, no. 3, p. 320, 2006.

[147] V. Ledent, O. Paquet, and M. Vervoort, "Phylogenetic analysis of the human basic helix-loop-helix proteins," *Genome biology*, vol. 3, no. 6, pp. research0030–1, 2002.

[148] L. de Pontual, Y. Mathieu, C. Golzio, M. Rio, V. Malan, N. Boddaert, C. Soufflet, C. Picard, A. Durandy, A. Dobbie, *et al.*, "Mutational, functional, and expression studies of the tcf4 gene in pitt-hopkins syndrome," *Human mutation*, vol. 30, no. 4, pp. 669–676, 2009.

[149] D. Hrckulak, L. Janeckova, L. Lanikova, V. Kriz, M. Horazna, O. Babosova, M. Vojtechova, K. Galuskova, E. Sloncova, and V. Korinek, "Wnt effector tcf4 is dispensable for wnt signaling in human cancer cells," *Genes*, vol. 9, no. 9, p. 439, 2018.

[150] M. Schoof, M. Hellwig, L. Harrison, D. Holdhof, M. C. Lauffer, J. Niesen, S. Virdi, D. Indenbirken, and U. Schüller, "The basic helix-loop-helix transcription factor tcf4 impacts brain architecture as well as neuronal morphology and differentiation," *European Journal of Neuroscience*, 2020.

[151] M. Gilliet, W. Cao, and Y.-J. Liu, "Plasmacytoid dendritic cells: sensing nucleic acids in viral infection and autoimmune diseases," *Nature Reviews Immunology*, vol. 8, no. 8, p. 594, 2008.

[152] B. Reizis, A. Bunin, H. S. Ghosh, K. L. Lewis, and V. Sisirak, "Plasmacytoid dendritic cells: recent progress and open questions," *Annual review of immunology*, vol. 29, pp. 163–183, 2011.

[153] V. Sisirak, D. Ganguly, K. L. Lewis, C. Couillault, L. Tanaka, S. Bolland, V. D'Agati, K. B. Elkon, and B. Reizis, "Genetic evidence for the role of plasmacytoid dendritic cells in systemic lupus erythematosus," *Journal of Experimental Medicine*, vol. 211, no. 10, pp. 1969–1976, 2014.

[154] S. M. Han, H. Y. Na, O. Ham, W. Choi, M. Sohn, S. H. Ryu, H. In, K.-C. Hwang, and C. G. Park, "Tcf4-targeting mir-124 is differentially expressed amongst dendritic cell subsets," *Immune network*, vol. 16, no. 1, pp. 61–74, 2016.

[155] N. Jain, K. Hartert, S. Tadros, W. Fiskus, O. Havranek, M. C. J. Ma, A. Bouska, T. Heavican, D. Kumar, Q. Deng, *et al.*, "Targetable genetic alterations of tcf4 (e2-2) drive immunoglobulin expression in diffuse large b cell lymphoma," *Science Translational Medicine*, vol. 11, no. 497, p. eaav5599, 2019.

[156] K. Takano, M. Lyons, C. Moyes, J. Jones, and C. Schwartz, "Two percent of patients suspected of having angelman syndrome have tcf4 mutations," *Clinical genetics*, vol. 78, no. 3, pp. 282–288, 2010.

[157] A. D. Torshizi, C. Armoskus, H. Zhang, M. P. Forrest, S. Zhang, T. Souaiaia, O. V. Evgrafov, J. A. Knowles, J. Duan, and K. Wang, "Deconvolution of transcriptional networks identifies tcf4 as a master regulator in schizophrenia," *BioRxiv*, p. 133363, 2019.

[158] H. Li, Y. Zhu, Y. M. Morozov, X. Chen, S. C. Page, M. D. Rannals, B. J. Maher, and P. Rakic, "Disruption of tcf4 regulatory networks leads to abnormal cortical development and mental disabilities," *Molecular psychiatry*, p. 1, 2019.

[159] M. Li, G. Santpere, Y. I. Kawasawa, O. V. Evgrafov, F. O. Gulden, S. Pochareddy, S. M. Sunkin, Z. Li, Y. Shin, Y. Zhu, *et al.*, "Integrative functional genomic analysis of human brain development and neuropsychiatric risks," *Science*, vol. 362, no. 6420, p. eaat7615, 2018.

[160] D. Steverding, "The history of african trypanosomiasis," *Parasites & vectors*, vol. 1, no. 1, p. 3, 2008.

Chapter 3

Computational Identification of Master Regulators Influencing Trypanotolerance in Cattle

Abirami Rajavel[1], Armin Otto Schmitt[1,2] and Mehmet Gültas[1,2,*]

[1] Breeding Informatics Group, Department of Animal Sciences, Georg-August University, Margarethe von Wrangell-Weg 7, 37075 Göttingen, Germany
[2] Center for Integrated Breeding Research (CiBreed), Albrecht-Thaer-Weg 3, Georg-August University, 37075 Göttingen, Germany
* Author to whom correspondence should be addressed.

Bibliographic Information

Rajavel, A.; Schmitt, A.O.; Gültas, M. Computational Identification of Master Regulators Influencing Trypanotolerance in Cattle. Int. J. Mol. Sci. 2021, 22(2), 562. https://doi.org/10.3390/ijms22020562.

Author contribution of Abirami Rajavel

The author contributed to the design of the study, and conducted computational analyses. Furthermore, the author interpreted the results and wrote a significant part of the final version of the manuscript.

3.1 Abstract

African Animal Trypanosomiasis (AAT) is transmitted by the tsetse fly which carries pathogenic trypanosomes in its saliva, thus causing debilitating infection to livestock health. As the disease advances, a multistage progression process is observed based on the progressive clinical signs displayed in the host's body. Investigation of genes expressed with regular monotonic patterns (known as Monotonically Expressed Genes (MEGs)) and of their master regulators can provide important clue for the understanding of the molecular mechanisms underlying the AAT disease. For this purpose, we analysed MEGs for three tissues (liver, spleen and lymph node) of two cattle breeds, namely trypanosusceptible Boran and trypanotolerant N'Dama. Our analysis revealed cattle breed-specific master regulators which are highly related to distinguish the genetic programs in both cattle breeds. Especially the master regulators *MYC* and *DBP* found in this study, seem to influence the immune responses strongly, thereby susceptibility and trypanotolerance of Boran and N'Dama respectively. Furthermore, our pathway analysis also bolsters the crucial roles of these master regulators. Taken together, our findings provide novel insights into breed-specific master regulators which orchestrate the regulatory cascades influencing the level of trypanotolerance in cattle breeds and thus could be promising drug targets for future therapeutic interventions.

3.2 Introduction

African Animal Trypanosomiasis (AAT), also known as 'nagana', is a parasitic disease of animals caused by the flagellated protozoan species of *Trypanosoma* which is transmitted primarily through the bite of infected tsetse flies [1]. This neglected disease is a threat to animal and human health, especially in sub-Saharan African countries [2, 3, 4]. It affects millions of livestock annually, leading to major economic loss of billions of US dollars every year and substantial decrease in agricultural productivity in Africa [5, 6].
Trypanosoma congolense, *Trypanosoma vivax*, and *Trypanosoma brucei* are the major causes of infection in livestock [7]. Trypanotolerance refers to the capability of the animal to control parasitaemia and anaemia and to remain productive despite the infection of the parasite [8, 9, 10]. It has become an important trait in the recent decade and this trait is widely found in some *Bos taurus* cattle breeds including N'Dama and West African shorthorn breeds [9, 11, 12, 13]. Although the aforementioned cattle breeds remain productive during the course of the disease, they are not desirable for farming due to their smaller size and lower draft power. On the other hand, Zebu (*Bos*

indicus) cattle are not particularly resistant to trypanosomiasis, unlike the native *Bos taurus* cattle. Farmers depend on the Zebu cattle breeds like Boran as they are more suitable for agricultural purposes due to their high draught power and agricultural productivity. However, susceptibility of these breeds to trypanosomiasis poses heavy economic constraints to the farmers [9, 10, 14, 15, 16, 17, 18, 19].

Until now, several studies have been performed in cattle by analysing genotype or gene expression data in order to understand the molecular mechanism underlying the genetic resistance to African trypanosomiasis [20, 21, 22, 23, 24, 25]. Hanotte *et al.* [24] identified quantitative trait loci (QTL) controlling trypanotolerance in a cross of tolerant West African N'Dama and susceptible East African Boran cattle. Moreover, Mekonnen *et al.* [20] surveyed the genome of the cattle breed Sheko to study the genotype-phentotype associations and identified genomic regions associated with trypanosomiasis. On the other hand, O'Gorman *et al.* [25] identified temporal changes in peripheral blood mononuclear cell (PBMC) gene expression in trypanotolerant N'Dama and trypanosusceptible Boran, by studying transcriptomic profiles during the disease progression. To this end, Noyes *et al.* [26] performed gene expression analysis and identified several candidate genes in pathways which responded to trypanosome infection in Boran and N'Dama.

Recently, by analysing the gene expression data set generated by Noyes *et al.* [26], we have deciphered the cattle breed-specific partner choice of transcription factors (TFs) during the disease progression [27]. For this purpose, we mainly considered the Monotonically Expressed Genes (MEGs) to capture the multistage progression process of the AAT disease in liver, spleen and lymph node tissues. Importantly, we highlighted the pivotal relevance of the preferential partner choice of the TF albumin D-site-Binding Protein (*DBP*) in these tissues.

It has been widely shown that the transcriptional regulation of *DBP* controls the circadian output/behaviour not only in the suprachiasmatic nucleus SCN [28], but also in the peripheral tissues [29, 30]. Taking this aspect into account, we addressed the association of *DBP* with circadian transcriptional regulation of tissue-specific processes in mammalian species, in the context of AAT. Especially our findings implicating the functional relationship of circadian control with the immune system are well substantiated by Solis *et al.* [31], Barik *et al.* [32] and Scheiermann *et al.* [33]. In reference to the study by Frank Hawking [34], in which he experimentally established the link between circadian rhythm and *T. congolense* infection in the blood of laboratory rodents, our findings also support the important role of the circadian rhythm in the AAT disease.

Exploring tissue-specific regulatory mechanisms is of utmost importance, especially in tissues such as liver, spleen and lymph nodes, which are likely the primary sites where anaemia occurs as extra-vascular haemolysis [12, 35, 36]. Therefore, similar to our previous studies [37, 38], taking advantage of the systems biology approaches, we attempted in this study to gain novel insights by unravelling the tissue- and breed-specific master regulators and over-represented signalling pathways that responded to trypanosome infection in Boran and N'Dama. Surprisingly, our results show *DBP* as a master regulator for liver tissue of N'Dama, emphasizing the role of the circadian rhythm in the hepatic metabolism and in the immune responses after trypanosome infection in this resistant breed. Altogether, our results highlight a striking feature of the circadian clock in trypanotolerance, especially in the regulatory role of *DBP* in the immunity of trypanotolerant cattle, which confirms our previous finding about the relevance of the clock-controlled gene *DBP* to AAT.

3.2.1 Master Regulators as Drug Targets

Recently omics technologies and computational approaches have become intriguing tools and approaches for drug discovery, making use of gene expression data. Exploiting the systems biology approaches in several studies [39, 40, 41, 42], master regulators have been reported as potential therapeutic targets. By definition, master regulators are molecules that are located at the top of the hierarchy involved in the transcriptomic regulation, where the nodes tend to converge after certain upstream steps [43]. In biological processes, master regulators specifically regulate the expression of downstream genes either directly or through cascades thereby leading to altered phenotypes. In cellular context, dendritic cells involved in antigen-specific responses are regarded as the master regulators which serve as a major link between the innate and the adaptive immune system [44]. A recent study from Vargas *et al.* on Alzheimer's disease has proposed several therapeutic molecular targets for drug development based on master regulator analysis [45]. Similar analyses revealed potential drug targets experimentally for anaplastic thyroid carcinoma in which few transcription factors were proposed as master regulators [46]. Few aforementioned examples demonstrate that transcription regulatory networks and master regulators could be promising drug candidates analysed for investigating complex diseases (including Alzheimer's disease and cancer) as they could be crucial drivers of the molecular mechanism of disease processes.

3.3 Materials and Methods

In this section, we provided an outline of the data set and the methods we used in this study. Figure 3.1 depicts the workflow of this study.

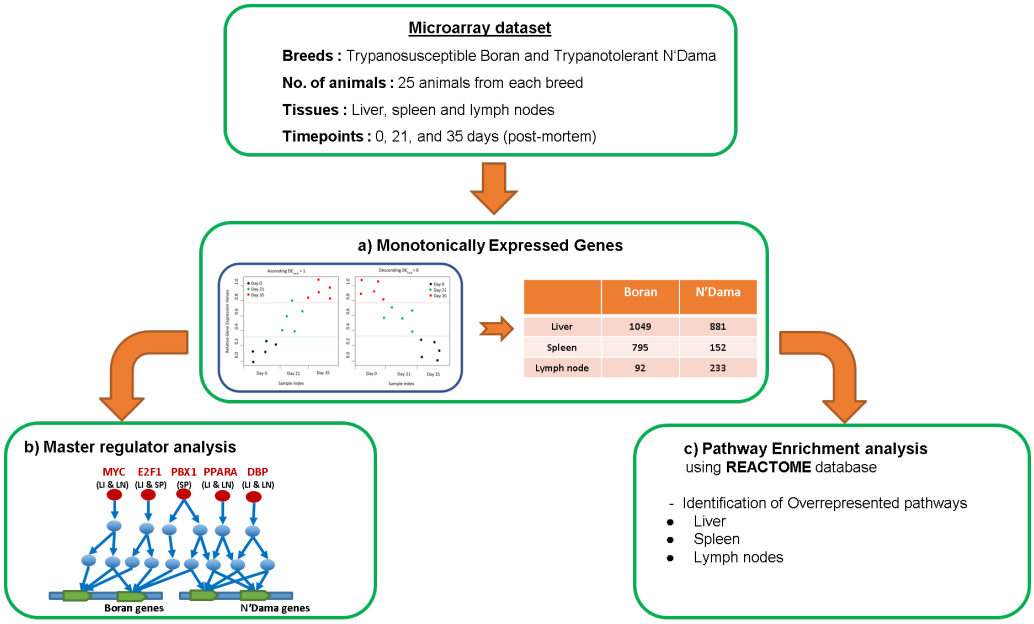

Figure 3.1: Flowchart of the analysis. (**a**) Monotonically Expressed Genes (MEGs) obtained from the analysis of microarray data set comprising gene expression profiles of two cattle breeds: trypanosusceptible Boran and trypanotolerant N'Dama (Boran Asc - MEGs expressed in the ascending pattern for Boran, Boran Des - MEGs expressed in the descending pattern for Boran, N'Dama Asc - MEGs expressed in the ascending pattern for N'Dama and N'Dama Des - MEGs expressed in the descending pattern for N'Dama; (**b**) Master regulator analysis (Red circles and texts in red represent the exemplarily selected master regulators from this study; LI, SP & LN stand for liver, spleen and lymph node tissues); (**c**) Pathway enrichment analysis performed using Reactome database.

3.3.1 Gene sets

In order to identify master regulators and over-represented pathways related to the genetic programs underlying AAT, we analysed six gene sets that exhibit regular monotonic expression patterns in liver-, spleen-, and lymph node tissues of the two cattle breeds, namely trypanosusceptible Boran and

trypanotolerant N'Dama, after being infected with *Trypanosoma congolense*. For this purpose, we took the gene sets from our previous study [27] in which we identified the genes based on the publicly available continuous transcription profiling time-series microarray data set (htttp://www.ebi.uk/arrayexpress/, accession no.E-MEXP-1778) [26]. In this study, we will mainly focus on the analysis of the gene sets. A brief summary about the microarray data set and the number of monotonically expressed genes (MEGs) is given below.

3.3.2 Microarray data set

In this section, we recapitulate the experimental procedure performed by Noyes *et al.* [26]. They performed a microarray experiment based on the cattle breeds Boran and N'Dama as per the following: In the animal experiment, 25 healthy trypanosomiasis-free animals from each breed (trypanosusceptible Boran and trypanotolerant N'Dama) were infected with *Trypanosoma congolense* IL1180 clone. To ensure the health of trypanosomiasis-free animals before experimental infection, the cattle were selected from herds in a tsetse fly-free and trypanosomiasis-free zone of the ILRI Kapiti Plains ranch and assessed negative for tick-borne parasites before transferring them to the ILRI research facility at Kabete. All procedures for handling the animals were performed according to the International Livestock Research Institute (ILRI) ethical review process. Liver, spleen and lymph node tissues were harvested from the cattle on day 0, day 21 and day 35. For control experiments, five animals from each breed were killed before infection and the gene expression readings were recorded for day 0. Tissue harvest was performed after infection. After infection of the cattle, the tissues were collected on day 21 and day 35 post-mortem. Additionally, needle biopsy method was applied only for the liver tissue sampling on day 0 (before infection), day 12, day 15, day 18, day 26, day 29, day 32 after infection. For each condition, extraction of RNA from tissues was done and hybridisation were performed on individual arrays.

3.3.3 Monotonically Expressed Genes

In our recent study [27], we identified the MEGs for each tissue of both cattle breeds by applying the monotonic feature selector (MFSelector) approach [47] to the microarray data set. The lists of MEGs for each tissue are provided in the Supplementary Table S1 and the numbers of MEGs are given in Table 3.1.

Table 3.1: Number of MEGs in ascending and descending order for liver-, spleen- and lymph node tissues for the cattle breeds Boran and N'Dama.

	Boran		**N'Dama**	
	Ascending	Descending	Ascending	Descending
Liver	741	308	757	124
Spleen	669	126	13	139
Lymph node	87	5	119	114

3.3.4 Finding Master Regulators and Over Represented Pathways

Similar to our previous studies [20, 37, 38], we applied well established systems biology approaches using the geneXplain platform [48] in order to identify master regulators and over-represented pathways.
For this purpose, we first run the "upstream analysis" workflow developed by [49] with the maximum radius of 10 steps upstream using the Reactome database [50]. The "upstream analysis" algorithm constructs a global signal transduction network and then identifies the master regulators based on the convergence points of these networks. In general, master regulators are located at the top of a regulatory hierarchy and control the downstream genes without their regulatory influence in signalling pathways [51].
Afterwards, we identified the over-represented pathways in order to unravel the functional properties of the MEGs. The knowledge about the over-represented pathways from Reactome database [50] provides mechanistic insight into the MEGs and helps to decipher novel biological functions underlying the AAT disease mechanisms.

3.4 Results

Mainly focusing on the regular monotonic changes of gene expression profiles in liver-, spleen-, and lymph node tissues during the AAT disease progression, we analyzed in this study for each tissue the related MEG set and identified master regulators as well as over-represented pathways.

3.4.1 Master Regulator Analysis

The "upstream analysis" workflow [48] has been employed using the MEG sets of the tissues in order to computationally identify a variety of master

regulators. As a result, we obtained altogether 10 unique master regulators for both breeds across all tissues as shown in Table 3.2. Remarkably, the vast majority of the master regulators are highly distinct between trypanosusceptible Boran and trypanotolerant N'Dama breeds, only *PBX1* is found common for the spleen tissue of both breeds.

Table 3.2: Master regulators of the breeds Boran and N'Dama.

	Boran	**N'Dama**
Liver	*MYC, E2F1, PPARG*	*DBP, PBX1, HOXA4, PPARA*
Spleen	*PITX2, E2F1, PBX1*	*PBX1*
Lymph node	*MYC, pSTAT1, PBX1*	*DBP, PPARA*

3.4.1.1 Master regulators in liver

Using the "upstream analysis" workflow, we identified three master regulators (*MYC, E2F1, PPARG*) for the liver tissue of Boran and four master regulators (*DBP, PBX1, HOXA4, PPARA*) for N'Dama.
MYC is a member of the basic helix-loop-helix (bHLH) transcription factor family. It regulates a wide range of biological processes including metabolism, apoptosis, cell cycle, cell growth, angiogenesis or reprogramming in several tissues [52, 53]. Importantly, *MYC* is highly pleiotropic [54] indicating that its deregulation is in close connection with all hematological malignancies, especially anaemia [55, 56] which is a prominent feature of the AAT disease and also with drug resistance [56, 57, 58, 59]. Furthermore, it has been reported as regulator of large networks of genes and has been associated with several cancer types, and is thus serving as a potential drug target [52, 60, 61, 62]. Furthermore, it has been reported in the host-parasite interaction, improving the survival rate of parasites in surpassing immune surveillance mechanisms [63, 64]. With regard to the parasite's survival, MYC could be playing pivotal roles in induction and manipulation of host cell's immunity in Animal African Trypanosomiasis as well. The master regulator *E2F1* found for Boran liver, plays a critical role in bile acid synthesis as per a study performed in mouse model [65]. *E2F1* inhibits the clearance of circulating cholesterol by regulating the expression of *PCSK9* [66], which might be related to the parasite's critical need of cholesterol-related metabolism from Boran's body, implicating the progressive conditions of hypocholesteraemia and hypolipidaemia after infection [67]. Peroxisome proliferator-activated receptor gamma (*PPARG*), found as the third master regulator for Boran liver,

belongs to the nuclear hormone receptor super family [68, 69]. In liver, induction of *PPARG* overexpression as a result of pathophysiological stress, has led to lipid accumulation. Interestingly, blocking of *PPARG* gene expression has reduced the accumulation of lipids and the expression of inflammatory genes [70, 71]. Therefore, *PPARG* is strongly associated not only with the lipid metabolism but also with inflammatory processes [72]. This suggests the role of *PPARG* in the induction of lipid metabolism by *T. congolense* to utilise a high amount of energy from Boran, resulting in weight loss and loss of body conditions in the cattle during the AAT disease.
Albumin D site-binding protein (*DBP*), found for the liver tissue of N'Dama, is a liver-enriched transcription factor [73] and plays important roles in circadian rhythm in the mammals [30, 74]. Specifically, it influences the circadian transcriptional regulation of several liver-specific genes namely P450 genes such as *Cyp2a4* and *Cyp2a5* [75, 76]. Belonging to the PAR bZIP basic leucine zipper family, *DBP* accumulates following a stringent circadian rhythm in liver cells [30]. The circadian control of the liver gene *CYP7* encoding the cholesterol 7α-hydroxylase enzyme, which catalyses the metabolism of cholesterol to bile acids [29, 77, 78], establishes the strong indispensable association of the circadian rhythm and *DBP* in cholesterol homeostasis. Remarkably, the expression of *DBP* was found upregulated in a tolerant mouse model after *T. congolense* infection, suggesting the strong link of *DBP* and trypanotolerance [79]. Our findings are further well-supported by the results in [34], in which he established the association of the circadian rhythm with the infection of *T. congolense* in rodents. Strikingly, the master regulator *PPARA* (Peroxisome proliferator-activated receptor α) is a ligand-induced nuclear receptor that is highly expressed in the liver of mammals [80, 81]. *PPARA* is well-known for its transcriptional regulatory role in metabolic and inflammatory pathways, making it a potential therapeutic target [82, 83, 84]. Particularly, it plays a crucial role in several metabolic processes, including bile and amino acid metabolism, transportation and metabolism of lipids, fatty acid beta-oxidation, ketogenesis and lipogenesis [81, 85, 86], which could contribute to the protection of the host from worsening conditions of AAT like weight loss and hypolipidemia.
Furthermore, *PBX1* (Pre-B-cell leukemia homeobox-1) is necessary for the maintenance of definitive hematopoiesis in the fetal liver, which indicates the host-protective role of *PBX1* from anaemia [87] and thus contributing to trypanotolerance of N'Dama. Another master regulator *HOXA4* from the homeobox family, is known for its role in hematopoiesis and B-cell progenitor population expansion [88], which implicates its importance in the production and maintenance of blood cells and immune cells, thus helping the cattle to control the major complications of AAT such as anaemia or parasitaemia.

3.4.1.2 Master regulators in spleen

The analysis of MEGs for the spleen tissues of the cattle breeds reveals that *PBX1* is a common master regulator between Boran and N'Dama. It is essential for the spleen tissue-specific function of hematopoiesis [89]. Another key regulator, *E2F1*, found for the spleen tissue of Boran, has been reported as a suppressor of dendritic cell maturation [90], therefore implicating its role as a transcription factor for the immunosuppression in the infected cattle Boran. Macrophages and dendritic cells play a significant role in the innate immune system. In particular, they are involved in the production of interferon γ (IFN-γ), which is important for resistance against *T. congolense* [91]. Inhibition of dendritic cell maturation inhibits IFN-γ secretion [92], thus resulting in the reduction of the immune response against *T. congolense* [93]. The regulator *E2F1* in Boran spleen may have a leading role in immune depression of Boran, thus contributing to the susceptibility of Boran to AAT. The remaining master regulator *PITX2*, found for the Boran spleen tissue, is a member of the bicoid transcription factor family, which is involved in a wide variety of developmental processes [94]. However, the reason for its importance and potential role with respect to host-pathogen interaction is still unclear.

3.4.1.3 Master regulators in lymph node

Investigation of the MEG sets of lymph node tissue unravelled three master regulators (*MYC*, p*STAT1*, and *PBX1*) for Boran and two master regulators (*DBP* and *PPARA*) for N'Dama. *MYC* plays an essential role in immune suppression and immune evasion mechanisms in assisting cancer cells to avoid the host's immunity, as suggested in cancer studies [95, 96]. It might play a role in helping the trypanosomes to escape the immune check points in host immune surveillance mechanisms, suggesting a major player in parasitaemia in Boran's body which is one of the major characteristics of the AAT disease. Strikingly, we identified p*STAT1* (signal transducer and activator of transcription 1) as the second master regulator. The role of STAT1 is strongly associated with the development of Th1 and Th17 responses which are CD4+ T-cell subsets [97, 98, 99]. This mainly implicates overproduction of pro-inflammatory cytokines (like IL-17) leading to cell death and inflammation, which connects the severity of anaemia in Boran [100, 101]. The master regulator *PBX1* has been studied in the homeostatic developmental programming of natural killer (NK) cells [102], which contributes to the main symptom of trypanosomiasis-associated acute anaemia as reported by [103].

DBP, identified as the master regulator in the lymph node of N'Dama, is a clock-controlled transcription factor and an important regulatory component of the circadian clock to ensure the 24 hour rhythm in mammalian species [104, 105]. Several studies have reported the rhythmic expression of clock genes in cells of the immune system such as macrophages, dendritic cells and B-cells [106, 107, 108, 109, 110, 111, 112, 113], representing the function of clock genes in immune responses. CD4+ T helper cells play crucial roles in the stimulation of effective antibody response and efficient isotype switching from IgM to IgG production [114, 115, 116], the critical features reported in N'Dama for its AAT tolerance [117]. These cells, being the significant regulators of adaptive immunity, harbor a circadian oscillator and generate cytokines such as IL-2, IL-4 and IFN-γ according to robust rhythms [118], implicating the tight connection of the circadian clock with adaptive immune responses during the AAT disease. Another interesting master regulator found for N'Dama is *PPARA* which has been reported to be expressed in B- and T-cells of the immune system [119, 120]. Importantly, *PPARA* is known as the crucial regulator of immune responses such as inflammation and cytokine production [121, 122, 123]. According to its biological function in the immune system, *PPARA* could be protecting the depletion of the host's cells from its own immune system during parasitic manipulation of immune responses, possibly controlling anaemia in N'Dama after trypanosome infection.

3.4.2 Pathway Analyses

To further decipher the specific biological functions of MEGs regarding AAT disease mechanisms, we investigated the over-represented pathways using the Reactome pathway database [50] for the three tissues of Boran and N'Dama. All the over-represented pathways obtained from the analysis are listed in the Supplementary Table S2 and the pathways unique for each tissue of the two breeds are shown in Tables 3.3–3.5. Mainly focusing on these pathways, we found that several of them are overlapping between the breeds (see Figure 3.2a–c). Interestingly, Figure 3.2b shows that there is only one pathway unique for the spleen tissue of Boran. Taking the liver tissue into account (Figure 3.2a), there are only 10 unique pathways obtained for each breed despite the big overlap. On the other hand, remarkably high numbers of pathways are found for the lymph node tissue of Boran, in comparison to N'Dama (Figure 3.2c).

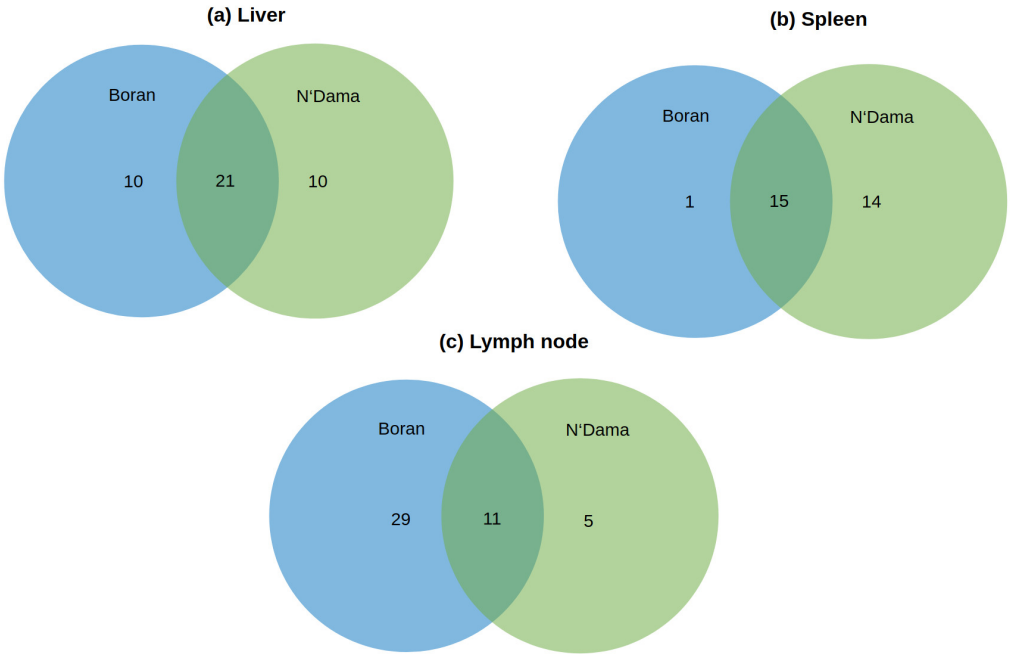

Figure 3.2: Venn diagram of over-represented pathways (p-adjusted < 0.05) obtained for liver-, spleen- and lymph node tissues of the two cattle breeds Boran and N'Dama after the infection of *Trypanosoma congolense*. Pathway enrichment analysis was performed based on the Reactome pathway database [50].

3.4.2.1 Over-represented pathways found for liver tissue

Analysis of over-represented pathways for the liver tissue of Boran and N'Dama uncovered ten pathways unique for each breed (see Table 3.3 and Figure 3.2a).
In Table 3.3, the top three over-represented pathways are associated with low oxygen environment and HIF accumulation due to parasitic infection [124, 125] in the liver tissue of the trypanosusceptible breed Boran. Furthermore, the TGF-β signaling pathway, which includes the TGF-β receptor complex, TGF-β family members and SMAD2/SMAD3:SMAD4 heterotrimer (shown in Figure 3.3), is reported as a critically important pathway for the parasite in mammalian cell invasion and to escape the host's immune system [126, 127]. Especially, this pathway is involved in the suppression of macrophages that are essential players against parasites [128, 129], implicating immunosuppression [130, 131] during AAT in Boran. Moreover, our findings also lend support to experimental studies on TGF-β on other species of *Trypanosoma* [132, 133]. Importantly, the interleukin-1 family signaling

Table 3.3: Significantly over-represented pathways found for the liver tissue of Boran and N'Dama (p-adjusted < 0.05).

Liver		
Pathway name	**Hit names**	**Adjusted p-value**
Boran		
Cellular responses to external stimuli	Arnt, Fos, Hif1a, Hsf1	4.64×10^{-4}
Regulation of beta-cell development	Foxo1, Hnf4g, Nkx2.2	0.0032
Regulation of Hypoxia-inducible Factor (HIF) by oxygen	Arnt, Hif1a	0.0033
Cellular response to hypoxia	Arnt, Hif1a	0.0033
Signaling by TGF-beta Receptor Complex	Myc, Smad3, Smad4	0.0036
Signaling by TGF-beta family members	Myc, Smad3, Smad4	0.0063
Signaling by NOTCH1	Hif1a, Myc	0.0105
Transcriptional activity of SMAD2/SMAD3:SMAD4 heterotrimer	Myc, Smad4	0.0147
Signaling by NOTCH	Hif1a, Myc	0.0349
Interleukin-1 family signaling	Nfkb1, Smad3	0.0371
Cellular Senescence	Fos, Jun	0.0416
N'Dama		
PTEN Regulation	Atf, Jun	0.0025
Activation of HOX genes during differentiation	Hoxa4, Jun, Meis1	0.0036
Activation of anterior HOX genes in hindbrain development during early embryogenesis	Hoxa4, Jun, Meis1	0.0036
BMAL1:CLOCK, NPAS2 activates circadian gene expression	Dbp, Ppara	0.0161
PIP3 activates AKT signaling	Atf, Jun	0.0229
Transcriptional regulation of pluripotent stem cells	Pbx1, Pou5f1	0.0229
Intracellular signaling by second messengers	Atf2, Jun	0.0280
Transcriptional regulation by RUNX2	Sox9, Stat1	0.0364
Transcriptional regulation of white adipocyte differentiation	Pparg, Rxra	0.0424
Circadian Clock	Dbp, Ppara	0.0488

pathway found for Boran has been reported in direct association with damaging inflammation [134], which could explain the development of anaemia in Boran.
Regarding N'Dama's liver tissue, the circadian clock and its components are inextricably in association with the transcriptional regulation of liver functions [135, 136, 137, 138], suggesting the integrity of the circadian rhythm in the trypanotolerance mechanisms of N'Dama as shown in Table 3.3 and Figure 3.4. This finding is consistent with our previous results [27] and also a recent study by Solis *et al.* [31]. Intriguingly, the second and third over-represented pathways found for N'Dama are emphasizing the activation of HOX genes, which have been implicated as master regulator genes in the process of haematopoiesis [139]. Haematopoietic cells, derived mainly from fetal liver and bone marrow, are important in self-renewal and long-term supply of blood cells, especially T cells, which play a crucial role in the immune system [139, 140, 141, 142]. These pathways strongly suggest their potential roles in protection of N'Dama from anaemia which is the most prominent feature of AAT disease.

3.4.2.2 Over-represented pathways found for spleen tissue

For the spleen tissue, I identified one unique significantly over-represented pathway for Boran and 14 significantly over-represented pathways for N'Dama (see Table 3.4 and Figure 3.2b).

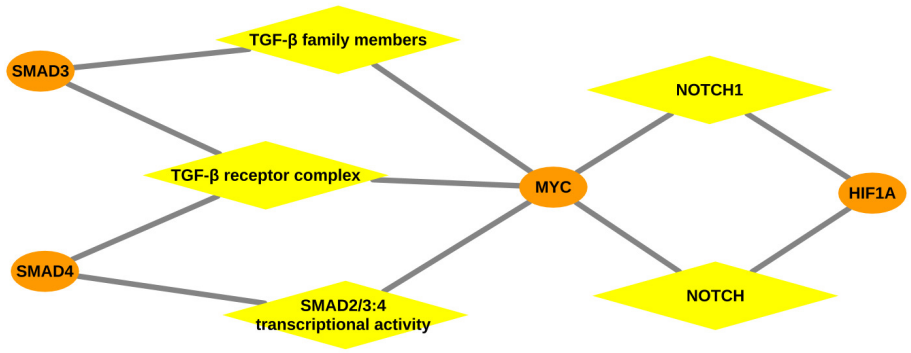

Figure 3.3: Gene-pathway network: Network representing the association between the over-represented pathways and transcription factors especially *MYC*, depicted for the liver tissue of Boran. Orange coloured ellipses in the network represent the transcription factor genes and yellow coloured rhombuses indicate the over-represented pathways. The transcription factors are connected to the pathways, thereby forming a network of interactions between them.

The over-represented pathway found for Boran's spleen tissue is related to the activation of genes during the proliferation by transcription factors *POU5F1*, *SOX2*, and *NANOG*. This pathway is also associated with the downstream processes related to self-renewal and pleuripotency in embryonic stem cells [143, 144] which appears to be the normal function of spleen tissue in mammals.

Circadian clock related pathways in Table 3.4 can be categorised as the most prominent pathways in the spleen tissue due to the circadian control of splenic macrophages and B-cell development [107, 145] (see Figure 3.4). In particular, these pathways are essential in the context of the trypanotolerance of N'Dama, since circadian regulation of immune responses are controlled at various levels in mammals [31, 73]. Fundamentally, as reported in previous studies [22, 25, 26, 146], MAPK family signalling cascades have been demonstrated to play critical roles in immune response through the production of pro-inflammatory cytokines in macrophages, suggesting their role in host defense against *T. congolense* in N'Dama [22, 25, 26, 146]. Another striking pathway (in Table 3.4 and Figure 3.4) is related to the regulation of lipid metabolism by *PPARA*, which strongly establishes the relationship of spleen tissue with cholesterol metabolism in trypanotolerant N'Dama. Splenomegaly and hypocholesteraemia [67], being clinical signs of AAT, are indications of high workload of the reticuloendothelial system [147, 148] for

Table 3.4: Significantly over-represented pathways found for the spleen tissue of Boran and N'Dama (p-adjusted < 0.05).

Spleen		
Pathway name	**Hit names**	**Adjusted p-value**
Boran		
POU5F1 (OCT4), SOX2, NANOG activate genes related to proliferation	Pou5f1, Stat3	0.0034
N'Dama		
Oxidative Stress Induced Senescence	Fos, Jun	0.0033
BMAL1:CLOCKNPAS2 activates circadian gene expression	Dbp, Ppara	0.0052
MAPK6/MAPK4 signaling	Foxo1, Jun	0.0052
Signaling by NOTCH3	Hes1, Pbx1	0.0052
Cellular responses to stress	Fos, Hsf1, Jun	0.0062
Fc epsilon receptor (FCERI) signaling	Fos, Jun	0.0067
Cellular responses to external stimuli	Fos, Hsf1, Jun	0.0135
MAPK family signaling cascades	Foxo1, Jun	0.0166
Circadian Clock	Dbp, Ppara	0.0166
Signaling by NOTCH	Hes1, Pbx1	0.0203
Generic Transcription Pathway	E2f1, Hes1, Sox9, Stat1, Tead1	0.0243
Cellular Senescence	Fos, Jun	0.0243
RNA Polymerase II Transcription	E2f1, Hes1, Sox9, Stat1, Tead1	0.0350
Regulation of lipid metabolism by Peroxisome proliferator-activated receptor alpha (PPARalpha)	PPara, Rxra	0.0432

the parasite clearance in the blood. This pathway could be controlling the AAT conditions of splenomegaly and hypocholesteraemia in trypanotolerant N'Dama.

3.4.2.3 Over-represented pathways found for lymph node tissue

I identified 29 and 5 unique significantly over-represented pathways for lymph node tissue of Boran and N'Dama, respectively(see Table 3.5 and Figure 3.2c).
Inspection of Table 3.5 shows that I obtained a list of immune-related pathways for the lymph node tissue of Boran, suggesting the activation of immune cells in response to trypanosome infection. The toll-like receptor- and MyD88-related signalling pathways, being the major pathways for pathogen recognition in the innate immune system, have been reported as activated after protozoan infection [149]. Furthermore, activation of MAPK family signalling cascades, cytokine signalling, Fc epsilon RI signalling, interleukin-17 (IL-17) and interleukin-1 (IL-1) signalling strongly suggest that the inflammatory responses are following the cytokine production in response to *T. congolense* infection in Boran, as reported in previous studies [22, 26, 150]. Conversely, it seems quite possible that few pathways are under the manipulation of the parasite leading to the hyperproduction of proinflammatory cytokines resulting in catastrophic inflammation of host cells in Boran, especially involving *MYC* (see Figure 3.5). Particularly CD4+ T cells, which secrete IL-17, are reported in autoimmunity wherein IL-17 development is promoted by cytokines IL-1 and TGF-β [151]. The autoimmune phenomena of Boran result in the chronic destruction of own cells mainly leading to se-

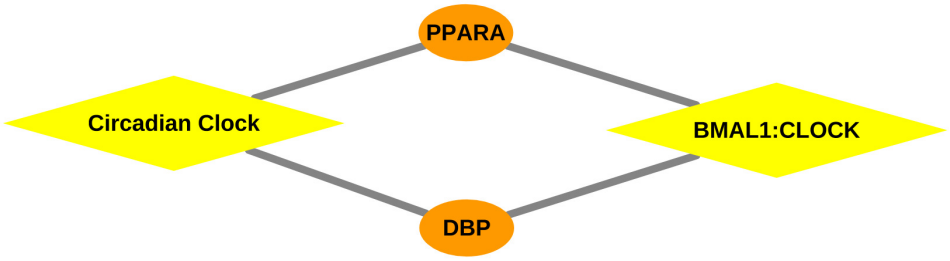

Figure 3.4: Gene-pathway network: Network representing the association between the over-represented pathways and transcription factors, especially *DBP*, depicted for the liver, spleen and lymph node tissue of N'Dama. Orange coloured ellipses in the network represent the transcription factor genes and yellow coloured rhombuses indicate the over-represented pathways. The transcription factors are connected to the pathways, thereby forming a network of interactions between them.

Table 3.5: Significantly over-represented pathways found for the lymph node tissue of Boran and N'Dama (p-adjusted < 0.05).

Lymph node		
Pathway name	**Hit names**	**Adjusted p-value**
Boran		
MAP kinase activation in TLR cascade	Atf1, Atf2, Fos, Jun, Nfkb1	$2.37*10^{-8}$
Interleukin-17 signaling	Atf1, Atf2, Fos, Jun, Nfkb1	1.83×10^{-7}
MAPK targets/ Nuclear events mediated by MAP kinases	Atf1, Atf2, Fos, Jun	3.66×10^{-7}
MyD88 cascade initiated on plasma membrane	Atf1, Atf2, Fos, Jun, Nfkb1	5.54×10^{-7}
MyD88 dependent cascade initiated on endosome	Atf1, Atf2, Fos, Jun, Nfkb1	7.62×10^{-7}
MyD88:Mal cascade initiated on plasma membrane	Atf1, Atf2, Fos, Jun, Nfkb1	1.03×10^{-6}
MyD88-independent TLR4 cascade	Atf1, Atf2, Fos, Jun, Nfkb1	2.27×10^{-6}
Toll Like Receptor 3 (TLR3) Cascade	Atf1, Atf2, Fos, Jun, Nfkb1	2.88×10^{-6}
Toll-Like Receptors Cascades	Atf1, Atf2, Fos, Jun, Nfkb1	4.39×10^{-5}
MAPK6/MAPK4 signaling	Foxo1, Jun, Myc	4.66×10^{-4}
Innate Immune System	Atf1, Atf2, Fos, Jun, Ltf, Nfkb1	9.27×10^{-4}
Signaling by Interleukins	Atf1, Atf2, Fos, Jun, Nfkb1, Stat1, Stat3	0.0010
PTEN Regulation	Atf2, Jun	0.0016
MAPK family signaling cascades	Foxo1, Jun, Myc	0.0028
Oxidative Stress Induced Senescence	Fos, Jun	0.0069
Cytokine Signaling in Immune system	Atf1, Atf2, Fos, Jun, Nfkb1, Stat1, Stat3	0.0081
Fc epsilon receptor (FCERI) signaling	Fos, Jun	0.0138
PIP3 activates AKT signaling	Atf2, Jun	0.0154
Transcriptional activity of SMAD2/SMAD3:SMAD4 heterotrimer	Myc, Smad4	0.0171
NGF signalling via TRKA from the plasma membrane	Atf1, Stat3	0.0171
Intracellular signaling by second messengers	Atf2, Jun	0.0189
Immune System	Atf1, Atf2, Fos, Jun, Ltf, Nfkb1, Relb, Stat1, Stat3	0.0241
Transcriptional regulation by the AP-2 (TFAP2) family of transcription factors	Mybl2, Myc	0.0311
Generic Transcription Pathway	E2f1, Mybl2, Myc, Smad4, Sox9, Stat1	0.0338
Mitotic G2-G2/M phases	Foxm1, Mybl2	0.0381
Mitotic G1-G1/S phases	E2f1, Mybl2	0.0381
Interleukin-1 family signaling	Nfkb1, Stat3	0.0430
Signaling by TGF-beta Receptor Complex	Myc, Smad4	0.0456
Cellular Senescence	Fos, Jun	0.0482
N'Dama		
POU5F1 (OCT4), SOX2, NANOG activate genes related to proliferation	Pou5f1, Stat3	0.0042
BMAL1:CLOCK, NPAS2 activates circadian gene expression	Dbp, Ppara	0.0078
Circadian Clock	Dbp, Ppara	0.0244
Factors involved in megakaryocyte development and platelet production	Irf1, Irf2	0.0354
Interleukin-12 family signaling	Stat1, Stat3	0.0375

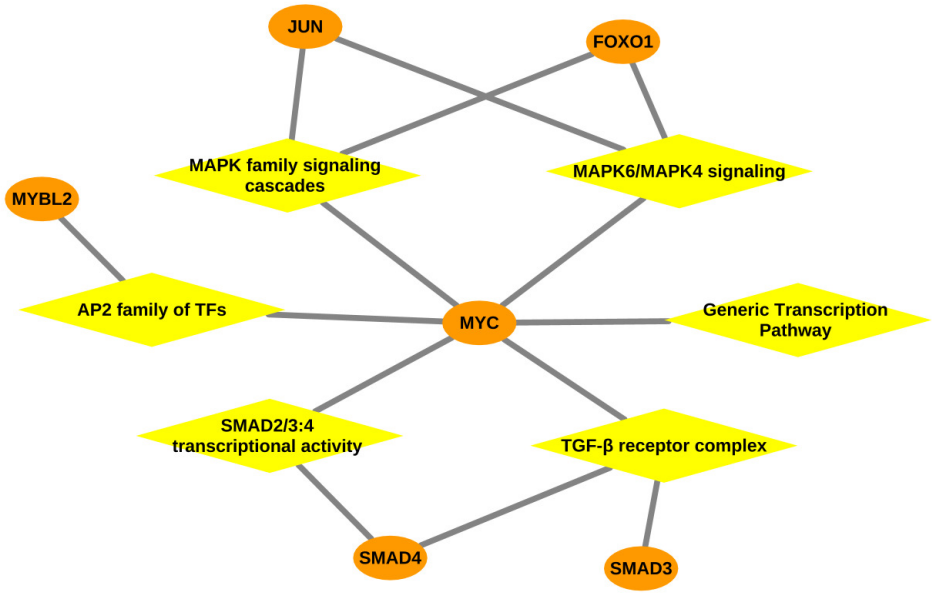

Figure 3.5: Gene-pathway network: Network representing the association between the over-represented pathways and transcription factors especially *MYC*, depicted for the lymph node tissue of Boran. Orange coloured ellipses in the network represent the transcription factor genes and yellow coloured rhombuses indicate the over-represented pathways. The transcription factors are connected to the pathways, thereby forming a network of interactions between them.

vere anaemia. Moreover, the major immune-related MAPK family signalling cascade has been demonstrated as the targeted pathway for manipulation by *T. congolense* in the host to escape the host immune responses [150].

For the lymph node tissue of N'Dama, two circadian clock related pathways have been found over-represented as shown in Table 3.5 and Figure 3.4, implicating the functioning of the immune system intimate accordance with the circadian clock as reported in several studies [31, 33, 152, 153]. Important functional aspects of the immune responses such as phagocytosis, antigen presentation and immune regulation are regulated by the circadian clock [107], suggesting the circadian control of immunity against *T. congolense* infection in trypanotolerant N'Dama. Furthermore, interleukin-12 family signalling harbors IL-12, which is mainly leading to IFN-γ production which has been reported to be responsible for resistance to trypanosomiasis [154, 155]. Remarkably, deletion of IL-12 in *T. b. brucei* and *T. evansi* infection models

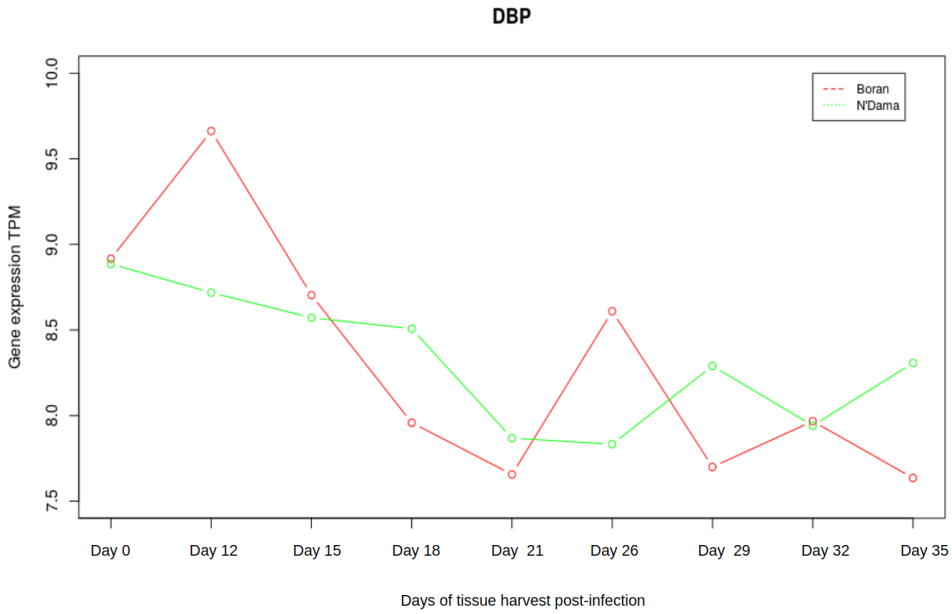

Figure 3.6: Expression of *DBP*. The red line in the plot corresponds to Boran whereas the green line corresponds to the breed N'Dama. In the early timepoints, expression of *DBP* reaches a peak in Boran and then starts to drop down. After the timepoint 5, the curve again increases slightly and at the later stages, it falls down again. On the other hand, the expression of *DBP* in N'Dama steadily drops until the timepoint 5, then it gradually increases.

has resulted in the deficient IFN-γ production required for controlling parasitaemia [154], emphasizing the critical role of interleukin-12 family signalling in establishing resistance to AAT in N'Dama.

3.4.2.4 Analysis of Gene Expression Profiles of Dbp and Myc

Taking the liver tissue into account, the microarray data set consisted of gene expression profiles for nine timepoints namely day 0, days 12, 15, 18, 21, 26, 29, 32, and 35. Emphasizing the crucial roles of the two master regulators *DBP* and *MYC*, I was interested in additionally investigating the changes in gene expression pattern of *DBP* and *MYC* of Boran and N'Dama during the disease progression. Interestingly, I observed slight antagonistic patterns of gene expression at different stages of the AAT disease, although both of these genes are not differentially expressed.
The expression of *DBP* is abruptly declining for Boran from timepoint 2 (see Figure 3.6). Then, *DBP* increases its expression slightly and again declines

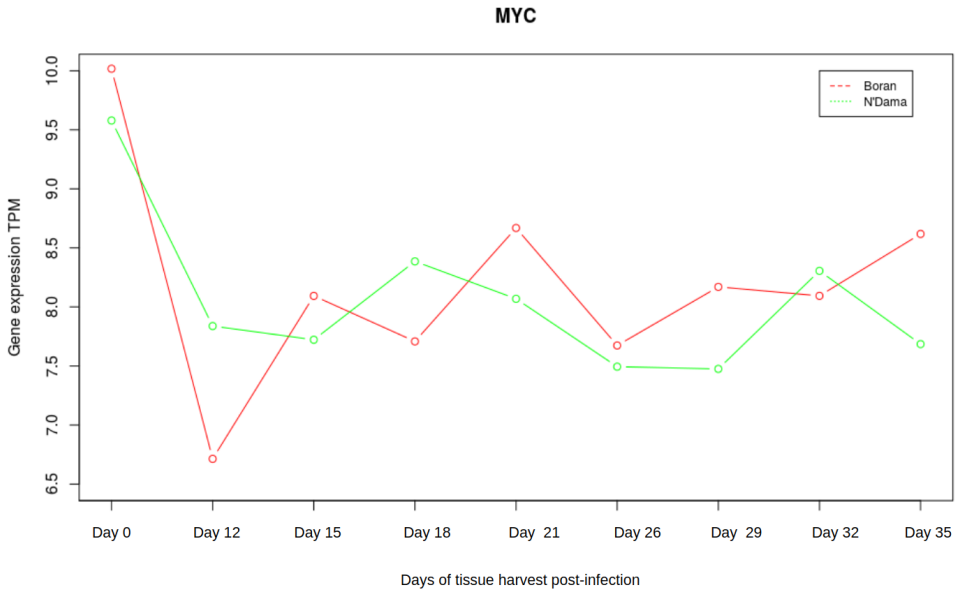

Figure 3.7: Expression of *MYC*. The red line in the plot corresponds to Boran whereas the green line corresponds to the breed N'Dama. In the early timepoints, expression of *MYC* decreases in both Boran and N'Dama. At the later stage, the expression of *MYC* increases for Boran whereas it decreases for N'Dama.

towards the later stages.
In contrast, the expression of *DBP* shows an increasing trend towards the later stages in N'Dama, despite the continuous steady decrease in its expression at earlier timepoints. It might be important to consider the expression of *DBP* during the later stages of the disease progression. AAT could be exhibiting similar effect in both cattle breeds during the earlier timepoints of infection. On the other hand, the expression of *MYC* seems quite clear for Boran with increasing tendency at the later timepoints (as shown in Figure 3.7). Whereas for N'Dama, *MYC* decreases its expression at several timepoints. These minor changes could be greatly contributing to susceptibility or tolerance mechanisms of both cattle breeds.

3.5 Discussion

African Animal Trypanosomiasis (AAT) is a vector-borne disease spread through the tsetse fly by carrying pathogenic African trypanosomes in its saliva. Clinical signs such as anaemia, hepatomegaly and splenomegaly are

displayed in the cattle during the course of the AAT disease, which gradually progresses in multiple steps. Based on its signs, the AAT disease shows a multi-stage progression process in the body of the animal. Previous studies have pointed out that the consideration of monotonic expression patterns of genes (MEGs) could reflect the stage-by-stage progression of the disease [47, 156]. Thus, we analysed several MEG sets in this study to identify master regulators which govern the transcriptional machinery of tissue-specific gene expression and thus influencing trypanosusceptibility and trypanotolerance of the breeds Boran and N'Dama, respectively. For this purpose, the consideration of the three tissues, liver, spleen, and lymph nodes are quintessential since they are the primary target sites of trypanosome infection. Inextricably, these tissues play crucial roles in generating host immune responses especially by increasing the number of macrophages, which results in the production of pro-inflammatory cytokines [9, 146].

Remarkably, our findings suggest that the master regulators *DBP* and *MYC* identified for liver and lymph node tissues, appear to be greatly influencing the genetic programs for trypanosusceptibility and trypanotolerance mechanisms in Boran and N'Dama. Notably, *DBP* could be supporting the regulation of immune responses [27, 31, 73, 152, 157] because of its function in the circadian oscillatory mechanism [104] thereby establishing trypanotolerance in N'Dama. On the other hand, *MYC* has been reported to be responsible for the disruption of the circadian clock in cancer cells [158, 159, 160], elucidating the possibility of a dysfunctional circadian rhythm in Boran. Furthermore, *MYC* has gained its importance as it directly programs inflammation and immune suppression [95], which are constantly reported conditions in trypanosome-infected Boran.

Kupffer cells, the largest immune cell population of macrophages resident in the liver tissue, play a critical role in the mononuclear phagocytic system mounting the first line of immune response to foreign antigens [161]. Immune responses in the liver tissue depend on the activation state of macrophages [161, 162]. M1 and M2 polarization of macrophages are known as two extremes in which M1 (classically activated) is characterised by expression of high pro-inflammatory cytokines and M2 (alternatively activated) by high anti-inflammatory cytokines [162, 163]. Surprisingly, all three master regulators *MYC*, *E2F1* and *PPARG* identified for the liver tissue of Boran in this study, have been reported as the regulators necessary for M2-like polarization of macrophages [164], which could be an advantage for the trypanosomes to increase their survival inside the host's body, and thereby contributing to enhanced parasitaemia in Boran.

In order to gain more mechanistic insights and to discover novel biological functions underlying the AAT disease progression of both breeds, the investi-

gation of over-represented pathways based on the MEG sets of tissues is crucial. Based on pathway analysis, we obtained a number of over-represented pathways (see Supplementary Table S2), several of which are in agreement with the results of previous studies [22, 26, 150] and are activated due to trypanosome infection in both breeds. Remarkably, the majority of these pathways were found to be common for both cattle breeds, while few of them are unique and breed-specific which could provide an important clue for distinguishing the biological processes controlling the mechanisms underlying trypanosusceptibility or trypanotolerance of the cattle breeds. Consequently, we focused in this study on outlining the potential roles of breed-specific unique pathways (see Tables 3.3–3.5) in association with the level of trypanotolerance in the respective cattle breeds. Although we reported the major immune-related pathways in Boran, these pathways could be leading to inflammation due to hyperproduction of pro-inflammatory cytokines in the host cells [165], thereby contributing to susceptibility of this breed. Despite the activation of immune signalling pathways, dysregulation causes the death of infected animals, especially dysregulated cytokine networks and overproduction of inflammatory cytokines (hallmark of African Trypanosomiasis) [146]. Contrarily, circadian clock related pathways (see Figure 3.4), interleukin-12 family signalling, regulation of lipid metabolism, and MAPK family signalling cascades might be properly regulated in N'Dama, indicating the underlying mechanism for trypanotolerance in N'Dama during the AAT disease. Especially, the pathways related to the circadian clock bolster our previous findings [27] in highlighting the important role of *DBP* and circadian rhythm in the coordination of the immune responses in trypanotolerant breed N'Dama.

Today it is well-known that the knowledge about master regulators is fundamental since they greatly control the TFs and the related genes [166, 167]. Further, it is also pivotal to understand the regulatory network of TFs that cooperatively regulate a large number of genes during a disease process [168]. In our previous study [27], by analysing the promoter regions of the MEGs, we reported the importance of several TFs and their preferential partner choices elucidating their roles during the AAT disease progression. The consideration of TFs and their cooperations only provides the information regarding the first level of the transcriptional regulatory hierarchy [168]. However, for gaining a proper understanding of the disease progress in both breeds, it is still necessary to establish the hierarchy of the transcriptomic regulation in order to identify the master regulators [166, 167, 168]. Thus, our main objective in this study was to identify the master regulators together with signal transduction pathways associated with the AAT disease as potential drug targets, to complement our previous study [27]. Our current study provides a com-

bined knowledge along with our previous findings. On one hand, it evidently indicates that *DBP* functions more as a master regulator of the circadian clock in peripheral tissues, supporting the trypanotolerance mechanisms in the cattle breed N'Dama. On the other hand, our analysis remarkably leads to the identification of novel master regulator *MYC* in association with the trypanosusceptibility mechanism of Boran.

Taken together, the systematic investigation of the upstream master regulators and over-represented pathways governing the regulatory mechanisms of the trypanotolerance level of two cattle breeds could provide novel mechanistic insights into the tissue- and breed-specific genetic programs. In particular, we identified *MYC* and *DBP* (as represented in Figures 3.3–3.5) as potential discriminators between the two cattle breeds, trypanosusceptible Boran and trypanotolerant N'Dama, which are likely to be promising therapeutic targets for future works and for the selective breeding of this trait.

Supplementary Information

The supplementary files can be accessed via the original publication (https://www.mdpi.com/1422-0067/22/2/562/s1). Table S1: Lists of Monotonically Expressed Genes obtained from the MFSelector approach, Table S2: Lists of all over-represented pathways obtained for liver-, spleen- and lymph node tissues of Boran and N'Dama. The common pathways between the breeds are highlighted in red.

Author contributions

M.G. designed and supervised the research. A.R. participated in the design of the study and conducted computational analyses together with M.G. and performed the literature survey. A.O.S. interpreted the results with AR and MG. A.R. and M.G. wrote the final version of the manuscript. M.G. conceived of and managed the project. All authors have read and agreed to the published version of the manuscript.

Data availability

Not applicable.

Acknowledgments

We acknowledge support by the German Research Foundation and the Open Access Publication Funds of the Göttingen University.

Conflicts of Interest

The authors declare no conflict of interest.

Bibliography

[1] D. Steverding, "Sleeping Sickness and Nagana Disease Caused by Trypanosoma," *Arthropod Borne Diseases*, p. 277, 2016.

[2] D. Steverding, "The history of African trypanosomiasis," *Parasites and Vectors*, 2008.

[3] B. Hursey and J. Slingenbergh, "The tsetse fly and its effects on agriculture in sub-Saharan Africa," *World Animal Review*, pp. 67–73, 1995.

[4] G. J. Losos and B. Ikede, "Review of pathology of diseases in domestic and laboratory animals caused by Trypanosoma congolense, T. vivax, T. brucei, T. rhodesiense and T. gambiense," *Veterinary Pathology*, vol. 9, no. 1_suppl, pp. 1–79, 1972.

[5] F. Giordani, L. J. Morrison, T. G. Rowan, H. P. De Koning, and M. P. Barrett, "The animal trypanosomiases and their chemotherapy: a review," *Parasitology*, vol. 143, no. 14, pp. 1862–1889, 2016.

[6] T. Angara, A. Ismail, and A. Ibrahim, "Research Paper Veterinary An Overview on the Economic Impacts of Animal Trypanosomiasis," 2014.

[7] L. J. Morrison, L. Vezza, T. Rowan, and J. C. Hope, "Animal African trypanosomiasis: time to increase focus on clinically relevant parasite and host species," *Trends in parasitology*, vol. 32, no. 8, pp. 599–607, 2016.

[8] J. Naessens, "Bovine trypanotolerance: A natural ability to prevent severe anaemia and haemophagocytic syndrome?," 2006.

[9] M. Murray, J. C. Trail, C. E. Davis, and S. J. Black, "Genetic resistance to African trypanosomiasis," 1984.

[10] M. Murray, W. I. Morrison, and D. Whitelaw, "Host susceptibility to African trypanosomiasis: trypanotolerance," vol. 21, pp. 1–68, 1982.

[11] D. Courtin, D. Berthier, S. Thevenon, G. K. Dayo, A. Garcia, and B. Bucheton, "Host genetics in African trypanosomiasis," 2008.

[12] J. Dargie, P. Murray, M. Murray, W. Grimshaw, and W. McIntyre, "Bovine trypanosomiasis: the red cell kinetics of Ndama and Zebu cattle infected with Trypanosoma congolense," *Parasitology*, vol. 78, no. 3, pp. 271–286, 1979.

[13] P. Starkey, "N'Dama cattle-a productive trypanotolerant breed," 1984.

[14] F. A. Hassan, "Climate and cattle in North Africa: a first approximation," *The origins and development of African livestock: Archaeology, genetics, linguistics and ethnography*, p. 61, 2006.

[15] R. Mattioli and R. Wilson, "Trypanosomes, tsetse and trypanotolerance: coevolution in tropical Africa.," *Parassitologia*, vol. 38, no. 3, pp. 531–535, 1996.

[16] R. Paling, S. Moloo, J. Scott, G. Gettinby, F. McOdimba, and M. Murray, "Susceptibility of N'Dama and Boran cattle to sequential challenges with tsetse-transmitted clones of Trypanosoma congolense," *Parasite immunology*, vol. 13, no. 4, pp. 427–445, 1991.

[17] M. Murray, J. Trail, G. d'Ieteren, *et al.*, "Trypanotolerance in cattle and prospects for the control of trypanosomiasis by selective breeding," *Rev Sci Tech*, vol. 9, no. 2, pp. 369–386, 1990.

[18] G. E. Roelants, "Natural resistance to African trypanosomiasis," 1986.

[19] H. Epstein, *The origin of the domestic animals of Africa*. Africana publishing corporation, 1971.

[20] Y. A. Mekonnen, M. Gültas, K. Effa, O. Hanotte, and A. O. Schmitt, "Identification of candidate signature genes and key regulators associated with trypanotolerance in the sheko breed," *Frontiers in Genetics*, vol. 10, p. 1095, 2019.

[21] S.-J. Kim, S. Ka, J.-W. Ha, J. Kim, D. Yoo, K. Kim, H.-K. Lee, D. Lim, S. Cho, O. Hanotte, *et al.*, "Cattle genome-wide analysis reveals genetic signatures in trypanotolerant N'Dama," *BMC genomics*, vol. 18, no. 1, p. 371, 2017.

[22] P. Fisher, C. Hedeler, K. Wolstencroft, H. Hulme, H. Noyes, S. Kemp, R. Stevens, and A. Brass, "A systematic strategy for large-scale analysis of genotype–phenotype correlations: identification of candidate genes involved in African trypanosomiasis," *Nucleic acids research*, vol. 35, no. 16, pp. 5625–5633, 2007.

[23] E. W. Hill, G. M. O'Gorman, M. Agaba, J. P. Gibson, O. Hanotte, S. J. Kemp, J. Naessens, P. M. Coussens, and D. E. MacHugh, "Understanding bovine trypanosomiasis and trypanotolerance: the promise of functional genomics," *Veterinary Immunology and Immunopathology*, vol. 105, no. 3-4, pp. 247–258, 2005.

[24] O. Hanotte, Y. Ronin, M. Agaba, P. Nilsson, A. Gelhaus, R. Horstmann, Y. Sugimoto, S. Kemp, J. Gibson, A. Korol, *et al.*, "Mapping of quantitative trait loci controlling trypanotolerance in a cross of tolerant West African N'Dama and susceptible East African Boran cattle," *Proceedings of the National Academy of Sciences*, vol. 100, no. 13, pp. 7443–7448, 2003.

[25] G. M. O'Gorman, S. D. Park, E. W. Hill, K. G. Meade, P. M. Coussens, M. Agaba, J. Naessens, S. J. Kemp, and D. E. MacHugh, "Transcriptional profiling of cattle infected with Trypanosoma congolense highlights gene expression signatures underlying trypanotolerance and trypanosusceptibility," *BMC genomics*, vol. 10, no. 1, p. 207, 2009.

[26] H. Noyes, A. Brass, I. Obara, S. Anderson, A. L. Archibald, D. G. Bradley, P. Fisher, A. Freeman, J. Gibson, M. Gicheru, *et al.*, "Genetic and expression analysis of cattle identifies candidate genes in pathways responding to Trypanosoma congolense infection," *Proceedings of the National Academy of Sciences*, vol. 108, no. 22, pp. 9304–9309, 2011.

[27] A. Rajavel, F. Heinrich, A. O. Schmitt, and M. Gültas, "Identifying cattle breed-specific partner choice of transcription factors during the african trypanosomiasis disease progression using bioinformatics analysis," *Vaccines*, vol. 8, no. 2, 2020.

[28] L. Lopez-Molina, F. Conquet, M. Dubois-Dauphin, and U. Schibler, "The DBP gene is expressed according to a circadian rhythm in the suprachiasmatic nucleus and influences circadian behavior," *The EMBO Journal*, vol. 16, no. 22, pp. 6762–6771, 1997.

[29] P. Fonjallaz, V. Ossipow, G. Wanner, and U. Schibler, "The two PAR leucine zipper proteins, TEF and DBP, display similar circadian and tissue-specific expression, but have different target promoter preferences.," *The EMBO journal*, vol. 15, no. 2, pp. 351–362, 1996.

[30] J. Wuarin and U. Schibler, "Expression of the liver-enriched transcriptional activator protein DBP follows a stringent circadian rhythm," *Cell*, vol. 63, no. 6, pp. 1257–1266, 1990.

[31] R. Orozco-Solis and L. Aguilar-Arnal, "Circadian Regulation of Immunity Through Epigenetic Mechanisms," *Frontiers in Cellular and Infection Microbiology*, vol. 10, p. 96, 2020.

[32] S. Barik, "Molecular Interactions between Pathogens and the Circadian Clock," *International journal of molecular sciences*, vol. 20, no. 23, p. 5824, 2019.

[33] C. Scheiermann, Y. Kunisaki, and P. S. Frenette, "Circadian control of the immune system," *Nature Reviews Immunology*, vol. 13, no. 3, pp. 190–198, 2013.

[34] F. Hawking, "Circadian rhythms of Trypanosoma congolense in laboratory rodents," *Transactions of the Royal Society of Tropical Medicine and Hygiene*, vol. 72, no. 6, pp. 592–595, 1978.

[35] K. A. Taylor, "Immune responses of cattle to African trypanosomes: protective or pathogenic?," *International journal for parasitology*, vol. 28, no. 2, pp. 219–240, 1998.

[36] A. G. Andrianarivo, P. Muiya, M. Opollo, and L. L. Loganhenfrey, "Trypanosoma congolense: Comparative Effects of a Primary Infection on Bone Marrow Progenitor Cells from N'Dama and Boran Cattle," *Experimental parasitology*, vol. 80, no. 3, pp. 407–418, 1995.

[37] F. Ramzan, S. Klees, A. O. Schmitt, D. Cavero, and M. Gültas, "Identification of Age-Specific and Common Key Regulatory Mechanisms Governing Eggshell Strength in Chicken Using Random Forests," *Genes*, vol. 11, no. 4, p. 464, 2020.

[38] D. Wlochowitz, M. Haubrock, J. Arackal, A. Bleckmann, A. Wolff, T. Beißbarth, E. Wingender, and M. Gültas, "Computational identification of key regulators in two different colorectal cancer cell lines," *Frontiers in genetics*, vol. 7, p. 42, 2016.

[39] M. R. Hosen, P. R. Goody, A. Zietzer, G. Nickenig, and F. Jansen, "Micrornas as master regulators of atherosclerosis: from pathogenesis to novel therapeutic options," *Antioxidants and Redox Signaling*, no. ja, 2020.

[40] W. Cai, W. Zhou, Z. Han, J. Lei, J. Zhuang, P. Zhu, X. Wu, and W. Yuan, "Master regulator genes and their impact on major diseases," *PeerJ*, vol. 8, p. e9952, 2020.

[41] Y. Ziegler, M. J. Laws, V. S. Guillen, S. H. Kim, P. Dey, B. P. Smith, P. Gong, N. Bindman, Y. Zhao, K. Carlson, *et al.*, "Suppression of foxm1 activities and breast cancer growth in vitro and in vivo by a

new class of compounds," *NPJ breast cancer*, vol. 5, no. 1, pp. 1–11, 2019.

[42] S. Picaud, K. Leonards, J.-P. Lambert, O. Dovey, C. Wells, O. Fedorov, O. Monteiro, T. Fujisawa, C.-Y. Wang, H. Lingard, *et al.*, "Promiscuous targeting of bromodomains by bromosporine identifies bet proteins as master regulators of primary transcription response in leukemia," *Science advances*, vol. 2, no. 10, p. e1600760, 2016.

[43] E. Barillot, L. Calzone, P. Hupe, J.-P. Vert, and A. Zinovyev, *Computational systems biology of cancer*. CRC Press, 2012.

[44] I. Mellman, "Dendritic cells: master regulators of the immune response," *Cancer immunology research*, vol. 1, no. 3, pp. 145–149, 2013.

[45] D. M. d. Vargas, M. A. De Bastiani, E. R. Zimmer, and F. Klamt, "Alzheimer's disease master regulators analysis: search for potential molecular targets and drug repositioning candidates," *Alzheimer's research & therapy*, vol. 10, no. 1, p. 59, 2018.

[46] Z. Pan, L. Li, Q. Fang, Y. Qian, Y. Zhang, J. Zhu, M. Ge, and P. Huang, "Integrated bioinformatics analysis of master regulators in anaplastic thyroid carcinoma," *BioMed research international*, vol. 2019, 2019.

[47] H.-W. Wang, H.-J. Sun, T.-Y. Chang, H.-H. Lo, W.-C. Cheng, G. C. Tseng, C.-T. Lin, S.-J. Chang, N. R. Pal, and I.-F. Chung, "Discovering monotonic stemness marker genes from time-series stem cell microarray data," in *BMC genomics*, vol. 16, p. S2, BioMed Central, 2015.

[48] E. Wingender and A. E. Kel, "geneXplain — eine integrierte Bioinformatik-Plattform," *BIOspektrum*, vol. 18, pp. 554–556, 2012.

[49] J. Koschmann, A. Bhar, P. Stegmaier, A. E. Kel, and E. Wingender, ""Upstream analysis": an integrated promoter-pathway analysis approach to causal interpretation of microarray data," *Microarrays*, vol. 4, no. 2, pp. 270–286, 2015.

[50] B. Jassal, L. Matthews, G. Viteri, C. Gong, P. Lorente, A. Fabregat, K. Sidiropoulos, J. Cook, M. Gillespie, R. Haw, F. Loney, B. May, M. Milacic, K. Rothfels, C. Sevilla, V. Shamovsky, S. Shorser, T. Varusai, J. Weiser, G. Wu, L. Stein, H. Hermjakob, and P. D'Eustachio, "The reactome pathway knowledgebase," *Nucleic Acids Research*, vol. 48, pp. D498–D503, 11 2019.

[51] S. S.-K. Chan and M. Kyba, "What is a master regulator?," *Journal of stem cell research & therapy*, vol. 3, 2013.

[52] L. Benetatos, A. Benetatou, and G. Vartholomatos, "Enhancers and MYC interplay in hematopoiesis," *Journal of Molecular Medicine*, pp. 1–11, 2020.

[53] S. C. Casey, L. Tong, Y. Li, R. Do, S. Walz, K. N. Fitzgerald, A. M. Gouw, V. Baylot, I. Gütgemann, M. Eilers, and D. W. Felsher, "MYC regulates the antitumor immune response through CD47 and PD-L1," *Science*, 2016.

[54] M. Conacci-Sorrell, L. McFerrin, and R. N. Eisenman, "An overview of MYC and its interactome," *Cold Spring Harbor Perspectives in Medicine*, 2014.

[55] M. D. Delgado and J. León, "Myc roles in hematopoiesis and leukemia," *Genes & cancer*, vol. 1, no. 6, pp. 605–616, 2010.

[56] Y. Guo, C. Niu, P. Breslin, M. Tang, S. Zhang, W. Wei, A. R. Kini, G. P. Paner, S. Alkan, S. W. Morris, *et al.*, "c-Myc–mediated control of cell fate in megakaryocyte-erythrocyte progenitors," *Blood, The Journal of the American Society of Hematology*, vol. 114, no. 10, pp. 2097–2106, 2009.

[57] M. Ohanian, U. Rozovski, R. Kanagal-Shamanna, L. V. Abruzzo, S. Loghavi, T. Kadia, A. Futreal, K. Bhalla, Z. Zuo, Y. O. Huh, *et al.*, "MYC protein expression is an important prognostic factor in acute myeloid leukemia," *Leukemia & lymphoma*, vol. 60, no. 1, pp. 37–48, 2019.

[58] S. Weng, S. Matsuura, C. T. Mowery, S. A. Stoner, K. Lam, D. Ran, A. G. Davis, M.-C. Lo, and D.-E. Zhang, "Restoration of MYC-repressed targets mediates the negative effects of GM-CSF on RUNX1-ETO leukemogenicity," *Leukemia*, vol. 31, no. 1, pp. 159–169, 2017.

[59] H. Matsushita, H. Nakajima, Y. Nakamura, H. Tsukamoto, Y. Tanaka, G. Jin, M. Yabe, S. Asai, R. Ono, T. Nosaka, *et al.*, "C/EBPα and C/EBPϵ induce the monocytic differentiation of myelomonocytic cells with the MLL-chimeric fusion gene," *Oncogene*, vol. 27, no. 53, pp. 6749–6760, 2008.

[60] E. H. Shroff, L. S. Eberlin, V. M. Dang, A. M. Gouw, M. Gabay, S. J. Adam, D. I. Bellovin, P. T. Trand, W. M. Philbrick, A. Garcia-Ocana,

S. C. Casey, Y. Li, C. V. Dang, R. N. Zare, and D. W. Felsher, "MYC oncogene overexpression drives renal cell carcinoma in a mouse model through glutamine metabolism," *Proceedings of the National Academy of Sciences of the United States of America*, 2015.

[61] A. Ventura, D. G. Kirsch, M. E. McLaughlin, D. A. Tuveson, J. Grimm, L. Lintault, J. Newman, E. E. Reczek, R. Weissleder, and T. Jacks, "Restoration of p53 function leads to tumour regression in vivo," *Nature*, 2007.

[62] B. Hoffman, A. Amanullah, M. Shafarenko, and D. A. Liebermann, "The proto-oncogene c-myc in hematopoietic development and leukemogenesis," *Oncogene*, 2002.

[63] M. Franco, A. J. Shastri, and J. C. Boothroyd, "Infection by toxoplasma gondii specifically induces host c-myc and the genes this pivotal transcription factor regulates," *Eukaryotic cell*, vol. 13, no. 4, pp. 483–493, 2014.

[64] F. Dessauge, S. Hilaly, M. Baumgartner, B. Blumen, D. Werling, and G. Langsley, "c-myc activation by theileria parasites promotes survival of infected b-lymphocytes," *Oncogene*, vol. 24, no. 6, pp. 1075–1083, 2005.

[65] Y. Zhang, N. Xu, J. Xu, B. Kong, B. Copple, G. L. Guo, and L. Wang, "E2F1 is a novel fibrogenic gene that regulates cholestatic liver fibrosis through the Egr-1/SHP/EID1 network," *Hepatology*, vol. 60, no. 3, pp. 919–930, 2014.

[66] Q. Lai, A. Giralt, C. Le May, L. Zhang, B. Cariou, P.-D. Denechaud, and L. Fajas, "E2F1 inhibits circulating cholesterol clearance by regulating Pcsk9 expression in the liver," *JCI insight*, vol. 2, no. 10, 2017.

[67] S. Biryomumaisho, E. Katunguka-Rwakishaya, and C. Rubaire-Akiiki, "Serum biochemical changes in experimental Trypanosoma congolense and Trypanosoma brucei infection in Small East Africa goats," *Veterinarski arhiv*, vol. 73, no. 3, pp. 167–180, 2003.

[68] S. Tyagi, P. Gupta, A. S. Saini, C. Kaushal, and S. Sharma, "The peroxisome proliferator-activated receptor: A family of nuclear receptors role in various diseases," *Journal of advanced pharmaceutical technology & research*, vol. 2, no. 4, p. 236, 2011.

[69] Y. Barak, M. C. Nelson, E. S. Ong, Y. Z. Jones, P. Ruiz-Lozano, K. R. Chien, A. Koder, and R. M. Evans, "PPARγ is required for placental, cardiac, and adipose tissue development," *Molecular cell*, vol. 4, no. 4, pp. 585–595, 1999.

[70] S. C. Hasenfuss, L. Bakiri, M. K. Thomsen, E. G. Williams, J. Auwerx, and E. F. Wagner, "Regulation of steatohepatitis and PPARγ signaling by distinct AP-1 dimers," *Cell metabolism*, vol. 19, no. 1, pp. 84–95, 2014.

[71] T. Yamazaki, S. Shiraishi, K. Kishimoto, S. Miura, and O. Ezaki, "An increase in liver PPARγ2 is an initial event to induce fatty liver in response to a diet high in butter: PPARγ2 knockdown improves fatty liver induced by high-saturated fat," *The Journal of nutritional biochemistry*, vol. 22, no. 6, pp. 543–553, 2011.

[72] E. Katunguka-Rwakishaya, M. Murray, and P. Holmes, "Pathophysiology of Trypanosoma congolense infection in two breeds of sheep, Scottish blackface and Finn Dorset," *Veterinary parasitology*, vol. 68, no. 3, pp. 215–225, 1997.

[73] K. Bozek, A. Relógio, S. M. Kielbasa, M. Heine, C. Dame, A. Kramer, and H. Herzel, "Regulation of clock-controlled genes in mammals," *PloS one*, vol. 4, no. 3, 2009.

[74] J. S. Takahashi, "Molecular components of the circadian clock in mammals," *Diabetes, Obesity and Metabolism*, vol. 17, pp. 6–11, 2015.

[75] H. Schrem, J. Klempnauer, and J. Borlak, "Liver-enriched transcription factors in liver function and development. Part II: the C/EBPs and D site-binding protein in cell cycle control, carcinogenesis, circadian gene regulation, liver regeneration, apoptosis, and liver-specific gene regulation," *Pharmacological reviews*, vol. 56, no. 2, pp. 291–330, 2004.

[76] D. J. Lavery, L. Lopez-Molina, R. Margueron, F. Fleury-Olela, F. Conquet, U. Schibler, and C. Bonfils, "Circadian expression of the steroid 15 α-hydroxylase (Cyp2a4) and coumarin 7-hydroxylase (Cyp2a5) genes in mouse liver is regulated by the PAR leucine zipper transcription factor DBP," *Molecular and Cellular Biology*, vol. 19, no. 10, pp. 6488–6499, 1999.

[77] Y.-H. Lee, J. A. Alberta, F. J. Gonzalez, and D. J. Waxman, "Multiple, functional DBP sites on the promoter of the cholesterol 7 alpha-hydroxylase P450 gene, CYP7. Proposed role in diurnal regulation of liver gene expression.," *Journal of Biological Chemistry*, vol. 269, no. 20, pp. 14681–14689, 1994.

[78] D. J. Lavery and U. Schibler, "Circadian transcription of the cholesterol 7 alpha hydroxylase gene may involve the liver-enriched bZIP protein DBP.," *Genes & development*, vol. 7, no. 10, pp. 1871–1884, 1993.

[79] S. Kierstein, H. Noyes, J. Naessens, Y. Nakamura, C. Pritchard, J. Gibson, S. Kemp, and A. Brass, "Gene expression profiling in a mouse model for African trypanosomiasis," *Genes & Immunity*, vol. 7, no. 8, pp. 667–679, 2006.

[80] S. Kersten, M. Rakhshandehroo, B. Knoch, and M. Müller, "Peroxisome proliferator-activated receptor alpha target genes," *PPAR Research*, 2010.

[81] S. Mandard, M. Müller, and S. Kersten, "Peroxisome proliferator-activated receptor α target genes," *Cellular and Molecular Life Sciences CMLS*, vol. 61, no. 4, pp. 393–416, 2004.

[82] T. Aoyama, J. M. Peters, N. Iritani, T. Nakajima, K. Furihata, T. Hashimoto, and F. J. Gonzalez, "Altered constitutive expression of fatty acid-metabolizing enzymes in mice lacking the peroxisome proliferator-activated receptor α (PPARα)," *Journal of Biological Chemistry*, 1998.

[83] T. Varga, Z. Czimmerer, and L. Nagy, "PPARs are a unique set of fatty acid regulated transcription factors controlling both lipid metabolism and inflammation," 2011.

[84] P. Gervois and R. M. Mansouri, "PPARα as a therapeutic target in inflammation-associated diseases," *Expert opinion on therapeutic targets*, vol. 16, no. 11, pp. 1113–1125, 2012.

[85] M. A. Abdelmegeed, K. H. Moon, J. P. Hardwick, F. J. Gonzalez, and B. J. Song, "Role of peroxisome proliferator-activated receptor-α in fasting-mediated oxidative stress," *Free Radical Biology and Medicine*, 2009.

[86] B. Desvergne and W. Wahli, "Peroxisome proliferator-activated receptors: nuclear control of metabolism," *Endocrine reviews*, vol. 20, no. 5, pp. 649–688, 1999.

[87] J. F. Dimartino, L. Selleri, D. Traver, M. T. Firpo, J. Rhee, R. Warnke, S. O'Gorman, I. L. Weissman, and M. L. Cleary, "The Hox cofactor and proto-oncogene Pbx1 is required for maintenance of definitive hematopoiesis in the fetal liver," *Blood*, 2001.

[88] M. Fournier, C.-E. Lebert-Ghali, G. Krosl, and J. J. Bijl, "HOXA4 induces expansion of hematopoietic stem cells in vitro and confers enhancement of pro-B-cells in vivo," *Stem cells and development*, vol. 21, no. 1, pp. 133–142, 2012.

[89] R. Zewdu, M. Risolino, A. Barbulescu, P. Ramalingam, J. M. Butler, and L. Selleri, "Spleen hypoplasia leads to abnormal stress hematopoiesis in mice with loss of Pbx homeoproteins in splenic mesenchyme," *Journal of anatomy*, vol. 229, no. 1, pp. 153–169, 2016.

[90] F. Fang, Y. Wang, R. Li, Y. Zhao, Y. Guo, M. Jiang, J. Sun, Y. Ma, Z. Ren, Z. Tian, *et al.*, "Transcription factor E2F1 suppresses dendritic cell maturation," *The Journal of Immunology*, vol. 184, no. 11, pp. 6084–6091, 2010.

[91] S. Magez, M. Radwanska, M. Drennan, L. Fick, T. N. Baral, F. Brombacher, and P. D. Baetselier, "Interferon-γ and nitric oxide in combination with antibodies are key protective host immune factors during Trypanosoma congolense Tc13 infections," *The Journal of infectious diseases*, vol. 193, no. 11, pp. 1575–1583, 2006.

[92] J. Pan, M. Zhang, J. Wang, Q. Wang, D. Xia, W. Sun, L. Zhang, H. Yu, Y. Liu, and X. Cao, "Interferon-γ is an autocrine mediator for dendritic cell maturation," *Immunology letters*, vol. 94, no. 1-2, pp. 141–151, 2004.

[93] H. Wu, G. Liu, and M. Shi, "Interferon gamma in African trypanosome infections: friends or foes?," *Frontiers in immunology*, vol. 8, p. 1105, 2017.

[94] C. J. Cox, H. M. Espinoza, B. McWilliams, K. Chappell, L. Morton, T. A. Hjalt, E. V. Semina, and B. A. Amendt, "Differential regulation of gene expression by PITX2 isoforms," *Journal of Biological Chemistry*, vol. 277, no. 28, pp. 25001–25010, 2002.

[95] R. M. Kortlever, N. M. Sodir, C. H. Wilson, D. L. Burkhart, L. Pellegrinet, L. B. Swigart, T. D. Littlewood, and G. I. Evan, "Myc cooperates with Ras by programming inflammation and immune suppression," *Cell*, vol. 171, no. 6, pp. 1301–1315, 2017.

[96] S. Casacuberta-Serra and L. Soucek, "Myc and Ras, the Bonnie and Clyde of immune evasion," *Translational cancer research*, vol. 7, no. Suppl 4, p. S457, 2018.

[97] A. V. Villarino, E. Gallo, and A. K. Abbas, "STAT1-activating cytokines limit Th17 responses through both T-bet–dependent and–independent mechanisms," *The Journal of Immunology*, vol. 185, no. 11, pp. 6461–6471, 2010.

[98] L. A. Tesmer, S. K. Lundy, S. Sarkar, and D. A. Fox, "Th17 cells in human disease," *Immunological reviews*, vol. 223, no. 1, pp. 87–113, 2008.

[99] S. J. Szabo, B. M. Sullivan, S. L. Peng, and L. H. Glimcher, "Molecular mechanisms regulating Th1 immune responses," *Annual review of immunology*, vol. 21, no. 1, pp. 713–758, 2003.

[100] I. Okwor, H. Muleme, P. Jia, and J. E. Uzonna, "Altered proinflammatory cytokine production and enhanced resistance to Trypanosoma congolense infection in lymphotoxin β-deficient mice," *Journal of Infectious Diseases*, vol. 200, no. 3, pp. 361–369, 2009.

[101] M. Shi, W. Pan, and H. Tabel, "Experimental African trypanosomiasis: IFN-γ mediates early mortality," *European journal of immunology*, vol. 33, no. 1, pp. 108–118, 2003.

[102] Z. Yonggang, F. Binqing, X. Xiuxiu, T. Xianhong, S. Rui, T. Zhigang, and W. Haiming, "Requirement for PBX1 in developmental programming of natural killer cells," 2017.

[103] J. Cnops, C. De Trez, B. Stijlemans, J. Keirsse, F. Kauffmann, M. Barkhuizen, R. Keeton, L. Boon, F. Brombacher, and S. Magez, "NK-, NKT-and CD8-Derived IFNγ drives myeloid cell activation and erythrophagocytosis, resulting in trypanosomosis-associated acute anemia," *PLoS pathogens*, vol. 11, no. 6, 2015.

[104] S. Yamaguchi, S. Mitsui, L. Yan, K. Yagita, S. Miyake, and H. Okamura, "Role of DBP in the circadian oscillatory mechanism," *Molecular and cellular biology*, vol. 20, no. 13, pp. 4773–4781, 2000.

[105] H. R. Ueda, S. Hayashi, W. Chen, M. Sano, M. Machida, Y. Shigeyoshi, M. Iino, and S. Hashimoto, "System-level identification of transcriptional circuits underlying mammalian circadian clocks," *Nature genetics*, vol. 37, no. 2, pp. 187–192, 2005.

[106] A. C. Silver, A. Arjona, M. E. Hughes, M. N. Nitabach, and E. Fikrig, "Circadian expression of clock genes in mouse macrophages, dendritic cells, and B cells," *Brain, behavior, and immunity*, vol. 26, no. 3, pp. 407–413, 2012.

[107] M. Keller, J. Mazuch, U. Abraham, G. D. Eom, E. D. Herzog, H.-D. Volk, A. Kramer, and B. Maier, "A circadian clock in macrophages controls inflammatory immune responses," *Proceedings of the National Academy of Sciences*, vol. 106, no. 50, pp. 21407–21412, 2009.

[108] S. N. Archer, A. U. Viola, V. Kyriakopoulou, M. von Schantz, and D.-J. Dijk, "Inter-individual differences in habitual sleep timing and entrained phase of endogenous circadian rhythms of BMAL1, PER2 and PER3 mRNA in human leukocytes," *Sleep*, vol. 31, no. 5, pp. 608–617, 2008.

[109] M. Hayashi, S. Shimba, and M. Tezuka, "Characterization of the molecular clock in mouse peritoneal macrophages," *Biological and Pharmaceutical Bulletin*, vol. 30, no. 4, pp. 621–626, 2007.

[110] A. Arjona and D. K. Sarkar, "The circadian gene mPer2 regulates the daily rhythm of IFN-γ," *Journal of interferon & cytokine research*, vol. 26, no. 9, pp. 645–649, 2006.

[111] A. Arjona and D. K. Sarkar, "Evidence supporting a circadian control of natural killer cell function," *Brain, behavior, and immunity*, vol. 20, no. 5, pp. 469–476, 2006.

[112] A. Arjona and D. K. Sarkar, "Circadian oscillations of clock genes, cytolytic factors, and cytokines in rat NK cells," *The Journal of Immunology*, vol. 174, no. 12, pp. 7618–7624, 2005.

[113] D. B. Boivin, F. O. James, A. Wu, P. F. Cho-Park, H. Xiong, and Z. S. Sun, "Circadian clock genes oscillate in human peripheral blood mononuclear cells," *Blood*, vol. 102, no. 12, pp. 4143–4145, 2003.

[114] Z. Sha and R. W. Compans, "Induction of CD4+ T-cell-independent immunoglobulin responses by inactivated influenza virus," *Journal of virology*, vol. 74, no. 11, pp. 4999–5005, 2000.

[115] A. Oxenius, R. M. Zinkernagel, and H. Hengartner, "CD4+ T-cell induction and effector functions: a comparison of immunity against soluble antigens and viral infections," in *Advances in immunology*, vol. 70, pp. 313–367, Elsevier, 1998.

[116] D. C. Parker, "T cell-dependent B cell activation," *Annual review of immunology*, vol. 11, no. 1, pp. 331–360, 1993.

[117] K. A. Taylor, V. Lutje, D. Kennedy, E. Authié, A. Boulangé, L. Logan-Henfrey, B. Gichuki, and G. Gettinby, "Trypanosoma congolense: B-lymphocyte responses differ between trypanotolerant and trypanosusceptible cattle," *Experimental parasitology*, vol. 83, no. 1, pp. 106–116, 1996.

[118] T. Bollinger, A. Leutz, A. Leliavski, L. Skrum, J. Kovac, L. Bonacina, C. Benedict, T. Lange, J. Westermann, H. Oster, *et al.*, "Circadian clocks in mouse and human CD4+ T cells," *PloS one*, vol. 6, no. 12, 2011.

[119] D. C. Jones, X. Ding, and R. A. Daynes, "Nuclear Receptor Peroxisome Proliferator-activated Receptor α (PPARα) Is Expressed in Resting Murine Lymphocytes. The PPARα in T and B lymphocytes is both transactivation and transrepression competent," *Journal of Biological Chemistry*, vol. 277, no. 9, pp. 6838–6845, 2002.

[120] R. Cunard, M. Ricote, D. DiCampli, D. C. Archer, D. A. Kahn, C. K. Glass, and C. J. Kelly, "Regulation of cytokine expression by ligands of peroxisome proliferator activated receptors," *The Journal of Immunology*, vol. 168, no. 6, pp. 2795–2802, 2002.

[121] Y. Yang, A. R. Gocke, A. Lovett-Racke, P. D. Drew, and M. K. Racke, "PPAR alpha regulation of the immune response and autoimmune encephalomyelitis," *PPAR research*, vol. 2008, 2008.

[122] R. Cunard, D. DiCampli, D. C. Archer, J. L. Stevenson, M. Ricote, C. K. Glass, and C. J. Kelly, "WY14, 643, a PPARα ligand, has profound effects on immune responses in vivo," *The Journal of Immunology*, vol. 169, no. 12, pp. 6806–6812, 2002.

[123] P. Delerive, K. De Bosscher, S. Besnard, W. V. Berghe, J. M. Peters, F. J. Gonzalez, J.-C. Fruchart, A. Tedgui, G. Haegeman, and B. Staels, "Peroxisome proliferator-activated receptor α negatively regulates the

vascular inflammatory gene response by negative cross-talk with transcription factors NF-κB and AP-1," *Journal of Biological Chemistry*, vol. 274, no. 45, pp. 32048–32054, 1999.

[124] J. Jantsch and J. Schödel, "Hypoxia and hypoxia-inducible factors in myeloid cell-driven host defense and tissue homeostasis," *Immunobiology*, vol. 220, no. 2, pp. 305–314, 2015.

[125] E. T. Arena, J.-Y. Tinevez, G. Nigro, P. J. Sansonetti, and B. S. Marteyn, "The infectious hypoxia: occurrence and causes during Shigella infection," *Microbes and infection*, vol. 19, no. 3, pp. 157–165, 2017.

[126] M. C. Waghabi, M. Keramidas, S. Bailly, W. Degrave, L. Mendonça-Lima, C. S. Maria de Nazaré, L. M. Maria de Nazareth, S. Paciornik, T. C. Araújo-Jorge, and J.-J. Feige, "Uptake of host cell transforming growth factor-β by Trypanosoma cruzi amastigotes in cardiomyocytes: potential role in parasite cycle completion," *The American journal of pathology*, vol. 167, no. 4, pp. 993–1003, 2005.

[127] K. R. Gantt, S. Schultz-Cherry, N. Rodriguez, S. M. Jeronimo, E. T. Nascimento, T. L. Goldman, T. J. Recker, M. A. Miller, and M. E. Wilson, "Activation of TGF-β by Leishmania chagasi: importance for parasite survival in macrophages," *The Journal of Immunology*, vol. 170, no. 5, pp. 2613–2620, 2003.

[128] V. Boutard, R. Havouis, B. Fouqueray, C. Philippe, J.-P. Moulinoux, and L. Baud, "Transforming growth factor-beta stimulates arginase activity in macrophages. Implications for the regulation of macrophage cytotoxicity.," *The Journal of Immunology*, vol. 155, no. 4, pp. 2077–2084, 1995.

[129] C. Bogdan and C. Nathan, "Modulation of Macrophage Function by Transforming Growth Factor β, Interleukin-4, and Interleukin-10 a," *Annals of the New York Academy of Sciences*, vol. 685, no. 1, pp. 713–739, 1993.

[130] C. J. Johnston, D. J. Smyth, D. W. Dresser, and R. M. Maizels, "TGF-β in tolerance, development and regulation of immunity," *Cellular immunology*, vol. 299, pp. 14–22, 2016.

[131] G. Hansen, J. J. McIntire, V. P. Yeung, G. Berry, G. J. Thorbecke, L. Chen, R. H. DeKruyff, and D. T. Umetsu, "CD4+ T helper cells

engineered to produce latent TGF-β1 reverse allergen-induced airway hyperreactivity and inflammation," *The Journal of clinical investigation*, vol. 105, no. 1, pp. 61–70, 2000.

[132] P. M. Ferrão, L. M. Nisimura, O. C. Moreira, M. G. Land, M. C. Pereira, L. de Mendonça-Lima, T. C. Araujo-Jorge, M. C. Waghabi, and L. R. Garzoni, "Inhibition of TGF-β pathway reverts extracellular matrix remodeling in T. cruzi-infected cardiac spheroids," *Experimental cell research*, vol. 362, no. 2, pp. 260–267, 2018.

[133] L. MacLean, J. E. Chisi, M. Odiit, W. C. Gibson, V. Ferris, K. Picozzi, and J. M. Sternberg, "Severity of human African trypanosomiasis in East Africa is associated with geographic location, parasite genotype, and host inflammatory cytokine response profile," *Infection and immunity*, vol. 72, no. 12, pp. 7040–7044, 2004.

[134] C. A. Dinarello, "Overview of the IL-1 family in innate inflammation and acquired immunity," *Immunological reviews*, vol. 281, no. 1, pp. 8–27, 2018.

[135] J. Bass and J. S. Takahashi, "Circadian integration of metabolism and energetics," *Science*, vol. 330, no. 6009, pp. 1349–1354, 2010.

[136] J. Wang, D. Mauvoisin, E. Martin, F. Atger, A. N. Galindo, L. Dayon, F. Sizzano, A. Palini, M. Kussmann, P. Waridel, *et al.*, "Nuclear proteomics uncovers diurnal regulatory landscapes in mouse liver," *Cell metabolism*, vol. 25, no. 1, pp. 102–117, 2017.

[137] D. Mauvoisin, J. Wang, C. Jouffe, E. Martin, F. Atger, P. Waridel, M. Quadroni, F. Gachon, and F. Naef, "Circadian clock-dependent and-independent rhythmic proteomes implement distinct diurnal functions in mouse liver," *Proceedings of the National Academy of Sciences*, vol. 111, no. 1, pp. 167–172, 2014.

[138] S. Panda, M. P. Antoch, B. H. Miller, A. I. Su, A. B. Schook, M. Straume, P. G. Schultz, S. A. Kay, J. S. Takahashi, and J. B. Hogenesch, "Coordinated transcription of key pathways in the mouse by the circadian clock," *Cell*, vol. 109, no. 3, pp. 307–320, 2002.

[139] T. Taghon, K. Thys, M. De Smedt, F. Weerkamp, F. Staal, J. Plum, and G. Leclercq, "Homeobox gene expression profile in human hematopoietic multipotent stem cells and T-cell progenitors: implications for human T-cell development," *Leukemia*, vol. 17, no. 6, pp. 1157–1163, 2003.

[140] T. Taghon, F. Stolz, M. De Smedt, M. Cnockaert, B. Verhasselt, J. Plum, and G. Leclercq, "HOX-A10 regulates hematopoietic lineage commitment: evidence for a monocyte-specific transcription factor," *Blood, The Journal of the American Society of Hematology*, vol. 99, no. 4, pp. 1197–1204, 2002.

[141] P. Res and H. Spits, "Developmental stages in the human thymus," in *Seminars in immunology*, vol. 11, pp. 39–46, Elsevier, 1999.

[142] M. C. Magli, C. Largman, and H. J. Lawrence, "Effects of HOX homeobox genes in blood cell differentiation," *Journal of cellular physiology*, vol. 173, no. 2, pp. 168–177, 1997.

[143] H. Niwa, "How is pluripotency determined and maintained?," *Development*, vol. 134, no. 4, pp. 635–646, 2007.

[144] H. Niwa, J. Miyazaki, and A. G. Smith, "Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells," *Nature genetics*, vol. 24, no. 4, pp. 372–376, 2000.

[145] Y. Sun, Z. Yang, Z. Niu, J. Peng, Q. Li, W. Xiong, A. N. Langnas, M. Y. Ma, and Y. Zhao, "MOP3, a component of the molecular clock, regulates the development of B cells," *Immunology*, vol. 119, no. 4, pp. 451–460, 2006.

[146] S. M. Kuriakose, R. Singh, and J. E. Uzonna, "Host intracellular signaling events and pro-inflammatory cytokine production in African trypanosomiasis," *Frontiers in immunology*, vol. 7, p. 181, 2016.

[147] M. Fatouros, K. Bourantas, E. Bairaktari, M. Elisaf, O. Tsolas, and D. Cassioumis, "Role of the spleen in lipid metabolism," *British journal of surgery*, vol. 82, no. 12, pp. 1675–1677, 1995.

[148] M. Aviram, J. Brook, I. Tatarsky, Y. Levy, and A. Carter, "Increased low-density lipoprotein levels after splenectomy: a role for the spleen in cholesterol metabolism in myeloproliferative disorders," *The American journal of the medical sciences*, vol. 291, no. 1, pp. 25–28, 1986.

[149] C. Egan, W. Sukhumavasi, B. Butcher, and E. Denkers, "Functional aspects of Toll-like receptor/MyD88 signalling during protozoan infection: focus on Toxoplasma gondii," *Clinical & Experimental Immunology*, vol. 156, no. 1, pp. 17–24, 2009.

[150] S. Kuriakose, C. Onyilagha, R. Singh, F. Olayinka-Adefemi, P. Jia, and J. E. UZONNA, "TLR-2-dependent activation of MAPK and STAT proteins regulates proinflammatory cytokine response and immunity to experimental Trypanosoma congolense infection," *Frontiers in Immunology*, vol. 10, p. 2673, 2019.

[151] C. E. Sutton, S. J. Lalor, C. M. Sweeney, C. F. Brereton, E. C. Lavelle, and K. H. Mills, "Interleukin-1 and IL-23 induce innate IL-17 production from gamma delta T cells, amplifying Th17 responses and autoimmunity," *Immunity*, vol. 31, no. 2, pp. 331–341, 2009.

[152] A. M. Curtis, M. M. Bellet, P. Sassone-Corsi, and L. A. O'Neill, "Circadian clock proteins and immunity," *Immunity*, vol. 40, no. 2, pp. 178–186, 2014.

[153] C. Scheiermann, J. Gibbs, L. Ince, and A. Loudon, "Clocking in to immunity," *Nature Reviews Immunology*, vol. 18, no. 7, pp. 423–437, 2018.

[154] M. Barkhuizen, S. Magez, R. A. Atkinson, and F. Brombacher, "Interleukin-12p70-dependent interferon-γ production is crucial for resistance in African trypanosomiasis," *The Journal of infectious diseases*, vol. 196, no. 8, pp. 1253–1260, 2007.

[155] Y. Liu, E. Ragaa, Z. Li, L. Nuortio, A. Mustafa, and M. Bakhiet, "Interferon-gamma and interleukin-12 genes are preferentially expressed during early experimental African trypanosomiasis and suppressed by denervation of the spleen.," *Scandinavian journal of immunology*, vol. 50, no. 5, pp. 485–491, 1999.

[156] S. Tian, "Identification of monotonically differentially expressed genes for non-small cell lung cancer," *BMC bioinformatics*, vol. 20, no. 1, p. 177, 2019.

[157] S. Hergenhan, S. Holtkamp, and C. Scheiermann, "Molecular interactions between components of the circadian clock and the immune system," *Journal of Molecular Biology*, 2020.

[158] B. J. Altman, A. L. Hsieh, A. Sengupta, S. Y. Krishnanaiah, Z. E. Stine, Z. E. Walton, A. M. Gouw, A. Venkataraman, B. Li, P. Goraksha-Hicks, *et al.*, "MYC disrupts the circadian clock and metabolism in cancer cells," *Cell metabolism*, vol. 22, no. 6, pp. 1009–1019, 2015.

[159] S. Gery and H. Koeffler, "The role of circadian regulation in cancer," in *Cold Spring Harbor symposia on quantitative biology*, vol. 72, pp. 459–464, Cold Spring Harbor Laboratory Press, 2007.

[160] S. Masri and P. Sassone-Corsi, "The emerging link between cancer, metabolism, and circadian rhythms," *Nature medicine*, vol. 24, no. 12, p. 1795, 2018.

[161] L. J. Dixon, M. Barnes, H. Tang, M. T. Pritchard, and L. E. Nagy, "Kupffer cells in the liver," *Comprehensive Physiology*, vol. 3, no. 2, pp. 785–797, 2013.

[162] R. D. Stout and J. Suttles, "Functional plasticity of macrophages: reversible adaptation to changing microenvironments," *Journal of leukocyte biology*, vol. 76, no. 3, pp. 509–513, 2004.

[163] S. Gordon and P. R. Taylor, "Monocyte and macrophage heterogeneity," *Nature Reviews Immunology*, vol. 5, no. 12, pp. 953–964, 2005.

[164] F. Hörhold, D. Eisel, M. Oswald, A. Kolte, D. Röll, W. Osen, S. B. Eichmüller, and R. König, "Reprogramming of macrophages employing gene regulatory and metabolic network models," *PLoS computational biology*, vol. 16, no. 2, p. e1007657, 2020.

[165] J. Uzonna, R. Kaushik, J. Gordon, and H. Tabel, "Cytokines and antibody responses during Trypanosoma congolense infections in two inbred mouse strains that differ in resistance," *Parasite immunology*, vol. 21, no. 2, pp. 57–71, 1999.

[166] A. E. Kel, "Search for master regulators in walking cancer pathways," in *Biological Networks and Pathway Analysis*, pp. 161–191, Springer, 2017.

[167] A. E. Kel, P. Stegmaier, T. Valeev, J. Koschmann, V. Poroikov, O. V. Kel-Margoulis, and E. Wingender, "Multi-omics "upstream analysis" of regulatory genomic regions helps identifying targets against methotrexate resistance of colon cancer," *EuPA Open Proteomics*, vol. 13, pp. 1–13, 2016.

[168] S. Sikdar and S. Datta, "A novel statistical approach for identification of the master regulator transcription factor," *BMC bioinformatics*, vol. 18, no. 1, pp. 1–11, 2017.

Chapter 4

Deciphering the Molecular Mechanism Underlying African Animal Trypanosomiasis by Means of the 1000 Bull Genomes Project Genomic Dataset

Abirami Rajavel [1,2,*], Selina Klees [1,2], Yuehan Hui [1], Armin Otto Schmitt [1,2] and Mehmet Gültas [2,3,*]

[1] Breeding Informatics Group, Department of Animal Sciences, Georg-August University, Margarethe von Wrangell-Weg 7, 37075 Göttingen, Germany
[2] Center for Integrated Breeding Research (CiBreed), Albrecht-Thaer-Weg 3, Georg-August University, 37075 Göttingen, Germany
[3] Faculty of Agriculture, South Westphalia University of Applied Sciences, Lübecker Ring 2, 59494 Soest, Germany
[*] Author to whom correspondence should be addressed.

Bibliographic Information

Rajavel, A.; Klees, S.; Hui, Y.; Schmitt, A.O.; Gültas, M. Deciphering the Molecular Mechanism Underlying African Animal Trypanosomiasis by Means of the 1000 Bull Genomes Project Genomic Dataset. Biology 2022, 11, 742.

https://doi.org/10.3390/biology11050742

Author contribution of Abirami Rajavel

The author contributed to the design of the study. Further, AR conducted computational analyses, prepared the datasets, implemented the framework and performed the literature survey. Furthermore, the author interpreted the results and wrote a significant part of the final version of the manuscript.

4.1 Abstract

African Animal Trypanosomiasis (AAT) is a neglected tropical disease and spreads by the vector tsetse fly, which carries the infectious *Trypanosoma* sp. in their saliva. Particularly, this parasitic disease affects the health of livestock, thereby imposing economic constraints on farmers, costing billions of dollars every year, especially in sub-Saharan African countries. Mainly considering the AAT disease as a multistage progression process, we previously performed upstream analysis to identify transcription factors (TFs), their co-operations, over-represented pathways and master regulators. However, downstream analysis, including effectors, corresponding gene expression profiles and their association with the regulatory SNPs (rSNPs), has not yet been established. Therefore, in this study, we aim to investigate the complex interplay of rSNPs, corresponding gene expression and downstream effectors with regard to the AAT disease progression based on two cattle breeds: trypanosusceptible Boran and trypanotolerant N'Dama. Our findings provide mechanistic insights into the effectors involved in the regulation of several signal transduction pathways, thereby differentiating the molecular mechanism with regard to the immune responses of the cattle breeds. The effectors and their associated genes (especially *MAPKAPK5*, *CSK*, *DOK2*, *RAC1* and *DNMT1*) could be promising drug candidates as they orchestrate various downstream regulatory cascades in both cattle breeds.

4.2 Introduction

Trypanosomiasis is a deadly neglected tropical disease that affects the health of several mammalian species, including cattle, horses and humans. When it affects the health of humans, this disease is commonly known as 'sleeping sickness' [1]. On the other hand, African Animal Trypanosomiasis (AAT), also known as nagana (which means 'useless' in the Zulu language), affects the health of livestock and it is spread by the tsetse fly carrying salivarian trypanosomes [2, 3, 4]. It prevails extensively in 40 sub-Saharan African countries and accounts for huge economic losses to farmers, particularly affecting meat and milk production [5, 6]. Therefore, it has gained socio-economic importance as it retards the agricultural development of several regions in those areas [7]. Particularly, AAT is caused by different *Trypanosoma* species, including *Trypanosoma congolense*, *Trypanosoma vivax* and *Trypanosoma brucei* spp. [7]. Out of them, *Trypanosoma congolense* is regarded as the most serious pathogen for livestock. In humans, these unicellular protozoans cause various diseases; for example, *T. brucei* causes sleeping sickness, which alters

the sleep-wake cycle by interfering the circadian rhythm [8, 9], whereas *T. cruzi* causes Chagas disease or American trypanosomiasis [10, 11].
Trypanosomes infect a wide range of hosts and are transmitted into the bloodstream of the mammalian host [12, 13, 14, 15]. When the tsetse fly transmits the trypanosomes into the body of the cattle, the parasite first infects the skin resulting in the lesions due to local host immune responses. Afterwards, it enters the blood circulation via lymphatic vessels [16, 17, 18, 19]. Important symptoms primarily observed in animals after being infected with the most pathogenic *T. congolense* include anaemia, loss of body conditions, thrombocytopenia [20], lymphopenia, immunosuppression [21, 22, 23] and other secondary infections [24].
Few West African cattle breeds like N'Dama can control the development of the disease AAT, in contrast to the other breeds such as Boran [25]. As a trait, trypanotolerance is the ability to control parasitemia (development of parasites) and the associated anaemia [12, 13, 14, 15]. Harnessing the genetic potential of trypanotolerant breeds like N'Dama, recent studies [26, 27, 28, 29] have focussed on investigating the trait of trypanotolerance.
Mainly considering the trait of trypanotolerance, several researchers [29, 30, 31, 32, 33, 34, 35] have performed different types of analysis based on either gene expression data sets or genotype $\times$ phenotype data sets from the cattle breeds, namely trypanosusceptible Boran and trypanotolerant N'Dama (for a short overview, see [26, 36]). Among these previous studies [29, 30, 31, 32, 33, 34, 35], especially, Hanotte *et al.* [32] performed genome-wide analyses and identified genomic regions to reveal the genetic differences between the cattle breeds related to the trait of trypanotolerance. In this regard, Noyes *et al.* [30] analysed the gene expression dataset to identify differentially expressed genes that responded to trypanosome infection to differentiate between the susceptible and tolerant cattle breeds. To this end, Mekonnen *et al.* [29] investigated the genetic background of N'Dama along with other cattle breeds. Moreover, O'Gorman et al. [34] and Gautier et al. [35] conducted the genetic and expression analyses to identify the significant chromosomal regions which could affect the susceptibility/tolerance of the cattle breeds.
To decipher the underlying regulatory mechanisms determining trypanosusceptibility/trypanotolerance of these cattle breeds, we have recently analysed a time-series gene expression data set of the two cattle breeds [37, 38]. Particularly, by considering the AAT disease development as a multi-stage progression process, we investigated Monotonically Expressed Genes (MEGs) to capture the complete progression process of the disease. As a result of our previous studies [37, 38], we reported several transcription factors (TFs), their co-operations and master regulators governing the upstream molecular

mechanism during the infection. Despite the rich literature on this disease, there is still a need for further investigation of genetic mechanisms of the regulatory processes addressing the complex interplay between regulatory SNPs, their corresponding gene expression and the downstream effectors in association with the AAT disease.
Recent progress in molecular biology created the opportunity to use heterologous animal models to investigate complex traits and genetics underlying the disease mechanisms [39, 40, 41]. Remarkably, integratomics is fast becoming the latest trend in omics research while integrating a variety of omics data (such as genomic, transcriptomic and proteomic data), irrespective of the species [42]. Access to genome sequences of species like cattle unlocked the potential for integrating transcriptomic and genomic data. The information about effectors, which are end products located several steps downstream and regulate the functioning of multiple signal transduction pathways, is pivotal for understanding the complex molecular mechanisms such as the response of the cell to an extracellular pathogen. In silico study of the candidate, MEGs were undertaken to identify the novel trypanotolerance-associated rSNPs and downstream effectors. The candidate MEGs from our analysis of effectors were analysed for their gene expression profiles.
To address this missing point of previous studies, we applied an integratomics approach to study the complex interplay of biological processes orchestrated by rSNPs, genes and downstream effectors during the AAT disease progression. For this purpose, we performed integrated systems biology and bioinformatics approaches while incorporating the transcriptomic data [30] and genomic data from the 1000 Bull Genome Project [43] for both cattle breeds. To examine the combinatorial interplay, we firstly identified the regulatory SNPs (rSNPs), which are located in the promoter regions of the MEGs and which, as per definition, exert a strong influence on the binding affinity of the TFs either by the deletion or the creation (gain/loss) of a transcription factor binding site (TFBS) [44, 45, 46]. In accordance with previous studies on the rSNPs [47, 48], it is today well-known that the rSNPs based on their consequences can influence and change individual steps of gene expression. Subsequently, we extracted for each tissue (liver, spleen and lymph node) the MEGs harbouring the regulatory SNPs in their promoters by manually studying their gene expression profiles during the AAT disease progression. Finally, we explored the corresponding downstream effectors that have a pronounced effect on the activation and regulation of a multitude of downstream signalling pathways. Taken together, our findings provide a multifaceted glimpse of (i) the regulatory SNPs governing the susceptibility/tolerance mechanism of the cattle breeds; (ii) downstream effectors associated with the MEGs harbouring rSNPs, and their biological and immune-related func-

tions, which could potentially distinguish the susceptibility/tolerance mechanism of cattle breeds to AAT disease; (iii) deciphering novel hypotheses and potential targets for breeding goals and therapeutic implications.

4.3 Materials and Methods

In this section, we illustrate an overview of the analyses to highlight the difference between our previous studies [37, 38] and the current study. Simultaneously this overview shows how this present study is complementing our previous studies. Figure 4.1 outlines the overview of our analyses.

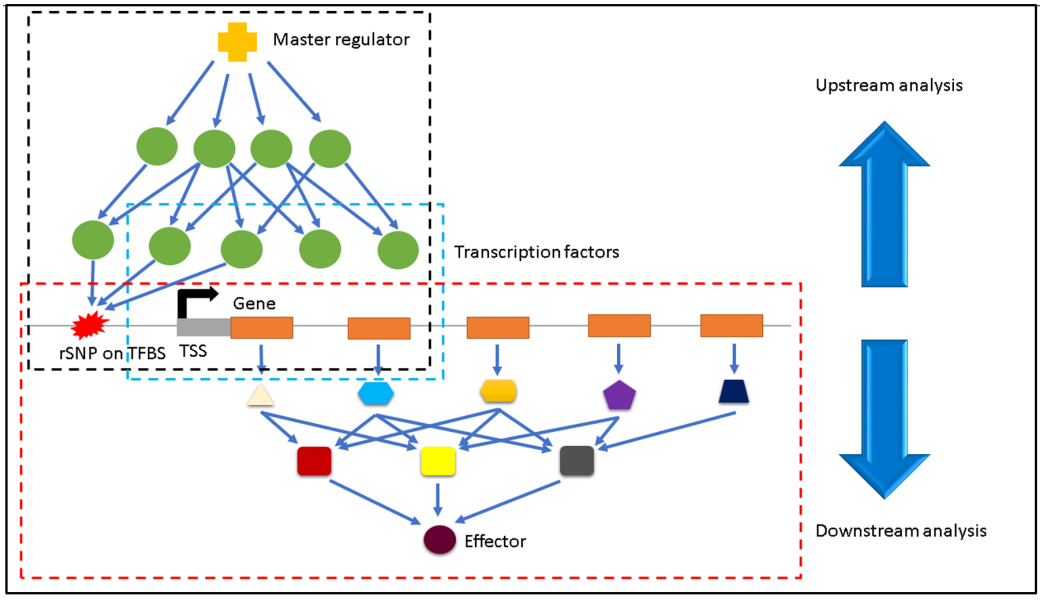

Figure 4.1: Overview of analyses. Our first computational study (middle box in blue-dotted lines) highlighted the transcription factor co-operations associated with the AAT disease progression [37]. In our second study (top box in black dashed lines), we performed an upstream analysis to detect master regulators and over-represented upstream pathways related to AAT [38]. In the current study (bottom box in red dashed lines), we focus on the downstream analysis to decipher the complex interplay of regulatory SNPs (rSNPs), their related gene expression and their corresponding downstream effectors which regulate a multitude of signal transduction pathways during the AAT disease progression.

4.3.1 Monotonically Expressed Genes

In this study, we investigate the complex interplay of regulatory SNPs (rSNPs), the related gene expression and their corresponding downstream effectors.
A time series microarray data set, originally published by Noyes *et al.* (http://www.ebi.ac.uk/arrayexpress/, accession no. E-MEXP-1778) [30], has been analysed in [37] to identify the Monotonically Expressed Genes (MEGs), which are expressed either with increasing or decreasing pattern during a biological process or a disease. The data set consisted of the gene expression recordings from three tissues (liver, spleen and lymph node) of two cattle breeds: trypanotolerant N'Dama and trypanosuceptible Boran. In this experiment, tissue harvest was performed on day 0, 21 and 35 day 21 and day 35. Only for the liver tissue, samples were collected at additional time points such as day 12, 15, 18, 26, 29 and 32 day 12, day 15, day 18, day 26, day 29, and day 32.
We use these identified MEGs in our further analysis. The numbers of MEGs are provided in Table 4.1 and the lists of MEGs are provided as Supplementary File S1.

Table 4.1: Numbers of statistically significant Monotonically Expressed Genes in ascending and descending order for liver-, spleen- and lymph node-tissues for the cattle breeds Boran and N'Dama.

	Boran		N'Dama	
	Ascending	Descending	Ascending	Descending
Liver	741	308	757	124
Spleen	669	126	13	139
Lymph node	87	5	119	114

4.3.2 Genotype data

The genotype-phenotype data set of the cattle breeds Boran and N'Dama used in this study are a part from the 1000 Bull Genomes Project [43].
The genotype data contains for 23 animals (11 Boran and 12 N'Dama) 783, 637 variants that are located in the promoter regions covering from -1000 bp to 0 bp relative to the transcription start sites of the MEGs. Furthermore, we considered the resistance of the cattle breed as qualitative phenotype and assigned '0' and '1' to represent the disease phenotypes for resistance and susceptibility, respectively. Similar to our previous studies [46, 49], for the purpose of quality control, filtering of genotype data was then carried out in

order to remove the SNPs with a minor allele frequency (MAF) less than 0.1, call rate less than 0.95 and which significantly deviated from Hardy–Weinberg Equilibrium ($p < 1 \times 10^{-8}$). After this filtering, the data set contained about 19,330 SNPs and 23 animals for further analyses. We performed a Genome-Wide Association analysis using PLINK 1.9 software [50]. The genotype-phenotype association was evaluated with PLINK by chi-squared allelic test. As suggested by Heinrich *et al.* [46], we used the false discovery rate (FDR) of 0.1 to control the type I error rate.

4.3.3 Identification of regulatory SNPs

In previous studies [44, 46], a SNP is defined to be a regulatory SNP (rSNP) if it is located in the promoter region of a gene and if it affects the binding affinity of one or more transcription factors (TFs) to their respective binding sites which leads to the gain/loss of TFBSs. According to the rSNP detection pipeline, we extracted the flanking sequence of ±25 bp for each selected SNP. Finally, we scanned the flanking sequences of the SNPs for both alternate and reference alleles using the MATCH^{TM} program [51] and thus classified a SNP as rSNP if it leads to gain and/or loss of a TFBS.

4.3.4 Finding the Effectors

Taking the rSNPs into account, we filtered the list of MEGs under study which harbour at least one rSNP within their promoter. Using the filtered list of MEGs for each tissue individually, we employed the systems biology platform geneXplain [52] in order to identify the effector molecules. Effectors are important signalling molecules that are end products located several steps downstream and regulate the functioning of a multitude of signalling cascades. With regard to AAT disease, the knowledge about the effectors could provide promising information to decipher their complex interplay with rSNPs and the corresponding MEGs. The identification of effectors was performed using the 'Effector search' function on the geneXplain platform, which utilises the TRANSPATH® database [53] for searching the downstream effectors regulated by the input set of MEGs.

4.4 Results and Discussion

By analysing regulatory SNPs (rSNPs), the related gene expression profiles of MEGs and their associated downstream effectors, we established their complex interplay involved in the AAT disease progression for both cattle

breeds. For this purpose, we firstly performed a genome-wide association analysis and obtained 19,330 significant SNPs, out of which 1849 SNPs have been further classified as rSNPs.
Uncovering disease-related SNPs is recently gaining utmost importance as they can have an impact on the disease progression and also on how the infected individual responds to the infection [54, 55, 56, 57, 58]. In particular, rSNPs are of great interest as they could be causal and thus alter the protein-DNA interaction. Afterwards, considering the MEGs of interest, which harbour at least one rSNP in their promoter regions, we created for each tissue a filtered list of monotonically expressed genes. Finally, using these lists of MEGs obtained for each tissue (liver, spleen and lymph node) for both cattle breeds, we identified the downstream effectors to investigate further the underlying molecular mechanisms that orchestrate differences in the level of tolerance of the cattle breeds to AAT. The numbers of rSNPs and MEGs of interest are given in Tables 4.2 and 4.3, respectively. The list of respective rSNPs and MEGs are provided as Supplementary Files S2 and S3.

Table 4.2: Numbers of regulatory SNPs found for liver-, spleen- and lymph node-tissues for the cattle breeds Boran and N'Dama.

	Boran		N'Dama	
	Gain of TFBS	Loss of TFBS	Gain of TFBS	Loss of TFBS
Liver	365	403	342	385
Spleen	152	154	3	8
Lymph node	10	12	3	12

Table 4.3: Numbers of MEGs under study harboring at least one rSNP in their promoter region, for liver-, spleen- and lymph node-tissues for the cattle breeds Boran and N'Dama.

	Boran	N'Dama
Liver	194	102
Spleen	157	9
Lymph node	13	5

4.4.1 Identification of Downstream Effectors

We employed the "Effector Search" algorithm from the geneXplain platform [52] using the tissue-based MEG sets of interest for the computational identification of downstream effectors. From this analysis, we obtained a total of 18 effectors that are unique for the breeds and the three tissues (given in Table 4.4). Remarkably, the effectors obtained are completely different between the susceptible and tolerant cattle breeds.

Table 4.4: Downstream effectors obtained for liver-, spleen- and lymph node-tissues for the cattle breeds Boran and N'Dama.

Cattle breed	Tissue	Effectors
Boran	Liver	Itk:Lck:PLCgamma1:SLP-76
Boran	Liver	PKCdelta
Boran	Liver	SRF
Boran	Spleen	histone H3:DNA-PKcs
Boran	Spleen	p53:HEXIM1
Boran	Spleen	HEXIM1:p53
Boran	Lymph node	LIMPII:Prpf8
Boran	Lymph node	VICKZ3:Prpf8
Boran	Lymph node	SNRPGP15:Prpf8
N'Dama	Liver	CHTOG:h3f3a
N'Dama	Liver	p85alpha
N'Dama	Liver	TFII-I
N'Dama	Spleen	LYZL2-isoform2:LRP11
N'Dama	Spleen	PON 2-isoform1:LRP11
N'Dama	Spleen	WSX-1:LRP11
N'Dama	Lymph node	Ssu72
N'Dama	Lymph node	MTMR4
N'Dama	Lymph node	Clathrin LCb

4.4.2 Downstream Effectors for Liver Tissue

The analysis of the MEGs for the liver tissue resulted in the detection of three effectors for Boran (namely SRF, PKCδ and a complex of proteins ITK, LCK, PLCγ and SLP76) and N'Dama (p85α, chTOG:H3F3A and TF2-1).
Serum response factor (c-fos serum response element-binding transcription factor) is a transcription factor belonging to the MADS (MCM1, Agamous, Deficiens and SRF) box superfamily. It is mainly involved in the regula-

tion of immediate-early genes and takes part in important cellular processes like cell differentiation, cell growth and apoptosis. The gene encoding this protein serves as the major target for several signalling pathways, in particular, the mitogen-activated protein kinase pathway (MAPK) that plays a significant role in the immune surveillance mechanism supporting the trypanosome infection [59]. Therefore, the SRF protein could be directing the immune evasion, thereby assisting susceptibility of the cattle breed in AAT disease progression.

The second effector, PKCδ, found in Boran's liver tissue, has been reported as the marker of inflammation and plays an essential role in tuberculosis disease progression in humans [60]. This could be an important hint for the AAT disease progression in the susceptible cattle breed Boran. Moreover, the third effector consists of four proteins, namely ITK, SLP 76, LCK and PLCγ1. Inducible T-cell kinase (ITK) belongs to the Tec family of non-receptor tyrosine kinases, which are expressed in immune cells like mast cells and T cells. It plays a critical role in T-lymphocyte development and functioning and is involved in regulating T-cell receptor signalling. Furthermore, it is activated with respect to antigen receptors, for example, T-cell receptor stimulation [61, 62, 63]. It is reported to be important for the replication of the virus inside the infected host cells [64], elucidating its role in supporting the pathogen infection in AAT. SH2-domain-containing leukocyte protein of 76 kDa (SLP 76) is one of the key adaptor proteins expressed only in the haematopoietic part of the immune cells such as monocyte, granulocyte and T lymphocyte lineage [65]. The protein SLP 76 plays a crucial role in the regulation of several signalling cascades [66]. Additionally, its expression is regulated during T cell maturation and activation [65]. This demonstrates the close association of the protein SLP 76 with the haematopoiesis and generation of immune responses relating to anaemia in AAT disease, an important hallmark of AAT. The association of LCK (lymphocyte-specific cytoplasmic protein-tyrosine kinase) to CD4 and CD8 is necessary for antigen-specific T cell development and activation [67]. Of particular interest, phospholipase C gamma 1 (PLCγ1) signalling is important for several physiological processes like cell differentiation [68, 69].

In our analysis, we found an effector as a complex of chTOG and H3F3A for the liver tissue of N'Dama. The chTOG is a human TOG protein, reported as a mitotic error correction factor playing an important role in accurate chromosome segregation during cell division [70]. Further, H3F3A belongs to the group of basic nuclear histone proteins supporting the structure of the chromosome, thereby maintaining the genome integrity [71]. Another effector, TF2-1, found in the liver tissue of N'Dama, is a non-infectious and intracellular retrotransposon [72]. However, both of these effectors were not

illustrated in relation to host-pathogen interaction, and thus, their potential roles in AAT disease progression are not studied. On the other hand, the third effector p85 α, is an adapter subunit of heterodimer phosphatidylinositol 3-kinase, which is involved in the production of phospholipids. By interacting with other proteins such as p110 α and PTEN, p85 α could regulate the PI3K pathway either in a positive or negative manner [73]. Due to the importance of the phosphatidylinositol 3-kinase (PI3K) signalling pathway in many diseases [74], the regulatory activity of p85 α is gaining importance in response to infections as well. This demonstrates the role of p85 α during AAT infection, which might play a crucial part in trypanotolerance of N'Dama by maintaining the lipid synthesis in the host's liver intact without interruption from the pathogenic attack.

4.4.3 Downstream Effectors for Spleen

The analysis of the effectors for spleen tissue unravelled p53:HEXIM1, HEXIM1:p53 and histone H3:DNA-PKcs for Boran and LYZL2 isoform 2:LRP11, PON-2 isoform 1:LRP11 and WSX-1:LRP11 for N'Dama.
The first two effectors are a complex of two proteins: HEXIM1 and p53. Hexamethylene bisacetamide-inducible protein 1 (HEXIM1) protein encoded by *HEXIM1* is known for its role in the regulation of gene expression, especially with regard to innate immunity [75]. Particularly, it has been reported in the *Trypanosoma cruzi* infection, in association with splenomegaly in the Hexim1$^{+/-}$ mice. It was shown how the downregulation of *HEXIM1* protects the host against *T. cruzi* infection [76]. This hints at the functioning of HEXIM1 during the infection process by increasing inflammation. Another part of the protein complex, p53, identified for the spleen tissue, acts as a tumour suppressor protein in humans, therefore called as "guardian of the genome" [77, 78]. In recent studies, it has been demonstrated that p53 regulates inflammation [79] which is highly associated with AAT. Especially in a study involving bacterial infection [80], deletion or inhibition of p53 resulted in the clearance of extracellular bacteria, which reveals the regulatory role of p53 in the defence against extracellular pathogens establishing the modulation of microbicidal function. Another effector found in the spleen tissue, DNA Protein Kinase, has been reported for its roles in regulating metabolic pathways, particularly in fatty acid synthesis [81]. It is one of the key players responding to DNA damage and in IRF-3-dependent innate immunity [82]. Especially, DNA Damage Response PK has been studied as a driver in evading host immunity [83] and in developmental transitions occurring between the vector and the host [84]. This effector could play a role in immune evasion, thereby supporting the trypanosome infection and

increasing the susceptibility of Boran.
For the spleen tissue of N'Dama, the identified effectors, including LYZL2 isoform 2, PON2 isoform 1 and WSX1 are complexes of LRP11 protein. LRP11 plays a key role in the development of stress responses in mice, as suggested by Xu et al. in [85]. It is well-known that through the activation of the stress response, the host's body provides energy immediately available for immune responses against the parasitic infection, therefore benefitting the host to recover earlier [86]. LYZL2 identified as one of the effectors, exhibits lysozyme activity, which functions as bacteriolytic factors [87] and they are mainly involved in the host defence. Their biological function in relation to parasitic infection has not been largely studied yet. Interestingly, we found Paraoxonase 2 (PON2) in the spleen tissue of N'Dama, which is an intracellular membrane protein exerting anti-oxidant functions [88]. Macrophages are key players against extracellular and intracellular pathogens. In this regard, PON2 has been studied for their expression in the macrophages [89]. In a study involving bacterial infection with *Pseudomonas aeroginosa*, the role of PON2 in the innate immune defence has been demonstrated [90]. The next effector, WSX1, is a class I cytokine receptor for IL27 and is predominantly expressed in lymphoid tissues like the spleen and lymph nodes [91]. Being the IL27 receptor, WSX1 has been studied to be associated with the IL27 signalling pathway. It is further involved in the regulation of Th1-type adaptive immune responses and also of the cells of the innate immune system [92]. Villarino et al. reported in their study [93] that WSX1 is necessary for resistance to parasitic infection from *Toxoplasma gondii*. Particularly, this could provide an important hint on the functioning of WSX1 in resistance of N'Dama to AAT disease.

4.4.4 Downstream Effectors for Lymph Node

The analysis of the MEGs of lymph node tissue reveals the effectors, namely LIMP-2:Prpf8, VICKZ3:Prpf8 and SNRPGP15:Prpf8, for Boran and the effectors Ssu72, MTMR4, Clathrin LCb for N'Dama. Considering the biological roles of effector LIMP-2, it is a type III glycoprotein belonging to the CD36 superfamily of scavenger proteins, facilitating the transport of the acid hydrolase β-glucocerebrosidase (GC) [94]. This protein provides a strong connection between cholesterol export and innate immunity [95, 96] as lipids play crucial roles in the multiplication of the trypanosome infection cycle. Therefore, the LIMP2 protein might be a strong candidate protein crucial for establishing the infection, thereby making the cattle breed Boran susceptible to AAT. Another effector, VICKZ3, for the lymph node issue of Boran, belongs to the family of RNA binding proteins and is expressed in the

developing central nervous system [97] during embryogenesis. This group of proteins are associated with the regulation of RNA and are involved in controlling the cellular processes like proliferation and translational repression [98]. Furthermore, the effector SNRPGP15 (Small Nuclear Ribonucleoprotein G-like protein 15) is a part of the spliceosome, which mainly takes part in RNA metabolism [99]. Finally, part of the protein complexes of all the three effectors is pre-mRNA processing factor 8 (Prpf8) is a highly conserved protein and known for its role in the pre-mRNA splicing process [100]. However, VICKZ3, SNRPGP15 and PRPF8 have not been largely studied in terms of host-pathogen interaction; therefore, their potential role in AAT disease progression is currently unknown. On the other hand, the effectors identified for the lymph node tissue of N'Dama suggest their crucial roles in immunity, bolstering the host's defence against the parasite. The effector Ssu72 is a dual protein phosphatase that plays a role in RNA processing. A recent study has associated the Ssu72 protein in macrophages with the process of immunometabolism [101], implicating a closer connection between immunity and trypanotolerance of N'Dama. The next effector, Myotubularin-related protein 4 (MTMR4), is an intracellular protein that exhibits lipid and protein phosphatase activities in several cellular functions. Especially MTMR4 is involved in the negative regulation of TGF-β signalling. During the infection of *Trypanosoma cruzi*, the role of TGF-β has been demonstrated to inhibit the functioning of immune effector cells and the production of interferon α, thereby resulting in the multiplication of the pathogen [102]. Therefore, MTMR4 indirectly assists the host in decreasing the pathogen numbers within the body, supporting the tolerance mechanism of the cattle breed N'Dama. Another effector, Clathrin, is a cytosolic protein made up of heavy and light chains. Clathrin light chains (LCb) are important components of clathrin-coated vesicles, especially necessary to uptake large foreign particles into the vesicles [103]. This effector found in lymph nodes could represent the engulfing of infectious parasites during the AAT disease in the body of N'Dama. In particular, the knowledge of these effectors provides essential information in distinguishing the downstream events underlying the susceptibility and tolerance mechanisms of the cattle breeds Boran and N'Dama, respectively. Further validation of the results from the molecular biology end is necessary to evaluate the biological importance of their functions in the AAT disease progression as well as to gain a comprehensive understanding of their roles in susceptibility/tolerance mechanisms of the cattle breeds.

4.4.5 Gene Expression Profile Analysis of MEGs Harbouring rSNPs

Using gene expression profiles, it is possible to gain insights into the differences in the expression levels under certain cellular conditions. Therefore, we were additionally interested in the gene expression profiles for the MEGs of interest to decipher their differentiation between the cattle breeds. For this purpose, we manually analysed and then annotated the gene expression profiles of MEGs for each tissue to investigate their expression patterns. A closer look at these gene expression profiles reveals the distinguishing expression patterns for five MEGs (namely *MAPKAPK5*, *CSK*, *DOK2*, *RAC1* and *DNTM1*) expressed over several time points in the liver tissue of both breeds Boran and N'Dama (see Supplementary File S4). Interestingly, these genes are key players in the detection of effectors found in liver tissue (see Supplementary File S4). Gene expression profiles of other MEGs of interest are provided in Supplementary File S5. Figure 4.2 exemplarily shows the changes in the gene expression profile of *MAPKAPK5* for liver tissue of both cattle breeds, harbouring rSNPs in its promoter region. Considering the biological roles, MAPKAPK5 (MAPK Activated Protein Kinase 5), encoded by the gene *MAPKAPK5*, is a serine/threonine protein kinase that plays a major role in the post-transcriptional regulation of MYC, [104, 105] which is intimately associated with immune evasion [106]. The protein encoded by the gene *CSK* plays a critical role in the activation of T-cells and is involved in several pathways, which include the regulation of Src family kinases [107]. Expression of *DOK2* has been reported to regulate the cell cycle of haematopoietic stem cells. Furthermore, the inactivation of *DOK2* also resulted in the aberrant activation of MAP kinase [108], implicating that their functional loss could exacerbate the AAT disease. The protein encoded by *RAC1* (Rac Family Small GTPase 1) is important in regulating cellular processes like phagocytosis of apoptotic cells and binds to effector proteins in their active state [109]. *DNMT1* plays a critical part in regulating the immune system and is regarded indispensable for the inhibition of Foxp3+Treg cells [110].

4.5 Conclusion

Transcription factors are involved in regulating transcription processes by binding to short DNA sequences called transcription factor binding sites (TFBSs). In particular, single nucleotide polymorphisms (SNPs) are widely studied with regard to the disease mechanisms as they can have direct con-

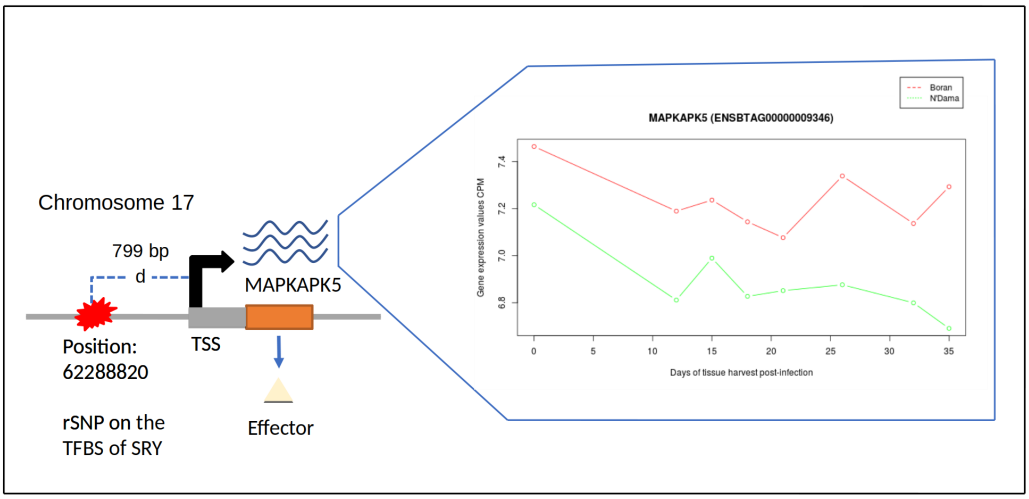

Figure 4.2: An overview of gene expression profile analysis. Schematic representation of rSNP at position 62,288,820 bp of chromosome 17 influencing the gene expression pattern of *MAPKAPK5*. 'd' refers to the distance of the rSNP from the transcription start site (TSS).

trol over the disease susceptibility (causal polymorphisms). Importantly, regulatory SNPs (rSNPs) that are located in the regulatory regions like promoters can significantly affect the gene expression, especially by modifying the binding sites of the TFs. Knowledge about the rSNPs and their complex interplay with the corresponding gene expression and downstream effectors could reveal multiple disease-associated polymorphisms, which can be further used as targets in drug design and breeding programs. Taking the importance of rSNPs and their combinatorial interplay into account, we performed a systematic investigation of genomic and transcriptomic data of two cattle breeds, Boran and N'Dama, to unravel the underlying genetic mechanisms of AAT disease progression. Our findings provide mechanistic insights into significant rSNPs, which are harboured within the promoter regions of MEGs. Moreover, our further investigation of effectors found in the liver, spleen and lymph node tissues of both cattle breeds enhanced our understanding of distinct mechanisms leading to either resistance or susceptibility of cattle breeds. Our current study complements our previous studies, which mainly focused on the upstream events, including TFs and their co-operations as well as master regulators. Taken together, our findings provide a multifaceted glimpse of (i) novel insights into the rSNPs governing the susceptibility/tolerance mechanism of the cattle breeds; (ii) downstream effectors, particularly LYZL2, WSX1 and MTMR4 and their biological roles related to

innate and adaptive immune responses during the AAT disease progression.

Supplementary Information

The supplementary files can be accessed via the original publication (https://www.mdpi.com/article/10.3390/biology11050742/s1). Supplementary File S1: Lists of Monotonically Expressed Genes obtained from the MFSelector approach. Supplementary File S2: Lists of regulatory SNPs obtained for liver-, spleen- and lymph node tissues of Boran and N'Dama. Supplementary File S3: List of breed-specific and tissue-specific MEGs harbouring regulatory SNPs in their promoter regions. Supplementary File S4: Gene expression profiles of monotonically expressed genes *MAPKAPK5*, *CSK*, *DOK2*, *RAC1* and *DNTM1* that harbour regulatory SNPs in their promoter regions. Supplementary File S5: Lists of QR codes of videos for gene expression profiles corresponding to the MEGs harbouring rSNPS associated with liver-, spleen- and lymph node tissues of Boran and N'Dama.

Author contributions

M.G. designed and supervised the research. A.R. together with M.G. participated in the design of the study. Further, A.R. conducted computational analyses, prepared the data sets, implemented the framework and performed the literature survey. S.K., Y.H. and A.O.S. interpreted the results with A.R. and M.G. A.R. and M.G. wrote the final version of the manuscript. M.G. conceived of and managed the project. All authors have read and agreed to the published version of the manuscript.

Data availability

Not applicable.

Acknowledgments

We acknowledge support by the German Research Foundation and the Open Access Publication Funds of the Göttingen University.

Conflicts of Interest

The authors declare no conflict of interest.

Bibliography

[1] R. Brun and J. Blum, "Human african trypanosomiasis.," *Infectious disease clinics of North America*, vol. 26, no. 2, pp. 261–273, 2012.

[2] M. P. Barrett, D. W. Boykin, R. Brun, and R. R. Tidwell, "Human african trypanosomiasis: pharmacological re-engagement with a neglected disease," *British journal of pharmacology*, vol. 152, no. 8, pp. 1155–1171, 2007.

[3] K. Kappmeier, E. Nevill, and R. Bagnall, "Review of tsetse flies and trypanosomosis in south africa," 1998.

[4] P. Van den Bossche, "Some general aspects of the distribution and epidemiology of bovine trypanosomosis in southern africa," *International journal for parasitology*, vol. 31, no. 5-6, pp. 592–598, 2001.

[5] A. Firesbhat and C. Desalegn, "Epidemiology and impacts of trypanosomosis in cattle," *Eur J Appl Sci*, vol. 7, no. 5, pp. 220–5, 2015.

[6] O. M. Radostits, C. C. Gay, K. W. Hinchcliff, and P. D. Constable, *Veterinary Medicine E-Book: A textbook of the diseases of cattle, horses, sheep, pigs and goats*. Elsevier Health Sciences, 2006.

[7] G. J. Losos and B. Ikede, "Review of pathology of diseases in domestic and laboratory animals caused by trypanosoma congolense, t. vivax, t. brucei, t. rhodesiense and t. gambiense," *Veterinary Pathology*, vol. 9, no. 1_suppl, pp. 1–79, 1972.

[8] A. Buguet, J. Bert, P. Tapie, F. Tabaraud, F. Doua, J. Lonsdorfer, P. Bogui, and M. Dumas, "Sleep-wake cycle in human african trypanosomiasis.," *Journal of clinical neurophysiology: official publication of the American Electroencephalographic Society*, vol. 10, no. 2, pp. 190–196, 1993.

[9] F. Rijo-Ferreira, T. Carvalho, C. Afonso, M. Sanches-Vaz, R. M. Costa, L. M. Figueiredo, and J. S. Takahashi, "Sleeping sickness is a circadian disorder," *Nature communications*, vol. 9, no. 1, pp. 1–13, 2018.

[10] A. Rassi Jr, A. Rassi, and J. A. Marin-Neto, "Chagas disease," *The Lancet*, vol. 375, no. 9723, pp. 1388–1402, 2010.

[11] A. Rassi and J. M. de Rezende, "American trypanosomiasis (chagas disease)," *Infectious Disease Clinics*, vol. 26, no. 2, pp. 275–291, 2012.

[12] G. d'Ieteren, E. Authié, N. Wissocq, and M. Murray, "Trypanotolerance, an option for sustainable livestock production in areas at risk from trypanosomosis.," *Revue scientifique et technique (International Office of Epizootics)*, vol. 17, no. 1, pp. 154–175, 1998.

[13] G. Roelants, "Natural resistance to african trypanosomiasis," *Parasite Immunology*, vol. 8, no. 1, pp. 1–10, 1986.

[14] É. Authie, "Contribution à l'étude des mécanismes immunologiques impliqués dans la trypanotolérance des taurins d'afrique," 1993.

[15] J. Ellis, J. Scott, N. D. MacHugh, G. Gettinby, and W. Davis, "Peripheral blood leucocytes subpopulation dynamics during trypanosoma congolense infection in boran and n'dama cattle: an analysis using monoclonal antibodies and flow cytometry," *Parasite immunology*, vol. 9, no. 3, pp. 363–378, 1987.

[16] H. W. Mulligan, W. Porrs, *et al.*, "The african trypanosomiases.," *The African trypanosomiases.*, 1970.

[17] A. Luckins and A. Gray, "An extravascular site of development of trypanosoma congolense," *Nature*, vol. 272, no. 5654, pp. 613–614, 1978.

[18] D. Emery and S. Moloo, "The dynamics of the cellular reactions elicited in the skin of goats by glossina morsitans morsitans infected with trypanosoma (nannomonas) congolense or t.(duttonella) vivax.," *Acta Tropica*, vol. 38, no. 1, pp. 15–28, 1981.

[19] G. Akol, M. Murray, H. Hirumi, K. Hirumi, and S. Moloo, "Infectivity to cattle of metacyclic forms of trypanosoma (nannomonas) congolense propagated in vitro. i. development of localized skin reactions following intradermal inoculation.," *Acta tropica*, vol. 43, no. 3, pp. 207–214, 1986.

[20] B. T. Wellde, R. M. Kovatch, D. A. Chumo, and D. E. Wykoff, "Trypanosoma congolense: thrombocytopenia in experimentally infected cattle," *Experimental Parasitology*, vol. 45, no. 1, pp. 26–33, 1978.

[21] F. Rurangirwa, H. Tabel, G. Losos, W. Masiga, and P. Mwambu, "Immunosuppressive effect of trypanosoma congolense and trypanosoma vivax on the secondary immune response of cattle to mycoplasma mycoides subsp mycoides.," *Research in veterinary science*, vol. 25, no. 3, pp. 395–397, 1978.

[22] F. Rurangirwa, H. Tabel, G. Losos, and I. Tizard, "Suppression of antibody response to leptospira biflexa and brucella abortus and recovery from immunosuppression after berenil treatment," *Infection and Immunity*, vol. 26, no. 3, pp. 822–826, 1979.

[23] F. Rurangirwa, A. Musoke, V. Nantulya, and H. Tabel, "Immune depression in bovine trypanosomiasis: effects of acute and chronic trypanosoma congolense and chronic trypanosoma vivax infections on antibody response to brucella abortus vaccine," *Parasite immunology*, vol. 5, no. 3, pp. 267–276, 1983.

[24] H. Tabel, G. Losos, and M. Maxie, "Experimental bovine trypanosomiasis (trypanosoma vivax and t. congolense). ii. serum levels of total protein, albumin, hemolytic complement, and complement component c3.," *Tropenmedizin und Parasitologie*, vol. 31, no. 1, pp. 99–104, 1980.

[25] P. H. Starkey, "N'dama cattle-a productive trypanotolerant breed," *World Animal Review*, 1984.

[26] D. Courtin, D. Berthier, S. Thevenon, G.-K. Dayo, A. Garcia, and B. Bucheton, "Host genetics in african trypanosomiasis," *Infection, Genetics and Evolution*, vol. 8, no. 3, pp. 229–238, 2008.

[27] H. Bahbahani, B. Salim, F. Almathen, F. Al Enezi, J. M. Mwacharo, and O. Hanotte, "Signatures of positive selection in african butana and kenana dairy zebu cattle," *PloS one*, vol. 13, no. 1, p. e0190446, 2018.

[28] A. Tijjani, Y. T. Utsunomiya, A. G. Ezekwe, O. Nashiru, and O. Hanotte, "Genome sequence analysis reveals selection signatures in endangered trypanotolerant west african muturu cattle," *Frontiers in genetics*, vol. 10, p. 442, 2019.

[29] Y. A. Mekonnen, M. Gültas, K. Effa, O. Hanotte, and A. O. Schmitt, "Identification of candidate signature genes and key regulators associated with trypanotolerance in the sheko breed," *Frontiers in genetics*, vol. 10, p. 1095, 2019.

[30] H. Noyes, A. Brass, I. Obara, S. Anderson, A. L. Archibald, D. G. Bradley, P. Fisher, A. Freeman, J. Gibson, M. Gicheru, *et al.*, "Genetic and expression analysis of cattle identifies candidate genes in pathways responding to trypanosoma congolense infection," *Proceedings of the National Academy of Sciences*, vol. 108, no. 22, pp. 9304–9309, 2011.

[31] P. Fisher, C. Hedeler, K. Wolstencroft, H. Hulme, H. Noyes, S. Kemp, R. Stevens, and A. Brass, "A systematic strategy for large-scale analysis of genotype–phenotype correlations: identification of candidate genes involved in african trypanosomiasis," *Nucleic acids research*, vol. 35, no. 16, pp. 5625–5633, 2007.

[32] O. Hanotte, Y. Ronin, M. Agaba, P. Nilsson, A. Gelhaus, R. Horstmann, Y. Sugimoto, S. Kemp, J. Gibson, A. Korol, *et al.*, "Mapping of quantitative trait loci controlling trypanotolerance in a cross of tolerant west african n'dama and susceptible east african boran cattle," *Proceedings of the National Academy of Sciences*, vol. 100, no. 13, pp. 7443–7448, 2003.

[33] E. W. Hill, G. M. O'Gorman, M. Agaba, J. P. Gibson, O. Hanotte, S. J. Kemp, J. Naessens, P. M. Coussens, and D. E. MacHugh, "Understanding bovine trypanosomiasis and trypanotolerance: the promise of functional genomics," *Veterinary Immunology and Immunopathology*, vol. 105, no. 3-4, pp. 247–258, 2005.

[34] G. M. O'Gorman, S. D. Park, E. W. Hill, K. G. Meade, P. M. Coussens, M. Agaba, J. Naessens, S. J. Kemp, and D. E. MacHugh, "Transcriptional profiling of cattle infected with trypanosoma congolense highlights gene expression signatures underlying trypanotolerance and trypanosusceptibility," *BMC genomics*, vol. 10, no. 1, pp. 1–21, 2009.

[35] M. Gautier, L. Flori, A. Riebler, F. Jaffrézic, D. Laloé, I. Gut, K. Moazami-Goudarzi, and J.-L. Foulley, "A whole genome bayesian scan for adaptive genetic divergence in west african cattle," *BMC genomics*, vol. 10, no. 1, pp. 1–18, 2009.

[36] M. Yaro, K. Munyard, M. Stear, and D. Groth, "Combatting african animal trypanosomiasis (aat) in livestock: the potential role of trypanotolerance," *Veterinary Parasitology*, vol. 225, pp. 43–52, 2016.

[37] A. Rajavel, F. Heinrich, A. O. Schmitt, and M. Gültas, "Identifying cattle breed-specific partner choice of transcription factors during the african trypanosomiasis disease progression using bioinformatics analysis," *Vaccines*, vol. 8, no. 2, p. 246, 2020.

[38] A. Rajavel, A. O. Schmitt, and M. Gültas, "Computational identification of master regulators influencing trypanotolerance in cattle," *International Journal of Molecular Sciences*, vol. 22, no. 2, p. 562, 2021.

[39] G. Shook, "Major advances in determining appropriate selection goals," *Journal of dairy science*, vol. 89, no. 4, pp. 1349–1361, 2006.

[40] M. Ron, G. Israeli, E. Seroussi, J. I. Weller, J. P. Gregg, M. Shani, and J. F. Medrano, "Combining mouse mammary gland gene expression and comparative mapping for the identification of candidate genes for qtl of milk production traits in cattle," *BMC genomics*, vol. 8, no. 1, pp. 1–11, 2007.

[41] J. Ogorevc, T. Kunej, A. Razpet, and P. Dovc, "Database of cattle candidate genes and genetic markers for milk production and mastitis," *Animal genetics*, vol. 40, no. 6, pp. 832–851, 2009.

[42] C. V. Cannistraci, J. Ogorevc, M. Zorc, T. Ravasi, P. Dovc, and T. Kunej, "Pivotal role of the muscle-contraction pathway in cryptorchidism and evidence for genomic connections with cardiomyopathy pathways in rasopathies," *BMC medical genomics*, vol. 6, no. 1, pp. 1–16, 2013.

[43] B. J. Hayes, I. M. MacLeod, H. D. Daetwyler, P. J. Bowman, A. Chamberlian, C. Vander Jagt, A. Capitan, H. Pausch, P. Stothard, X. Liao, *et al.*, "Genomic prediction from whole genome sequence in livestock: the 1000 bull genomes project," in *10. World Congress of Genetics Applied to Livestock Production*, p. np, 2014.

[44] S. Klees, T. M. Lange, H. Bertram, A. Rajavel, J.-S. Schlüter, K. Lu, A. O. Schmitt, and M. Gültas, "In silico identification of the complex interplay between regulatory snps, transcription factors, and their related genes in brassica napus l. using multi-omics data," *International Journal of Molecular Sciences*, vol. 22, no. 2, p. 789, 2021.

[45] S. Klees, F. Heinrich, A. O. Schmitt, and M. Gültas, "agreg-snpdb: A database of regulatory snps for agricultural animal species," *Biology*, vol. 10, no. 8, p. 790, 2021.

[46] F. Heinrich, M. Wutke, P. P. Das, M. Kamp, M. Gültas, W. Link, and A. O. Schmitt, "Identification of regulatory snps associated with vicine and convicine content of vicia faba based on genotyping by sequencing data using deep learning," *Genes*, vol. 11, no. 6, p. 614, 2020.

[47] P. R. Buckland, "The importance and identification of regulatory polymorphisms and their mechanisms of action," *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, vol. 1762, no. 1, pp. 17–28, 2006.

[48] H.-J. Jin, S. Jung, A. R. DebRoy, and R. V. Davuluri, "Identification and validation of regulatory snps that modulate transcription factor chromatin binding and gene expression in prostate cancer," *Oncotarget*, vol. 7, no. 34, p. 54616, 2016.

[49] F. Ramzan, S. Klees, A. O. Schmitt, D. Cavero, and M. Gültas, "Identification of age-specific and common key regulatory mechanisms governing eggshell strength in chicken using random forests," *Genes*, vol. 11, no. 4, p. 464, 2020.

[50] C. C. Chang, C. C. Chow, L. C. Tellier, S. Vattikuti, S. M. Purcell, and J. J. Lee, "Second-generation plink: rising to the challenge of larger and richer datasets," *Gigascience*, vol. 4, no. 1, pp. s13742–015, 2015.

[51] E. Wingender, X. Chen, E. Fricke, R. Geffers, R. Hehl, I. Liebich, M. Krull, V. Matys, H. Michael, R. Ohnhaeuser, *et al.*, "Match-a tool for searching transcription factor binding sites in dna sequences," *Nucl. Acids Res*, vol. 29, pp. 281–283, 2001.

[52] P. Stegmaier, A. Kel, and E. Wingender, "genexplainr: An r interface for the genexplain platform," *Journal of Open Source Software*, vol. 2, no. 18, p. 412, 2017.

[53] M. Krull, S. Pistor, N. Voss, A. Kel, I. Reuter, D. Kronenberg, H. Michael, K. Schwarzer, A. Potapov, C. Choi, *et al.*, "Transpath®: an information resource for storing and visualizing signaling pathways and their pathological aberrations," *Nucleic acids research*, vol. 34, no. suppl_1, pp. D546–D551, 2006.

[54] C. Smith, "Genomics: Snps and human disease.," *Nature*, vol. 435, no. 7044, pp. 993–993, 2005.

[55] M. Singh, P. Singh, P. K. Juneja, S. Singh, and T. Kaur, "Snp–snp interactions within apoe gene influence plasma lipids in postmenopausal osteoporosis," *Rheumatology international*, vol. 31, no. 3, pp. 421–423, 2011.

[56] Y. Sapkota, J. R. Mackey, R. Lai, C. Franco-Villalobos, S. Lupichuk, P. J. Robson, K. Kopciuk, C. E. Cass, Y. Yasui, and S. Damaraju, "Assessing snp-snp interactions among dna repair, modification and metabolism related pathway genes in breast cancer susceptibility," *PloS one*, vol. 8, no. 6, p. e64896, 2013.

[57] V. Ü. Onay, L. Briollais, J. A. Knight, E. Shi, Y. Wang, S. Wells, H. Li, I. Rajendram, I. L. Andrulis, and H. Ozcelik, "Snp-snp interactions in breast cancer susceptibility," *BMC cancer*, vol. 6, no. 1, pp. 1–16, 2006.

[58] A. Moszyńska, M. Gebert, J. F. Collawn, and R. Bartoszewski, "Snps in microrna target sites and their potential role in human disease," *Open biology*, vol. 7, no. 4, p. 170019, 2017.

[59] M. Soares-Silva, F. F. Diniz, G. N. Gomes, and D. Bahia, "The mitogen-activated protein kinase (mapk) pathway: role in immune evasion by trypanosomatids," *Frontiers in microbiology*, vol. 7, p. 183, 2016.

[60] S. Parihar, M. Ozturk, M. Marakalala, D. Loots, R. Hurdayal, D. B. Maasdorp, M. Van Reenen, D. Zak, F. Darboe, A. Penn-Nicholson, *et al.*, "Protein kinase c-delta (pkcδ), a marker of inflammation and tuberculosis disease progression in humans, is important for optimal macrophage killing effector functions and survival in mice," *Mucosal immunology*, vol. 11, no. 2, pp. 496–511, 2018.

[61] L. J. Berg, L. D. Finkelstein, J. A. Lucas, and P. L. Schwartzberg, "Tec family kinases in t lymphocyte development and function," *Annu. Rev. Immunol.*, vol. 23, pp. 549–600, 2005.

[62] M. Felices, M. Falk, Y. Kosaka, and L. J. Berg, "Tec kinases in t cell and mast cell signaling," *Advances in immunology*, vol. 93, pp. 145–184, 2007.

[63] J. A. Readinger, K. L. Mueller, A. M. Venegas, R. Horai, and P. L. Schwartzberg, "Tec kinases regulate t-lymphocyte development and function: new insights into the roles of itk and rlk/txk," *Immunological reviews*, vol. 228, no. 1, pp. 93–114, 2009.

[64] K. Fan, Y. Jia, S. Wang, H. Li, D. Wu, G. Wang, and J.-L. Chen, "Role of itk signalling in the interaction between influenza a virus and t-cells," *Journal of general virology*, vol. 93, no. 5, pp. 987–997, 2012.

[65] J. L. Clements, S. E. Ross-Barta, L. T. Tygrett, T. J. Waldschmidt, and G. A. Koretzky, "Slp-76 expression is restricted to hemopoietic cells of monocyte, granulocyte, and t lymphocyte lineage and is regulated during t cell maturation and activation," *The Journal of Immunology*, vol. 161, no. 8, pp. 3880–3889, 1998.

[66] G. A. Koretzky, F. Abtahian, and M. A. Silverman, "Slp76 and slp65: complex regulation of signalling in lymphocytes and beyond," *Nature Reviews Immunology*, vol. 6, no. 1, pp. 67–78, 2006.

[67] H. Xu and D. R. Littman, "A kinase-independent function of lck in potentiating antigen-specific t cell activation," *Cell*, vol. 74, no. 4, pp. 633–643, 1993.

[68] R. Lattanzio, M. Iezzi, G. Sala, N. Tinari, M. Falasca, S. Alberti, S. Buglioni, M. Mottolese, L. Perracchio, P. G. Natali, *et al.*, "Plc-gamma-1 phosphorylation status is prognostic of metastatic risk in patients with early-stage luminal-a and-b breast cancer subtypes," *BMC cancer*, vol. 19, no. 1, pp. 1–10, 2019.

[69] Y. R. Yang, J. H. Choi, J.-S. Chang, H. M. Kwon, H.-J. Jang, S. H. Ryu, and P.-G. Suh, "Diverse cellular and physiological roles of phospholipase c-γ1," 2012.

[70] J. A. Herman, M. P. Miller, and S. Biggins, "chtog is a conserved mitotic error correction factor," *Elife*, vol. 9, p. e61773, 2020.

[71] C.-W. Jang, Y. Shibata, J. Starmer, D. Yee, and T. Magnuson, "Histone h3. 3 maintains genome integrity during mammalian development," *Genes & development*, vol. 29, no. 13, pp. 1377–1392, 2015.

[72] D. C. Weaver, G. V. Shpakovski, E. Caputo, H. L. Levin, and J. Bocke, "Sequence analysis of closely related retrotransposon families from fission yeast," *Gene*, vol. 131, no. 1, pp. 135–139, 1993.

[73] P. Mellor, L. A. Furber, J. N. Nyarko, and D. H. Anderson, "Multiple roles for the p85α isoform in the regulation and function of pi3k signalling and receptor trafficking," *Biochemical Journal*, vol. 441, no. 1, pp. 23–37, 2012.

[74] R. B. Chagpar, P. H. Links, M. C. Pastor, L. A. Furber, A. D. Hawrysh, M. D. Chamberlain, and D. H. Anderson, "Direct positive regulation of pten by the p85 subunit of phosphatidylinositol 3-kinase," *Proceedings of the National Academy of Sciences*, vol. 107, no. 12, pp. 5471–5476, 2010.

[75] M. Morchikh, A. Cribier, R. Raffel, S. Amraoui, J. Cau, D. Severac, E. Dubois, O. Schwartz, Y. Bennasser, and M. Benkirane, "Hexim1 and neat1 long non-coding rna form a multi-subunit complex that regulates

dna-mediated innate immune response," *Molecular cell*, vol. 67, no. 3, pp. 387–399, 2017.

[76] E. Mascareno, R. Gupta, L. A. Martello, M. Dhar-Mascareno, L. Salciccioli, D. Beckles, M. G. Walsh, F. S. Machado, H. B. Tanowitz, and M. Haseeb, "Rapidly progressive course of trypanosoma cruzi infection in mice heterozygous for hexamethylene bis-acetamide inducible 1 (hexim1) gene," *Microbes and infection*, vol. 20, no. 1, pp. 25–36, 2018.

[77] K. H. Vousden and C. Prives, "Blinded by the light: the growing complexity of p53," *Cell*, vol. 137, no. 3, pp. 413–431, 2009.

[78] A. Efeyan and M. Serrano, "p53: guardian of the genome and policeman of the oncogenes," *Cell cycle*, vol. 6, no. 9, pp. 1006–1010, 2007.

[79] E. A. Komarova, V. Krivokrysenko, K. Wang, N. Neznanov, M. V. Chernov, P. G. Komarov, M.-L. Brennan, T. V. Golovkina, O. Rokhlin, D. V. Kuprash, *et al.*, "p53 is a suppressor of inflammatory response in mice," *The FASEB journal*, vol. 19, no. 8, pp. 1030–1032, 2005.

[80] J. H. Madenspacher, K. M. Azzam, K. M. Gowdy, K. C. Malcolm, J. A. Nick, D. Dixon, J. J. Aloor, D. W. Draper, J. J. Guardiola, M. Shatz, *et al.*, "p53 integrates host defense and cell fate during bacterial pneumonia," *Journal of Experimental Medicine*, vol. 210, no. 5, pp. 891–904, 2013.

[81] J. H. Chung, "The role of dna-pk in aging and energy metabolism," *The FEBS journal*, vol. 285, no. 11, pp. 1959–1972, 2018.

[82] B. J. Ferguson, D. S. Mansur, N. E. Peters, H. Ren, and G. L. Smith, "Dna-pk is a dna sensor for irf-3-dependent innate immunity," *elife*, vol. 1, p. e00047, 2012.

[83] J. A. Black, K. Crouch, L. Lemgruber, C. Lapsley, N. Dickens, L. R. Tosi, J. C. Mottram, and R. McCulloch, "Trypanosoma brucei atr links dna damage signaling during antigenic variation with regulation of rna polymerase i-transcribed surface antigens," *Cell reports*, vol. 30, no. 3, pp. 836–851, 2020.

[84] N. Baker, C. M. Catta-Preta, R. Neish, J. Sadlova, B. Powell, E. V. Alves-Ferreira, V. Geoghegan, J. B. Carnielli, K. Newling, C. Hughes, *et al.*, "Systematic functional analysis of leishmania protein kinases identifies regulators of differentiation or survival," *Nature communications*, vol. 12, no. 1, pp. 1–15, 2021.

[85] J. Xu, R. Cai, L. Lu, C. Duan, X. Tao, D. Chen, Y. Liu, X. Wang, M. Cao, and Y. Chen, "Genetic regulatory network analysis reveals that low density lipoprotein receptor-related protein 11 is involved in stress responses in mice," *Psychiatry research*, vol. 220, no. 3, pp. 1131–1137, 2014.

[86] M. Seguel, D. Perez-Venegas, J. Gutierrez, D. E. Crocker, and E. J. DeRango, "Parasitism elicits a stress response that allocates resources for immune function in south american fur seals (arctocephalus australis)," *Physiological and Biochemical Zoology*, vol. 92, no. 3, pp. 326–338, 2019.

[87] A. Fleming, "On a remarkable bacteriolytic element found in tissues and secretions," *Proceedings of the Royal Society of London. Series B, Containing Papers of a Biological Character*, vol. 93, no. 653, pp. 306–317, 1922.

[88] C. J. Ng, D. J. Wadleigh, A. Gangopadhyay, S. Hama, V. R. Grijalva, M. Navab, A. M. Fogelman, and S. T. Reddy, "Paraoxonase-2 is a ubiquitously expressed protein with antioxidant properties and is capable of preventing cell-mediated oxidative modification of low density lipoprotein," *Journal of biological chemistry*, vol. 276, no. 48, pp. 44444–44449, 2001.

[89] N. Bourquard, C. J. Ng, and S. T. Reddy, "Impaired hepatic insulin signalling in pon2-deficient mice: a novel role for the pon2/apoe axis on the macrophage inflammatory response," *Biochemical Journal*, vol. 436, no. 1, pp. 91–100, 2011.

[90] A. Devarajan, N. Bourquard, V. R. Grijalva, F. Gao, E. Ganapathy, J. Verma, and S. T. Reddy, "Role of pon2 in innate immune response in an acute infection model," *Molecular genetics and metabolism*, vol. 110, no. 3, pp. 362–370, 2013.

[91] C. A. Sprecher, F. J. Grant, J. W. Baumgartner, S. R. Presnell, S. K. Schrader, T. Yamagiwa, T. E. Whitmore, P. J. O'Hara, and D. F. Foster, "Cloning and characterization of a novel class i cytokine receptor," *Biochemical and biophysical research communications*, vol. 246, no. 1, pp. 82–90, 1998.

[92] S. Pflanz, L. Hibbert, J. Mattson, R. Rosales, E. Vaisberg, J. F. Bazan, J. H. Phillips, T. K. McClanahan, R. de Waal Malefyt, and

R. A. Kastelein, "Wsx-1 and glycoprotein 130 constitute a signal-transducing receptor for il-27," *The Journal of Immunology*, vol. 172, no. 4, pp. 2225–2231, 2004.

[93] A. Villarino, L. Hibbert, L. Lieberman, E. Wilson, T. Mak, H. Yoshida, R. A. Kastelein, C. Saris, and C. A. Hunter, "The il-27r (wsx-1) is required to suppress t cell hyperactivity during infection," *Immunity*, vol. 19, no. 5, pp. 645–655, 2003.

[94] D. Reczek, M. Schwake, J. Schröder, H. Hughes, J. Blanz, X. Jin, W. Brondyk, S. Van Patten, T. Edmunds, and P. Saftig, "Limp-2 is a receptor for lysosomal mannose-6-phosphate-independent targeting of β-glucocerebrosidase," *Cell*, vol. 131, no. 4, pp. 770–783, 2007.

[95] D. Neculai, M. Schwake, M. Ravichandran, F. Zunke, R. F. Collins, J. Peters, M. Neculai, J. Plumb, P. Loppnau, J. C. Pizarro, *et al.*, "Structure of limp-2 provides functional insights with implications for sr-bi and cd36," *Nature*, vol. 504, no. 7478, pp. 172–176, 2013.

[96] S. Heybrock, K. Kanerva, Y. Meng, C. Ing, A. Liang, Z.-J. Xiong, X. Weng, Y. A. Kim, R. Collins, W. Trimble, *et al.*, "Lysosomal integral membrane protein-2 (limp-2/scarb2) is involved in lysosomal cholesterol export," *Nature communications*, vol. 10, no. 1, pp. 1–12, 2019.

[97] H. Mori, S.-i. Sakakibara, T. Imai, Y. Nakamura, T. Iijima, A. Suzuki, Y. Yuasa, M. Takeda, and H. Okano, "Expression of mouse igf2 mrna-binding protein 3 and its implications for the developing central nervous system," *Journal of neuroscience research*, vol. 64, no. 2, pp. 132–143, 2001.

[98] Y. Natkunam, G. Vainer, J. Chen, S. Zhao, R. J. Marinelli, A. S. Hammer, S. Hamilton-Dutoit, E. Pikarsky, G. Amir, R. Levy, *et al.*, "Expression of the rna-binding protein vickz in normal hematopoietic tissues and neoplasms," *Haematologica*, vol. 92, no. 2, pp. 176–183, 2007.

[99] S. B. Patel and M. Bellini, "The assembly of a spliceosomal small nuclear ribonucleoprotein particle," *Nucleic acids research*, vol. 36, no. 20, pp. 6482–6493, 2008.

[100] H. R. LUO, G. A. MOREAU, N. LEVIN, and M. J. MOORE, "The human prp8 protein is a component of both u2-and u12-dependent spliceosomes," *Rna*, vol. 5, no. 7, pp. 893–908, 1999.

[101] S.-Y. Lee, E.-J. Park, S. Kim, and C.-W. Lee, "Ssu72 phosphatase is involved in immunometabolism," 2018.

[102] J. S. Silva, D. R. Twardzik, and S. G. Reed, "Regulation of trypanosoma cruzi infections in vitro and in vivo by transforming growth factor beta (tgf-beta).," *The Journal of experimental medicine*, vol. 174, no. 3, pp. 539–545, 1991.

[103] L. Wang, Q. Li, S. Ni, Y. Huang, J. Wei, J. Liu, Y. Yu, S. Wang, and Q. Qin, "The roles of grouper clathrin light chains in regulating the infection of a novel marine dna virus, singapore grouper iridovirus," *Scientific reports*, vol. 9, no. 1, pp. 1–14, 2019.

[104] H. Ni, X. S. Wang, K. Diener, and Z. Yao, "Mapkapk5, a novel mitogen-activated protein kinase (mapk)-activated protein kinase, is a substrate of the extracellular-regulated kinase (erk) and p38 kinase," *Biochemical and biophysical research communications*, vol. 243, no. 2, pp. 492–496, 1998.

[105] T. R. Kress, I. G. Cannell, A. B. Brenkman, B. Samans, M. Gaestel, P. Roepman, B. M. Burgering, M. Bushell, A. Rosenwald, and M. Eilers, "The mk5/prak kinase and myc form a negative feedback loop that is disrupted during colorectal tumorigenesis," *Molecular cell*, vol. 41, no. 4, pp. 445–457, 2011.

[106] S. Casacuberta-Serra and L. Soucek, "Myc and ras, the bonnie and clyde of immune evasion," *Translational cancer research*, vol. 7, no. Suppl 4, p. S457, 2018.

[107] M. Okada, "Regulation of the src family kinases by csk," *International journal of biological sciences*, vol. 8, no. 10, p. 1385, 2012.

[108] M. Niki, A. Di Cristofano, M. Zhao, H. Honda, H. Hirai, L. Van Aelst, C. Cordon-Cardo, and P. P. Pandolfi, "Role of dok-1 and dok-2 in leukemia suppression," *The Journal of experimental medicine*, vol. 200, no. 12, pp. 1689–1695, 2004.

[109] S.-Y. Kim, S. Kim, D.-J. Bae, S.-Y. Park, G.-Y. Lee, G.-M. Park, and I.-S. Kim, "Coordinated balance of rac1 and rhoa plays key roles in determining phagocytic appetite," *PLoS One*, vol. 12, no. 4, p. e0174603, 2017.

[110] L. Wang, Y. Liu, U. H. Beier, R. Han, T. R. Bhatti, T. Akimova, and W. W. Hancock, "Foxp3+ t-regulatory cells require dna methyltransferase 1 expression to prevent development of lethal autoimmunity," *Blood, The Journal of the American Society of Hematology*, vol. 121, no. 18, pp. 3631–3639, 2013.

Chapter 5

Unravelling the Complex Interplay of Transcription Factors Orchestrating Seed Oil Content in *Brassica napus* L.

Abirami Rajavel [1], Selina Klees [1], Johanna-Sophie Schlüter [1], Hendrik Bertram [1], Kun Lu [2,3,4], Armin Otto Schmitt [1,5] and Mehmet Gültas [1,5,*]

[1] Breeding Informatics Group, Department of Animal Sciences, Georg-August University, Margarethe von Wrangell-Weg 7, 37075 Göttingen, Germany
[2] College of Agronomy and Biotechnology, Southwest University, Beibei, Chongqing 400715, China
[3] Academy of Agricultural Sciences, Southwest University, Beibei, Chongqing 400715, China
[4] State Cultivation Base of Crop Stress Biology, Southern Mountainous Land of Southwest University, Beibei, Chongqing 400715, China
[5] Center for Integrated Breeding Research (CiBreed), Albrecht-Thaer-Weg 3, Georg-August University, 37075 Göttingen, Germany
[*] Author to whom correspondence should be addressed.

Bibliographic Information

Rajavel, A.; Klees, S.; Schlüter, J.S.; Bertram, H.; Lu, K.; Schmitt, A.O.; Gültas, M. Unravelling the complex interplay of transcription factors orches-

trating the seed oil content in Brassica napus. Int. J. Mol. Sci. 2021, 21. https://doi.org/10.3390/ijms22031033.

Author contribution of Abirami Rajavel

The author contributed to the design of the study, and conducted computational analyses. Furthermore, the author interpreted the results and wrote a significant part of the final version of the manuscript.

5.1 Abstract

Transcription factors (TFs) and their complex interplay are essential for directing specific genetic programs, such as responses to environmental stresses, tissue development, or cell differentiation by regulating gene expression. Knowledge regarding TF–TF cooperations could be promising in gaining insight into the developmental switches between the cultivars of *Brassica napus* L., namely Zhongshuang11 (ZS11), a double-low accession with high-oil-content, and Zhongyou821 (ZY821), a double-high accession with low-oil-content. In this regard, we analysed a time series RNA-seq data set of seed tissue from both of the cultivars by mainly focusing on the monotonically expressed genes (MEGs). The consideration of the MEGs enables the capturing of multi-stage progression processes that are orchestrated by the cooperative TFs and, thus, facilitates the understanding of the molecular mechanisms determining seed oil content. Our findings show that TF families, such as NAC, MYB, DOF, GATA, and HD-ZIP are highly involved in the seed developmental process. Particularly, their preferential partner choices as well as changes in their gene expression profiles seem to be strongly associated with the differentiation of the oil content between the two cultivars. These findings are essential in enhancing our understanding of the genetic programs in both cultivars and developing novel hypotheses for further experimental studies.

5.2 Introduction

Oil crops have been gaining great economic importance in agriculture as well as in the trade world during the past years [1, 2], and the consumption of vegetable oil is anticipated to double by the year 2030 [3]. *Brassica napus* L. (rapeseed or canola) is the third largest source of oilseed crop, which is widely cultivated across the globe [4, 5, 6]. The seeds of *B. napus* are rich in oil content and fatty acids, which include primarily oleic and linoleic acid [7]. However, erucic acid and glucosinolates are anti-nutritive compounds that are present in the *B. napus* seeds that are not desirable for human consumption or as fodder for animal consumption [8, 9]. Therefore, enhancing the seed quality with improved oil content has become the major selective trait for rapeseed breeding programs due to the growing global demand for oil production, for their use as bio-fuel, animal fodder, and vegetable oil [10, 11, 12].

The seeds of *B. napus* are an excellent reservoir of triacylglycerol (TAG), the primary storage form of oil that is essential for the seedling growth fol-

lowed by seedling germination [13, 14, 15, 16, 17, 18]. Recent studies have proposed that the oil content of the seeds could be enhanced by varying the expression levels of individual or a combination of genes encoding transcription factors/enzymes that are associated with TAG metabolism [14, 19]. To this end, transcriptome studies have been extensively carried out in order to understand the underlying molecular mechanism regulating the seed oil content of *B. napus* [20, 21, 22, 23]. For this purpose, Xiao *et al.* [24] identified candidate genes that are involved in regulating the oil content by combining genome-wide association studies and transcriptome analysis in *B. napus*. They performed a comparison study between two extremely high-oil-content lines and one extremely low-oil-content line and identified differentially expressed genes (DEGs) between the lines, contributing to seed oil content. On the other hand, Qu *et al.* [25] analysed the metabolic profiles of genes that are involved in the flavanoid synthesis in both yellow- and black-seeded rapeseed accessions at the early, middle, and late stages of seed development, and compared the transcriptional differences between them by RNA-sequencing. Moreover, Niu *et al.* [26] performed pairwise comparisons and also identified DEGs regulating seed size, colour, and oil content in *Brassica rapa*, by taking the seven developmental stages of the seeds into account. In this regard, a recent study conducted by Lu *et al.* [27] integrated genome-wide association studies and transcriptome analyses, and mainly focused on the identification of DEGs that are related to environmental adaptation, oil content, seed quality, and ecotype improvement for two cultivars of *B. napus*: a double-low accession with high-oil-content and a double-high accession with low-oil-content. Several of the aforementioned studies specifically focused on the GO categories and pathway enrichment anaylses based on the identified DEG sets, while primarily investigating the biological functions of the DEGs regarding seed oil content.

However, today it is well known that transcription factors (TFs) and their complex interplay ccontrol gene expression. Until now, several studies showed that TFs in plants are key regulatory elements controlling the expression of several genes, thereby regulating a variety of essential biological processes, like growth, tissue development, differentiation, metabolism, homeostasis, and adaptation to the environment [28, 29, 30, 31, 32, 33, 34]. Especially in terms of developmental switches and specifying cellular fate in eukaryotes, the orchestration of cell differentiation changes its direction, depending on the specific gene regulatory events that are governed by TFs and their preferential partner choices (for a review see [35]). Thus, the knowledge regarding TFs and their cooperations could provide new insight into the genetic programs regulating various biological processes.

Despite the rich literature on TFs and their cooperations, today there is still

a need to unravel the complex interplay of transcription factors orchestrating the seed oil content in *B. napus*. For this purpose, in this study we analysed a time series RNA-seq data set of seed tissue of two *B. napus* cultivars: Zhongshuang11 (ZS11), a double-low accession with high-oil- content and Zhongyou821 (ZY821), a double-high accession with low-oil-content. Unlike the previous studies [24, 25, 26, 27], we investigated the genes with monotonic expression patterns, known as Monotonically Expressed Genes (MEGs), in order to capture the multi-stage progression during seed development. The consideration of the MEGs is promising for capturing the multi-stage progression processes that are directed by the combinatorial interplay of the TFs [36] and, thus, facilitates the understanding of the molecular mechanisms determining the seed oil content. We computationally identified the interplay between the TFs for both cultivars in order to decipher the gene regulatory mechanisms controlling the specific expression pattern of MEGs. Our results show that a vast majority of the TFs are overlapping in both cultivars, while few TFs changes their partners, which could be controlling the switches of developmental programs regarding the oil content of both cultivars.

5.3 Materials and Methods

Our analysis follows the structure, as presented in Figure 5.1.

5.3.1 *B. napus* Data Set and Data Preparation

In this study, we use a publicly available time series transcriptomic data set of the seeds that were obtained from two *B. napus* cultivars with different seed oil content, namely Zhongshuang11 (ZS11), a double-low accession with high-oil-content, and Zhongyou821 (ZY821), a double-high accession with low-oil-content. The RNA-sequencing of the seeds from both cultivars at four different time points, namely day 7, day 10, day 15, and day 45 after flowering, with two biological replicates each were generated by Lu *et al.* [27]. The raw sequencing data are available from the BIG Data Center under the BioProject accession code PRJCA001246.
Following the original study [27], we mapped the filtered reads to the *B. napus* reference genome version 4.1 (obtained from [39] and available under https://wwwdev.genoscope.cns.fr/brassicanapus/data/) while using STAR 2.4.4a [40]. We then obtained the gene counts from the aligned sequence reads by applying the htseq-count program [41]. In total, the data set comprises raw count values for 80,927 genes and 16 samples (four time points with two biological replicates each for two cultivars). Finally, the raw counts

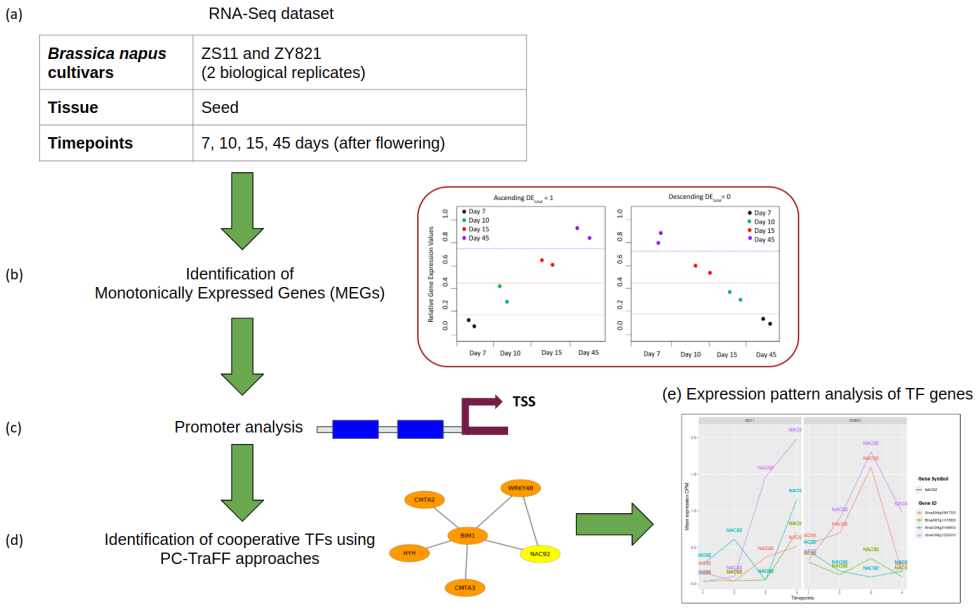

Figure 5.1: Flowchart of analysis. (**a**) Processing of the RNA-Seq dataset; (**b**) identification of Monotonically Expressed Genes using the MFSelector approach [37]; (**c**) promoter analysis (TSS: transcription start site); (**d**) identification of cooperative transcription factors (TFs) using the PC-TraFF approach [38]; and, (**e**) Expression pattern analysis of TF genes.

were normalized while using the R function `voom` with normalization method "cyclicloess" of the package `limma` (version 3.40.6) [42] in order to obtain the counts-per-million (CPM) normalized values.

5.3.2 Identification of Monotonically Expressed Genes

By applying the MFSelector (monotonic feature selector) approach [37] to the time series data set obtained from RNA-seq consisting of four time points, we identified Monotonically Expressed Genes (MEGs), whose expression patterns are closely linked to the development of the seeds over time in both cultivars. The underlying algorithm of MFSelector compares the expression values of each gene observed for all of the samples between time points in order to assess whether these values follow a strong monotonic pattern. In addition, a permutation test is performed to determine the significance level of these patterns. Consequently, it provides two sets of MEGs with corresponding p-values: while the first set contains the genes with significantly monotonically increasing patterns, the second set only includes genes with significantly monotonically decreasing expression values.

In MFSelector, the parameters *permut*, *svdetimes*, and *svdenoise* have to be specified in order to define the level of stringency for monotonicity. Following our previous study [36], we chose *permut* = 100, *svdetimes* = 100, and *svdenoise* = 0.1. Finally, we considered, in our further analysis, only the genes as significant MEGs that have a FDR ≤ 0.1 and sample variance for discriminating error value ≤ 1 .

5.3.3 Gene Ontology Enrichment Analysis

While using PlantGSEA (http://systemsbiology.cau.edu.cn/PlantGSEA/) [43], we performed the gene set enrichment analysis in order to obtain GO (Gene Ontology) terms on "biological process" regarding the ascending and descending set of MEGs for both ZS11 and ZY821 cultivars. Fisher's exact test was applied and a GO term was considered to be statistically significant if its corresponding FDR value ≤ 0.05. The enriched GO terms were visualised as tree map while using REVIGO [44].

5.3.4 Identification of Transcription Factor Cooperations

For the identification of transcription factors (TFs) that cooperatively bind to promoter regions of genes, we applied the PC-TraFF approach [38]. The PC-TraFF is an information theory based methodology that applies the pointwise mutual information (PMI) metric in order to quantify the cooperation level of TFs, according to the co-occurrence of their binding sites in the promoters of the MEGs [38]. Its underlying algorithm consists of six phases and it requires the following input parameters:

- Promoter sequences: Similar to our study [28], we extracted, for each MEG, its corresponding promoter sequence ranging from -500 bp to $+100$ bp regions relative to a transcription start site while using the reference genome version 4.1 and gene annotation given in [39].

- Transcription Factor Binding Site (TFBS) detection: For the detection of putative TFBSs in the promoter sequences, we employed the MATCH™ program [45] with a non-redundant plant position weight matrix (PWM) library from the JASPAR database [46].

- Pre-defined distances: For the identification of the regular cooperative binding pattern of TFs, the underlying PC-TraFF algorithm requires the pre-defined minimum and maximum distance thresholds between

TFBSs. In our analysis, we used the recommended values of a distance ≤ 20 for the maximum and ≥ 5 for the minimum distance.

Consequently, the algorithm of the PC-TraFF approach assigns each TF-pair (T_a and T_b) a PMI $(T_a; T_b)$-score and it transforms the PMI $(T_a; T_b)$-score into the *z-score* as a final step. A cooperation between any T_a and T_b is considered to be significant if they have a *z-score* ≥ 3.

5.3.5 Expression Pattern Analysis of TF Genes

Following the analysis strategy of Zeidler *et al.* [47], we analysed the changes in the expression values of TF genes during the seed developmental stages to gather knowledge on the combinatorial gene regulation underlying the differentiation of the oil content between the two cultivars. For this purpose, we determined, for each TF, whose gene symbol is often defined in *Arabidopsis thaliana* or *Zea mays*, its orthologous genes in *B. napus* while using the Ensembl plant database [48, 49].

5.4 Results and Discussion

5.4.1 Identification and Analysis of MEGs

Analysing the time series data set of *B. napus* seeds from the two cultivars ZS11 and ZY821, we have obtained the MEGs for each cultivar, which are monotonically expressed either in ascending or descending patterns during the seed development. Table 5.1 provides the numbers of MEGs obtained for each cultivar and the obtained MEGs are listed in Supplementary Table S1. Table 5.1 shows that there is a comparatively large number of MEGs (both ascending and descending MEGs) for the ZS11 cultivar in comparison to these of the ZY821 cultivar.

Table 5.1: Numbers of significant Monotonically Expressed Genes in ascending and descending order for the seeds of two cultivars of *B. napus* namely ZS11 and ZY821.

	ZS11		**ZY821**	
	Ascending	Descending	Ascending	Descending
Genes	4310	2344	1734	1021

Further analysis of the MEGs reveals that, while the vast majority of MEGs are primarily unique to a particular cultivar, only a small number of genes are overlapping between the gene sets of ZS11 and ZY821 (see Figure 5.2).

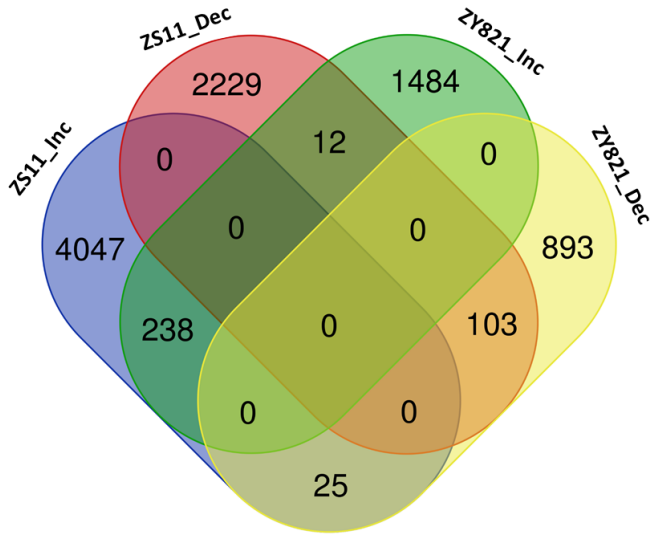

Figure 5.2: Venn-diagram of the MEGs expressed in ascending and descending orders for the seeds of two cultivars ZS11 and ZY821. (ZS11_Inc: Ascendings MEGs for ZS11; ZS11_Dec: Descending MEGs for ZS11 ; ZY821_Inc: Ascendings MEGs for ZY821; ZY821_Dec: Descending MEGs for ZY821) (visualised with http://bioinformatics.psb.ugent.be/webtools/Venn/)

Moreover, we performed a gene set enrichment analysis [43] while using the MEGs to obtain deeper insight into their crucial biological functions and clustered these functions based on the GO terms.

The GO enrichment results regarding the MEGs of ZS11 cultivar revealed that the ascending MEGs are significantly enriched mainly in the term "fatty acid metabolism" (see Figure S1), which is highly associated with oil content. Other enriched GO terms, including "sucrose biosynthetic process", "glycerophospholipid synthetic process", "sphingolipid metabolic process", and "galactose metabolic process" are greatly relevant to fatty acid synthesis and accumulation processes. In contrast, for the descending MEGs of the ZS11 cultivar (see Figure S2) , the GO term "protein phosphorylation" is significantly enriched that represents particularly the inverse correlation of oil and protein levels [50] in rapeseed. The GO terms indicate that the increasing pattern of MEGs in fatty acid metabolism could contribute to seed oil content of the rapeseed in ZS11 cultivar.

On the other hand, the enrichment analysis of ascending MEGs that were obtained for ZY821 cultivar showed that several significantly enriched GO terms are linked to the "thiamine metabolism", which is associated with

the adaptation to biotic and abiotic stress [43] (see Figure S3). Reversely, for the descending MEGs, GO terms that are related to biological processes, such as "lipid metabolism", "carbohydrate metabolism", "GDP-mannose metabolism", and "sulphur compound metabolism" show that genes that are involved in multiple metabolic processes are following the decreasing pattern in the seeds of ZY821 cultivar during the germination process, especially of lipids (see Figure S4). These results imply that the seeds of ZY821 cultivar might be involved in stress responses while diminishing the other metabolic processes, especially those that are related to oil synthesis.

5.4.2 Cooperative TFs Regulating Seed Developmental Processes of ZS11 and ZY821

For the better understanding of the transcriptional regulation underlying oil synthesis in the seeds of *B. napus* and, thereby deciphering specific regulatory programs differentiating between double low and double high cultivars of *B. napus*, we analysed the promoter regions of the MEGs of both cultivars of *B. napus*. Consequently, we identified cooperative transcription factors (TFs) using the algorithm of PC-TraFF, similar to our other studies [38, 47, 52, 53]. The obtained cooperative TF pairs are depicted as cooperation networks in which the elliptical nodes represent the individual TFs, whereas the edges (grey lines in the cooperation networks) represent the cooperation between the TFs (see Figures 5.3 and 5.4). The cooperation networks comprise 42 and 54 cooperative TF pairs for ZS11 and ZY821, respectively. The overlapping TFs between the two cooperation networks are shaded in orange and TFs of the cooperation network unique to a particular cultivar are shaded in yellow. Remarkably, a brief analysis of the cooperation networks reveals that there are only five and 17 TFs unique to ZS11 and ZY821, respectively.

5.4.3 Cooperation Network of ZS11 and ZY821

Taking a closer look into the cooperation networks of ZS11 and ZY821 reveals that the majority of the single TFs are overlapping in the networks of both cultivars, while a few of them change their partners. Among several transcription factors in the cooperation networks of ZS11 and ZY821, a majority of them belong predominantly to five TF families: NAC, MYB, DOF, GATA, and the HD-ZIP family. Hence, in our further analysis, we mainly focused on the members of these TF families and their preferential partner choices in order to explain, in detail, their relevance for fatty acid synthesis, transport, and accumulation in the seed tissue of *B. napus* in the two cultivars. The ex-

pression profiles of transcription factors that are present in the cooperation networks are provided in Supplementary Information Figures S1– S36.

5.4.3.1 NAC Family of Transcription Factors

The NAC family members NAC92 and ANAC050 are found with different partners in the cooperation networks of ZS11 and ZY821 (see Figures 5.3 and 5.4). This family encodes NAC (NAM, ATAF1, -2, and CUC2) domain transcription factors that are exclusively found in plants and well-studied for their functioning in abiotic stress responses [54, 55, 56, 57, 58] and defense mechanisms [54, 55, 58]. In Figure 5.3, we observed that, while NAC92 cooperates with the transcription factors BIM1 and WRKY48, the factor ANAC050 cooperates with AT3G24120. Particularly, NAC92 has been reported in several plant species in controlling the age-dependent senescence and seed germination processes [59]. Therefore, it could play a crucial part in determining the seed oil content, as senescence directly affects the quality of seeds. On the other hand, ANAC050 has been studied in transcriptional repression and flowering time control [60] and, thus, its cooperation with AT3G24120 (see Figure 5.4) could potentially play negative roles in fatty acid accumulation.
Moreover, when considering the orthologous genes of NAC92 in *B. napus*, we identified four gene IDs (*BnaA04g09470D*, *BnaA07g14730D*, *BnaC04g31690D*, and *BnaC06g12550D*) while using the Ensembl plant database [48, 49]. The gene expression values of these genes show different patterns in both cultivars, as shown in Figure 5.5. Especially, *BnaC06g12550D* is clearly showing an increasing trend after 10 days of flowering in ZS11 (time point 2 in Figure 5.5), while its expression level, together with that of *BnaA04g09470D*, is decreasing in ZY821 during the late stage (day 45 after flowering) of seed development. However, the analysis of ANAC050 orthologous genes in *B. napus* reveals that their expression patterns are similar in both of the cultivars (see Supplemetary Information Figure S1).

5.4.3.2 GATA Transcription Factors

Another interesting family of transcription factors observed in the networks of ZS11 and ZY821 is the GATA family of transcription factors (namely GATA8, GATA12, and GATA15), which contain type-IV zinc finger motifs. GATA transcription factors have been identified in the regulatory regions of the light-responsive genes [61], especially genes that are associated with photosynthesis e.g., the chlorophyll a/b binding protein [62, 63]. They play

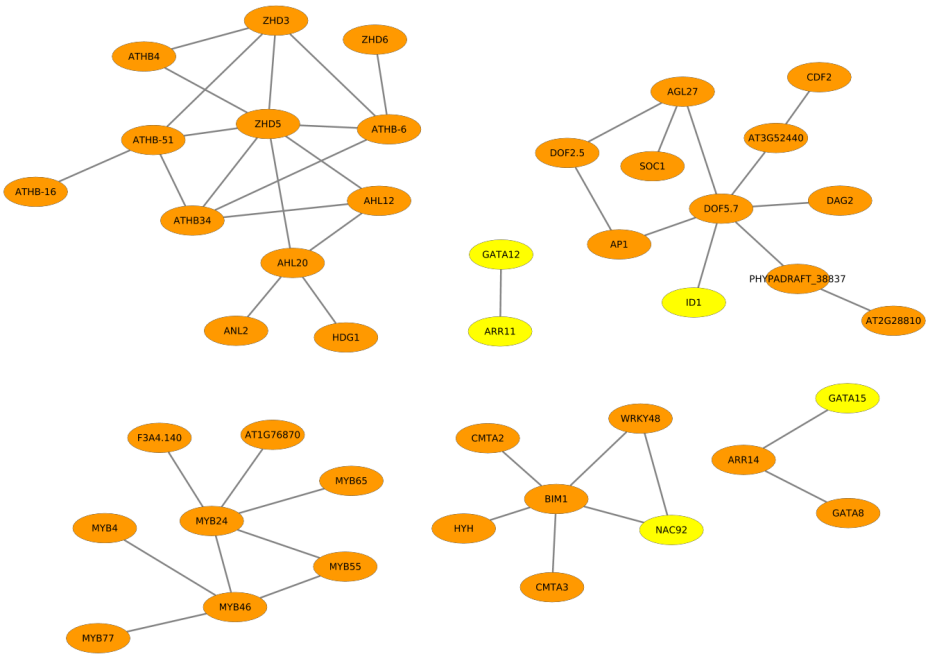

Figure 5.3: Cooperation network of TF pairs identified for the double low accession ZS11 cultivar. The orange shaded nodes represent the overlapping TFs between ZS11 and ZY821, whereas the yellow shaded nodes represent the TFs that are unique for the ZS11 cultivar.

a pivotal role as regulators that are involved in the nitrogen assimilation process in plants [64, 65]. Additionally, few members of the GATA family have also been identified as a differentially expressed TF family while comparing high- and low-yielding oil palm [66], which could also explain their major role in rapeseed. In particular, we identified GATA8 with its cooperation partner ARR14 for both cultivars, while GATA12 as well as GATA15 with their cooperation partners ARR11 and ARR14, have only been found for the ZS11 cultivar (see Figures 5.3 and 5.4). This finding suggests that, during the seed germination processs, these three GATA family members form dimers with ARR family TFs, which play essential roles in stress responses (involving triacylglycerol) [67]. These TF cooperations in the seed tissue of ZS11 might be directing the regulatory processes that are involved in fatty acid synthesis and accumulation processes.
Furthermore, we observed a strong increment in the expression levels of *BnaA07g16490D* and *BnaA09g34590D* until day 15 after flowering in the ZS11 cultivar in comparison to ZY821 while taking the expression profiles of five GATA8 orthologous genes (*BnaA07g16490D*, *BnaC08g25560D*,

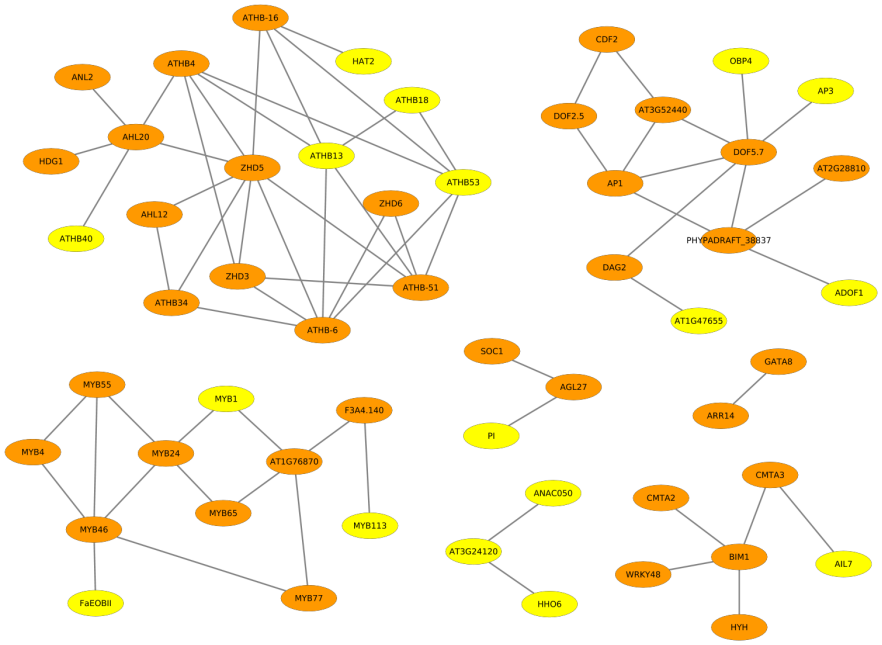

Figure 5.4: Cooperation network of TF pairs identified for the double high accession ZY821 cultivar. The orange shaded nodes represent the overlapping TFs between ZS11 and ZY821, whereas the yellow shaded nodes represent the TFs that are unique for the ZY821 cultivar.

BnaC04g25920D, *BnaC06g15420D*, and *BnaA09g34590D*) into account (see Figure 5.6). Interestingly, in both cultivars, the gene expression values of all five orthologous genes abruptly decreases during the late stage of seed development. Similar patterns of the changes in the expression values during the seed development have been obtained for GATA12 and GATA15 orthologous genes (see Supplementary Information Figures S2 and S3).

5.4.3.3 DOF Family of Transcription Factors

Importantly, we identified several DOF family members (DOF2.5, DOF5.7, DAG2, CDF2, AT2G28810, AT3G52440, OBP4, ADOF1, AT1G47655) in the cooperation networks of ZS11 and ZY821 (see Figures 5.3 and 5.4). This family of transcription factors, encoding zinc finger protein, is specific for plants, and it is not found in other eukaryotes [68]. They have been found to be particularly implicated in controlling seed germination, seed storage, and the mobilisation of proteins and fatty acids [69]. Interestingly, DOF2.5/DAG2 acts as a positive regulator of seed germination [70]. Although the functions of DOFs are not well-characterised with regard to seed

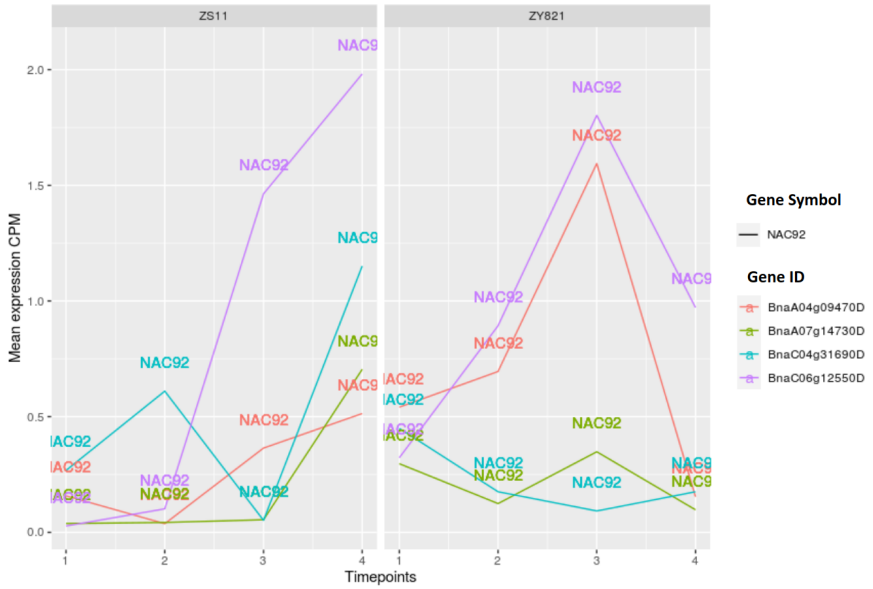

Figure 5.5: Expression values of NAC92 orthologous genes. Time points 1 to 4 represents day 7, day 10, day 15, and day 45 after flowering.

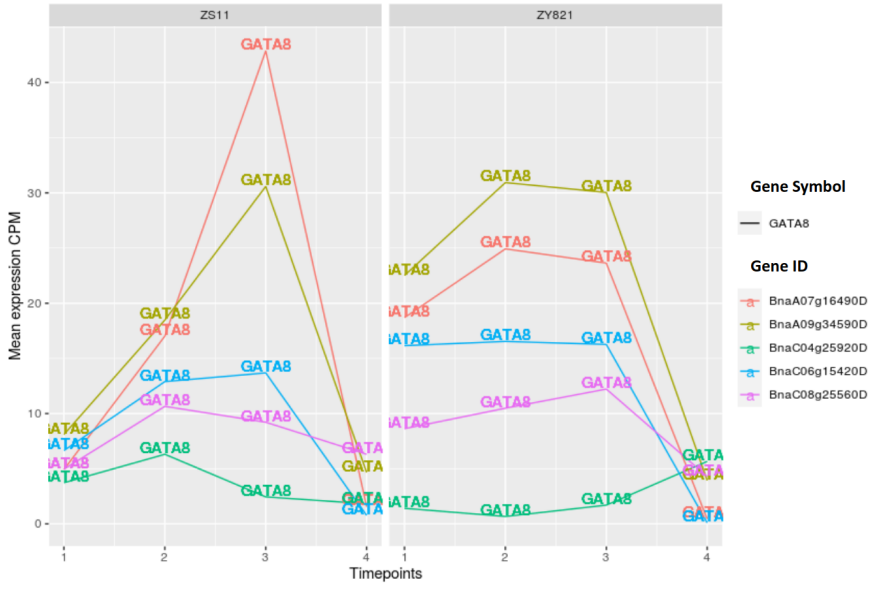

Figure 5.6: Expression values of GATA8 orthologous genes. Time points 1 to 4 represents day 7, day 10, day 15, and day 45 after flowering.

oil content in *Arabidopsis*, the soybean genes *GmDof4* and *GmDof11* are shown to increase the seed oil content by directly inducing the acetyl CoA carboxylase and long-chain-acyl CoA synthetase synthesis genes [71] that are involved in fatty acid synthesis and metabolism [71, 72]. Likewise, the overexpression of *GhDOF1* (*Gossypium hirsutum*) leads to increased lipid levels of cotton [73].
In the cooperation network for ZS11 (see Figure 5.3), there is unique cooperation between the TFs DOF5.7 and ID1. On the other hand, the factor DOF5.7 with its cooperation partners OBP4 and AP3 have been found to be unique for the network for ZY821 (see Figure 5.4). Furthermore, we found that the factor DAG2 cooperates with AT1G47655 and ADOF1 (see Figure 5.4). Given the importance of the DOF family of TFs in influencing the oil content, these preferential partner choices of DOF5.7 could be playing important roles in differentiating the regulatory processes in the seeds of the ZS11 and ZY821 cultivars.
Further, regarding the expression profiles of DOF5.4 orthologous genes (*BnaA06g24490D*, *BnaA02g43890D*, *BnaA09g07030D*), even though the gene *BnaA06g24490D* is absent for ZS11 in early stages (< day 10)), its gene expression value is strongly increasing until day 15 after flowering and decreasing after day 15 in both of the cultivars (see Figure 5.7). The expression profiles of other DOF family members are given in the Supplementary Information Figures S4–S8.

5.4.3.4 HD-ZIP Family of Transcription Factors

Multiple transcription factors from the cooperation networks, including HDG1, HDG11, ANL2, ATHB-6, ATHB-13, ATHB-40, and ATHB-53, belong to the homeodomain zipper family (see Figures 5.3 and 5.4). From this family of proteins, GL2 was the first identified protein that is responsible for the outgrowth of trichome in the epidermis [74] and for contributing to seed oil content in *Arabidopsis* [75]. Additionally, other members of this family have similar functions that are associated with the epidermis [76]. Moreover, the factor ANL2 is implicated in the regulation of anthocyanin accumulation in the shoot and also in the development of root [77]. Several studies have reported the functioning of the HD-ZIP family of transcription factors in the cuticle development. HDG11 has been implicated as a positive regulator of drought stress tolerance [78]. Furthermore, the overexpression of *OCL1* in maize belonging to HD-ZIP family highly influenced the expression levels of various genes that are associated with lipid metabolism [79]. Regarding the functioning of ATHB-6, it has been reported as a regulator of ABA hormone responses and it is also regarded as a target of protein phos-

phatase ABI1, which is a negative regulator of TAG accumulation and ABA signalling [80, 81, 82]. Moreover, ATHB-53 is regarded as an auxin-inducible protein and it plays a prominent role in the auxin/cytokinin pathway during root development [83]. However, several members of this family are related to the epidermis development [76], which is also an integral part of the seed coat.
We identified four unique HD-ZIP family members (ATHB-13, ATHB-18, ATHB-40 and ATHB-53) for the ZY821 cultivar while comparing the cooperation networks of the ZS11 and ZY821 cultivars (see Figures 5.3 and 5.4). Interestingly, our comparitive analysis reveals that the transcription factors ANL2 and ATHB-6 cooperate with the same partners in both networks. Figures 5.8 and 5.9 show the changes in the expression levels of the orthologous genes of both TFs. When considering the expression profiles of ANL2 orthologous genes (*BnaA03g27000D*, *BnaC03g31960D*), the gene expression values of both genes show higher expression levels until day 15 after flowering and they are decreasing after day 15 in ZS11. Whereas, the ANL2 orthologous gene is expressed at high expression levels in an early stage and is declined afterwards during subsequent stages (see Figure 5.8). In contrary to ANL2, the ATHB-6 orthologous genes (*BnaA09g42630D*, and *BnaC08g35090D*) show higher expression levels in ZS11 in the early stage (day 7) and lower expression level in ZY821. Intriguingly, these expression levels are continuously decreasing in ZS11 during the seed developmental stages. On the other hand, *BnaA09g42630D* and *BnaC08g35090D* are antagonistically expressed to each other in the ZY821 cultivar.

5.4.3.5 MYB Family of Transcription Factors

In both cooperation networks, we identified the transcription factors MYB1, MYB4, MYB24, MYB46, MYB55, MYB65, MYB77, MYB113, and FaEOBII, which belong to the MYB superfamily. A large number of MYB TFs play central roles in a variety of physiological processes, especially growth, development, synthesis of secondary metabolites, metabolism, and responses to biotic and abiotic stresses [84, 85, 86, 87, 88].
Interestingly, the factor MYB24 cooperates with MYB46, MYB55, and MYB65 during seed development in both of the cultivars. Taking its unique cooperations into account, MYB24 forms dimers with F3A4.140 and AT1G76870 in the ZS11 cultivar (see Figure 5.3). On the other hand, MYB24 is interacting with MYB1 in ZY821 (see Figure 5.4). Regarding the function of MYB1, it has been characterised as a pivotal positive regulator of the anthocyanin synthesis in onion, thus representing an important player in the flavanoid synthesis pathway on the transcriptional level [89].

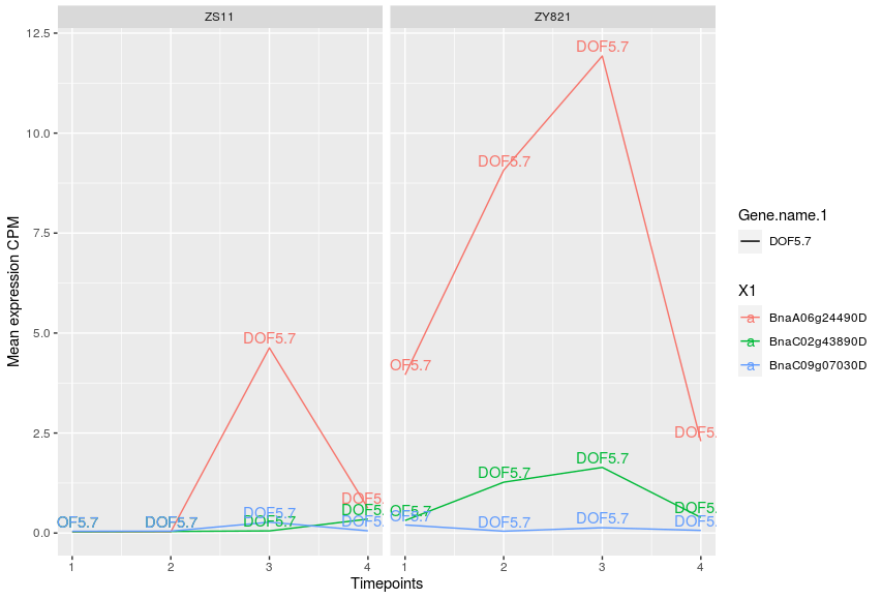

Figure 5.7: Expression values of DOF5.7 orthologous genes. Time points 1 to 4 represents day 7, day 10, day 15, and day 45 after flowering.

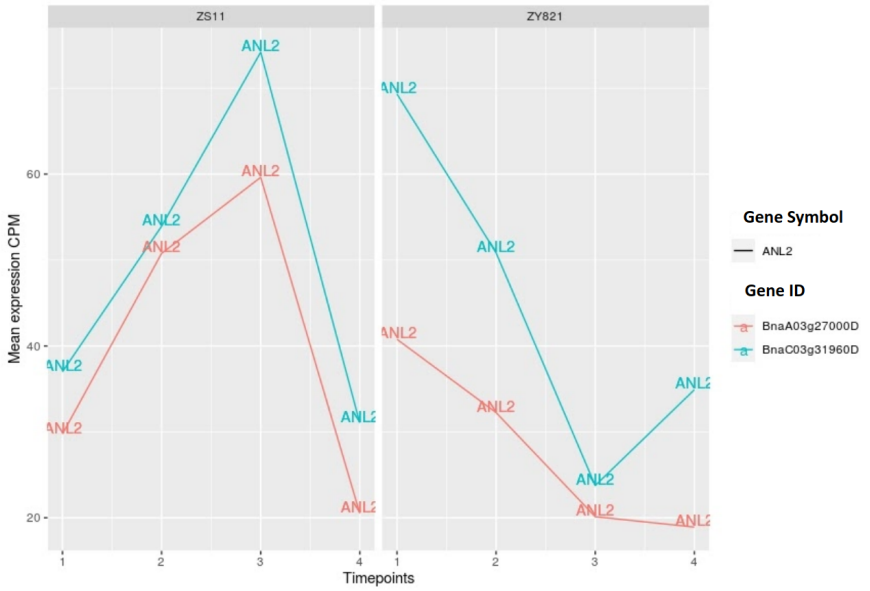

Figure 5.8: Expression values of ANL2 orthologous genes. Time points 1 to 4 represents day 7, day 10, day 15, and day 45 after flowering.

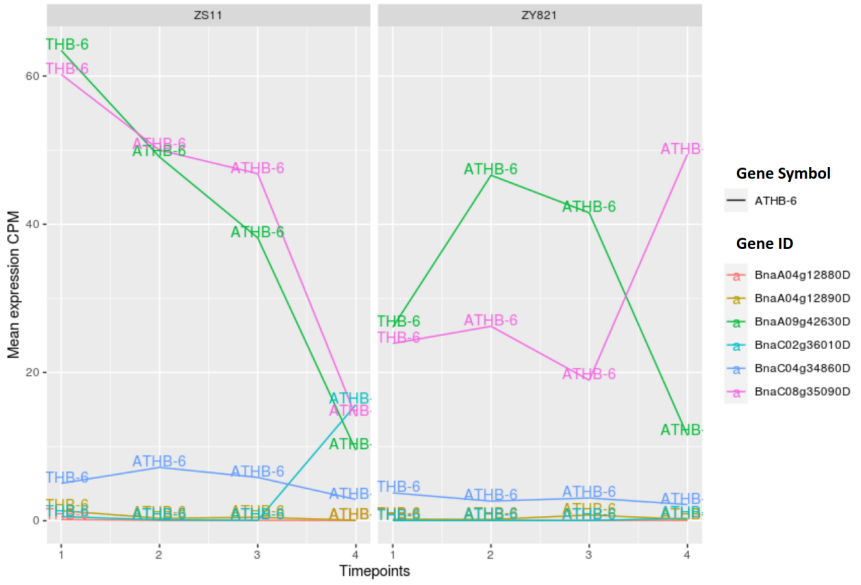

Figure 5.9: Expression values of ATHB-6 orthologous genes. Time points 1 to 4 represents day 7, day 10, day 15, and day 45 after flowering.

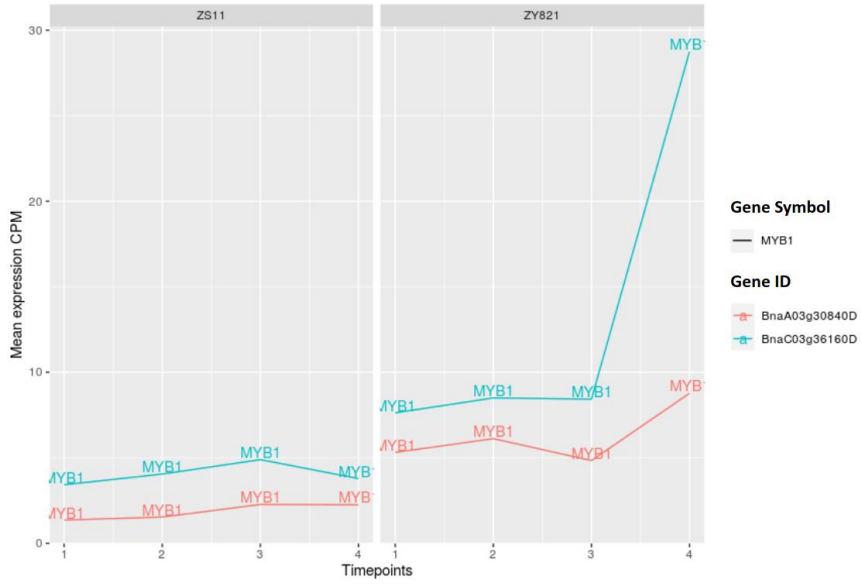

Figure 5.10: Expression values of MYB1 orthologous genes. Time points 1 to 4 represents day 7, day 10, day 15, and day 45 after flowering.

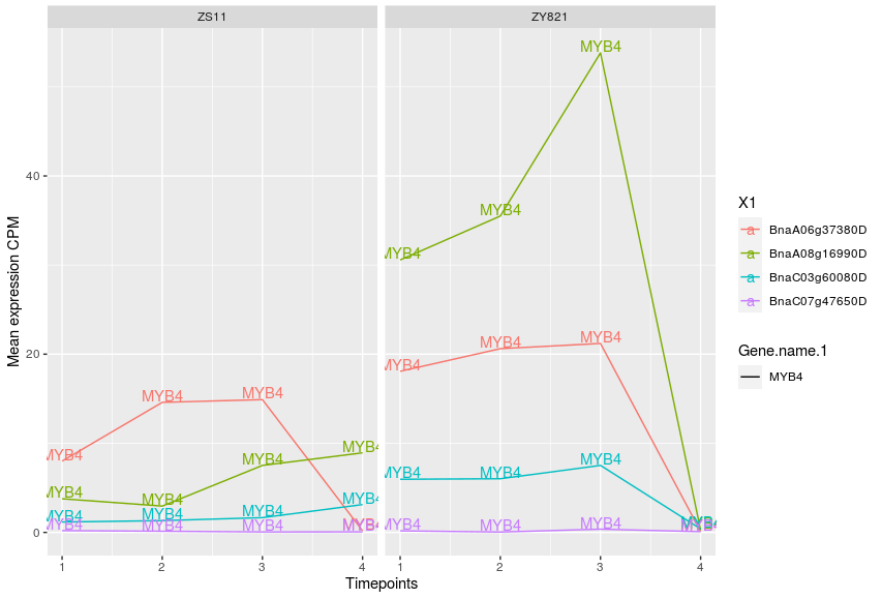

Figure 5.11: Expression values of MYB4 orthologous genes. Time points 1 to 4 represents day 7, day 10, day 15, and day 45 after flowering.

A similar role has been studied for MYB113 in purple cauliflower [90] and for FaEOBII in the phenypropanoid pathway in strawberries [91]. Likewise, MYB4 has been reported in *Arabidopsis* for its dual role in the flavanoid biosynthetic pathway, describing the precise regulation of anthocyanin and phenylalanine synthesis [92]. The role of flavonoid snythesis pathway genes in contributing to seed colour that differentiates between high-oil content and low-oil content accessions has been well-reported in [26]. More importantly, the transcription factor MYB46 functions as a master regulator in the secondary wall synthesis, regulating the genes that are involved in the synthesis of three major components (cellulose, hemi cellulose, and lignin) of secondary cell walls [93]. Therefore, it is a modulator in the regulation of defense responses to the fungus *Botrytis cinerea* [94].

The gene expression analysis of MYB1 and MYB4 orthologous genes show that the corresponding expression levels are clearly different between ZS11 and ZY821 cultivars. In particular, there is a remarkable increase in the expression level of the MYB1 orthologous genes *BnaC03g63160D* during the late stage after flowering in the ZY821 cultivar (day 15) (see Figure 5.10). Another interesting pattern has been observed for the MYB4 orthologous gene *BnaA08g16990D*: While its expression level increases until day 15 after flowering in ZY821 cultivar, it sharply decreases at the late stage (day 45) (see Figure 5.11).

5.4.3.6 Other Transcription Factors

There are other crucial transcription factors, like CAMTA2 and CAMTA3, found for both cultivars or ARR11 as well as ID1 found only for the ZS11 cultivar. The roles of these TFs are well studied in biotic and abiotic stress responses [44, 45, 46, 47]. For example, CAMTA2/3 are calmodulin binding transcription factors linking calcium signalling to the induction of defense response genes during abiotic and biotic stress conditions [44, 45, 46]. They are involved in low temperature and freezing tolerance in *Arabidopsis* [44]. Further, the factor ARR11, encoding the *Arabidopsis* response regulator 11, is implicated for its essential role in mediating abscisic acid and cytokinin signalling pathways and tolerance to drought, thereby it is involved in the generation of drought stress responses [47]. Figure S5 in Supplementary Information gives the expression profile of ARR11. In contrary, the ID1 (Indeterminate Domain) transcription factor found in the cooperation network of ZS11 (see Figure 5.3) encodes a zinc finger, which has been reported in the regulation of seed development in maize [98, 99]. In *Arabidopsis*, it is regarded to activate or inhibit seed germination, with respect to gibberellic acid [98]. Because there is a close association between the stress responses and the contribution of triacylglycerol, which is a major lipid reserve [100], this could explain the contribution of TFs to multiple processes, including fatty acid accumulation, seed germination, and the generation of stress responses.

5.5 Conclusions

Transcriptional regulation in plants plays a pivotal role in governing a variety of physiological processes. In oil crops, like *B. napus*, a deeper knowledge regarding TFs and their combinatorial interplay sheds light into the regulatory mechanisms that underlie seed oil content, particularly in the accumulation of fatty acids. In our study, by analysing a RNA-seq data set of seed tissue from two *B.napus* accessions, a double-low accession with high-oil-content and a double-high accession with low-oil-content, we identified several TFs and their preferential partner choices, which are likely to influence the quality of seed oil content. Interestingly, some of the TFs have the same cooperation partners in both cultivars, whereas the gene expression patterns of their orthologous genes clearly show distinguishing patterns between the cultivars during the seed development process. To the best of our knowledge, this is the first study performing a systematic analysis to decipher the complex interplay of the TFs that are linked with developmental switches resulting in a higher oil content. Our findings could be promising for deepening the ex-

isting knowledge on the transcriptional regulation governing seed oil content notwithstanding the absence of experimental validation. Therefore, further progress from the molecular plant biology end is needed, not only to validate the functions of these TFs, but also for a future perspective on generating novel hypotheses in genetic programs that involve seed oil improvement.

Supplementary Information

The supplementary files can be accessed via the original publication (https://www.mdpi.com/1422-0067/22/3/1033/s1). Figure S1: REVIGO treemap for monotically expressed genes in the ascending pattern obtained for double-low accession Zhongshuang11 illustrating over-represented gene ontology biological process categories; Figure S2: REVIGO treemap for monotically expressed genes in the descending pattern obtained for double-low accession Zhongshuang11 illustrating over-represented gene ontology biological process categories; Figure S3: REVIGO treemap for monotically expressed genes in the ascending pattern obtained for double-high accession Zhongyou821 illustrating over-represented gene ontology biological process categories; Figure S4: REVIGO treemap for monotically expressed genes in the descending pattern obtained for double-high accession Zhongyou821 illustrating over-represented gene ontology biological process categories; Table S1: Lists of Monotonically Expressed Genes obtained from the MFSelector approach for the seed tissue of Zhongshuang11 (ZS11) and Zhongyou821 (ZY821); Table S2: Gene ontology terms enriched for ascending and descending MEGs for ZS11 and ZY821 cultivars; Supplementary Information: Gene expression profiles of genes present in the cooperation networks for the seed tissue of Zhongshuang11 (ZS11) and Zhongyou821 (ZY821).

Author contributions

M.G. designed and supervised the research. A.R. participated in the design of the study and conducted computational analyses together with M.G. and performed the literature survey. S.K., J.S.S., H.B., K.L. and A.O.S. interpreted the results with A.R. and M.G.. A.R. and M.G. wrote the final version of the manuscript. M.G. conceived and managed the project. All authors read and approved the final manuscript.

Data availability

Not applicable.

Acknowledgments

We acknowledge support by the German Research Foundation and the Open Access Publication Funds of the Göttingen University. We would like to thank Faisal Ramzan, Martin Wutke, Thomas Martin Lange, and Felix Heinrich for proofreading the manuscript and providing helpful advices and discussions.

Conflicts of Interest

The authors declare no conflict of interest.

Bibliography

[1] F. D. Gunstone, J. L. Harwood, and A. J. Dijkstra, *The lipid handbook with CD-ROM*. CRC press, 2007.

[2] F. USDA, "Oilseeds: world markets and trade," *Available at:(accessed 10 December 2014)*, 2014.

[3] F. USDA, "Unites states department of agriculture foreign agricultural service, global agricultural trade system (gats), forest products trade statistics," *Access: http://www. fas. usda. gov/gats/default. aspx*, 2011.

[4] H. Woodfield and J. Harwood, "Oilseed crops: Linseed, rapeseed, soybean, and sunflower," 2017.

[5] A. Gracka, H. H. Jeleń, M. Majcher, A. Siger, and A. Kaczmarek, "Flavoromics approach in monitoring changes in volatile compounds of virgin rapeseed oil caused by seed roasting," *Journal of Chromatography A*, vol. 1428, pp. 292–304, 2016.

[6] S. Liu, C. Fan, J. Li, G. Cai, Q. Yang, J. Wu, X. Yi, C. Zhang, and Y. Zhou, "A genome-wide association study reveals novel elite allelic variations in seed oil content of *Brassica napus*," *Theoretical and Applied Genetics*, vol. 129, no. 6, pp. 1203–1215, 2016.

[7] L. Velasco, C. Möllers, and H. C. Becker, "Estimation of seed weight, oil content and fatty acid composition in intact single seeds of rapeseed (*Brassica napus* l.) by near-infrared reflectance spectroscopy," *Euphytica*, vol. 106, no. 1, pp. 79–85, 1999.

[8] B. Wittkop, R. Snowdon, and W. Friedt, "Status and perspectives of breeding for enhanced yield and quality of oilseed crops for europe," *Euphytica*, vol. 170, no. 1-2, p. 131, 2009.

[9] R. Snowdon, W. Lühs, and W. Friedt, "Oilseed rape," in *Oilseeds*, pp. 55–114, Springer, 2007.

[10] A. Abbadi and G. Leckband, "Rapeseed breeding for oil content, quality, and sustainability," *European Journal of Lipid Science and Technology*, vol. 113, no. 10, pp. 1198–1206, 2011.

[11] A. Hannoufa, B. V. Pillai, and S. Chellamma, "Genetic enhancement of *Brassica napus* seed quality," *Transgenic research*, vol. 23, no. 1, pp. 39–52, 2014.

[12] A. L. Harper, M. Trick, J. Higgins, F. Fraser, L. Clissold, R. Wells, C. Hattori, P. Werner, and I. Bancroft, "Associative transcriptomics of traits in the polyploid crop species *Brassica napus*," *Nature biotechnology*, vol. 30, no. 8, pp. 798–802, 2012.

[13] S. Knauer, A. L. Holt, I. Rubio-Somoza, E. J. Tucker, A. Hinze, M. Pisch, M. Javelle, M. C. Timmermans, M. R. Tucker, and T. Laux, "A protodermal mir394 signal defines a region of stem cell competence in the arabidopsis shoot meristem," *Developmental cell*, vol. 24, no. 2, pp. 125–132, 2013.

[14] H. Tan, X. Yang, F. Zhang, X. Zheng, C. Qu, J. Mu, F. Fu, J. Li, R. Guan, H. Zhang, *et al.*, "Enhanced seed oil production in canola by conditional expression of *Brassica napus* leafy cotyledon1 and lec1-like in developing seeds," *Plant physiology*, vol. 156, no. 3, pp. 1577–1588, 2011.

[15] N. Nesi, R. Delourme, M. Brégeon, C. Falentin, and M. Renard, "Genetic and molecular approaches to improve nutritional value of *Brassica napus* l. seed," *Comptes rendus biologies*, vol. 331, no. 10, pp. 763–771, 2008.

[16] D. J. Murphy, I. Cummins, and A. S. Kang, "Synthesis of the major oil-body membrane protein in developing rapeseed (brassica napus) embryos. integration with storage-lipid and storage-protein synthesis and implications for the mechanism of oil-body formation," *Biochemical Journal*, vol. 258, no. 1, pp. 285–293, 1989.

[17] R. B. Goldberg, G. De Paiva, and R. Yadegari, "Plant embryogenesis: zygote to seed," *Science*, vol. 266, no. 5185, pp. 605–614, 1994.

[18] A. H. Huang, "Oil bodies and oleosins in seeds," *Annual review of plant biology*, vol. 43, no. 1, pp. 177–200, 1992.

[19] H. Van Erp, A. A. Kelly, G. Menard, and P. J. Eastmond, "Multigene engineering of triacylglycerol metabolism boosts seed oil content in arabidopsis," *Plant Physiology*, vol. 165, no. 1, pp. 30–36, 2014.

[20] L. Zhu, X. Zhao, Y. Xu, Q. Wang, H. Wang, D. Wu, and L. Jiang, "Effect of germination potential on storage lipids and transcriptome changes in premature developing seeds of oilseed rape (*Brassica napus* l.)," *Theoretical and Applied Genetics*, pp. 1–14, 2020.

[21] M. Shahid, G. Cai, F. Zu, Q. Zhao, M. U. Qasim, Y. Hong, C. Fan, and Y. Zhou, "Comparative transcriptome analysis of developing seeds and silique wall reveals dynamic transcription networks for effective oil production in *Brassica napus* l.," *International journal of molecular sciences*, vol. 20, no. 8, p. 1982, 2019.

[22] J. Jiang, S. Zhu, Y. Yuan, Y. Wang, L. Zeng, J. Batley, and Y.-P. Wang, "Transcriptomic comparison between developing seeds of yellow-and black-seeded *Brassica napus* reveals that genes influence seed quality," *BMC plant biology*, vol. 19, no. 1, p. 203, 2019.

[23] L.-N. Ding, X.-J. Guo, M. Li, Z.-L. Fu, S.-Z. Yan, K.-M. Zhu, Z. Wang, and X.-L. Tan, "Improving seed germination and oil contents by regulating the *GDSL* transcriptional level in *Brassica napus*," *Plant cell reports*, vol. 38, no. 2, pp. 243–253, 2019.

[24] Z. Xiao, C. Zhang, F. Tang, B. Yang, L. Zhang, J. Liu, Q. Huo, S. Wang, S. Li, L. Wei, *et al.*, "Identification of candidate genes controlling oil content by combination of genome-wide association and transcriptome analysis in the oilseed crop *Brassica napus*," *Biotechnology for biofuels*, vol. 12, no. 1, p. 216, 2019.

[25] C. Qu, N. Yin, S. Chen, S. Wang, X. Chen, H. Zhao, S. Shen, F. Fu, B. Zhou, X. Xu, *et al.*, "Comparative analysis of the metabolic profiles of yellow-versus black-seeded rapeseed using uplc–hesi–ms/ms and transcriptome analysis," *Journal of Agricultural and Food Chemistry*, vol. 68, no. 10, pp. 3033–3049, 2020.

[26] Y. Niu, L. Wu, Y. Li, H. Huang, M. Qian, W. Sun, H. Zhu, Y. Xu, Y. Fan, U. Mahmood, *et al.*, "Deciphering the transcriptional regulatory networks that control size, color, and oil content in *Brassica rapa* seeds," *Biotechnology for Biofuels*, vol. 13, pp. 1–20, 2020.

[27] K. Lu, L. Wei, X. Li, Y. Wang, J. Wu, M. Liu, C. Zhang, Z. Chen, Z. Xiao, H. Jian, *et al.*, "Whole-genome resequencing reveals *Brassica napus* origin and genetic loci involved in its improvement," *Nature communications*, vol. 10, no. 1, pp. 1–12, 2019.

[28] S. Klees, H. Lange, Thomas M.and Bertram, A. Rajavel, J.-S. Schlüter, K. Lu, A. O. Schmitt, and M. Gültas, "Integrated multi-omics analysis for the investigation of the complex interplay between regulatory snps, transcription factors, and their related genes in *Brassica napus* l.," *International Journal of Molecular Sciences (Under Review)*, 2020.

[29] K. Kaufmann, F. Wellmer, J. M. Muiño, T. Ferrier, S. E. Wuest, V. Kumar, A. Serrano-Mislata, F. Madueno, P. Krajewski, E. M. Meyerowitz, *et al.*, "Orchestration of floral initiation by apetala1," *science*, vol. 328, no. 5974, pp. 85–89, 2010.

[30] C. Gómez-Mena, S. de Folter, M. M. R. Costa, G. C. Angenent, and R. Sablowski, "Transcriptional program controlled by the floral homeotic gene agamous during early organogenesis," *Development*, vol. 132, no. 3, pp. 429–438, 2005.

[31] E. M. Meyerowitz, "Plants compared to animals: the broadest comparative study of development," *Science*, vol. 295, no. 5559, pp. 1482–1485, 2002.

[32] S. Yanagisawa, "Transcription factors in plants: physiological functions and regulation of expression," *Journal of Plant Research*, vol. 111, no. 3, pp. 363–371, 1998.

[33] K. B. Singh, "Transcriptional regulation in plants: the importance of combinatorial control," *Plant Physiology*, vol. 118, no. 4, pp. 1111–1120, 1998.

[34] T. J. Guilfoyle, "The structure of plant gene promoters," in *Genetic engineering*, pp. 15–47, Springer, 1997.

[35] K. Kaufmann, A. Pajoro, and G. C. Angenent, "Regulation of transcription in plants: mechanisms controlling developmental switches," *Nature Reviews Genetics*, vol. 11, no. 12, pp. 830–842, 2010.

[36] A. Rajavel, F. Heinrich, A. O. Schmitt, and M. Gültas, "Identifying cattle breed-specific partner choice of transcription factors during the african trypanosomiasis disease progression using bioinformatics analysis," *Vaccines*, vol. 8, no. 2, p. 246, 2020.

[37] H.-W. Wang, H.-J. Sun, T.-Y. Chang, H.-H. Lo, W.-C. Cheng, G. C. Tseng, C.-T. Lin, S.-J. Chang, N. R. Pal, and I.-F. Chung, "Discovering monotonic stemness marker genes from time-series stem cell microarray data," in *BMC genomics*, vol. 16, p. S2, BioMed Central, 2015.

[38] C. Meckbach, R. Tacke, X. Hua, S. Waack, E. Wingender, and M. Gültas, "Pc-traff: identification of potentially collaborating transcription factors using pointwise mutual information," *BMC bioinformatics*, vol. 16, no. 1, p. 400, 2015.

[39] B. Chalhoub, F. Denoeud, S. Liu, I. A. Parkin, H. Tang, X. Wang, J. Chiquet, H. Belcram, C. Tong, B. Samans, *et al.*, "Early allopolyploid evolution in the post-neolithic *Brassica napus* oilseed genome," *science*, vol. 345, no. 6199, pp. 950–953, 2014.

[40] A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, and T. R. Gingeras, "Star: ultrafast universal rna-seq aligner," *Bioinformatics*, vol. 29, no. 1, pp. 15–21, 2013.

[41] S. Anders, P. T. Pyl, and W. Huber, "Htseq – a python framework to work with high-throughput sequencing data," *bioRxiv*, 2014.

[42] G. K. Smyth, M. Ritchie, N. Thorne, J. Wettenhall, W. Shi, and Y. Hu, "limma: linear models for microarray and rna-seq data user's guide," 2002.

[43] X. Yi, Z. Du, and Z. Su, "Plantgsea: a gene set enrichment analysis toolkit for plant community," *Nucleic acids research*, vol. 41, no. W1, pp. W98–W103, 2013.

[44] F. Supek, M. Bošnjak, N. Škunca, and T. Šmuc, "Revigo summarizes and visualizes long lists of gene ontology terms," *PloS one*, vol. 6, no. 7, p. e21800, 2011.

[45] A. E. Kel, E. Gossling, I. Reuter, E. Cheremushkin, O. V. Kel-Margoulis, and E. Wingender, "Matchtm: a tool for searching transcription factor binding sites in dna sequences," *Nucleic acids research*, vol. 31, no. 13, pp. 3576–3579, 2003.

[46] A. Khan, O. Fornes, A. Stigliani, M. Gheorghe, J. A. Castro-Mondragon, R. Van Der Lee, A. Bessy, J. Chèneby, S. R. Kulkarni, G. Tan, *et al.*, "Jaspar 2018: update of the open-access database of transcription factor binding profiles and its web framework," *Nucleic acids research*, vol. 46, no. D1, pp. D260–D266, 2018.

[47] S. Zeidler, C. Meckbach, R. Tacke, F. S. Raad, A. Roa, S. Uchida, W.-H. Zimmermann, E. Wingender, and M. Gültas, "Computational detection of stage-specific transcription factor clusters during heart development," *Frontiers in Genetics*, vol. 7, p. 33, 2016.

[48] A. D. Yates, P. Achuthan, W. Akanni, J. Allen, J. Allen, J. Alvarez-Jarreta, M. R. Amode, I. M. Armean, A. G. Azov, R. Bennett, *et al.*, "Ensembl 2020," *Nucleic acids research*, vol. 48, no. D1, pp. D682–D688, 2020.

[49] D. M. Bolser, D. M. Staines, E. Perry, and P. J. Kersey, "Ensembl plants: integrating tools for visualizing, mining, and analyzing plant genomic data," in *Plant Genomics Databases*, pp. 1–31, Springer, 2017.

[50] S. Kambhampati, J. A. Aznar-Moreno, C. Hostetler, T. Caso, S. R. Bailey, A. H. Hubbard, T. P. Durrett, and D. K. Allen, "On the inverse correlation of protein and oil: Examining the effects of altered central carbon metabolism on seed composition using soybean fast neutron mutants," *Metabolites*, vol. 10, no. 1, p. 18, 2020.

[51] M. Rapala-Kozik, N. Wolak, M. Kujda, and A. K. Banas, "The upregulation of thiamine (vitamin b 1) biosynthesis in *Arabidopsis thaliana* seedlings under salt and osmotic stress conditions is mediated by abscisic acid at the early stages of this stress response," *BMC plant biology*, vol. 12, no. 1, p. 2, 2012.

[52] L. Steuernagel, C. Meckbach, F. Heinrich, S. Zeidler, A. O. Schmitt, and M. Gültas, "Computational identification of tissue-specific transcription factor cooperation in ten cattle tissues," *PLOS ONE*, vol. 14, pp. 1–21, 05 2019.

[53] C. Meckbach, E. Wingender, and M. Gültas, "Removing background co-occurrences of transcription factor binding sites greatly improves the prediction of specific transcription factor cooperations," *Frontiers in genetics*, vol. 9, 2018.

[54] M. Collinge and T. Boller, "Differential induction of two potato genes, *St*prx2 and *St*nac, in response to infection by *Phytophthora infestans* and to wounding," *Plant molecular biology*, vol. 46, no. 5, pp. 521–529, 2001.

[55] D. Hegedus, M. Yu, D. Baldwin, M. Gruber, A. Sharpe, I. Parkin, S. Whitwill, and D. Lydiate, "Molecular characterization of *Brassica napus* nac domain transcriptional activators induced in response to biotic and abiotic stress," *Plant molecular biology*, vol. 53, no. 3, pp. 383–397, 2003.

[56] L.-S. P. Tran, K. Nakashima, Y. Sakuma, S. D. Simpson, Y. Fujita, K. Maruyama, M. Fujita, M. Seki, K. Shinozaki, and K. Yamaguchi-Shinozaki, "Isolation and functional analysis of arabidopsis stress-inducible nac transcription factors that bind to a drought-responsive cis-element in the early responsive to dehydration stress 1 promoter," *The Plant Cell*, vol. 16, no. 9, pp. 2481–2498, 2004.

[57] M. Fujita, Y. Fujita, K. Maruyama, M. Seki, K. Hiratsu, M. Ohme-Takagi, L.-S. P. Tran, K. Yamaguchi-Shinozaki, and K. Shinozaki, "A dehydration-induced nac protein, rd26, is involved in a novel aba-dependent stress-signaling pathway," *The Plant Journal*, vol. 39, no. 6, pp. 863–876, 2004.

[58] A. N. Olsen, H. A. Ernst, L. L. Leggio, and K. Skriver, "Nac transcription factors: structurally distinct, functionally diverse," *Trends in plant science*, vol. 10, no. 2, pp. 79–87, 2005.

[59] S. Balazadeh, H. Siddiqui, A. D. Allu, L. P. Matallana-Ramirez, C. Caldana, M. Mehrnia, M.-I. Zanor, B. Köhler, and B. Mueller-Roeber, "A gene regulatory network controlled by the nac transcription factor anac092/*At*nac2/ore1 during salt-promoted senescence," *The Plant Journal*, vol. 62, no. 2, pp. 250–264, 2010.

[60] Y.-Q. Ning, Z.-Y. Ma, H.-W. Huang, H. Mo, T.-t. Zhao, L. Li, T. Cai, S. Chen, L. Ma, and X.-J. He, "Two novel nac transcription factors regulate gene expression and flowering time by associating with the histone demethylase jmj14," *Nucleic acids research*, vol. 43, no. 3, pp. 1469–1484, 2015.

[61] G. Argüello-Astorga and L. Herrera-Estrella, "Evolution of light-regulated plant promoters," *Annual review of plant biology*, vol. 49, no. 1, pp. 525–555, 1998.

[62] K. Koch, "Carbohydrate-modulated gene expression in plants," *Annual review of plant biology*, vol. 47, no. 1, pp. 509–540, 1996.

[63] W. B. Terzaghi and A. R. Cashmore, "Photomorphenesis: Seeing the light in plant development," *Current Biology*, vol. 5, no. 5, pp. 466–468, 1995.

[64] R. Rastogi, N. J. Bate, S. Sivasankar, and S. J. Rothstein, "Footprinting of the spinach nitrite reductase gene promoter reveals the preservation of nitrate regulatory elements between fungi and higher plants," *Plant molecular biology*, vol. 34, no. 3, pp. 465–476, 1997.

[65] I. C. Oliveira and G. M. Coruzzi, "Carbon and amino acids reciprocally modulate the expression of glutamine synthetase in arabidopsis," *Plant physiology*, vol. 121, no. 1, pp. 301–310, 1999.

[66] W.-C. Yeap, F.-C. Lee, D. K. Shabari Shan, H. Musa, D. R. Appleton, and H. Kulaveerasingam, "Wri 1-1, abi 5, nf-ya 3 and nf-yc 2 increase

oil biosynthesis in coordination with hormonal signaling during fruit development in oil palm," *The Plant Journal*, vol. 91, no. 1, pp. 97–113, 2017.

[67] X. Huang, L. Hou, J. Meng, H. You, Z. Li, Z. Gong, S. Yang, and Y. Shi, "The antagonistic action of abscisic acid and cytokinin signaling mediates drought stress response in arabidopsis," *Molecular plant*, vol. 11, no. 7, pp. 970–982, 2018.

[68] S. Yanagisawa, "The dof family of plant transcription factors," *Trends in plant science*, vol. 7, no. 12, pp. 555–560, 2002.

[69] M. Ahmad, Y. Rim, H. Chen, and J. Kim, "Functional characterization of arabidopsis dof transcription factor *At*dof4. 1," *Russian journal of plant physiology*, vol. 60, no. 1, pp. 116–123, 2013.

[70] S. Santopolo, A. Boccaccini, R. Lorrai, V. Ruta, D. Capauto, E. Minutello, G. Serino, P. Costantino, and P. Vittorioso, "*DOF AFFECTING GERMINATION 2* is a positive regulator of light-mediated seed germination and is repressed by *DOF AFFECTING GERMINATION 1*," *BMC plant biology*, vol. 15, no. 1, p. 72, 2015.

[71] H.-W. Wang, B. Zhang, Y.-J. Hao, J. Huang, A.-G. Tian, Y. Liao, J.-S. Zhang, and S.-Y. Chen, "The soybean dof-type transcription factor genes, *Gm*dof4 and *Gm*dof11, enhance lipid content in the seeds of transgenic arabidopsis plants," *The Plant Journal*, vol. 52, no. 4, pp. 716–729, 2007.

[72] P. R. Fobert, "Role of transcription factors in storage lipid accumulation in plants," 2012.

[73] Y. Su, W. Liang, Z. Liu, Y. Wang, Y. Zhao, B. Ijaz, and J. Hua, "Overexpression of *Gh*dof1 improved salt and cold tolerance and seed oil content in gossypium hirsutum," *Journal of plant physiology*, vol. 218, pp. 222–234, 2017.

[74] W. G. Rerie, K. A. Feldmann, and M. D. Marks, "The glabra2 gene encodes a homeo domain protein required for normal trichome development in arabidopsis.," *Genes & development*, vol. 8, no. 12, pp. 1388–1399, 1994.

[75] B. Shen, K. W. Sinkevicius, D. A. Selinger, and M. C. Tarczynski, "The homeobox gene glabra2 affects seed oil content in arabidopsis," *Plant molecular biology*, vol. 60, no. 3, pp. 377–387, 2006.

[76] F. D. Ariel, P. A. Manavella, C. A. Dezar, and R. L. Chan, "The true story of the hd-zip family," *Trends in plant science*, vol. 12, no. 9, pp. 419–426, 2007.

[77] H. Kubo, A. J. Peeters, M. G. Aarts, A. Pereira, and M. Koornneef, "*ANTHOCYANINLESS2*, a homeobox gene affecting anthocyanin distribution and root development in arabidopsis," *The Plant Cell*, vol. 11, no. 7, pp. 1217–1226, 1999.

[78] H. Yu, X. Chen, Y.-Y. Hong, Y. Wang, P. Xu, S.-D. Ke, H.-Y. Liu, J.-K. Zhu, D. J. Oliver, and C.-B. Xiang, "Activated expression of an arabidopsis hd-start protein confers drought tolerance with improved root system and reduced stomatal density," *The Plant Cell*, vol. 20, no. 4, pp. 1134–1151, 2008.

[79] M. Javelle, V. Vernoud, N. Depege-Fargeix, C. Arnould, D. Oursel, F. Domergue, X. Sarda, and P. M. Rogowsky, "Overexpression of the epidermis-specific homeodomain-leucine zipper iv transcription factor outer cell layer1 in maize identifies target genes involved in lipid metabolism and cuticle biosynthesis," *Plant physiology*, vol. 154, no. 1, pp. 273–286, 2010.

[80] Å. Grimberg, I. Lager, N. R. Street, K. M. Robinson, S. Marttila, N. Mähler, P. K. Ingvarsson, and R. P. Bhalerao, "Storage lipid accumulation is controlled by photoperiodic signal acting via regulators of growth cessation and dormancy in hybrid aspen," *New Phytologist*, vol. 219, no. 2, pp. 619–630, 2018.

[81] E. Lechner, N. Leonhardt, H. Eisler, Y. Parmentier, M. Alioua, H. Jacquet, J. Leung, and P. Genschik, "Math/btb crl3 receptors target the homeodomain-leucine zipper athb6 to modulate abscisic acid signaling," *Developmental cell*, vol. 21, no. 6, pp. 1116–1128, 2011.

[82] J. Leung, S. Merlot, and J. Giraudat, "The arabidopsis *ABSCISIC ACID-INSENSITIVE2* (*ABI2*) and *ABI1* genes encode homologous protein phosphatases 2c involved in abscisic acid signal transduction.," *The Plant Cell*, vol. 9, no. 5, pp. 759–771, 1997.

[83] O. Son, H.-Y. Cho, M.-R. Kim, H. Lee, M.-S. Lee, E. Song, J. H. Park, K. H. Nam, J.-Y. Chun, H.-J. Kim, *et al.*, "Induction of a homeodomain–leucine zipper gene by auxin is inhibited by cytokinin in arabidopsis roots," *Biochemical and biophysical research communications*, vol. 326, no. 1, pp. 203–209, 2004.

[84] H. Jin and C. Martin, "Multifunctionality and diversity within the plant myb-gene family," *Plant molecular biology*, vol. 41, no. 5, pp. 577–585, 1999.

[85] S. Ambawat, P. Sharma, N. R. Yadav, and R. C. Yadav, "Myb transcription factor genes as regulators for plant responses: an overview," *Physiology and Molecular Biology of Plants*, vol. 19, no. 3, pp. 307–321, 2013.

[86] R.-L. Mu, Y.-R. Cao, Y.-F. Liu, G. Lei, H.-F. Zou, Y. Liao, H.-W. Wang, W.-K. Zhang, B. Ma, J.-Z. Du, *et al.*, "An r2r3-type transcription factor gene *At*myb59 regulates root growth and cell cycle progression in arabidopsis," *Cell research*, vol. 19, no. 11, pp. 1291–1304, 2009.

[87] Y. Liao, H.-F. Zou, H.-W. Wang, W.-K. Zhang, B. Ma, J.-S. Zhang, and S.-Y. Chen, "Soybean *Gm*myb76, *Gm*myb92, and *Gm*myb177 genes confer stress tolerance in transgenic arabidopsis plants," *Cell research*, vol. 18, no. 10, pp. 1047–1060, 2008.

[88] S. Araki, M. Ito, T. Soyano, R. Nishihama, and Y. Machida, "Mitotic cyclins stimulate the activity of c-myb-like factors for transactivation of g2/m phase-specific genes in tobacco," *Journal of Biological Chemistry*, vol. 279, no. 31, pp. 32979–32988, 2004.

[89] K. E. Schwinn, H. Ngo, F. Kenel, D. A. Brummell, N. W. Albert, J. A. McCallum, M. Pither-Joyce, R. N. Crowhurst, C. Eady, and K. M. Davies, "The onion (*Allium cepa* l.) r2r3-myb gene myb1 regulates anthocyanin biosynthesis," *Frontiers in plant science*, vol. 7, p. 1865, 2016.

[90] L.-W. Chiu and L. Li, "Characterization of the regulatory network of *Bo*myb2 in controlling anthocyanin biosynthesis in purple cauliflower," *Planta*, vol. 236, no. 4, pp. 1153–1164, 2012.

[91] L. Medina-Puche, F. J. Molina-Hidalgo, M. Boersma, R. C. Schuurink, I. López-Vidriero, R. Solano, J.-M. Franco-Zorrilla, J. L. Caballero, R. Blanco-Portales, and J. Muñoz-Blanco, "An r2r3-myb transcription factor regulates eugenol production in ripe strawberry fruit receptacles," *Plant Physiology*, vol. 168, no. 2, pp. 598–614, 2015.

[92] X.-C. Wang, J. Wu, M.-L. Guan, C.-H. Zhao, P. Geng, and Q. Zhao, "Arabidopsis myb4 plays dual roles in flavonoid biosynthesis," *The Plant Journal*, vol. 101, no. 3, pp. 637–652, 2020.

[93] J.-H. Ko, H.-W. Jeon, W.-C. Kim, J.-Y. Kim, and K.-H. Han, "The myb46/myb83-mediated transcriptional regulatory programme is a gatekeeper of secondary wall biosynthesis," *Annals of botany*, vol. 114, no. 6, pp. 1099–1107, 2014.

[94] V. Ramírez, J. García-Andrade, and P. Vera, "Enhanced disease resistance to botrytis cinerea in myb46 arabidopsis plants is associated to an early down-regulation of cesa genes," *Plant signaling & behavior*, vol. 6, no. 6, pp. 911–913, 2011.

[95] Y. Kim, S. Park, S. J. Gilmour, and M. F. Thomashow, "Roles of camta transcription factors and salicylic acid in configuring the low-temperature transcriptome and freezing tolerance of a rabidopsis," *The Plant Journal*, vol. 75, no. 3, pp. 364–376, 2013.

[96] C. J. Doherty, H. A. Van Buskirk, S. J. Myers, and M. F. Thomashow, "Roles for arabidopsis camta transcription factors in cold-regulated gene expression and freezing tolerance," *The Plant Cell*, vol. 21, no. 3, pp. 972–984, 2009.

[97] L. Du, G. S. Ali, K. A. Simons, J. Hou, T. Yang, A. Reddy, and B. Poovaiah, "Ca 2+/calmodulin regulates salicylic-acid-mediated plant immunity," *Nature*, vol. 457, no. 7233, pp. 1154–1158, 2009.

[98] M. Kumar, D. T. Le, S. Hwang, P. J. Seo, and H. U. Kim, "Role of the *INDETERMINATE DOMAIN* genes in plants," *International journal of molecular sciences*, vol. 20, no. 9, p. 2286, 2019.

[99] J. Colasanti, R. Tremblay, A. Y. Wong, V. Coneva, A. Kozaki, and B. K. Mable, "The maize *INDETERMINATE1* flowering time regulator defines a highly conserved zinc finger protein family in higher plants," *Bmc Genomics*, vol. 7, no. 1, p. 158, 2006.

[100] J. Lu, Y. Xu, J. Wang, S. D. Singer, and G. Chen, "The role of triacylglycerol in plant stress response," *Plants*, vol. 9, no. 4, p. 472, 2020.

Chapter 6

General Discussion

In this chapter, I discuss the results obtained from the application of different systems biology approaches to cattle datasets with regard to the investigation of the AAT disease progression. Further, I elaborate the discussion on the results highlighting the importance of TF co-operations in influencing the seed oil content of *Brassica napus*. The context of this chapter is mainly based on my publications [1, 2, 3, 4].

6.1 Deciphering the molecular mechanism underlying the AAT disease progression

In eukaryotes, gene expression is regulated by the combinatorial interplay of several transcription factor (TF) proteins. This process involves multiple signaling pathways orchestrating a complex gene regulatory network which is essential for the regulation of cellular processes. A large number of genes (more than 30,000) is regulated by a relatively small number of regulatory proteins which range from 2000 to 3000. In particular, TF proteins regulate their expression with respect to the interaction partners and DNA enhancer site. Therefore, describing the cooperative TFs is crucial for understanding the regulatory mechanism underpinning a biological process/disease [5]. Especially, knowledge about TFs and their complex interplay is pivotal to understand the regulation of genetic programs which maintain the adaptation of the animal to different pathophysiological stresses like parasitic infections.

Concerning the parasitic diseases, AAT is a vector-borne disease which spreads through the tsetse fly by carrying pathogenic trypanosomes in its saliva. Clinical signs such as anaemia, hepatomegaly and splenomegaly are exhibited by the cattle during the course of the AAT disease, which gradually

progresses in multiple steps. Based on its signs, the AAT disease shows a multi-stage progression process in the body of the animal. Previous studies have pointed out that the consideration of monotonic expression patterns of genes could reflect the stage-by-stage progression of the disease [6, 7]. Thus, I analysed several MEG sets (see Figure 2.2) to investigate the regulatory mechanisms which govern the tissue-specific gene expression and are thus influencing trypanosusceptibility and trypanotolerance of the breeds Boran and N'Dama, respectively. For this purpose, the consideration of the three tissues, liver, spleen, and lymph nodes are quintessential since they are the primary target sites of trypanosome infection. Inextricably, these tissues play crucial roles in generating host immune responses especially by increasing the number of macrophages, which results in the production of pro-inflammatory cytokines [8, 9].

Despite the rich literature on trypanotolerance and molecular studies addressing the identification of candidate markers in AAT, its underlying molecular mechanism in cattle has not been well studied. As of now, there has been no study performed on AAT, especially in cattle, with the aspect of examining the upstream regulatory processes involved in gene regulation, for example, TFs and their cooperations (see Figure 6.1 middle box in blue dashed lines). Today it is well known that TFs and their complex interplay have critical roles in the progression of a disease [10, 11]. In order to address the importance of cooperative TFs in the AAT disease, I analysed the dataset published by Noyes *et al.* [12]. In Chapter 2, I focussed on the identification as well as on the analysis of MEGs to completely capture the multistage progression process of the AAT disease (Table 2.1). Further, I applied our previously published computational PC-TraFF approach [10] to the promoter regions of the MEGs in order to identify specific cooperative TFs in different tissues of Boran and N'Dama cattle breeds.

Especially with regard to the AAT-disease, results of this study (see Figure 2.3) suggest that the preferential partner choice of TFs could be related to the gene regulatory mechanisms determining the level of AAT-tolerance of the cattle breeds. Particularly, the partner choice of the transcription factor DBP is likely to orchestrate the genetic programs governing the molecular mechanism of the level of trypanotolerance of both cattle breeds.

Regarding the results, our findings in Chapter 2 indicate that, given the AAT disease progression, the preferential partner choice of TFs is strongly related to the tissue type and the susceptibility/resistance of the cattle breeds. Especially the results (see Figure 2.3) emphasised the higher relevance of the factor DBP along with its partners to the circadian rhythm and lipid metabolism, which could be associated with the pathogenesis of AAT in trypanotolerant

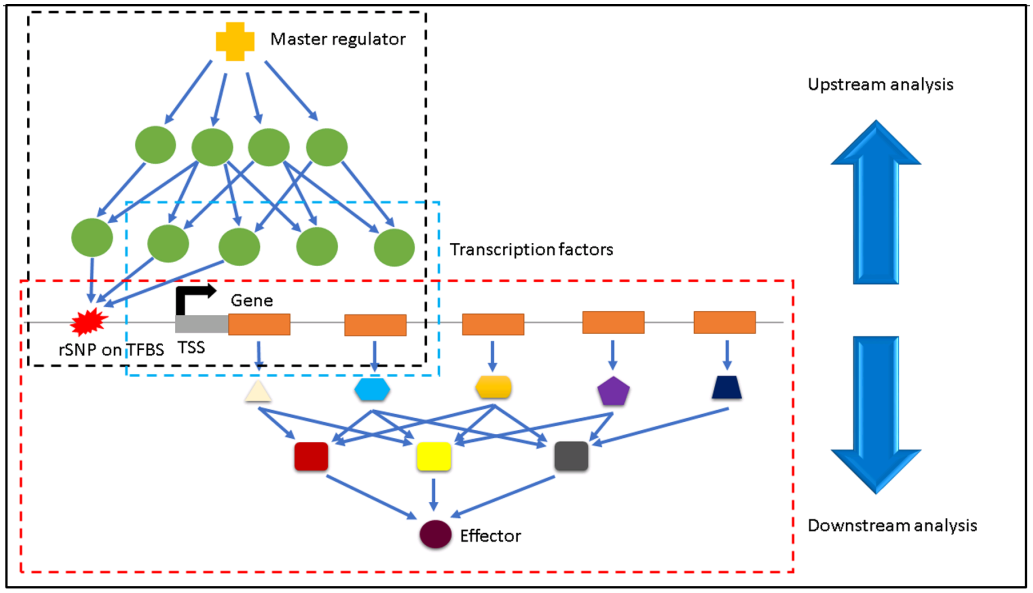

Figure 6.1: Overview of the analyses. My first computational study (middle box in blue-dotted lines) highlights the transcription factor co-operations associated with the AAT disease progression [1]. In the second study (top box in black dashed lines), I perform an upstream analysis to detect master regulators and over-represented upstream pathways related to AAT [4]. In the third study (bottom box in red dashed lines), I focussed on the downstream analysis to decipher the complex interplay of regulatory SNPs (rSNPs), their related gene expression and their corresponding downstream effectors which regulate a multitude of signal transduction pathways during the AAT disease progression [2]. (Figure is taken from [2] and presented in Chapter 4, Page 108)

N'Dama and trypanosusceptible Boran. Focusing on DBP's function in association with the circadian rhythm, I attempted to highlight the significant role of the circadian transcriptional program in the regulation of immune responses to the pathogen infection at the tissue level. Importantly, the recent study of Solis *et al.* [13] on the crucial role of circadian regulation for the coordination of the immune functions lends support to our findings that the circadian control of the immune system influenced by host-pathogen interaction might have resulted in the transcriptional reprogramming of the regulation determining the level of AAT-tolerance in cattle.

Furthermore, knowledge about master regulators is fundamental since they greatly control the TFs and the related genes [14, 15]. The consideration of

TFs and their cooperations only provides the information regarding the first level of the transcriptional regulatory hierarchy [16]. However, for gaining a proper understanding of the disease progression in both breeds, it is necessary to establish the hierarchy of the transcriptomic regulation in order to identify the master regulators [14, 15, 16]. Thus, our main objective of Chapter 3 is to identify the master regulators together with signal transduction pathways associated with the AAT disease as potential drug targets, to complement our previous study [17]. Chapter 3 (see Figure 6.1 top box in black dashed lines) provides essential knowledge on master regulators along with our previous findings in Chapter 2. On the one hand, it strongly indicates that *DBP* functions more as a master regulator of the circadian clock in peripheral tissues, supporting the trypanotolerance mechanisms in the cattle breed N'Dama. On the other hand, this analysis leads to the identification of the novel master regulator *MYC* in association with the trypanosusceptibility mechanism of Boran.

Remarkably, findings in Chapter 3 suggest that the master regulators *DBP* and *MYC* identified for liver and lymph node tissues, appear to be greatly influencing the genetic programs responsible for trypanosusceptibility and trypanotolerance in Boran and N'Dama. Notably, *DBP* could be supporting the regulation of immune responses [13, 18, 19, 20] because of its function in the circadian oscillatory mechanism [21] thereby establishing trypanotolerance in N'Dama. On the other hand, *MYC* has been reported to be responsible for the disruption of the circadian clock in cancer cells [22, 23, 24], elucidating the possibility of a dysfunctional circadian rhythm in Boran. Furthermore, *MYC* has gained its importance as it directly programs inflammation and immune suppression [25], which are constantly reported conditions in trypanosome-infected Boran.

Kupffer cells, the largest immune cell population of macrophages resident in the liver tissue, play a critical role in the mononuclear phagocytic system mounting the first line of immune response to foreign antigens [26]. Immune responses in the liver tissue depend on the activation state of macrophages [26, 27]. M1 and M2 polarization of macrophages are known as two extremes in which M1 (classically activated) is characterised by the expression of the high pro-inflammatory cytokines and M2 (alternatively activated) by high anti-inflammatory cytokines [27, 28]. Surprisingly, all three master regulators *MYC*, *E2F1* and *PPARG* identified for the liver tissue of Boran, have been reported as the regulators necessary for M2-like polarization of macrophages [29], which could be an advantage for the trypanosomes to increase their survival inside the host's body, and thereby contributing to enhanced parasitaemia in Boran.

In order to gain more mechanistic insights and to discover novel biological functions underlying the AAT disease progression of both breeds, the investigation of over-represented pathways based on the MEG sets of tissues is crucial. Based on pathway analysis, we obtained a number of over-represented pathways (see Supplementary File S2 of Chapter 3), several of which are in agreement with the results of previous studies [12, 30, 31] and are activated due to trypanosome infection in both breeds. Remarkably, the majority of these pathways were found to be common for both cattle breeds, while few of them are unique and breed-specific which could provide an important clue for distinguishing the biological processes controlling the mechanisms underlying trypanosusceptibility or trypanotolerance of the cattle breeds. Consequently, I focused on outlining the potential roles of breed-specific unique pathways in association with the level of trypanotolerance in the respective cattle breeds (see Tables 3.3–3.5). Although I reported the major immune-related pathways in Boran, these pathways could be leading to inflammation due to hyperproduction of pro-inflammatory cytokines in the host cells [32], thereby contributing to the susceptibility of this breed. Despite the activation of immune signaling pathways, dysregulation causes the death of infected animals, especially through dysregulated cytokine networks and overproduction of inflammatory cytokines (hallmark of African Trypanosomiasis) [8]. Contrarily, circadian clock related pathways (see Figure 3.4), interleukin-12 family signaling, regulation of lipid metabolism, and MAPK family signaling cascades might be properly regulated in N'Dama, indicating the underlying mechanism for trypanotolerance in N'Dama during the AAT disease. Especially, the pathways related to the circadian clock bolster our previous findings [17] in highlighting the important role of *DBP* and circadian rhythm in the coordination of the immune responses in the trypanotolerant breed N'Dama.

Taken together, I specifically identified *MYC* and *DBP* (see Figures 3.3–3.5) as potential discriminators between the two cattle breeds, trypanosusceptible Boran and trypanotolerant N'Dama, which are likely to be promising therapeutic targets for future works and for the selective breeding of this trait.

Importantly, single nucleotide polymorphisms (SNPs) are widely studied with regard to the disease mechanisms as they can have direct control over the disease susceptibility (causal polymorphisms). Importantly, rSNPs that are located in the regulatory regions like promoters can significantly affect the gene expression, especially by modifying the binding sites of the TFs. Knowledge about the rSNPs and their complex interplay with the corresponding gene expression and downstream effectors could reveal multiple disease-associated polymorphisms, which can be further used as targets in drug design and breeding programs. Taking the importance of rSNPs and their combinatorial

interplay into account by the combined usage of transcriptomic and genomic datasets of Boran and N'Dama cattle breeds, we attempted in Chapter 4 to unravel the complex interplay of rSNPs, MEGs and downstream effectors in orchestrating the molecular mechanism of the AAT disease progression. For this purpose, we mainly used the genomic data from the 1000 Bull Genomes Project (version Run7), which contained genotypes of 12 N'Dama and 11 Boran. In agreement with our previous studies [1, 4] based on the multi-stage progression process of the AAT disease, I focused on the MEGs found for three tissues of both cattle breeds and 783,637 variants (SNPs) that are located in the promoter regions of the MEGs. Despite several molecular studies available on the mechanisms underlying the AAT disease, the influence of downstream regulatory events involving the effector molecules and their complex interplay with the rSNPs and gene expression have not yet been established. To complement our previous studies, we were further interested to gain mechanistic insights into the combinatorial interplay of effector molecules, rSNPs and gene expression which differentiates the molecular mechanisms leading to susceptibility/tolerance of cattle breeds to the AAT disease (see Figure 6.1 bottom box in red dashed lines).

By manually analyzing and annotating the gene expression profiles of MEGs for each tissue, we found highly interesting gene expression profiles which reveal their distinguishing expression patterns for five MEGs (namely *MAPKAPK5*, *CSK*, *DOK2*, *RAC1* and *DNTM1*). These genes are harbouring key rSNPs and are expressed over several time points in the liver tissue of both Boran and N'Dama. Particularly, these genes are key players for the detection of effectors found for liver tissue.

Taken together, we deciphered the rSNPs in association with MEGs (see Table 4.2) and their associated downstream effectors (see Table 4.4) which control the complex genetic mechanisms involved in the regulation of AAT disease progression unravelling the candidate targets for selective breeding for trypanotolerance.

6.2 Unravelling the complex interplay of transcription factors orchestrating seed oil content in *Brassica napus* L.

Brassica napus L. (rapeseed or canola) is the third largest oilseed crop, which is widely cultivated across the globe [33, 34, 35]. The seeds of *B. napus* are rich in oil content and fatty acids, which include primarily oleic and linoleic acid [36]. However, erucic acid and glucosinolates are anti-nutritive compounds that are present in the *B. napus* seeds that are not desirable for human consumption or as fodder for animal consumption [37, 38]. Therefore, enhancing the seed quality with improved oil content has become the major selective trait for rapeseed breeding programs due to the growing global demand for oil production, for their use as bio-fuel, animal fodder, and vegetable oil [39, 40, 41].

Transcriptional regulation in plants plays a pivotal role in governing a variety of physiological processes which include responses to environmental stresses, tissue development, and cell differentiation. In oil crops, like *B. napus*, a deeper knowledge regarding TFs and their combinatorial interplay sheds light into the regulatory mechanisms that underlie seed oil content, particularly in the accumulation of fatty acids. Knowledge about TF–TF cooperations could be promising in gaining insight into the developmental switches between the cultivars of *B. napus*, namely Zhongshuang11 (ZS11), a double-low accession with high-oil-content, and Zhongyou821 (ZY821), a double-high accession with low-oil-content.

In this regard, we analysed in Chapter 5 a time series RNA-seq data set of seed tissue from the aforementioned cultivars by mainly focusing on the MEGs. The consideration of the MEGs with regular ascending or descending monotonic expression pattern during the seed development process enables the capturing of multi-stage progression processes that are orchestrated by the cooperative TFs and, thus, facilitates the understanding of the molecular mechanisms determining seed oil content. Firstly, I performed a gene set enrichment analysis [42] to obtain deeper insight into their crucial biological functions and clustered these functions based on the GO terms. Furthermore, by the application of the PC-TraFF approach to the promoters of the obtained MEGs, I identified several TFs and their preferential partner choices, which are likely to influence the quality of seed oil content.

The GO enrichment analysis regarding the MEGs of the ZS11 cultivar revealed that the ascending MEGs are significantly enriched mainly in the term "fatty acid metabolism“ (see Figure S1 in Chapter 5), which is highly asso-

ciated with the seed oil content of the rapeseed in ZS11 cultivar. On the other hand, the enrichment analysis of ascending MEGs that were obtained for the ZY821 cultivar showed that several significantly enriched GO terms are linked to the "thiamine metabolism", which is associated with the adaptation to biotic and abiotic stress [43] (see Figure S3 in Chapter 5). These results imply that the seeds of the ZY821 cultivar might be involved in stress responses while decreasing the other metabolic processes, especially those that are related to oil synthesis.

Results from this study show that five major TF families, namely NAC, MYB, DOF, GATA, and HD-ZIP are highly involved in the seed developmental process. There are other crucial transcription factors, like CAMTA2 and CAMTA3, found for both cultivars or ARR11 as well as ID1 found only for the ZS11 cultivar. The roles of these TFs are well studied in biotic and abiotic stress responses [44, 45, 46, 47]. In particular, their preferential partner choices as well as the changes in their gene expression profiles seem to be strongly associated with the differentiation of the oil content between the two cultivars. Interestingly, some of the TFs have the same cooperation partners in both cultivars (see Figures 5.3 and 5.4), whereas the gene expression patterns of their orthologous genes (in *Arabidopsis thaliana* or *Zea mays*) clearly show distinguishing patterns between the cultivars during the seed development process (see Figures 5.5 - 5.11). These results are essential for enhancing our understanding of the genetic programs in both cultivars and developing novel hypotheses for further experimental studies. To the best of my knowledge, this is the first study performing a systematic analysis to decipher the complex interplay of the TFs that are linked with developmental switches resulting in a higher oil content. The findings from this study could be promising for deepening the existing knowledge on the transcriptional regulation governing seed oil content notwithstanding the absence of experimental validation. Therefore, further progress from the molecular plant biology end is needed, not only to validate the functions of these TFs, but also for a future perspective on generating novel hypotheses in genetic programs that involve seed oil improvement.

Bibliography

[1] A. Rajavel, F. Heinrich, A. O. Schmitt, and M. Gültas, "Identifying cattle breed-specific partner choice of transcription factors during the african trypanosomiasis disease progression using bioinformatics analysis," *Vaccines*, vol. 8, no. 2, p. 246, 2020.

[2] A. Rajavel, S. Klees, Y. Hui, A. O. Schmitt, and M. Gültas, "Deciphering the molecular mechanism underlying african animal trypanosomiasis by means of the 1000 bull genomes project genomic dataset," *Biology*, vol. 11, no. 5, 2022.

[3] A. Rajavel, S. Klees, J.-S. Schlüter, H. Bertram, K. Lu, A. O. Schmitt, and M. Gültas, "Unravelling the complex interplay of transcription factors orchestrating seed oil content in brassica napus l.," *International journal of molecular sciences*, vol. 22, no. 3, p. 1033, 2021.

[4] A. Rajavel, A. O. Schmitt, and M. Gültas, "Computational identification of master regulators influencing trypanotolerance in cattle," *International Journal of Molecular Sciences*, vol. 22, no. 2, p. 562, 2021.

[5] A. Reményi, H. R. Schöler, and M. Wilmanns, "Combinatorial control of gene expression," *Nature structural & molecular biology*, vol. 11, no. 9, pp. 812–815, 2004.

[6] S. Tian, "Identification of monotonically differentially expressed genes for non-small cell lung cancer," *BMC bioinformatics*, vol. 20, no. 1, p. 177, 2019.

[7] H.-W. Wang, H.-J. Sun, T.-Y. Chang, H.-H. Lo, W.-C. Cheng, G. C. Tseng, C.-T. Lin, S.-J. Chang, N. R. Pal, and I.-F. Chung, "Discovering monotonic stemness marker genes from time-series stem cell microarray data," in *BMC genomics*, vol. 16, p. S2, BioMed Central, 2015.

[8] S. M. Kuriakose, R. Singh, and J. E. Uzonna, "Host intracellular signaling events and pro-inflammatory cytokine production in african trypanosomiasis," *Frontiers in immunology*, vol. 7, p. 181, 2016.

[9] K. A. Taylor, "Immune responses of cattle to African trypanosomes: protective or pathogenic?," *International journal for parasitology*, vol. 28, no. 2, pp. 219–240, 1998.

[10] C. Meckbach, R. Tacke, X. Hua, S. Waack, E. Wingender, and M. Gültas, "Pc-traff: identification of potentially collaborating transcription factors using pointwise mutual information," *BMC bioinformatics*, vol. 16, no. 1, p. 400, 2015.

[11] D. Wlochowitz, M. Haubrock, J. Arackal, A. Bleckmann, A. Wolff, T. Beißbarth, E. Wingender, and M. Gültas, "Computational identification of key regulators in two different colorectal cancer cell lines," *Frontiers in genetics*, vol. 7, p. 42, 2016.

[12] H. Noyes, A. Brass, I. Obara, S. Anderson, A. L. Archibald, D. G. Bradley, P. Fisher, A. Freeman, J. Gibson, M. Gicheru, *et al.*, "Genetic and expression analysis of cattle identifies candidate genes in pathways responding to Trypanosoma congolense infection," *Proceedings of the National Academy of Sciences*, vol. 108, no. 22, pp. 9304–9309, 2011.

[13] R. Orozco-Solis and L. Aguilar-Arnal, "Circadian regulation of immunity through epigenetic mechanisms," *Frontiers in Cellular and Infection Microbiology*, vol. 10, p. 96, 2020.

[14] A. E. Kel, "Search for master regulators in walking cancer pathways," in *Biological Networks and Pathway Analysis*, pp. 161–191, Springer, 2017.

[15] A. E. Kel, P. Stegmaier, T. Valeev, J. Koschmann, V. Poroikov, O. V. Kel-Margoulis, and E. Wingender, "Multi-omics "upstream analysis" of regulatory genomic regions helps identifying targets against methotrexate resistance of colon cancer," *EuPA Open Proteomics*, vol. 13, pp. 1–13, 2016.

[16] S. Sikdar and S. Datta, "A novel statistical approach for identification of the master regulator transcription factor," *BMC bioinformatics*, vol. 18, no. 1, pp. 1–11, 2017.

[17] A. Rajavel, F. Heinrich, A. O. Schmitt, and M. Gültas, "Identifying cattle breed-specific partner choice of transcription factors during the african trypanosomiasis disease progression using bioinformatics analysis," *Vaccines*, vol. 8, no. 2, 2020.

[18] K. Bozek, A. Relógio, S. M. Kielbasa, M. Heine, C. Dame, A. Kramer, and H. Herzel, "Regulation of clock-controlled genes in mammals," *PloS one*, vol. 4, no. 3, 2009.

[19] A. M. Curtis, M. M. Bellet, P. Sassone-Corsi, and L. A. O'Neill, "Circadian clock proteins and immunity," *Immunity*, vol. 40, no. 2, pp. 178–186, 2014.

[20] S. Hergenhan, S. Holtkamp, and C. Scheiermann, "Molecular interactions between components of the circadian clock and the immune system," *Journal of Molecular Biology*, 2020.

[21] S. Yamaguchi, S. Mitsui, L. Yan, K. Yagita, S. Miyake, and H. Okamura, "Role of DBP in the circadian oscillatory mechanism," *Molecular and cellular biology*, vol. 20, no. 13, pp. 4773–4781, 2000.

[22] B. J. Altman, A. L. Hsieh, A. Sengupta, S. Y. Krishnanaiah, Z. E. Stine, Z. E. Walton, A. M. Gouw, A. Venkataraman, B. Li, P. Goraksha-Hicks, *et al.*, "MYC disrupts the circadian clock and metabolism in cancer cells," *Cell metabolism*, vol. 22, no. 6, pp. 1009–1019, 2015.

[23] S. Gery and H. Koeffler, "The role of circadian regulation in cancer," in *Cold Spring Harbor symposia on quantitative biology*, vol. 72, pp. 459–464, Cold Spring Harbor Laboratory Press, 2007.

[24] S. Masri and P. Sassone-Corsi, "The emerging link between cancer, metabolism, and circadian rhythms," *Nature medicine*, vol. 24, no. 12, p. 1795, 2018.

[25] R. M. Kortlever, N. M. Sodir, C. H. Wilson, D. L. Burkhart, L. Pellegrinet, L. B. Swigart, T. D. Littlewood, and G. I. Evan, "Myc cooperates with Ras by programming inflammation and immune suppression," *Cell*, vol. 171, no. 6, pp. 1301–1315, 2017.

[26] L. J. Dixon, M. Barnes, H. Tang, M. T. Pritchard, and L. E. Nagy, "Kupffer cells in the liver," *Comprehensive Physiology*, vol. 3, no. 2, pp. 785–797, 2013.

[27] R. D. Stout and J. Suttles, "Functional plasticity of macrophages: reversible adaptation to changing microenvironments," *Journal of leukocyte biology*, vol. 76, no. 3, pp. 509–513, 2004.

[28] S. Gordon and P. R. Taylor, "Monocyte and macrophage heterogeneity," *Nature Reviews Immunology*, vol. 5, no. 12, pp. 953–964, 2005.

[29] F. Hörhold, D. Eisel, M. Oswald, A. Kolte, D. Röll, W. Osen, S. B. Eichmüller, and R. König, "Reprogramming of macrophages employing gene regulatory and metabolic network models," *PLoS computational biology*, vol. 16, no. 2, p. e1007657, 2020.

[30] P. Fisher, C. Hedeler, K. Wolstencroft, H. Hulme, H. Noyes, S. Kemp, R. Stevens, and A. Brass, "A systematic strategy for large-scale analysis of genotype–phenotype correlations: identification of candidate genes involved in African trypanosomiasis," *Nucleic acids research*, vol. 35, no. 16, pp. 5625–5633, 2007.

[31] S. Kuriakose, C. Onyilagha, R. Singh, F. Olayinka-Adefemi, P. Jia, and J. E. UZONNA, "TLR-2-dependent activation of MAPK and STAT proteins regulates proinflammatory cytokine response and immunity to experimental Trypanosoma congolense infection," *Frontiers in Immunology*, vol. 10, p. 2673, 2019.

[32] J. Uzonna, R. Kaushik, J. Gordon, and H. Tabel, "Cytokines and antibody responses during Trypanosoma congolense infections in two inbred mouse strains that differ in resistance," *Parasite immunology*, vol. 21, no. 2, pp. 57–71, 1999.

[33] H. Woodfield and J. Harwood, "Oilseed crops: Linseed, rapeseed, soybean, and sunflower," 2017.

[34] A. Gracka, H. H. Jeleń, M. Majcher, A. Siger, and A. Kaczmarek, "Flavoromics approach in monitoring changes in volatile compounds of virgin rapeseed oil caused by seed roasting," *Journal of Chromatography A*, vol. 1428, pp. 292–304, 2016.

[35] S. Liu, C. Fan, J. Li, G. Cai, Q. Yang, J. Wu, X. Yi, C. Zhang, and Y. Zhou, "A genome-wide association study reveals novel elite allelic variations in seed oil content of *Brassica napus*," *Theoretical and Applied Genetics*, vol. 129, no. 6, pp. 1203–1215, 2016.

[36] L. Velasco, C. Möllers, and H. C. Becker, "Estimation of seed weight, oil content and fatty acid composition in intact single seeds of rapeseed (*Brassica napus* l.) by near-infrared reflectance spectroscopy," *Euphytica*, vol. 106, no. 1, pp. 79–85, 1999.

[37] B. Wittkop, R. Snowdon, and W. Friedt, "Status and perspectives of breeding for enhanced yield and quality of oilseed crops for europe," *Euphytica*, vol. 170, no. 1-2, p. 131, 2009.

[38] R. Snowdon, W. Lühs, and W. Friedt, "Oilseed rape," in *Oilseeds*, pp. 55–114, Springer, 2007.

[39] A. Abbadi and G. Leckband, "Rapeseed breeding for oil content, quality, and sustainability," *European Journal of Lipid Science and Technology*, vol. 113, no. 10, pp. 1198–1206, 2011.

[40] A. Hannoufa, B. V. Pillai, and S. Chellamma, "Genetic enhancement of *Brassica napus* seed quality," *Transgenic research*, vol. 23, no. 1, pp. 39–52, 2014.

[41] A. L. Harper, M. Trick, J. Higgins, F. Fraser, L. Clissold, R. Wells, C. Hattori, P. Werner, and I. Bancroft, "Associative transcriptomics of traits in the polyploid crop species *Brassica napus*," *Nature biotechnology*, vol. 30, no. 8, pp. 798–802, 2012.

[42] X. Yi, Z. Du, and Z. Su, "Plantgsea: a gene set enrichment analysis toolkit for plant community," *Nucleic acids research*, vol. 41, no. W1, pp. W98–W103, 2013.

[43] M. Rapala-Kozik, N. Wolak, M. Kujda, and A. K. Banas, "The upregulation of thiamine (vitamin b 1) biosynthesis in *Arabidopsis thaliana* seedlings under salt and osmotic stress conditions is mediated by abscisic acid at the early stages of this stress response," *BMC plant biology*, vol. 12, no. 1, p. 2, 2012.

[44] Y. Kim, S. Park, S. J. Gilmour, and M. F. Thomashow, "Roles of camta transcription factors and salicylic acid in configuring the low-temperature transcriptome and freezing tolerance of a rabidopsis," *The Plant Journal*, vol. 75, no. 3, pp. 364–376, 2013.

[45] C. J. Doherty, H. A. Van Buskirk, S. J. Myers, and M. F. Thomashow, "Roles for arabidopsis camta transcription factors in cold-regulated gene expression and freezing tolerance," *The Plant Cell*, vol. 21, no. 3, pp. 972–984, 2009.

[46] L. Du, G. S. Ali, K. A. Simons, J. Hou, T. Yang, A. Reddy, and B. Poovaiah, "Ca 2+/calmodulin regulates salicylic-acid-mediated plant immunity," *Nature*, vol. 457, no. 7233, pp. 1154–1158, 2009.

[47] X. Huang, L. Hou, J. Meng, H. You, Z. Li, Z. Gong, S. Yang, and Y. Shi, "The antagonistic action of abscisic acid and cytokinin signaling mediates drought stress response in arabidopsis," *Molecular plant*, vol. 11, no. 7, pp. 970–982, 2018.

Chapter 7

General Conclusion

This study investigated the regulatory mechanisms involved in complex traits of agricultural species, namely the trait of trypanotolerance in two cattle breeds and the seed oil content of *Brassica napus*. Both studies were mainly based on the identification of monotonically expressed genes and transcription factor co-operations.

Regarding the trypanotolerance of cattle breeds, I performed a comprehensive systems biology analysis to investigate the genetic mechanisms and regulatory processes of two different cattle breeds, namely N'Dama and Boran focussing on their susceptibility and tolerance mechanisms. By analyzing time-series gene expression data of Boran and N'Dama, I demonstrated that the AAT disease displays the symptoms of a multi-stage progression process. As a result, the activity of transcription factors, upstream signaling pathways as well as master regulators are differentiating between both cattle breeds, based on their immune responses against the disease. Particularly with regard to the AAT-disease, we discovered that the partner choice of the transcription factor *DBP* is likely to orchestrate the genetic programs governing the molecular mechanism of the level of trypanotolerance of both cattle breeds. Especially, focusing on *DBP*'s function in association with the circadian rhythm, we attempted to highlight the significant role of the circadian transcriptional program in the regulation of immune responses to the pathogen infection at the tissue level.
In this regard, by combining these gene expression data with genomic data from the 1000 Bull Genomes Project, I applied integrated systems biology and bioinformatics approaches to examine the combinatorial interplay among regulatory SNPs (rSNPs), their related gene expression and their corresponding downstream effectors which regulate a multitude of signal transduction pathways during the AAT disease progression. Moreover, further investigation of downstream effectors found for the liver, spleen and lymph node

tissues of both cattle breeds enhanced our understanding of distinct mechanisms leading to either resistance or susceptibility of cattle breeds.

With regard to the seed oil content of the oil crop *Brassica napus*, I performed a systematic analysis to decipher the complex interplay of the TFs that are linked with developmental switches resulting in a higher oil content. Results from this study show that TF families, such as NAC, MYB, DOF, GATA, and HD-ZIP are highly involved in the seed developmental process. Particularly, their preferential partner choices as well as changes in their gene expression profiles seem to be strongly related to the differentiation of the oil content between the two cultivars. These findings are essential for enhancing our understanding of the genetic programs governing seed oil content in both cultivars and for developing novel hypotheses for further experimental studies.

Transcriptional regulation is a complex process which involves several cis- and trans- factors to regulate the expression of target genes. In this thesis, I mainly considered TF pairs for both studies on cattle as well as rapeseed. As a follow-up study, it could be interesting to investigate TFs of higher order combinations as TFs being integral component in the broad network of protein-protein interactions, are involved in combinatorial regulation of tissue-specific gene expression. Investigation of the related protein interactions and understanding their regulatory influence in governing the biological processes will be beneficial to a greater extent. As an outlook, progress from the molecular biology end will be necessary, not only to validate the functions of molecular targets proposed in this study, but also for generating novel hypotheses in genetic programs that govern complex traits.

Chapter 8

Appendix

A. Curriculum Vitae

Personal Information

Name	Abirami
Surname	Rajavel
Date of birth	24.06.1993
Place of birth	Trichy, India
E-mail	abirami.rajavel@uni-goettingen.de

Education

07/2019 - Present	International Ph.D. Program for Agricultural Sciences in Göttingen (IPAG) Georg-August-Universität, Göttingen, Germany Thesis: "*Deciphering the molecular mechanisms underlying complex traits using bioinformatics and computational biology approaches*" Supervisors: Prof. Dr. Armin Schmitt, Prof. Dr. Mehmet Gültas, Prof. Dr. Klaus Jung
10/2016 - 03/2019	M.Sc. Molecular Life Sciences Georg August University, Göttingen, Germany Thesis: "*Validation of short random peptides as inhibitors of -synuclein aggregation and toxicity in a yeast model of Parkinson's disease*" Supervisors: Prof. Dr. Gerhard Braus, Dr. Blaga Popova
07/2011 – 04/2015	B.Tech. Biotechnology Anna University, Tiruchirappalli, India Thesis: "Identification of metabolites in the hairy roots of *Tephrosia tinctoria*" Supervisor: Prof. Dr. Suresh Kumar

Teaching

Tutor for the following courses (2019 - 2022):

- Statistical Genetics, Breeding Informatics and experimental design

- Data Analysis with R
- Applied Bioinformatics with R
- Breeding Informatics

Work Experience

07/2019 - 09/2021	Student assistant Breeding Informatics group Georg August University, Göttingen, Germany
04/2019 - 06/2019	Intern at Breeding Informatics group Georg August University, Göttingen, Germany
09/2018 - 11/2018	Intern at NanoTag Biotechnologies, Göttingen, Germany
05/2017 - 07/2017	Intern at Max Planck Institute for Biophysical Chemistry, Göttingen, Germany

Supervision of projects

1. Yuehan Hui: *Genome-wide association studies on African Animal Trypanosomiasis disease progression in cattle.* Scientific project, **2020**.
   Duration: 10 weeks
2. Muhammad Jawad: *Investigation of molecular mechanism underlying mastitis using systems biology approaches.* Scientific project, **2022**.
   Duration: 16 weeks

Skills

Languages	Tamil: Native speaker English: Advanced user (Level C1) German: Intermediate user (Level B2)
Programming	R and Python programming for data analysis Advanced skills of scientific writing in LaTeX UNIX; Linux; Shell-scripting Good knowledge of bioinformatics platforms like geneXplain
Operating System	Proficient user of Linux, Microsoft windows and macOS

Bioinformatics	Extensive experience in the analysis of large omics datasets using R and Linux command line Extensive experience in the usage of biological databases
Microbiology	Cell culture techniques Fluorescence microscopy Immunohistochemistry Molecular cloning Western blotting Protein expression and purification techniques

Scientific posters and talks

November 14, 2022	Seminar series on Infection Biology, University of Veterinary Medicine Hannover Foundation, Hannover, Germany (Invited talk) Title: "Deciphering the molecular mechanism underlying the African Animal Trypanosomiasis disease progression", Oral presentation
October 13-14, 2022	CiBreed workshop Georg-August University, Göttingen, Germany (1) "Unraveling the molecular mechanism underlying the African Animal Trypanosomiasis disease progression", Poster presentation (2) "Revealing the complex interplay of transcription factors orchestrating seed oil content in *Brassica napus* L.", Poster presentation
September 22, 2022	DGfZ/GfT Jahrestagung, Kiel, Germany Title: "Unravelling the molecular mechanism underlying the African Animal Trypanosomiasis disease progression", Oral presentation
September 12-21, 2022	European student council symposium and European conference on computational biology (ECCB), Sitges, Barcelona, Spain Title: "Deciphering the molecular mechanism underlying the African Animal Trypanosomiasis disease progression", Oral and poster presentation
September 6-8, 2022	German Conference on Bioinformatics (GCB), Halle(Saale), Germany (1) "Unraveling the molecular mechanism underlying the African Animal Trypanosomiasis disease progression", Poster presentation (2) "Revealing the complex interplay of transcription factors orchestrating seed oil content in *Brassica napus* L.", Poster presentation

July 4-8, 2022	Systems biology workshop, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, Cambridge, United Kingdom Title: "Unraveling the molecular mechanism underlying the African Animal Trypanosomiasis disease progression", Flash talk and poster presentation
September 29, 2020	CiBreed workshop Georg-August University, Göttingen, Germany Title: "Deciphering master regulators and cattle breed-specific partner choice of transcription factors during the African Animal Trypanosomiasis disease progression", Oral presentation
September 9-10, 2019	CiBreed workshop Georg-August University, Göttingen, Germany Title: "The role of preferential partner switches of transcription factors in animal African Trypanosomiasis disease", Poster presentation

List of publications

1. **Rajavel, A.**; Klees, S.; Hui, Y.; Schmitt, A.O.; Gültas, M. 'Deciphering the Molecular Mechanism Underlying African Animal Trypanosomiasis by Means of the 1000 Bull Genomes Project Genomic Dataset.' *Biology*, 2022, 11, 742. https://doi.org/10.3390/biology11050742
2. **Rajavel, A.**; Klees, S.; Schlüter, J.S.; Bertram, H.; Lu, K.; Schmitt, A.O.; Gültas, M. 'Unravelling the complex interplay of transcription factors orchestrating the seed oil content in *Brassica napus*.' *Int. J. Mol. Sci.*, 2021, 22(3), 1033. https://doi.org/10.3390/ijms22031033
3. **Rajavel, A.**; Schmitt, A.O.; Gültas, 'Computational Identification of Master Regulators Influencing Trypanotolerance in Cattle.' *Int. J. Mol. Sci.*, 2021, 22(2), 562. https://doi.org/10.3390/ijms22020562
4. **Rajavel, A.**; Heinrich, F.; Schmitt, A.O.; Gültas, M. 'Identifying Cattle Breed-Specific Partner Choice of Transcription Factors during the African Trypanosomiasis Disease Progression Using Bioinformatics Analysis.' *Vaccines*, 2020, 8(2), 246. https://doi.org/10.3390/vaccines8020246
5. Heinrich, F.; Ramzan, F.; **Rajavel, A.**; Schmitt, A.O.; Gültas, M. 'MIDESP: Mutual Information-Based Detection of Epistatic SNP Pairs for Qualitative and Quantitative Phenotypes.' *Biology*, 2021, 10(9), 921. https://doi.org/10.3390/biology10090921
6. Klees, S.; Lange, T.M.; Bertram, H.; **Rajavel, A.**; Schlüter, J.S.; Lu, K.; Schmitt, A.O.; Gültas, M. 'In Silico Identification ofthe Complex Interplay between Regulatory SNPs, Transcription Factors, and Their Related Genes in *Brassica napus* L. Using Multi-Omics Data.' *Int. J. Mol. Sci.*, 2021, 22 (2), 789. https://doi.org/10.3390/ijms22020789

7. Popova, B.; Wang, D.; **Rajavel, A.**; Dhamotharan, K.; Lázaro, D. F.; Gerke, J.; ... Braus, G. H. 'Identification of two novel peptides that inhibit -synuclein toxicity and aggregation.' *Frontiers in Molecular Neuroscience*, 2021, 14, 659926. https://doi.org/10.3389/fnmol.2021.659926

8. Götzke, H.; Kilisch, M.; Martínez-Carranza, M.; Sograte-Idrissi, S.; **Rajavel, A.**; Schlichthaerle, T.; ... Frey, S. 'The ALFA-tag is a highly versatile tool for nanobody-based bioscience applications.' *Nat commun*, 2019, 10, 4403. https://doi.org/10.1038/s41467-019-12301-7

B. Declaration of academic integrity

I confirm that I have composed the present scientific thesis independently using no other sources and resources than those stated. I have accepted the assistance of third parties only in a scope that is scientifically justifiable and compliant with the legal statutes of the examinations. In particular, I have completed all parts of the dissertation myself. I have neither, nor will I, accept unauthorized outside assistance either free of charge or subject to a fee. Furthermore, I have not applied for an equivalent doctoral examination elsewhere and submitted the present thesis as a whole or in parts at another university. I am aware of the fact that untruthfulness with respect to the above declaration repeals the admission to complete the doctoral studies and/or subsequently entitles termination of the doctoral process or withdrawal of the title attained.

.....

Abirami Rajavel