\documentclass[a4paper,12pt,twoside]{article}

\usepackage[a4paper,,left=25mm,right=25mm,top=25mm,bottom=25mm]{geometry}

\usepackage{chngcntr}

\counterwithin{figure}{section} % make figure names match section number

\counterwithin{table}{section} % same for tables

\usepackage{float} % place figures exactly where desired. Important to use [H]

\usepackage{float}

\restylefloat{table} % put table caption bellow the table. It is automatically placed bellow

%from me

\usepackage{mathptmx} %font to TMNR

\linespread{1.5} %select linespread

\usepackage[labelfont=bf]{caption} % make names of figures bold

\usepackage{afterpage}

\usepackage{fancyhdr} %make cool headers

\usepackage{multirow} %make tables values expand over multiple rows

%from José

\usepackage[utf8]{inputenc}

\usepackage{graphics}

\usepackage{float}

%\usepackage[ngerman]{babel} % Necessary for german Zeug

\usepackage[utf8]{inputenc} %

\usepackage{graphicx} % Better stuff for graphics

\usepackage{color} % Probably colors

\usepackage{colortbl} % Colored tables

\usepackage{amsmath} % Useful math

\usepackage{amssymb} % Useful math

\usepackage{enumerate} % Lists

\usepackage{enumitem} % More lists

\usepackage{siunitx} % Units

\usepackage{hyperref} % Hyperlinks and references

\usepackage{pdfpages} % Include pdfs, haven't used it

\usepackage{pgfplots} % no idea

\usepackage{parskip} % Formatting

\usepackage{verbatim} % Quoting and stuff

\usepackage[normalem]{ulem} % Underlining, why do I have most of these?

\usepackage{url} % Formatting links

\usepackage{mdframed} % Boxes

\usepackage{lmodern} % Pagestyle I think

\usepackage{physics} % Physics amen

\usepackage{titlesec} % Sections and so

\usepackage{bm} % Bold math

\usepackage{fancyhdr} % Required for modifying headers and footers

\usepackage{multicol} % Required for splitting text into multiple columns

\usepackage{glossaries} % Glossary

\usepackage{lastpage} % For page # of ##

\usepackage{stackengine} % I think this is for stacking stuff

\usepackage{caption} % captions inside minipages

\usepackage{newfloat} % new types of caption environments

\usepackage{makecell} % different sized table cells

\usepackage{float} % new types of caption environments

\title{Bachelor Thesis}

\author{Ilya Schneider}

\date{July 2022}

\begin{document}

\begin{titlepage}

\begin{center}

\begin{minipage}{0.75\linewidth}

\centering

{\normalsize Ruprecht-Karls-Universität Heidelberg

Fakultät für Biowissenschaften

Bachelorstudiengang Molekulare Biotechnologie\par}

\vspace{3cm}

{\LARGE \textbf{Dynamic Modelling of Gene Regulatory Networks Describing CD8+ T-Cell Exhaustion}\par}

\vspace{2cm}

{\Large Bachelorarbeit\par}

\vspace{1cm}

{\normalsize Ilya Schneider

geboren in Moskau, Russland

\par}

\vspace{5cm}

{\normalsize Juli

2022}

\end{minipage}

\end{center}

\end{titlepage}

\newcommand\myemptypage{

\null

\thispagestyle{empty}

\addtocounter{page}{-1}

\newpage

}

\myemptypage

\pagestyle{fancy}

\fancyhf{}

\fancyfoot[LE,RO]{\thepage}

\pagenumbering{roman}

Die vorliegende Bachelorarbeit wurde im Institut für Quantitative Analysis of Molecular and

Cellular Biosystems der Universität Heidelberg in der Health Data

Science Unit an der Medizinischen Fakultät Heidelberg in der Zeit vom 23.05.2022 bis 29.07.2022

angefertigt.

\vspace{1cm}

Gutachter der Arbeit:\null\hfill PD Dr. Carl Herrmann

\newline\null\hfill Health Data Science Unit

\newline\null\hfill Medizinische Fakultät Heidelberg

\vspace{1cm}

\hrulefill{}

\vspace{1cm}

Ich erkläre hiermit ehrenwörtlich, dass:

\begin{enumerate}

\item ich die vorliegende Bachelorarbeit selbständig unter Anleitung verfasst und keine anderen

als die angegebenen Quellen und Hilfsmittel benutzt habe;

\item die Übernahme wörtlicher Zitate aus der Literatur/Internet sowie die Verwendung der

Gedanken anderer Autoren an den entsprechenden Stellen innerhalb der Arbeit gekennzeichnet wurde;

\item ich meine Bachelorarbeit bei keiner anderen Prüfungvorgelegt habe.

\end{enumerate}

Ich bin mir bewusst, dass eine falsche Erklärung rechtliche Folgen haben wird.

\newcommand{\namesigdate}[2][5cm]{%

\begin{minipage}{#1}

#2 \vspace{1cm}\dotfill\smallskip

\end{minipage}

}

\vspace{1cm}\smallskip \\ \\

\noindent \namesigdate{Ort, Datum} \ \hfill \namesigdate[5cm]{Unterschrift}

\newpage

\section\*{Acknowledgements}

\newpage

\section\*{Abstract}

\newpage

\section\*{Zusammenfassung}

\newpage

\pagenumbering{arabic}

\tableofcontents

\newpage

\addcontentsline{toc}{section}{List of Abbreviations}

\section\*{List of Abbreviations}

\textbf{ATAC-seq}: Assay for Transposase-Accessible Chromatin using sequencing\\

\textbf{AUC}: area under the curve\\

\textbf{BN}: boolean network\\

\textbf{BTLA}: B- and T-lymphocyte attenuator\\

\textbf{CD8}: cluster of differentiation 8\\

\textbf{CTLA-4}: cytotoxic T lymphocyte antigen 4\\

\textbf{DAAT}: direct-acting antiviral therapy\\

\textbf{GRN}: genetic regulatory network\\

\textbf{HCV}: hepatitis C virus\\

\textbf{IS}: interaction score\\

\textbf{MHCI}: major histocompatibility complex 1\\

\textbf{NIS}: normalized interaction score\\

\textbf{NFATC1}: nuclear factor of activated T-cells cytoplasmic 1\\

\textbf{PD-1}: programmed cell death protein 1\\

\textbf{pySCENIC}: python implementation of the Single-Cell Regulatory Network Inference and Clustering\\

\textbf{RF}: Random Forest\\

\textbf{scRNA-seq}: single cell RNA sequencing\\

\textbf{TCR}: T-cell receptor\\

\textbf{TF}: transcription factor\\

\newpage

\section{Introduction}

\fancyhead[LE,RO]{\textit{Introduction}}

\subsection{CD8+ T-cells}

Cytotoxic T-cells also known as CD8+ T-cells belong to lymphoid cell line. Originally, they develop from hematopoietic stem cells and later on relocate to thymus as T-cell progenitors. There further development stages occur, till the cells express receptors during differentiation, for instance T-cell receptor (TCR) and cluster of differentiation 8 (CD8). TCR is specific for each antigen and is located on the surface of a cytotoxic T-cells, where it interacts with major histocompatibility complex class one (MHCI) proteins. This molecule presents the antigen to the T-killer cells, which in its turn destroys the infected cell. CD8 as well as CD3 and CD28, act as a coreceptor and bind to MHCI to support T-cell activation (Vohr, 2016).

Hence, CD8+ T-cells act antigen specifically during adaptive immune response, for example in cases of viral infection or tumors. Characterisation of CD8+ T-cells is based upon their interaction with the antigens. Naïve T-cells are the ones that have not had contact with antigen yet. Consequently, they are a precursor of effector T-cells, which are the cells that when presented to an antigen develop characteristics like cytotoxicity and ability to produce cytokines. Finally, after the antigen has been eliminated, effector T-cells differentiate into a less active, memory state. These cells can be reactivated in case there is need to eliminate the antigen to which their TCRs are specific to. Thus, a quicker immune response takes place in case of a secondary antigen exposure (Murphy, 2017).

\subsection{Chronic and Acute Viral Infections}

Viral infections can be classified according to antigen exposure time, which has potential influence on development stages of cytotoxic T-cells. During acute infection a relatively short exposure to an antigen occurs. In that case CD8+ T-cells undergo classic maturing stages from naïve to effector and after eliminating the source of infection the few surviving effector cells transform to memory cells (Figure 1.1). Analogous behavior occurs after vaccination (McLane \textit{et al.}, 2019).

However, during chronic infections (or cancer) the antigen exposure is more prolongated and is excessive. A greater expression of inhibitory receptors like programmed cell death protein 1 (PD-1) and cytotoxic T lymphocyte antigen 4 (CTLA-4) are a result of these continuous antigen signals. This strongly alters maturation of CD8+ T-cells (Figure 1.1). Thereupon, T-Cells do not develop into fully functional effector or memory state. This phenomenon is known as exhaustion, a cellular state that results in inability of T-Cells to fully eliminate the antigen and to differentiate into memory cells (Kurachi, 2019).

\begin{figure}[htp]

\centering

\includegraphics[scale=0.4]{Figure 1 Exhaustion.jpg}

\caption{\textbf{Variations in the course of T-cell maturation.} On the left a usual T-cell development process is represented. On the right is a case of unending antigen exposure, which causes T-cells enter exhaustion (McLane \textit{et al.}, 2019).}

\label{Figure 1.1}

\end{figure}

It has been already demonstrated that the effects of the T-cell exhaustion can be reversed, therefore reviving functional CD8+ T-Cells (Barber \textit{et al.}, 2006). This provides grounds for investigating new therapeutic possibilities for cancer and viral infection treatment.

\subsection{Chronic Hepatitis C Virus Infection}

Hepatitis C virus (HCV) is an RNA virus from \textit{Flaviviridae} family, which mainly disturbs liver functioning. Chronic HCV infection is a consequence of a failure in direct virus elimination. It is estimated that between 75 and 85\% of people infected by this virus will develop this chronic infection and possibly further related complications (Chen \textit{et al.}, 2006). Nowadays this condition is being linked with a reduced effectivity of CD8+ T-cells during exhaustion (Hofmann \textit{et al.}, 2021). Left unsupervised or untreated it could lead to fibrosis, cirrhosis and hepatocellular carcinoma (Zaltron \textit{et al.}, 2012).

Recently, exhausted CD8+ T-cells have been described in three states: memory-like, transient and terminally exhausted. For simplicity these states will be referred to as memory, transient and exhausted throughout this project. These states are characterized by differences in expression of genetic markers, specifically CD127, which is more expressed in memory cells. Moreover, it has been demonstrated that memory cells must go through transient state to become exhausted. There are various up- and down-regulated genes, that are known to be common in exhausted cytotoxic T-cells. However, distinguishable signatures for each of the above-mentioned states are less clear (Hensel \textit{et al.}, 2021).

Direct-acting antiviral therapy (DAAT) treatment appears to be highly effective against HCV infection. Even though, it is possible to achieve virus elimination, an exhaustion epigenetic footprint is left (Hensel \textit{et al.}, 2021). This highlights the need for more rigorous investigation into therapeutic approaches and a better understanding of exhaustion states. Particularly, changes in gene expression and activity, that take place during this process.

\subsection{Gene Regulatory Networks}

A gene regulatory network (GRN) is a collection of entities (genes, proteins or molecules) that interact with each other and describe a state and/or changes between multiple states. The network itself and the interactions between the entities are based on experimental data. Each network subunit is represented as a node that can be linked to further subunits. A node can have multiple incoming and outgoing interactions (Kaern \textit{et al.}, 2003).

The purposes of these networks can vary a lot. They could only have a descriptive character and visualize interactions within a specific cell state. Others may be used to create in silico knock outs or upregulations to simulation system perturbations, make predictions and guide lab experiments. Besides that, they can also provide insight into biological processes, by highlighting network's architecture and dynamics, and be a method to combine empirically driven data and theory to find answers on biological questions (Karlebach and Shamir, 2008).

A Boolean network (BN) is a specific example of a GRN. It discards all the intermediate activity states by making a simplification of considering only two node sates, active (1) and inactive (0) (Ramírez and Mendoza, 2017). As the result, interactions, that determine the state of each node, are also simplified, to strictly being activating or inhibitory. The binarization of nodes scores is usually performed by setting a threshold. Expression or activity values (depending on the network) above the threshold are set to 1, while the ones below to 0. Logical laws are used to formalise and describe regulatroy dependencies between the nodes (Ristevski, 2015).

BNs are relatively simple and reduce the network dynamics. Therefore, they cannot describe slight gene activity or expression changes at all. They are also more suited for smaller networks, as the number of states for such networks grows exponentially: $2^n$ (n being the number of nodes in the network) (Ristevski, 2015).

\subsection{Predicting Transcription Factor Activity}

Python implementation of the Single-Cell Regulatory Network Inference and Clustering (\textit{pySCENIC}) was utilized to obtain targets of the transcription factors (TFs), that are present in CD8+ T-cells. \textit{pySCENIC} is a tool that works with single cell RNA sequencing (scRNA-seq) data and represents a combination of 3 other tools, that analyze the input in a stepwise manner (Aibar \textit{et al.}, 2017).

In the first step, \textit{GENIE3} or \textit{GRNBoost}, is used to determine co-expression between genes from scRNA-seq data, creating a regulon of targets for each TF. Each identified interaction is given an importance score (Van de Sande \textit{et al.}, 2020). However, some of this co-expression may be result of not cis-regulatory elements, that directly interact with neighboring genes, but of trans-regulatory elements, which act on more distant genes. For that reason, \textit{RcisTarget} is implemented to reduce regulons and keep only cis-interactions (direct targets). This is achieved by conducting enrichment analysis of TF binding sequence and the target gene location. Finally, activity of each TF is quantified in each cell by calculating the area under the curve (AUC) using \textit{AUCell} (Aibar \textit{et al.}, 2017).

\newpage

\subsection{Objective and Premise of the Thesis}

The focus of this project lies in investigating chronic viral infection caused by HCV and its effects on the CD8+ T-cell activity. The objective is to create a GRN to explain CD8+ T-cell exhaustion. This GRN should describe interactions among TFs and bring forward a coherent outline of processes on the intracellular level. The ultimate goal is to further comprehend and represent the genetic differences between the three cell states (memory, transient and exhausted).

Two other studies were used as a reference in this project. Bolouri \textit{et al.} has already created a network to describe CD8+ T-cell exhaustion based on the avaliable literature sources and information from other models (Figure 1.2). The Hérault \textit{et al} has computed a BN to describe early hematopoiesis aging, based on scRNA-seq data and \textit{pySCENIC} implementation (Hérault \textit{et al.}, 2022).

This strudy aims to use computationally obtained regulons to create a GRN and curate this network with data from Bolouri \textit{et al.}, thus using it as literature source for higher integrity. The workflow and idea behind this project were inspired by the Hérault \textit{et al.}.

\newpage

\begin{figure}[htp]

\vspace{+5cm}

\centering

\includegraphics[scale=0.45]{Figure 2 Bolouri network.jpg}

\caption{\textbf{Bolouri \textit{et al.} network.} Contains 120 interactions among 57 nodes (Bolouri \textit{et al.}, 2020).}

\label{Figure 2.2}

\end{figure}

\newpage

\section{Materials}

\fancyhead[LE,RO]{\textit{Materials}}

\subsection{Data Description}

The entire data used during this project is provided by University Hospital Freiburg, Faculty of Medicine and was published in Hensel \textit{et al.}, 2021.

The raw scRNA-seq data of the CD8+ T-cells contains count values of 43 779 genes among 1 258 cells. ATAC-seq analysis results of 4 676 genes were contributed. Additionally, cell state annotations of 784 cells were contributed.

\subsection{Documentation}

The code for this project can be accessed on

\url{https://github.com/IlyaSchneider10/GRN\_Tcell\_Exh}.A more detailed file description is given there. The input data, figures and intermediate results are avaliable on the \textit{curry0} cluster of BioQuant Heidelberg (\url{curry0.bioquant.uni-heidelberg.de}).

\subsection{Software}

\begin{itemize}

\item R, v4.0.2

\item Python, v 3.10.4

\item pySCENIC, v0.10.0 (Van de Sande \textit{et al.}, 2020)

\item Bonesis, v0.0a0 (https://github.com/bioasp/bonesis)

\item Cytoscape, v3.9.1

\item GenomicFeatures, R package v1.42.3

\item GenomicRanges, R package v1.42.0

\item TxDb.Hsapiens.UCSC.hg38.knownGene, R package v 3.10.0

\item Homo.sapiens, R package v1.3.1

\item plyranges, R package v1.10.0

\item AUCell, R package v1.12.0

\item GSEABase, R package v1.52.1

\item GEOquery, R package v1.52.0

\item openxlsx, R package v4.2.5

\item stringr, R package v1.4.0

\item gplots, R package v3.1.3

\item igraph, R package v1.3.2

\item RColorBrewer, R package v1.1.3

\item ComplexHeatmap, R package v2.6.2

\item randomForest, R package v4.6.14

\item mousetrap, R package v3.2.1

\item LaplacesDemon, R package v16.1.6

\item devtools, R package v2.3.2

\item dorothea, R package v1.2.2

\item Seurat, R package v4.1.1

\item ggpubr, R package v0.4.0

\item ggplot2, R package v3.3.6

\end{itemize}

\newpage

\section{Methods}

\fancyhead[LE,RO]{\textit{Methods}}

\subsection{Data Preprocessing for pySCENIC Pipeline}

From the raw scRNA-seq data only expression information in cells that had one of the three cell states (memory, transient, exhausted) assigned was kept. Genes were selected according to a following criteria, where $g\_i$ is an expression value of a gene in a cell $i$ and $n$ is the total number of cells:

\begin{equation\*}

\sum\_{i=0}^{n}g\_i>2 \times 0.01 \times n

\end{equation\*}

Some differences in nomenclature between our scRNA-seq data and Boulori \textit{et al.} network were detected and unified. Out of 57 nodes, 46 were already added during the gene filtering step. To maximize the amount of genes from Bolouri \textit{et al.} network in the \textit{pySCENIC} input 7 genes that did not meet filtering criteria were added manually. As a result, a table containing expression data of 10 245 genes among 784 cells was used as \textit{pySCENIC} input.

\subsection{Transcription Factor Activity Prediction}

One \textit{pySCENIC} run consists of these 3 steps, described in detail in the introduction. For the purpose of this project \textit{pySCENIC} was iterated 50 times to obtain more robust regulons and interactions. All the TFs that were given AUC score were kept as reference TFs subset. Therefore, two version of the regulons were kept. The first one, without any filtering, containing all possible genes identified as targets for a particular TF. The second one, had the targets reduced to only TF-TF interactions from the TF reference list.

To reduce the scope of interactions to be analyzed even further, only interactions that appeared in at least 80\% of the \textit{pySCENIC} runs were kept. For these interactions $NIS$ score was calculated, where for a target gene $t$ with $n$ regulators $r$ all $n$ regulators are considered.

\begin{equation\*}

NIS(r\_t; t) = \frac{IS(r\_t; t)}{\sum\_{i=0}^{n}IS(r\_i; t)}

\end{equation\*}

$IS(r\_t, t)$ is an average of all the importance scores for this interaction, generated by \textit{GRNBoost}, multiplied by the number of \textit{pySCENIC} runs, in which interaction from $r\_t$, to $t$ was identified. NIS allows to give more relevancy to interactions that are recovered more frequently. It is a mean to weight the interactions and to identify communities inside of a network.

\subsection{Verification of the Interactions}

To validate \textit{pySCENIC} interaction databases \textit{Cistrome} and \textit{DoRothEA} were chosen. Additionally, assay for transposase-accessible chromatin using sequencing (ATAC-seq) data from patients who had HCV infection and underwent DAAT was used as another source of confirmation.

Cistrome is based around ChIP-seq experimental data. Only bed files with TFs that were classified as “Blood” in the tissue type category were selected for further analysis. For each TF peaks were arranged according to their score and 6000 peaks with the highest values were kept. Each of the peaks was expanded 1 kb in both directions in order to overlap it with human genome. The genes that were identified in this TF extended window were considered its targets. \textit{TxDb.Hsapiens.UCSC.hg38.knownGene} package was used as reference for transcripts. To conduct the overlap between extended peaks and the transcripts \verb|intersect()| from \textit{GenomicRanges} was implemented. Finally, to get genes for the overlapped transcripts, package \textit{Homo.sapiens} was utilized as reference genome and \verb|find\_overlaps()| from \textit{plyranges} was applied to extract gene names.

\textit{DoRothEA} database is supported on transcriptional data and has a confidence score for every interaction from E to A, E being the lowest score and A the highest. Out of all overlapping interactions only the ones that had A confidence were kept.

ATAC-seq data originates from the same cells as the scRNA-seq data and thereby is of high relevance. Open chromatin regions that were identified were mapped with TF and compared with \textit{pySCENIC} interactions. The overlaps was marked as confirmed interactions.

\subsection{Network Modelling}

TFs that were left after 80\% frequency cut were further reduced using Random Forest (RF) approach. The idea was to use AUC scores from TFs as explanatory variables to describe the belonging of the cells to one or another cell type (response variable). TFs were ranked according to their importance scores and top 100 were kept for the network modulation: \verb|randomForest(cell\_type~.,data = random\_forest\_input, importance=TRUE)|.

Interactions between these 100 TFs were selected from the ones retained after 80\% filtering and further reduced to only interactions that appeared in at least 90\% of \textit{pySCENIC} runs.

\subsection{Logical Rules Formulation}

To study network dynamics and logical rules behind it we implemented \textit{Bonesis} (\url{https://github.com/bioasp/bonesis}). \textit{Bonesis} was used as the tool to establish rules under which nodes interact with each other and switch between cellular states. The main premise was that cells must start in the memory state and go through transient configuration to finally end exhausted. Exhausted state was additionally defined as a fixed point, meaning that once the network reached this state it gets trapped in it and cannot switch to other states. Consequently, the attractor states describe T-cell exhaustion. Besides definition of state transitions, this tool also requires characteristic information about each of them. After normalizing and scaling our raw scRNA-seq data using \textit{Seurat} package, gene expression data was used to create cellular sate profiles.

\newpage

\section{Results}

\fancyhead[LE,RO]{\textit{Results}}

\subsection{pySCENIC Output Analysis}

In total, \textit{pySCENIC} identified 353 different TFs among 50 runs, with 6 254 unique interactions. After the reduction to the most robust interactions, present in 80\% of the runs, 641 interactions among 251 TFs were left.

29\% (73 TFs) of these TFs were found in the \textit{Cistrome} database. In total 20\% of interactions that made it through the frequency filter were confirmed. \textit{DoRothEA} contributed to confirming 2\% of most frequent interactions. ATAC-Seq data verified 15\% of interactions, that passed the frequency filter.

To, conclude, across the 641 most frequent interactions from 50 \textit{pySCENIC} runs 33\% (210 interactions) were confirmed by at least one of mentioned validation sources.

\subsection{Binarizing Transcription Factor Activity}

To quantify activity of TFs in the cells \textit{AUCell} was used. Full lists of targets (not only restricting to TF-TF interactions) were used as input regulons for each TF. Small regulons ($<$ 15 targets) were removed, leaving 222 regulons. After normalizing the \textit{AUCell} scores, an activity difference among cell states within a TF could be observed (Figure 4.1).

Following TF reduction to 100 most relevant ones using RF an average of activity within each cell type was calculated (Figure 6.1). The distributions of averages were bimodal, as it was confirmed by bimodality tests. That being the case, a cutoff value for each cell type was determined, which were used to binarize the average \textit{AUCell} scores from Figure 6.1.

A percentage of active cells was calculated within each cell state for every TF (Figure 4.2). The executed binarization was not conducted based on activity fluctuations within TFs as it was initially intended, but was done according to the activity differences in each cell type among all 100 TFs from RF. Finally, each TF was assigned to a cell type, where it had most active cells (Table 6.1). To conclude, 43 TFs are classified as memory, 40 as exhausted and 17 as transient.

\begin{figure}[htp]

\vspace{0}

\centering

\includegraphics[scale=0.45]{Figure 4.1.jpg}

\caption{\textbf{Normalized \textit{AUCell} Scores.} Representation of 50 out of 222 \textit{AUCell} scores. Activity difference can be well recognized in \textit{JUN}, \textit{KLF} and \textit{NFAT} families.}

\label{Figure 4.1}

\end{figure}

\newpage

\begin{figure}[H]

\includegraphics[scale=0.55]{Figure 4.2.jpg}

\hspace{-1cm}

\centering

\caption{\textbf{Percentage of active cells in each cell state for every TF.} The cell state with highest percentage of active cells was determined for each TF and assigned to it as its characteristic state. Detailed results of this classification can be observed in Table 6.1.}

\label{Figure 4.2}

\end{figure}

\newpage

\docum

\subsection{Network of 100 Transcription Factors from Random Forest}

To visualize the TFs, interactions were reduced to the once that appeared in 90\% of \textit{pySCENIC} runs. Further restriction of interactions to only between 100 TFs results in a network with 209 interactions and 82 TFs (Figure 4.3). Communities 2, 3 and 4 have a dominant cell type presence among their TFs, community 2 being dominantly exhausted, meanwhile both communities 3 and 4 dominantly memory. Community 1 consists only of two TFs and is therefore too small to make conclusions about its belonging. Community 5 has a more heterogeneous composition of TFs.

\begin{figure}[htp]

\hspace{-0cm}

\centering

\includegraphics[scale=0.6]{Figure 4.3.png}

\caption{\textbf{\textit{pySCENIC} network.} 38 nodes assigned to exhausted state (filled red), 32 to memory sate (filled blue) and 12 to transient state (filled purple). Out of 209 interactions 36\% were confirmed by at least one of our validation sources. They are marked in red. Louvain clustering identified 5 comunities using $NIS$ (thickness of the lines) scores. Community 1 (C1) consists of 2 nodes: \textit{KL9} and \textit{E2F4}.}

\label{Figure 4.3}

\end{figure}

When driving comparison with Bolouri \textit{et al.} network, \textit{pySCENIC} network shares 9 TFs with it, them being: \textit{EOMES}, \textit{EZH2}, \textit{FOS}, \textit{JUN}, \textit{MYC}, \textit{NFATC2}, \textit{NFKB1}, \textit{PRDM1} (\textit{BLIMP1} in Bolouri \textit{et al.} network) and \textit{RUNX3}. All but 2 overlapping TFs have the same cell type assigned to. The discrepancies being: \textit{EZH2} and \textit{NFATC2}, in \textit{pySCENIC} network are assigned to exhausted cell state.

To reduce number of nodes and interactions to a greater extend, the network was reduced to only these 9 TFs and their direct interacting neighbors. As a consequence, this size reduction led to a network with 41 nodes and 106 interactions (Figure 6.2). Two communities were identifiable after Louvain clustering, one of them being dominantly exhausted and another one dominantly memory.

\subsection{Fusion of reduced \textit{pySCENIC} network and Bolouri \textit{et al.} network}

Prior to overlapping reduced \textit{pySCENIC} network and Bolouri \textit{et al.} network a nomenclature unification had to be conducted. \textit{FOS}, \textit{FOSB}, \textit{JUN}, \textit{JUNB} and \textit{JUND} nodes were combined into one node in order to match the \textit{AP1} heterodimer node in Bolouri paper network (in the network appears as \textit{JUN-FOS}). Similarly, a new node for a cooperative regulation by a protein complex \textit{NFATC1:JUN-FOS:IRF4:BATF} was added, since the model in this project only considers TF-TF regulations and Bolouri \textit{et al.} network takes into more complex features, like this cooperative action of four genes. Furthermore, the same procedure of extracting 9 overlapping nodes and their direct interacting neighbors was conducted.

The resulting fusion network consists of 66 nodes and 170 interactions (Figure 4.4). Communities 1 and 3 are dominantly exhausted and memory-like respectively. Communities 1 and 4 do not show any homogeneity.

\begin{figure}[htp]

\hspace{-0cm}

\centering

\includegraphics[scale=0.7]{Figure 4.4.png}

\caption{\textbf{Fusion network between \textit{pySCENIC} network and Bolouri \textit{et al.} network.} The fusion was based on 9 common nodes, which are marked with dark purple border. 28 TFs belong to exhausted group (filled red), 27 to memory group (filled blue) and 2 to transient group (filled purple). Additionally, 8 cell receptors (filled yellow) and 1 protein complex (filled green) were incorporated from Bolouri \textit{et al.} Out of 170 interactions 64 came from Bolouri \textit{et al.} (marked with dashed lines) and 106 from \textit{pySCENIC} network (marked with full lines), out of which 37\% are confirmed by at least one of our validation sources (marked in red). Louvain clustering revealed 4 TF communities using $NIS$ score (thickness of the lines), which was conserved for the \textit{pySCENIC} interactions and was set equal to 1 for the ones coming from Bolouri \textit{et al.} network. Interactions between these two models do not overlap.}

\label{Figure 4.4}

\end{figure}

\subsection{Reduction of the Fusion Network to Only Computationally Relevant Nodes}

Not all the nodes from the fusion network are relevant for the next step, which is determining the logical rules between TFs. Nodes that exclusively possess incoming or outgoing interactions can be discarded, since they do not bring any additional information when it comes to determining logical rules to decide if another TF is active or not. A good example of such nodes can be seen in community three grouped around \textit{JUN-FOS} (Figure 4.4).

After this manual curation the resulting network has 38 nodes and 103 interactions, 18 exhausted TFs, 14 memory TFs, 1 transient TF, as well as 4 cell receptors and 1 protein complex (Figure 6.3).

Even though, the curated fusion network became significantly reduced it is still too complex to define logical rules. That being the case,

network was drastically simplified. First, exclusively interactions from Bolouri \textit{et al.} network were kept. Second, as in previous network curation, only nodes that are relevant for logical rules were kept. The resulting network contains 13 nodes with 23 interactions among them (Figure 4.5).

\begin{figure}[htp]

\hspace{-0cm}

\centering

\includegraphics[scale=0.8]{Figure 4.5.png}

\caption{\textbf{Reduced fusion network.} 5 to memory sate (filled blue), 4 nodes assigned to exhausted state (filled red), 3 receptors (filled yellow) and 1 protein complex (filled green). All interactions come from Bolouri \textit{et al.} network}.

\label{Figure 4.5}

\end{figure}

\newpage

\subsection{Establishing Logical Rules}

The size of this network was suitable to begin analyzing it with \textit{Bonesis}. A dot plot of gene expression from the \textit{pySCENIC} input data was computed to create cellular sate profiles (Figure 6.4). Based on this plot and the Figure 4.2 cellular configurations for each state were created and used as \textit{Bonesis} input (Table 6.2). As the result, further nodes have adopted a different state, when compared to Bolouri \textit{et al.} network. \textit{NFATC1} describes memory state instead, meanwhile \textit{FOXO1} switched to exhausted cell type. Therefore, the colors of the nodes had to be adjusted (Figure 4.6).

\begin{figure}[htp]

\hspace{-0cm}

\centering

\includegraphics[scale=0.8]{Figure 4.6.png}

\caption{\textbf{Reduced fusion network with adjusted gene activity and expression.} 5 to memory sate (filled blue), 4 nodes assigned to exhausted state (filled red), 3 receptors (filled yellow) and 1 protein complex (filled green). All interactions come from Bolouri \textit{et al.} network. The amount of TFs characteristic for each state is stayed the same, as in Figure 4.5}

\label{Figure 4.6}

\end{figure}

Ultimately, after conducting simulation for logical rules and reducing solutions to the ones that can reach exhaustion from the memory state 355 806 networks that fulfill such constrains were obtained. Nodes for \textit{TCRs (CD3, TCR, CD8, CD28)},\textit{ CTLA-4} and \textit{ PD-1} did not have any logical solutions and were described as constantly turned on during the exhaustion, suggesting that these nodes do not have enough inputs to describe changes in their activity. Consequently, this model in total appears to be too simple to describe nodes’ activity, so further modifications are required.

\subsection{Splitting the Model in Two Regions}

As the scalability of Bonesis is limited in number of nodes and interactions it takes in, the network was split in two regions. The top half, consisting of 6 nodes (\textit{NFATC1}, \textit{CTLA-4}, \textit{PD-1}, \textit{NFATC1:JUN-FOS:IRF4:BATF}, \textit{TCRs} and \textit{EZH2}) and the bottom half with 7 nodes (\textit{TCF-1}, \textit{BCL6}, \textit{BACH2}, \textit{PRDM1}, \textit{ID3}, \textit{IL2-R} and \textit{FOXO1}) (Figure 4.7).

\begin{figure}[H]

\hspace{-0,3cm}

\centering

\includegraphics[scale=0.6]{Figure 4.7.jpg}

\caption{\textbf{Reduced fusion network split in two regions.} The right side represents the top segment; the left one the bottom one. The split is justified by the architecture of the network since the top half does not receive any input signals from the bottom part, therefore being the driving portion of the network.}

\label{Figure 4.7}

\end{figure}

The idea behind this approach is to look for informative input nodes and interactions to explain the activity switches of the three nodes from the top half (\textit{TCRs}, \textit{CTLA-4} and \textit{PD-1}), without drastically increasing running time and amount of solutions provided by \textit{Bonesis}. Consequently network from Figure 6.3 was examined for the interactions, that would have one of three problematic nodes as a target. The goal was to add minimum number of TFs, while also try and include some interactions from \textit{pySCENIC}.

\raggedbottom %remove weird spaces between paragraphs

As an additional input to the \textit{CD3, TCR, CD8, CD28} node \textit{BTLA} was introduced. In order to connect \textit{BTLA} with the rest of the network \textit{NFKB1} and \textit{JUN-FOS} were incorporated. Interaction between \textit{JUN-FOS} and \textit{NFKB1}, as well as between \textit{EZH2} and \textit{JUN-FOS} originate from \textit{pySCENIC} and have no description of the interaction being activating or inhibiting. To determine that, the activities of these TFs were compared in each cell state using the heatmap (Figure 4.2). Interaction from \textit{JUN-FOS} was defined as activating, meantime from \textit{EZH2} to \textit{JUN-FOS} inhibiting.

\textit{NFATC2} was selected to address \textit{PD-1} and \textit{CTLA-4}. As in the previous case, by incorporating only this node it would have no input interactions and subsequently it is not possible to switch it on. \textit{STAT1} was picked as it targets \textit{NFATC2} and it is targeted by \textit{EZH2}. The sign of the interaction between these nodes was determined as formerly with interactions originating from \textit{pySCENIC}, in this case it being activated. By adding \textit{NFATC2} to this network an activating interaction to \textit{NFATC1}.

The resulting network consists of 11 nodes and 17 interactions (Figure 4.8). It was tested to determine the logical rules and verify that the issue with the nodes having one fixed state solution was eliminated. The cellular profiles in the (Table 4.1) were determined as earlier using a dot plot (Figure 4.9) and heatmap (Figure 4.2) 25 792 networks that could describe transition from memory to exhausted state were identified.

\begin{figure}[htp]

\hspace{-8,3cm}

\includegraphics[scale=0.6]{Figure 4.8.png}

\caption{\textbf{Top region of reduced fusion network with additional interactions.} 4 TFs belong to exhausted group (filled red), 4 are cell receptors (filled yellow), 3 belong to memory group (filled blue) and 1 node is protein complex (filled green). Out of 17 interactions 13 come from Bolouri \textit{et al.} (marked with dashed lines) and 4 from \textit{pySCENIC} network (marked with full lines), out of which 50\% are confirmed by at least one of our validation sources (marked in red).}

\label{Figure 4.8}

\end{figure}

\begin{table}[H]

\begin{center}

\label{Tabel 4.1}

\begin{tabular}{c|ccc}

node&memory&transient&exhausted\\\hline

EZH2&0&-&1\\

CD3,TCR,CD8,CD28&-&0&1\\

CTLA-4&-&0&1\\

NFATC1:JUN-FOS:IRF4:BATF&-&-&-\\

NFATC1&1&-&0\\

PD-1&-&0&1\\

BTLA&-&-&-\\

JUN-FOS&1&0&0\\

NFKB1&1&-&0\\

NFATC2&0&0&1\\

STAT1&0&0&1\\

\hline

\smallskip

\caption{\textbf{Input table for \textit{Bonesis} to enumerate solutions for the network from Figure 4.8.} 1 means node is active in the state, 0 is interpreted as inactive. - stands for cases where no clear state could be determined or information was absent.}

\end{tabular}

\end{center}

\end{table}

\begin{figure}[H]

\hspace{0cm}

\centering

\includegraphics[scale=0.27]{Figure 4.9.jpg}

\caption{\textbf{Dot plot characterising gene expression of nodes from Figure 4.8.} No expression data was found for IL2-R. CD28 represents the node of TCRs and PDCD1 stands for PD-1. Protein complex components' expression is visualized separately.}

\label{Figure 4.9}

\end{figure}

The bottom part of the split network was brought back and incorporated with the expanded top part. However, after performing Bonesis simulation the increase in number of solutions was so drastic, that it was not possible to exactly determine it. The inclusion of only 2 nodes of the 7 bottom nodes, \textit{BCL6} and \textit{TCF-1}, mounted the total amount of solutions networks to 1 049 908. This output was considered too big and not analyzable. For that reason, it was decided to continue the project with the network from Figure 4.10.

\begin{figure}[H]

\hspace{0cm}

\centering

\includegraphics[scale=0.63]{Figure 4.10.jpg}

\caption{\textbf{Final network.} Different representation of network in Figure 4.8. Additionally, it is indicated if the interaction is an inhibition ($-$) or an activation ($+$).}

\label{Figure 4.10}

\end{figure}

\subsection{Reduction of Network Solutions}

Table 4.2 represents exhibits unique rules for each node and their share within every node. Even though, all nodes posses a conditional solution, all of them also have a fixed value (0, 1 or both) as well.

\begin{table}[H]

\begin{center}

%\renewcommand\arraystretch{1.5} %bigger spaces between the rows

\caption{\textbf{Rules for nodes from Figure 4.10 network.} Each node has at least one conditional solution. Numbers in the brackets represent frequency of this solution within a node. $\land$ represents and; $\lor$ represents or.}

\label{Tabel 4.2}

\resizebox{\textwidth}{!}{\begin{tabular}{c|cccccc}

Nodes &\multicolumn{6}{c}{Rules}\\\hline

\\EZH2 & CD3,TCR,CD8,CD28 & 1 \\

& (58\%) & (42\%)\\

\\CD3,TCR,CD8,CD28 & !BTLA $\lor$ !CTLA-4 $\lor$ !PD-1 & \begin{tabular}{@{}c@{}} (!CTLA-4 $\land$ !PD-1)\\$\lor$ !BTLA\end{tabular}& !BTLA $\lor$ !PD-1 & BTLA $\lor$ !CTLA-4 & !BTLA & 1\\

& (15\%) & (17\%) & (16\%) & (16\%) & (17\%) & (20\%)\\

\\CTLA-4 & NFATC1 $\lor$ NFATC2 & NFATC2 & 1\\

& (34\%) & (34\%) & (32\%)\\

\\\begin{tabular}{@{}c@{}}NFATC1:JUN-FOS:\\IRF4:BATF\end{tabular} & !JUN-FOS $\lor$ NFATC1 & !JUN-FOS $\land$ NFATC1 & NFATC1 & !JUN-FOS & 1 & 0\\

& (12,5\%) & (12,5\%) & (12,5\%) & (12,5\%) & (12,5\%) & (37,5\%)\\

\\NFATC1 & \begin{tabular}{@{}c@{}}NFATC1:JUN-FOS:IRF4:BATF \\ $\land$ NFATC2\end{tabular} & \begin{tabular}{@{}c@{}}NFATC1:JUN-FOS:\\IRF4:BATF\end{tabular}& 0\\

& (12,5\%) & (12,5\%) & (37,5\%)\\

\\PD-1 & NFATC1 $\lor$ NFATC2 & NFATC2 & 1\\

& (34\%) & (34\%) & (32\%)\\

\\BTLA & NFKB1 & 1 & 0\\

& (45\%) & (7\%) & (48\%)\\

\\JUN-FOS & !EZH2 & 0\\

& (42\%) & (58\%) \\

\\NFKB1 & JUN-FOS & 0\\

& (48\%) & (52\%)\\

\\NFATC2 & STAT1 & 1\\

& (50\%) & (50\%)\\

\\STAT1 & EZH2 & 1\\

& (50\%) & (50\%)\\

\hline

\end{tabular}}

\end{center}

\end{table}

For \textit{TCRs}, \textit{CTLA-4}, \textit{NFATC1:JUN:FOS:IRF4:BATF, NFATC2}, PD-\textit{1} and \textit{STAT1} fixed value/values represented 50\% or less of the total node solutions. Thereupon, the solution networks containing these values in these nodes were removed, leaving a total of 1 264 networks.

The same procedure was done for \textit{BTLA}, \textit{EZH2}, \textit{JUN-FOS} and \textit{NFKB1}, since they now also qualify according to the former criteria. Finally, 64 networks were left. For these 64 solutions the attractor states were extracted from the \textit{Bonesis} output (Table 4.3). The final set of rules together with the frequencies can be seen in Table 4.4.

\begin{table}[H]

\begin{center}

\caption{\textbf{Attractor states for the network from Figure 4.10.} 1 means node is active in the state, 0 is interpreted as inactive. Both attractors appear with the same frequency.}

\label{Tabel 4.3}

\begin{tabular}{c|cc}

Node& \multicolumn{2}{c}{Attractors}\\\hline

EZH2&1&1\\

CD3,TCR,CD8,CD28&1&1\\

CTLA-4&1&1\\

NFATC1:JUN-FOS:IRF4:BATF&1&0\\

NFATC1&0&0\\

PD-1&1&1\\

BTLA&0&0\\

JUN-FOS&0&0\\

NFKB1&0&0\\

NFATC2&1&1\\

STAT1&1&1\\

\hline

\end{tabular}

\end{center}

\end{table}

\begin{table}[H]

\begin{center}

\caption{\textbf{Rules for nodes from Figure 4.10 network after curation.} Each node has at least one conditional solution. Numbers in the brackets represent frequency of this solution within a node.}

\label{Tabel 4.4}

\resizebox{\textwidth}{!}{\begin{tabular}{c|cccc}

Nodes &\multicolumn{4}{c}{Rules}\\\hline

\\EZH2 & CD3,TCR,CD8,CD28 \\

& (100\%)\\

\\CD3,TCR,CD8,CD28 & \begin{tabular}{@{}c@{}}!BTLA $\lor$ !CTLA-4\\$\lor$ !PD-1\end{tabular}& \begin{tabular}{@{}c@{}}(!CTLA-4 $\land$ !PD-1)\\$\lor$ !BTLA \end{tabular} &!BTLA $\lor$ !PD-1 & BTLA $\lor$! CTLA-4 \\

& (25\%) & (25\%) & (25\%) & (25\%)\\

\\CTLA-4&NFATC1 $\lor$ NFATC2 & NFATC2\\

& (50\%) & (50\%) &\\

\\\begin{tabular}{@{}c@{}}NFATC1:JUN-FOS:\\IRF4:BATF \end{tabular} &!JUN-FOS & !JUN-FOS $\lor$ NFATC1 &!JUN-FOS $\land$ NFATC1&NFATC1\\

& (25\%) & (25\%) & (25\%) & (25\%)\\

\\NFATC1& 0 \\

& (100\%) \\

\\PD-1&NFATC2 & NFATC1 $\lor$ NFATC2 \\

& (50\%)&(50\%)\\

\\BTLA& NFKB1\\

& (100\%)\\

\\JUN-FOS& !EZH2\\

& (100\%)\\

\\NFKB1 & JUN-FOS\\

& (100\%)\\

\\NFATC2& STAT1\\

& (100\%)\\

\\STAT1& EZH2 \\

& (100\%)\\

\hline

\end{tabular}}

\end{center}

\end{table}

\section{Discussion}

\newpage

\section{Supplementary Materials}

\begin{figure}[htp]

\hspace{0cm}

\centering

\includegraphics[scale=0.3]{Figure 6.1.jpg}

\caption{\textbf{Averaged \textit{AUCell} Scores.}}

\label{Figure 6.1}

\end{figure}

\begin{table}[H]

\begin{center}

\caption{100 Transcription factors from Random Forest analysis and their assigned cell state.}

\label{Tabel 6.1}

\begin{tabular}{c c|c c|c c|c c}

TF & Cell State & TF & Cell State & TF & Cell State & TF & Cell State\\ \hline

ATF2&mem & ATF3&mem&ATF4&mem&ATF5&mem\\

ATF6&mem&BPTF&exh&CEBPB&mem&CEBPZ&mem\\

CLOCK&mem&CREM&mem&CTCF&exh&CUX1&mem\\

E2F2&exh&E2F4&exh&E2F6&trans&E4F1&mem\\

EGR1&mem&ELF1&exh&ELF4&exh&EOMES&exh\\

ETS1&exh&ETS2&exh&ETV1&trans&ETV6&exh\\

ETV7&trans&EZH2&exh & FLI1& exh &FOS& mem\\

FOSB& mem &FOXK1& mem &FOXN2&mem &FOXN3&mem\\

FOXO3& mem &FOXP1& mem &FOXP2& trans&HIVEP2&mem\\

HIVEP3& trans &HKR1& mem &HMGA1&exh &HMGA2& trans\\

HSF1& exh &JUN& mem&JUNB& mem &JUND&mem\\

KLF13&exh &KLF2&mem &KLF3&mem &KLF6&mem\\

KLF7& mem &KLF9& trans &KMT2A& trans &MAFF&mem\\

MAX& exh &MBD1& mem &MEF2D& exh &MXI1& trans\\

MYC& mem &NFATC1& mem &NFATC2& exh &NFE2L2&mem\\

NFIA& trans &NFIC& trans &NFKB1& mem &NR2C2&trans\\

NR3C1& mem &NRF1& exh &PLAGL1& mem &POU6F1&mem\\

PRDM1& exh &REL&mem &RELA& exh &REST&exh\\

RUNX1& exh &RUNX3& exh &SETDB1& exh &SOX6&trans\\

SP2& trans &SP3& mem &SREBF1& mem &SRF&exh\\

STAT1&exh &STAT3& exh&STAT4& trans&STAT5A& mem\\

TBP& trans &TBX21& trans &TFDP1& exh &THAP1& exh \\

THAP11& exh &THRA& mem &TRPS1& exh &VEZF1&mem\\

YY1&exh & ZBTB7A& exh &ZFP90& exh &ZNF121&exh\\

ZNF274& exh &ZNF281& exh&ZNF362& exh &ZNF84& exh\\

\hline

\end{tabular}

\end{center}

\end{table}

\begin{figure}[htp]

\hspace{-0,3cm}

\vspace{1cm}

\includegraphics[scale=0.3]{Figure 6.2.png}

\caption{\textbf{Curated \textit{pySCENIC} network}. 22 nodes assigned to exhausted state (filled red),

17 to memory sate (filled blue) and 2 to transient state (filled purple). Out of 106

interactions 37\% were confirmed by at least one of our validation sources. They are

marked in red. Louvain clustering identified 2 comunities using $NIS$ (thickness of the

lines) scores.}

\label{Figure 6.2}

\end{figure}

\begin{figure}[htp]

\hspace{0,3cm}

\centering

\includegraphics[scale=0.7]{Figure 6.3.png}

\caption{\textbf{Curated fusion network.} 18 TFs belong to exhausted group (filled red), 14 to memory group (filled blue), 4 cell receptors (filled yellow), 1 to transient group (filled purple) and 1

protein complex (filled green) were incorporated from Bolouri et al. Common nodes between networks are marked with dark purple border. Out of 103 interactions

34 came from Bolouri et al. (marked with dashed lines) and 69 from pySCENIC network

(marked with full lines), out of which 30\% are confirmed by at least one of our validation

sources (marked in red). Louvain clustering revealed 4 TF communities using $NIS$ score

(thickness of the lines), which was conserved for the pySCENIC interactions and was set

equal to 1 for the ones coming from Bolouri et al. network.}

\label{Figure 6.3}

\end{figure}

\begin{figure}[htp]

\hspace{0cm}

\centering

\includegraphics[scale=0.75]{Figure 6.4.jpg}

\caption{\textbf{Dot plot characterising gene expression of nodes from Figure 4.5.} No expression data was found for IL2-R. CD28 represents the node of TCRs, PDCD1 stands for PD-1 and HNF1A illustrates TCF-1. Protein complex components' expression is visualized separately.}

\label{Figure 6.4}

\end{figure}

\begin{table}[H]

\begin{center}

\caption{\textbf{Input table for \textit{Bonesis} to enumerate solutions for Figure 6.3.} 1 means node is active in the state, 0 is interpreted as inactive. - stands for cases where no clear state could be determined or information was absent.}

\label{Tabel 6.2}

\begin{tabular}{c|ccc}

node&memory&transient&exhausted\\\hline

EZH2&0&-&1\\

CD3,TCR,CD8,CD28&-&0&1\\

CTLA-4&-&0&1\\

IL2-R&-&-&-\\

NFATC1:JUN-FOS:IRF4:BATF&-&-&-\\

NFATC1&1&-&0\\

PD-1&-&0&1\\

PRDM1&0&-&1\\

BACH2&1&-&-\\

FOXO1&-&0&1\\

ID3&-&-&-\\

BCL6&-&-&-\\

TCF-1&-&-&-\\

\hline

\end{tabular}

\end{center}

\end{table}

\end{document}

%\thispagestyle{empty} for the supplementary mateirlas and etc

\begin{table}[H]

\begin{center}

\caption{\textbf{Rules for nodes from Figure 4.10 network.} Each node has at least one conditional solution. Numbers in the brackets represent frequency of this solution within a node.}

\label{Tabel 4.2}

\resizebox{\textwidth}{!}{\begin{tabular}{c|ccccccc}

Nodes & \multicolumn{7}{c}{Rules}\\\hline

\multirow{2}{\*}{EZH2}&\multirow{2}{\*}{1 (48\%)}&&CD3,TCR,\\

&&&CD8,CD28 (52\%)\\\\

\multirow{2}{\*}{CD3,TCR,CD8,CD28} & \multirow{2}{\*}{1 (20\%)}&&!BTLA $\lor$ !CTLA-4 & !BTLA $\lor$ & (!CTLA-4 $\land$ !PD-1)& BTLA $\lor$&\multirow{2}{\*}{!BTLA (17\%)}\\

&&&$\lor$ !PD-1 (15\%)&!PD-1 (16\%) &$\lor$ !BTLA (17\%)&!CTLA-4 (16\%)\\\\

CTLA-4&1&&NFATC1 $\lor$ NFATC2&NFATC2\\\\

NFATC1:JUN-FOS:&\multirow{2}{\*}{1}&\multirow{2}{\*}{0} & !JUN-FOS &!JUN-FOS $\lor$ NFATC1 & !JUN-FOS $\land$ NFATC1& NFATC1 & !JUN-FOS\\

IRF4:BATF&&&\\\\

\hline

\end{tabular}}

\end{center}

\end{table}