NTNU
Norwegian University of
Science and Technology

From gene lists to interaction networks for biological interpretation

Reinterpretation of the results from a diet intervention study in light of a new statistical analysis

Maren Svanem

Master of Science
Submission date: June 2017
Supervisor: Martin Tremen R. Kuiper, IBI
Co-supervisor: Berit Johansen, IBI

Norwegian University of Science and Technology
Department of Biology

Preface

This master thesis marks the end of an era, a final task at the Norwegian University of Science and Technology (NTNU) before life as a student suddenly becomes a happy memory and life as a working adult becomes reality. Life as a teacher awaits. The thesis is the end of a master's degree project in cell and molecular biology at the Department of Biology at NTNU. In 2010, Berit Johansen's research group, with the PhD students Ingerid Arbo and Hans-Richard Brattbakk in front, completed a study, here referred to as "the diet intervention study". Subsequently, a new statistical analysis of the results has been conducted, and a need for a system understanding of the biology behind both the new and the old results arose. This thesis is an attempt to address this need.

I want to thank my main supervisor Martin Kuiper, who has been patient with me and always helpful. In addition, I have had valuable help from my co supervisor Berit Johansen, who made me believe in my part of this project. Special thanks must be granted to Rafel Riudavets, who programmed for me in R Studio – I could not have produced my results without you.

Finally, a huge thank you to my family and my boyfriend for all the support throughout these years. Most importantly is it, however, to thank my fellow students through five years of Teacher Education with Master of Science. You have all contributed to making the hard times easier and filled my days with laughter and happiness. I have made friends for life. Now, let's go and inspire young souls to embrace science.

Maren Svanem

Abstract

The results of a diet intervention study aiming to enlighten dietary carbohydrates role in proinflammatory responses have been reviewed in light of a new statistical analyses conducted on the microarray data. Two diets, a high-carb (AHC) diet and a moderate-carb (BMC) diet have been studied. The resulting gene expression data have been analyzed internally at NTNU, and subsequently by an external partner, KUL. Overlaps and differences between the diets and between the results of the two analyses performed have been addressed using a system-approach for biological interpretation. The analysis conducted in this project was carried out using a variety of software tools and is based on an already existing data sets. The gene sets were used for building of a regulatory network and for further analysis, specifically with respect to changes in proinflammatory pathways.

The genes affected by the diets and the processes they influence can indeed be related to proinflammatory processes. There have been induced some changes on a transcriptional level in the participants of the diet intervention study, even though the changes barely are perceived as considerable. Every gene that has been studied shows similar change in both diets, if they are upregulated in AHC, they are also upregulated in BMC. The same pattern is also observed for downregulation. After taking KUL's data into consideration and interpreting the results in a new manner, the connection to a proinflammatory response is weakened compared to what was presented in the initial study. The tendencies are, however, existing.

Sammendrag

I lys av en ny statistisk analyse som har blitt utført, har resultatene fra et tidligere diettbasert studium blitt vurdert på nytt. To dietter, en høykarbohydratsdiett (AHC) og en moderatkarbohydratsdiett (BMC) har blitt studert. De resulterende genuttrykksdataene har tidligere vært analysert internt hos NTNU, og i etterkant av en ekstern partner, KUL. Likheter og ulikheter mellom diettene, samt mellom resultatene fra de to analysene som er utført, har blitt adressert via en systemtilnærming for biologisk tolkning. Analysen i dette prosjektet ble utført på allerede eksisterende datasett ved hjelp av en rekke programvareverktøy. Det ble valgt ut noen gener som ble videre brukt i byggingen av et regulatorisk nettverk og for videre analyse, da spesielt med hensyn til endring i proinflammatorisk respons.

Genene som påvirkes av diettene og de prosessene gene påvirker viser seg å kunne være relatert til proinflammatoriske prosesser. Det har skjedd noen endringer på transkripsjonsnivå hos deltakerne studien, selv om endringene er minimale. Hvert gen som har blitt studert viser tilsvarende forandring i begge dietter. Oppregulerte gener i AHC er også oppregulerte i BMC, og det samme gjelder for nedregulering. Etter å ha tatt med KULs data i betraktningen og tolket resultatene på en ny måte, har forbindelsen til en pro-inflammatorisk respons blitt svekket sammenlignet med det som ble presentert i den opprinnelige studien som var basert kun på NTNUs data. De samme trekkene kan likevel observeres til en viss grad.

Contents

List of Figures

List of Tables

Abbreviations

AHC	Diet A – High levels of carbohydrates
BMC	Diet B – Moderate levels of carbohydrates
CAT	Correspondence at the top
cDNA	Complementary deoxyribonucleic acid
DDO	Data-Driven Objective
FC	Fold change
FDR	False Discovery Rate
GO	Gene Ontology
IKK	I$\varkappa$B kinase, leads to activation of NF-$\varkappa$B
I$\varkappa$B	NF-kappaB inhibitor
KEGG	Kyoto Encyclopedia of Genes and Genomes
KUL	Katholieke Universiteit Leuven, here mainly used to refer to the analysis conducted at the university
NF-$\varkappa$B	Nuclear Factor Kappa B
NTNU	Norwegian University for Technology and Science, here mainly used to refer to the analysis conducted at the university
TF	Transcription factor
TLR	Toll-like receptor
TNF	Tumor necrosis factor
UniProtKB	Universal Protein Resource Knowledgebase

1 Introduction

1.1 Diet and health

Many chronic diseases such as obesity, type 2 diabetes and cardiovascular disease are largely determined by lifestyle. The last few decades, several different diets have been suggested to achieve a healthier lifestyle, although many of them are contradictory (Malik & Hu, 2007). Fat quality and quantity has received a fair amount of attention, whereas carbohydrates' role has been less studied. The group of Prof. Berit Johansen has for some years been focusing on the relationship between diet and health. They have found evidence that several proinflammatory markers are elevated when a diet relatively high in carbohydrate content is consumed, whereas a so-called balanced diet (approximately equal amounts of calories from the major nutrient groups carbohydrates, protein and fat) can alleviate these symptoms (Arbo et al., 2010).

1.2 Inflammatory responses on a pathway level

It has been acknowledged that the key role in inflammatory diseases is played by NF-kappaB/Rel transcription family (Tak & Firestein, 2001). The NF-κB/Rel family includes NFKB1 (p50/p105), NFKB2 (p52/p100), p65 (RelA), RelB, and c-Rel (Chen, Castranova, Shi, & Demers, 1999). Nuclear factor kappa B (NF-κB) is a dimer, either a homodimer or a heterodimer, which acts as transcription factor (TF) for several genes in response to inflammatory signals (Barnes & Karin 1997). The dimer is most frequently consisting of either a p50 or a p52 subunit together with p65, in which the latter contains the transactivation domain (Tak & Firestein, 2001). The NF-κB dimer exists in an inhibited state in the cytoplasm, physically bound to a NF-κB inhibitory protein (IκB). Specific IκB kinases (IKKs) respond to certain activation signals, and do hence phosphorylate the IκB protein bound to the NF-κB complex, leading to proteolytic degradation of the inhibitory protein. The free NF-κB can migrate into the nucleus and contribute to the transcription of genes encoding proinflammatory proteins (Barnes & Karin 1997). Proinflammatory proteins include cytokines, chemokines, adhesion molecules, matrix metalloproteinases (MMPs), Cox-2 (UniProt ID: Q05769), and inducible nitric oxide (iNOS) (Tak & Firestein, 2001). The IKK/NF-κB signaling

pathway described here is triggered by certain members of the tumor necrosis factor (TNF) cytokine family, such as TNF-α (gene *TNF*, UniProt ID: P01375), which elicits NF-κB activation (Luo, Kamata, & Karin, 2005).

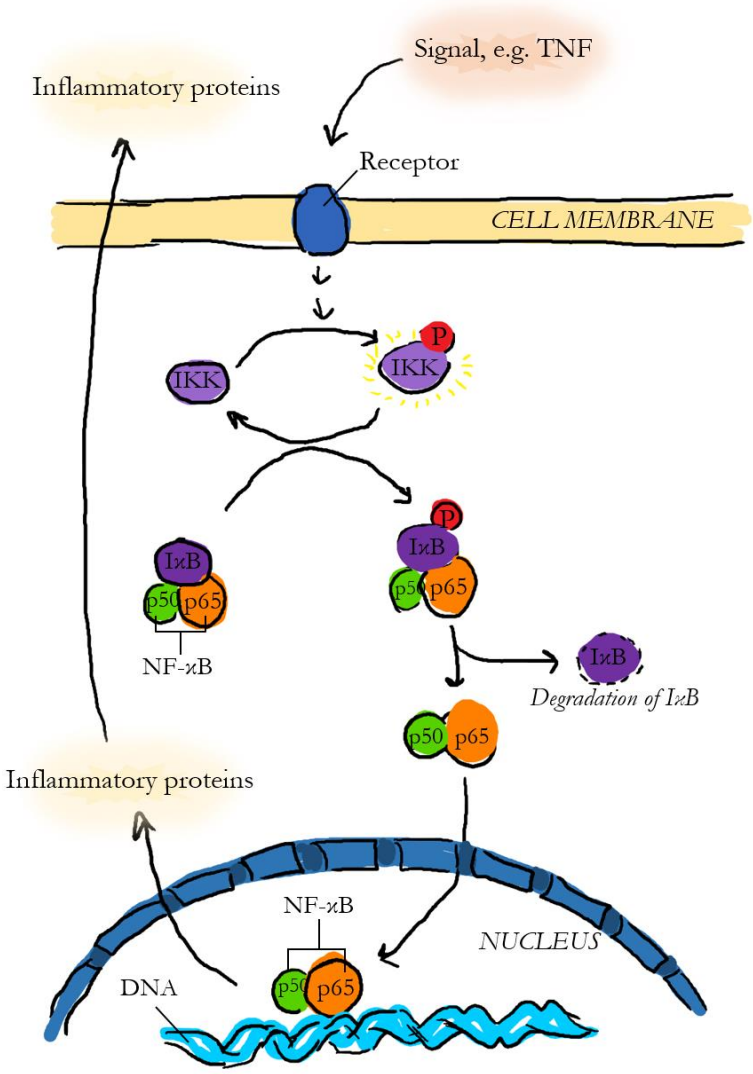


Figure 1. Illustration of the IKK/NF-κB signaling pathway produced in Microsoft Word 2016. The illustration is inspired by Barnes and Karin (1997). An extracellular signal, e.g. TNF, initiates the activation of IKKs, which phosphorylate IκB in the IκB:NF-κB complex and thus releases NF-κB from its inhibitor. NF-κB migrates to the nucleus where it acts as a transcription factor for a variety of inflammatory proteins.

1.3 The diet intervention study

In the Johansen Group's project (Arbo et al., 2010), a small cohort of slightly overweight, but otherwise healthy men and women in the age range 18-30 participated in the study here referred to as 'the diet intervention study'. 32 of the participants completed the study. Each participant completed two diets of different nutrient composition (carbohydrates:proteins:fats): the AHC diet (65:15:20) and BMC diet (27:30:43). Both diets lasted for 6 days each, and there was an 8-days wash-out period between the two diets. Data were collected from the subjects at four time points, before and after each of the two diet periods. The data consists of fasting blood samples, used for analyzing blood markers and leukocyte gene expression.

The collected samples underwent a biochemical analysis to measure levels of triglycerides, total cholesterol, HDL cholesterol, glucose, hemoglobin, total leukocytes, differential count of leukocytes, platelets, hsCRP and uric acid. A protein analysis was also implemented to determine twelve diabetes related biomarkers (in the classes of cytokines, adipokines, gut hormones and incretins, and glucose disposal hormones). A microarray analysis was performed on cDNA. All data were statistically analyzed, including the microarray data. The data was considered significant at $P<0.05$. The analysis was performed by Mette Langaas at the Norwegian University of Science and Technology (NTNU) and is described in Arbo et al. (2010). Langaas' analysis will from this point forward be referred to as 'NTNU' analysis.

Brattbakk (Arbo et al., 2010) concluded that the AHC diet induced changes in gene expression to a much larger extent than the BMC diet, including both up- and downregulation of genes within the same pathways. The AHC diet resulted in expression of 1370 genes, whereas 843 genes overlapped with the BMC diet. All except 10 genes changed in the same direction. Few genes differed among the two diets, but among them were two growth factors and a regulator of DNA methylation. Both diets induced stimulation of genes related to apoptosis, proliferation and cancer. However, genes with relevance to stress and immunity were upregulated by the AHC diet, but downregulated by the BMC diet (Arbo et al., 2010).

Subsequently, the microarray data was statistically analyzed by an external partner, Wim de Mulder, from the Katholieke Universiteit Leuven (KUL), from now on referred to as 'KUL' analysis. NTNU and KUL both considered the data significant at $P<0.05$, but they used two slightly different statistical methods to analyze the microarray data. The difference was mainly in the way correction for multiple testing was performed, thus yielding slightly different results. Both methods were based on a Linear Mixed model approach using either the R statistical software package or the SAS package.

1.4 A system's approach for biological understanding

Knowledge and progress in molecular biology has improved a lot until today. The use of genetic, molecular, and biochemical approaches during the past decades has led to most of the current knowledge (Kim & Ren, 2006). Different techniques and approaches allow genome sequencing and high-throughput measurements, enabling collection of comprehensive data and information regarding the underlying molecules of systems performance. However, the identification of genes and proteins in an organism is not sufficient to understand its complexity. A system-level understanding requires a change in mindset, shifting the focus from genes and proteins, to structure and dynamics (Kitano, 2002). Systems biology should explain a system on several levels at once; from molecular pathways and regulatory pathways, through cells and organs, and ultimately to the level of the whole organism (Wierling, Herwig, & Lehrach, 2007).

The availability of genome sequences has also contributed to the rise of other technologies: the 'omics' technologies. 'Omics', like transcriptomics and proteomics, are helpful to identify genes and gene products, as well as the relationships between them. These results should however be viewed with caution due to a wide occurrence of false-positive and false-negative results. 'Omics' depend on annotations, and single annotations are not adequate for a full description of a gene's function. To get data that are more informative on relationships and interactions, data from several separate experiments should be combined and integrated. By systematically identifying interactions between protein-protein, protein-DNA or protein-RNA, interaction networks could emerge (Ge, Walhout, & Vidal, 2003). The building of a biological network requires understanding of structure, function, and dynamics of the individual components, as well as their effect on each other. Studies of biological networks require mapping of information regarding thousands of proteins, RNAs, promoter sites, and other macromolecules, all at once. The information is further used to make network maps, which is generally visualized as nodes representing the biological components, and edges representing the interaction that connect them (Brasch, Hartley, & Vidal, 2004).

Data integration is an important part of systems biology. Access to databases and public repositories that store functional high-throughput data and annotations of protein function and biological pathways is important in the most fundamental step towards a biological network. Among the many databases that exist, the ones dedicated particularly to pathways could be an importance resource (Wierling et al., 2007). By using the information acquired from databases, network models can be set up to summarize all relevant reactions, interactions, and processes. Models of biological networks are cornerstones of systems biology (Shannon et al., 2003). There are different software tools available for modeling, such as Cytoscape (Shannon et al., 2003; Wierling et al., 2007).

1.5 Aim of master project

Starting with the microarray data from the diet intervention study and the P values produced by both NTNU and KUL, the aim in this thesis was to identify possible differences based on the two statistical analyses, and the new data was used to produce annotated networks in Cytoscape. The networks were built by connecting the different genes together by gene-protein interactions and protein-protein interaction. The genes and the networks were analyzed using different analysis tools to allow a biological system interpretation of the previous and the new data. The findings were compared to those of Hans-Richard Brattbakk, who took part in the initial study with the Berit Johansen Group. His results can be viewed in Arbo et al. (2010). Possible extensions that can either confirm the initial conclusion or identify of discrepancies were searched for.

2

Materials and methods

The analysis conducted in this master thesis was carried out using a variety of software tools and is based on an already existing data set. The data set is the original gene expression data set produced using Illumina microarray for the samples from the diet intervention study conducted by the Johansen Group (Arbo et al., 2010). Two different methods for statistical analysis have been used on the microarray data, one was conducted at NTNU and the other was performed by KUL. The results of both statistical approaches were taken into consideration in this thesis, and an effort to explain the differences have been performed. The initial gene lists ranked by P values were compared using a Correspondence at the Top (CAT) analysis (section 2.1). The two statistical data sets were, together with the original microarray data, used to compare gene sets and pick out smaller subsets of genes based on different criteria (section 2.2), for analysis of the gene list subsets with respect to processes and terms related to the genes (section 2.3), and for interpretation of biological significance using network building approaches (section 2.4 and 2.5). The gene sets were used for building of a regulatory network using a data-driven objective (DDO) and further analysis, specifically the response with respect to changes to proinflammatory pathways. A flowchart illustrating the work done can be viewed in **Figure 2**.

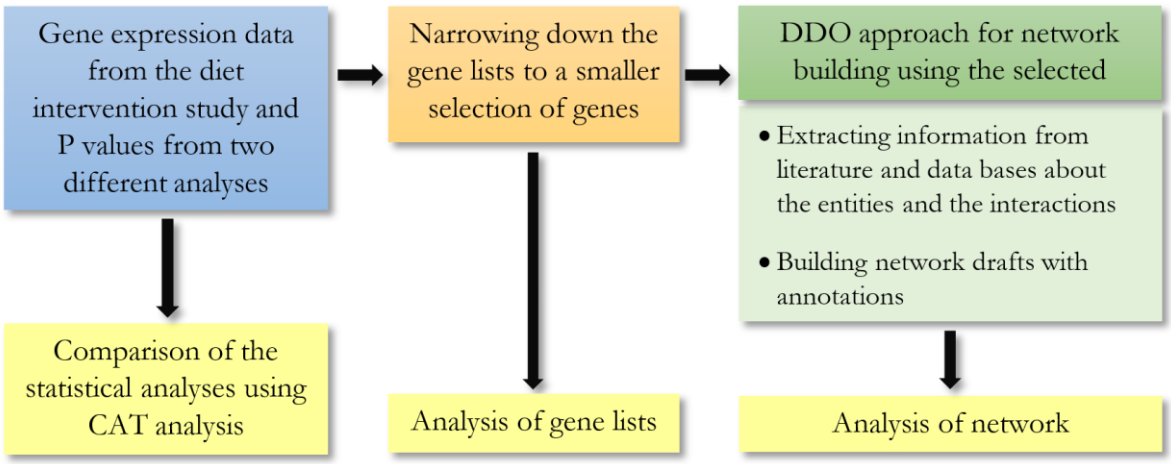


Figure 2. Flowchart overview of the work conducted in this master project.

2.1 Comparing the initial gene lists using CAT analysis

In an attempt to address the similarities and differences between the two statistical analyses conducted on the data produced in the diet intervention study, a CAT analysis was carried out to compare the gene lists with respect to the P values associated to each gene. The initial gene lists were ranked purely by P value, from lowest to highest. A CAT analysis compares the correspondence at the top by plotting the proportion of genes in common between two lists against the lists size, yielding proportion of agreement measures (Irizarry et al., 2005). CAT analysis is often used for comparing differential gene expression results retrieved from different microarray platforms (Gupta & Marchionni, 2012), such as the microarray data from the diet intervention study.

The CAT analysis was conducted in RStudio (download from Rstudio.com) using the `matchbox` R package (Marchionni & Gupta, 2013). Two different CAT analyses were conducted, using both the 'equalRank' parameter, which compare gene ranks only, and the 'equalStat' parameter, which take the genes' assigned P values into consideration.

2.2 Analysis for selecting genes for the final gene list

The initial data sets contain a large quantity of genes of different statistical significance. **Table 1** shows the number of genes in each data set used. To narrow down the selection of genes to an attainable size, different approaches for comparing the gene lists have been used. By comparing the most significant genes in both diets (AHC and BMC) based on both statistical analyses (KUL and NTNU), the aim for this part of the thesis is to get a list of the approximately hundred most significantly up- or downregulated genes.

To further guide the selection of genes, the lists were compared with respect to both the adjusted P value for the change in gene expression and the fold change (FC) value describing the quantity of change. Both statistical analyses were used when selecting the genes, thereby increasing the possibility of including the most significant genes. It is, however, important to keep in mind that the different statistical approaches can introduce more false positives, which is why the main selection of genes will be based on the genes of significance in both statistical data sets. Results with $P>0.05$ have a 5% chance of being a false positive, and in this thesis, genes with a $P>0.05$ are not considered. Nonetheless, there are a lot of genes to consider with $P<0.05$ (**Table 1**). To narrow down the number of genes even further, the magnitude of fold change (FC) was taken into account using a Volcano plot code ran in RStudio.

Table 1. Number of genes in the initial datasets.

Dataset	All genes	Genes w/ P<0.05
Gene expression data	27372	-
AHC KUL	3717	3443
BMC KUL	3717	3583
AHC NTNU	3379	3353
BMC NTNU	630	602

2.2.1 Volcano plot

As an initial step to creating a Volcano plot, it is necessary to calculate the fold change (FC) for the genes analyzed in the diet intervention study. FC is a value which gives information regarding whether a gene is either upregulated or downregulated between two different experimental groups and how much the expression levels have changed. By using gene expression data from microarray (or other approaches yielding expression data), expression values for a control sample and an experimental sample can be used to calculate the FC. In the diet intervention study, microarray analysis was used to address the genes' reponses to the different diets in the participants. In the data set used here, the gene expression data was presented as log2 values. To calculate the FC, the log2 expression data for each gene in each participant were used by subtracting the initial 'control' data (day 0, d0) from the final 'experimental' data (day 7, d7) (**Equation 1**).

$$\log2\ FC = d7 - d0 \qquad \textbf{(Equation 1)}$$

Log2 FC values were calculated for each microarray probe for all individual participants. The mean log2 FC value was calculated for each probe by adding the log2 FC values for the particular probe for each participant and dividing the sum on number of participants.

By using the calculated log2 FC values in combination with the P values from the statistical analyses conducted by KUL and NTNU, a Volcano plot can be created. A Volcano plot is a graphical representation in the shape of a scatter plot, where the dots represent genes scattered in two dimensions (Cui & Churchill, 2003). The y axis is a negative log10-transformed axis for P values, thus placing the genes with the lowest P values in the upper area of the graphical plot. A horizontal

threshold can be set, placing the genes that are considered statistically significant above the threshold line. Along the x axis, the genes are separated based on the log2 FC – downregulated genes on the negative axis, and downregulated genes on the positive axis. A pair of vertical threshold lines is used to delineate the genes with a large enough FC to be included in further analysis. In this way, genes of statistical significance and a considerably large FC (of your own choice), will be located in the upper left and/or upper right parts of the plot, making it more intuitive to see which genes to study further (Cui & Churchill, 2003).

In this thesis, the Volcano plot was produced in RStudio. The code used can be reviewed in **Appendix 1**. Instead of using solid lines to separate the dots in the scatter plot, colors (**Table 2**) were used to identify which genes fit the different criteria.

Table 2. Color interpretation in the Volcano plots. Genes with *-marked values were not part of the output file produced by the Volcano plot code.

Color		P value	Log2 FC value
Black		> 0.05*	< 0.38*
Orange		> 0.05*	> 0.38
Red		< 0.05	< 0.38*
Green		< 0.05	> 0.38
Blue		< 0.05	> 0.50
Turquoise		< 0.05	> 0.68

2.2.2 Cross-ranking of genes based on P value and log2FC value

In addition to the graphical representation the Volcano plot provides, the R code produced files containing a list with the top genes in each Volcano plot. The top genes were the genes with P<0.05 and a log2 FC>0.38 when writing the code which extracted them from the original data set. However, in the produced files, the lists were ranked by P value only, and did consequently not give the full information the Volcano plots represented. Nevertheless, by manually cross-ranking the lists based on both P value and log2 FC value, the lists will provide similar information to the Volcano plot. By first ranking the genes from lowest to highest P value (the lowest P value gets rank 1, and the rank increases by one with each increase in P value), and thereby ranking the genes from highest

to lowest log2 FC value (giving the highest value a rank 1 etc.), a total rank based on adding the P value rank to the log2 FC value rank can give information about which genes are both statistically significant and have a relatively high fold change. This was executed for both AHC and BMC based on the statistical analysis of both KUL and NTNU.

2.2.3 Selection of statistically significant observations

The four gene lists received after the cross-ranking were used in further comparison (**Figure 3**), giving information regarding common and unique genes between the lists. To compare them, a bioinformatics and research tool (Whitehead Institute for Biomedical Research, 2013) was used. The lists were compared in different manners, and the result for each single comparison was three lists of genes: one with the genes unique to the first entry list, one with the genes unique to the second entry list, and one with the genes common to both entry lists. The comparison was conducted for the combination of the two AHC diets, for the two BMC diets, for AHC and BMC based on KUL's statistical data, and for AHC and BMC based on NTNU's statistical data. In addition, further comparisons were made to identify the genes unique to each diet.

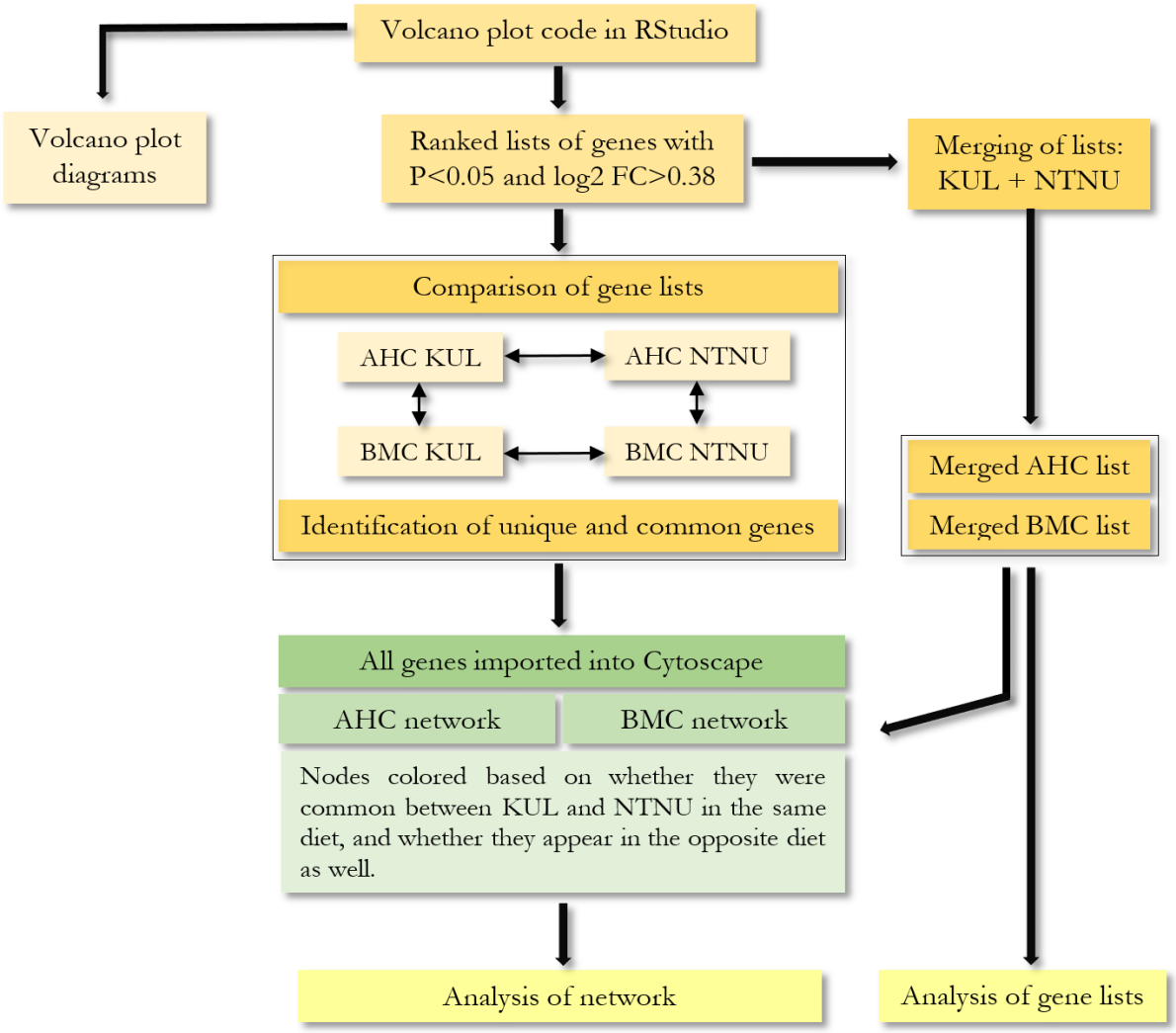


Figure 3. Flowchart illustrating how the selected genes for the final gene lists were fed into the downstream analysis.

2.3 Analysis of gene lists

The selected gene lists (**Appendix 3**) were analyzed to gain information regarding the genes' function as a set. First, an overrepresentation analysis was conducted to identify enriched biological terms connected to the gene sets. Subsequently, a pathway-based analysis was done, in which the gene sets were analyzed with respect to biological pathways.

2.5.1 Overrepresentation analysis

As a first step to gaining information about the gene sets, two different overrepresentation analyses were conducted. An overrepresentation analysis compares a gen set to a random reference set with the intention of discovering GO terms connected to the genes and that appear more frequently in the input gene set compared to the reference set. Elevated terms are referred to as overrepresented terms in a gene set. In this project, the Gene Ontology tools BiNGO and ClueGO were used to achieve overrepresented GO terms in AHC and BMC gene lists. The gene lists were at this point merged together, thus including both statistical analyses. The two analyses were chosen due to the differentially organized output they produce.

BiNGO is a Cytoscape plug-in (Cytoscape is described in section 2.4.1) which assesses the overrepresentation of GO categories in a set of genes (Maere, Heymans, & Kuiper, 2005). The genes in the test set are connected to relevant GO annotations throughout the GO hierarchy, and the test set is subsequently compared to a random reference set. Assuming a hypergeometric distribution, GO terms that appear more frequently in the test set compared to the random reference set are presented in the results. The results provided by BiNGO contain both a visual and text-based aspect. The visual representation is a hierarchical rendering of the GO tree and the nodes are labeled with GO terms and are colored based on the P value, which should give an idea of the relevance of the specific GO term. The text-file is tab-delimited and contains more detailed results, including analysis options, adjusted P value for each significantly overrepresented GO class, the number and identities of the test set genes which are annotated to the specific GO classes, as well as the number of genes annotated to the classes in the reference set (Maere et al., 2005).

ClueGO, also a Cytoscape plug-in, integrates GO terms and Kyoto Encyclopedia of Genes and Genomes (KEGG)/BioCarta pathways into a network. GO annotates genes to different biological, cellular, and/or molecular terms in a hierarchical way, while KEGG and BioCarta assign the genes to different functional pathways. Instead of using the hierarchical ontology tree to link overrepresented GO terms together, which is the case for BiNGO, ClueGO uses kappa statistics, which

indicated the extent to which two GO terms annotate the same genes in the test set, to link the terms to each other. The result is a functionally organized GO/pathway network, with nodes representing the terms, linked together based on a predefined kappa score level (Bindea et al., 2009).

2.5.2 Pathway-based analysis

As a tool for retrieving information regarding the pathways involved in the network, the knowledge base Reactome (Reactome.org) was used. Reactome is an online curated resource for human pathway data and analysis tools (Vastrik et al., 2007), and therefore a useful resource to retrieve the knowledge sought in this thesis. By uploading gene lists in Reactome, the genes are cross-referenced with the Reactome database, which is manually curated, as well as to several external databases, such as UniProtKB. In addition to being a knowledge base, Reactome provides a computational tool which can aid in the interpretation of microarray data (Vastrik et al., 2007). Uploading gene lists with both gene identifiers and their respective FC value provide intuitive information of whether processes and pathways have been affected by the conditions studied in a microarray experiment such as the one performed in the diet intervention study.

The web interface was used in the gathering of pathway knowledge in this project. The selected gene lists were analyzed separately for subsequent comparison. In addition to analyzing the genes considered of interest, the full gene lists from the initial data were analyzed as a basis for further assessment. The pathways connected to the diets was compared between the diets with the intention of finding possible patterns in common or unique pathways between the different diets as well as between the lists based on different statistical analyses. The different AHC diets were compared to each other, and so were the different BMC diets. In addition, the AHC and BMC diets were compared to each other.

2.4 Networking

With gene lists as a starting point, a DDO approach to network building was performed. DDO is an approach used for generating information regarding the relationships between genes and/or proteins identified in an experiment, such as a microarray study, in which the relationships are typically not well understood (Viswanathan, Seto, Patil, Nudelman, & Sealfon, 2008). In this thesis, the final gene lists will serve as the identified genes. The first step in DDO pathway construction is to retrieve information from relevant sources by text-mining. The information gained is further

used for the construction of a pathway prototype (Viswanathan et al., 2008). The gathered information was assembled into a pathway prototype using the pathway building tool Cytoscape version 3.4.0 (download from Cytoscape.org).

2.4.1 Cytoscape

Cytoscape provide visualization, modeling and analysis of molecular and genetic interaction networks (Cline et al., 2007). Biomolecular interaction networks can be integrated in Cytoscape with high-throughput expression data and other molecular states. Cytoscape is especially useful in conjunction with large databases of protein-protein, protein-DNA, and genetic interactions for humans and model organisms. The Cytoscape software allows several different plug-in modules which extends the use of Cytoscape (Shannon et al., 2003). Networks built in Cytoscape contains nodes that represent biological entities, such as genes or proteins. These nodes are connected via edges which represent pairwise interactions. The nodes and edges can be visually modified for properties such as color, shape and size, which contributes to the visual aspect of Cytoscape (Cline et al., 2007).

As a first step in the networking process, all genes from the final gene lists were imported into Cytoscape as nodes. To connect the nodes with edges, which are representing interactions, different tools were used. To get a quick idea of which genes to connect, both gene lists were loaded into GeneMANIA (Genemania.org), a web interface that uses a large resource of available genomics and proteomics data to create interactive functional association network that can aid in the search for gene function and relationships. By entering a query gene list, GeneMANIA connects and extends the list by adding functionally similar genes from publicly available databases (Warde-Farley et al., 2010). The result from GeneMANIA was considered in the research of protein-protein and gene-protein interactions.

In addition, two genes were added purely for analysis intentions. *RELA* (UniProt ID: Q04206) and *NFKB1* (UniProt ID: P19838) are two of the most common subunits of the NF-κB dimer, and by attempting to connect them to the network, possible connections might come to light. The genes were added together with the gene lists in GeneMANIA, and they were considered as a dimer. In the network, they are referred to as 'NF-kB complex', and are connected to all genes with a connection to either *RELA* and/or *NFKB1* in GeneMANIA. *TNF* (UniProt ID: P01375) was added as well, due to its potential impact on NF-κB activation.

2.4.2 Text-mining

As an initial step, all genes in the final gene lists were researched as individual genes. The purpose was to gain an initial and superficial view of the protein products and the processes they are involved in. The task was conducted using the Universal Protein Resource Knowledgebase (UniProtKB). UniProtKB is a database containing integrated protein information with cross-references to multiple sources. There are two sections in UniProtKB: Swiss-Prot and TrEMBL, both containing information extracted from literature and computational analysis, making UniProtKB a rich knowledgebase. In Swiss-Prot, the information is a manually and continuously annotated by an expert team of biologists (UniProt Consortium, 2009), which contributes to UniProtKB's credibility. It is useful for finding information regarding for example the transcribed protein, synonyms, subcellular location, and biological processes. The information gained in this step was mainly used for annotating nodes in Cytoscape.

To gain more specific knowledge of the interactions between the genes, two text mining tools were used for literature research: Information Hyperlinked over Proteins (iHOP) and LitInspector for text mining. LitInspector is a search tool for literature within the NCBI PubMed database (Frisch, Klocke, Haltmeier, & Frech, 2009). It is a useful tool in gene and signal transduction pathway mining, and yields results containing PubMed abstracts where the genes, transcription factors and key words are highlighted. The highlights are color-coded, making it the reading easier and more efficient. LitInspector's ability to consider search results for all synonyms of a gene is advantageous. LitInspector provides a high gene recognition quality due to the strategies for homonym resolution and rejection of 'non-gene' abbreviations. The gene recognition is based on the comprehensive gene synonym list of NCBI's Entrez Gene. LitInspector also allows a search of three genes at a time, where OR or AND functions can help narrow down a search. Because LitInspector is an automatic pathway mining tool and not manually curated, the results are always up to date. The results provide an overview of all possible pathway associations and potential interactions of the query gene(s). Literature references are provided, so the user can verify the results for him-/herself (Frisch et al., 2009). However, due to LitInspector's license demand, it was used in a limited trial time only. The free option iHOP was therefore more frequently used.

iHOP structures and links together biomedical literature from PubMed by using genes and proteins as hyperlinks, thus making it possible to navigate through the sea of existing literature in one continuously updated workspace (Hoffmann & Valencia, 2005). iHOP does however only allow one gene to be searched at a time. To narrow down the search, the connections produced by GeneMANIA were prioritized. The information mined through LitInspector and iHOP were used to

gain information about the nodes and in the attempt to annotate the edges in the Cytoscape networks.

2.5 Network-based analysis

A network understanding of a biological system can give insight into connections that are challenging to discover in any other way. By analyzing the genes and their connections, new insight to their function may be discovered. The completed networks were analyzed mainly by using different Cytoscape plug-ins.

2.5.3 Graph-based analysis

NetworkAnalyzer is a Cytoscape plug-in that performs analysis of biological networks and calculates network topology parameters. These parameters include the diameter of a network, the average number of neighbors, and the number of connected pairs of nodes. In addition, it calculates more complex parameters, including node degrees, average clustering coefficients, topological coefficients, and shortest path lengths (Smoot, Albrecht, & Assenov, 2016). NetworkAnalyzer was used to analyze both the AHC and the BMC network. To interpret the results provided by NetworkAnalyzer, information regarding each parameter was retrieved from the NetworkAnalyzer Online Help (Max-Planck-Institut für Informatik). NetworkAnalyzer provides a summary of simple parameters in a list, as well as more complex parameters which can be reviewed as graphical presentations.

The simple parameters include the total *number of nodes* in the network, as well as how many of them that are *isolated nodes*, meaning that they have zero neighbors. The nodes' *average number of neighbors* and the *characteristic path length*: the expected distance between the nodes (measured in number of nodes that act as bridges between two specific nodes) are also provided. The *clustering coefficient* of the network describes the average cohesiveness of all nodes' neighborhoods by quantifying the different node neighborhoods' chance of being part of a clique where every node is connected to each other (Albert, 2005). Nodes with less than two neighbors have a clustering coefficient of zero (Max-Planck-Institut für Informatik). The nodes are connected through paths of edges, and all nodes that are connected in pairs are thus *connected components* of the network. The number of connected components calculated by NetworkAnalyzer give information regarding the network's connectivity, in which a lower number indicates a stronger connectivity. The *network density* is a value between 0 and 1 which indicates the density of edges in the network – zero edges gives a density

of 0, while a clique where all nodes are connected to each other gives a density of 1(Max-Planck-Institut für Informatik). *Network heterogeneity* is a measurement of hub tendencies in the network (Dong & Horvath, 2007).

The more complex parameters involve e.g. betweenness centrality and node degree distribution. The *betweenness centrality* of a node reflects the influence the node have on the interactions of other nodes in the network (Yoon, Blumer, & Lee, 2006). The *node degree distribution* shows the number of edges linked to the nodes in the network. The node degree distribution can be used to distinguish between random and scale-free networks (Barabasi & Oltvai, 2004).

2.5.4 Superimposing of data from the microarray onto the networks

To visualize the gene expression data in Cytoscape, the genes' respective log2 FC values were superimposed onto the two networks. By adjusting the setting in the 'style' section in Cytoscape, the nodes were colored by a color gradient ranging from red (log2 FC = -1.0) to green (log2 FC = 1.0), with white as a zero-point color (**Figure 4**). Because several genes had multiple probes on the microarray, and therefore multiple log2 FC values, the mean log2 FC for these genes were calculated and used in the data overlay. In addition, the nodes' size was adjusted based on P value from both the KUL and the NTNU analysis. The node size is conversely proportional to the P value, yielding a bigger node as the P value decreases. The biggest nodes are thus the statistically most significant. The smallest node size was set at P=0.5.

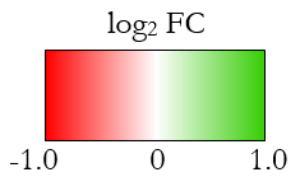


Figure 4. Log2 FC color gradient used in the network data overlay. A log2 FC of -1.0 indicates a 1x downregulation in gene expression (halving the number of transcripts), whereas a log2 FC of 1.0 indicates a 1x upregulation in gene expression (doubling the number of transcripts).

3

Results

In the light of a new statistical analysis of the diet intervention data produced by Berit Johansen's group, previous and new results have been addressed and compared. The differences between the resulting gene lists based on different correction for multiple testing by KUL and NTNU was attempted enlightened by a CAT analysis. The data were further used in a DDO approach with the aim of gaining a system understanding of the set of genes that have been affected by the diets.

3.1 CAT analysis

To gain insight into the differences between the two statistical approaches, two different CAT analyses were performed on the four initial gene lists with respect to both rank ('equalRank') and P values ('equalStat'). Graphical representation of the results from the CAT analysis can be viewed in **Appendix 2**. If the lists compared are completely identical with respect to the genes and their P values, the correspondence in P values and thus also the ranking by P value should be identical. That is, 'equalRank' analysis and the 'equalStat' analysis should yield identical results.

There is a notable difference between the two CAT analyses. The 'equalRank' results (**Table 3**) show a correspondence that is overall lower than for the 'equalStat' results (**Table 4**) for all four comparisons. There are, however, observed some similar trends between the 'equalRank' and the 'equalStat' results. The BMC diets are more similar to each other that the AHC diets, but neither have a high correspondence, especially when considering the 'equalRank' results. The BMC diets do, correspond considerably based on the 'equalStat' results. When reaching top 200, the AHC diets have a notable correspondence as well. When looking at the top 50 genes in all comparisons, the correspondence is smallest for 'KUL AHC-BMC'. In the NTNU analysis, the correspondence between AHC and BMC ('NTNU AHC-BMC') is the highest of all comparisons. There are no data for the top 1000 in any comparison that include NTNU's BMC due to the lower number of genes in the list. In the 'NTNU AHC-BMC 'comparisons, no results are presented for top 500 either, the reason behind this is still unknown.

Table 3. Results from the CAT analysis performed in RStudio using the 'matchBox' package and the 'equalRank' parameter. 'AHC KUL-NTNU' is the comparison of the two different AHC gene lists based on the two different statistical analyses performed by KUL and NTNU. 'BMC KUL-NTNU' is the comparison of the two different BMC gene lists based on the two different statistical analyses performed by KUL and NTNU. 'KUL AHC-BMC' is the comparison of AHC and BMC based on the KUL analysis, while 'NTNU AHC-BMC' compares AHC and BMC based on the NTNU analysis.

Top # genes	AHC KUL-NTNU	BMC KUL-NTNU	KUL AHC-BMC	NTNU AHC-BMC
50	0.020	0.080	0.020	0.100
100	0.040	0.150	0.070	0.300
200	0.085	0.360	0.140	0.565
500	0.218	0.896	0.260	-
1000	0.403	-	0.384	-

Table 4. Results from the CAT analysis performed in RStudio using the 'matchBox' package and the 'equalStat' parameter. 'AHC KUL-NTNU' is the comparison of the two different AHC gene lists based on the two different statistical analyses performed by KUL and NTNU. 'BMC KUL-NTNU' is the comparison of the two different BMC gene lists based on the two different statistical analyses performed by KUL and NTNU. 'KUL AHC-BMC' is the comparison of AHC and BMC based on the KUL analysis, while 'NTNU AHC-BMC' compares AHC and BMC based on the NTNU analysis.

Top # genes	AHC KUL-NTNU	BMC KUL-NTNU	KUL AHC-BMC	NTNU AHC-BMC
50	0.604	0.867	0.494	0.982
100	0.866	0.927	0.633	0.983
200	0.946	0.950	0.752	0.975
500	0.972	0.986	0.864	-
1000	0.984	-	0.922	-

3.2 Analysis for selection of final gene list

To narrow down the selection of genes, the lists were compared with respect to both the adjusted P value for the change in gene expression and the fold change (FC) value describing the quantity of change.

3.2.1 Volcano Plots

By using the calculated log2 FC values in combination with the P values from the statistical analyses conducted by both KUL and NTNU, four different Volcano plots were created – one for each diet, whereas both diets were analyzed using both statistical datasets. Genes of statistical significance and log2 FC > 0.38 are located in the upper left and/or upper right areas of the plots, colored in green, blue and turquoise (**Table 2**).

The Volcano plots for AHC (**Figure 5**) and BMC (**Figure 6**) show that the majority of the genes have a log2 FC close to zero. The gene dots are evenly distributed in the plot with respect to the P values produced by KUL for both AHC and BMC, whereas NTNU's data have a distinct cut-off at -log10 P≈1.5, corresponding to P=0.05. Overall, there is no remarkable change in gene expression, regardless of P value. A few genes exceeded the log2 FC>0.68 limit, corresponding to a 70% change (**Table 5**). *DEFA3* (UniProt ID: P59666) is downregulated in AHC, and *HMGA1* (UniProt ID: P17096) is upregulated in AHC. *PKM2* (UniProt ID: P14618) and *PKD1P1* (no UniProt ID) are upregulated in BMC. *NRGN* (UniProtID: Q92686) and *GRINA* (UniProt ID: Q7Z429) are upregulated in both diets.

Table 5. The genes with a log2 FC>0.68 and P>0.05 in the gene lists.

Gene list		Downregulated genes	Upregulated genes
AHC	NTNU	-	*NRGN, HMGA1*
	KUL	*DEFA3*	*NRGN, HMGA1, GRINA*
BMC	NTNU	-	*NRGN, GRINA, PKM2, LOC339047 (PKD1P1)*
	KUL	-	*NRGN, GRINA, PKM2, LOC339047 (PKD1P1)*

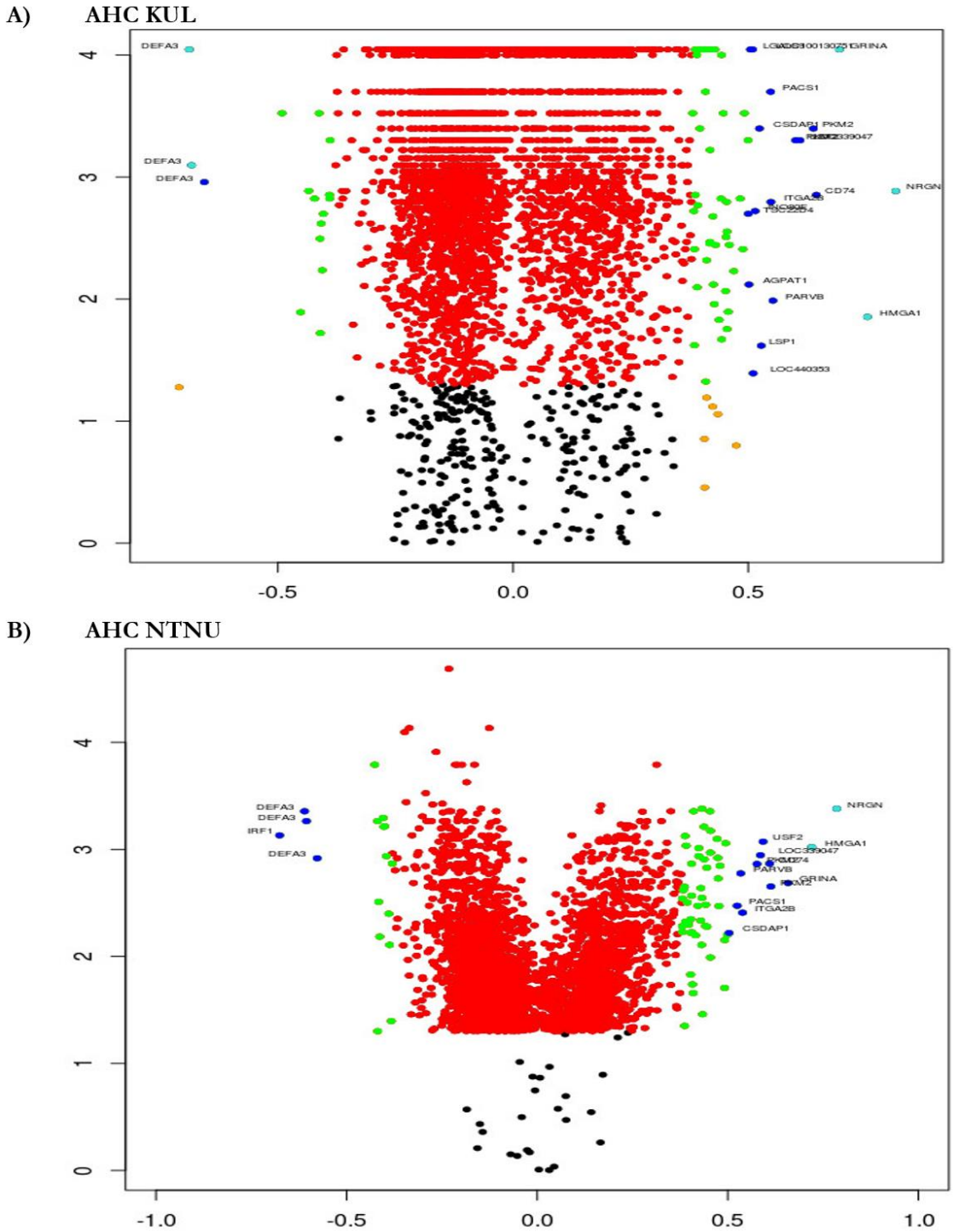


Figure 5. Volcano plot produced in R studio using: **A)** KUL's statistical data for AHC, which initially contained 3717 genes. 79 of the input genes are colored in either green, blue or turquoise, and these are the genes considered of interest. **B)** NTNU's statistical data for AHC, which contained 3379 genes initially. 78 of the input genes are colored either in green, blue or turquoise. The log2 FC can be viewed along the X axes, while the Y axes show the -log10 of adjusted P values from the respective diet.

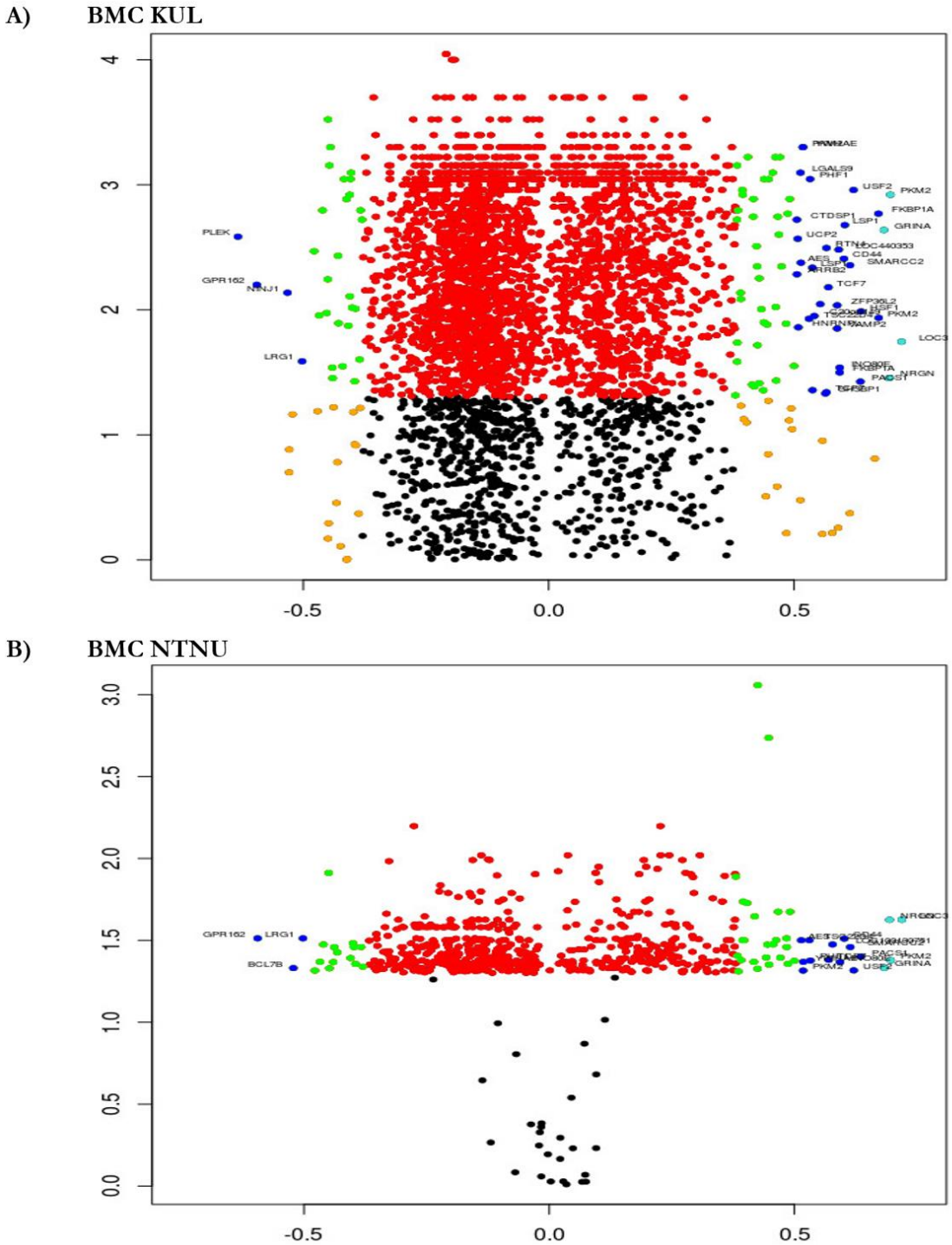


Figure 6. Volcano plot produced in R studio using **A)** KUL's data for BMC, which initially contained 3717 genes. 104 of the input genes are colored in green, and thus considered of interest. No genes are blue or turquoise. **B)** NTNU's data for BMC, which contained 630 genes initially. 60 of the input genes are colored in either green, blue or turquoise, and are thus of interest. The log2 FC can be viewed along the X axes, while the Y axes show the -log10 of adjusted P values from the respective diet.

3.2.2 Cross-ranking and comparison of lists

The four gene lists produced by the Volcano plot code were cross-ranked, resulting in four lists in which the genes were ranked based on the combination of low P value and high log2 FC (**Appendix 3**). The gene lists were compared with respect to the genes they contained to gain information about unique and common genes between the diets. The number of genes in each category is summarized in the Venn diagrams shown in **Figure 7**. The genes involved in each category can be reviewed in **Appendix 4**. The two AHCs have more genes in common than the two BMCs. The BMCs have a distribution similar to the AHC vs. BMC comparisons. The two statistical analyses could thus seem to agree in which genes that are of significance in AHC, but the new analysis by KUL adds 50 new genes to BMC. In total, there are now more genes assigned to BMC than AHC. Many genes are common between the two diets.

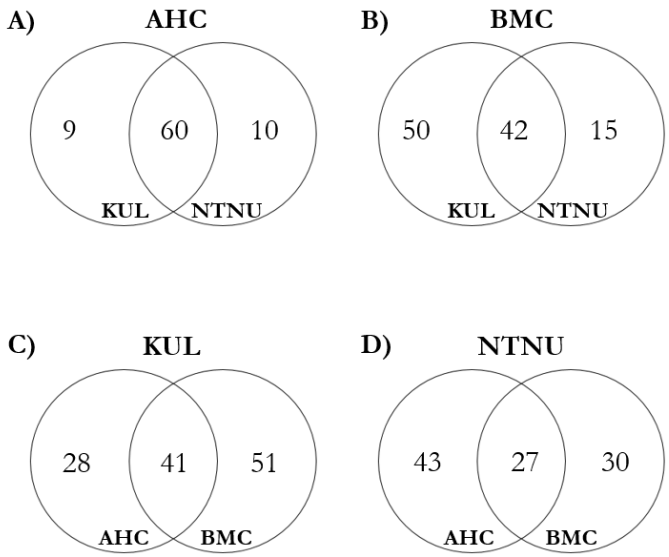


Figure 7. Venn diagrams showing overlap between the different gene lists after cross-ranking based on both P value and log2 FC. **A)** AHC based on the P values produced by KUL (left) vs. AHC based on the P values produced by NTNU (right). The respective genes can be viewed in **Table 18**. **B)** BMC based on the P values produced by KUL (left) vs. BMC based on the P values produced by NTNU (right). The respective genes can be viewed in **Table 19**. **C)** AHC (left) vs. BMC (right) based on the P values produced by KUL. The respective genes can be viewed in **Table 20**. **D)** AHC (left) vs. BMC (right) based on the P values produced by NTNU. The respective genes can be viewed in **Table 21**.

3.3 Analysis of gene lists

The selected gene lists (**Appendix 3**) were analyzed to gain information regarding the genes' function as a set. An overrepresentation analysis was conducted to identify enriched biological terms connected to the gene sets using both ClueGO and BiNGO.

3.3.1 Overrepresentation analysis

The results from the ClueGO analysis (**Table 6**) show that different GO terms which are featured in the two diets. For AHC, the overrepresented terms include homotypic cell-cell adhesion, platelet aggregation, response to gamma radiation, negative regulation of response to DNA damage stimulus, and negative regulation of intrinsic apoptotic signaling pathway in response to DNA damage. The genes connected to these terms are upregulated, except for the protein kinase C substrate-encoding gene *PLEK* (UniProt ID: P08567), as well as the Serine/threonine-protein kinase gene *PRKDC* (UniProt ID: P78527). The overrepresented terms in BMC are connected to type I interferon production, interleukin-12 production, interferon-beta production, regulation of cytokine biosynthetic process, and regulation of toll-like receptor signaling pathway. The genes associated with the overrepresented terms of highest significance in BMC were mostly downregulated, except from the genes *ARRB2* (UniProt ID: P32121), *LGALS9* (UniProt ID: O00182) and *CD4* (UniProt ID: P01730), which were upregulated.

The full list of overrepresented terms for AHC is shown in **Figure 8** and for BMC in **Figure 9**, and the associated GO IDs, P values and genes can be viewed in **Appendix 5**. The term group colored in the colder purple is the most prominent in AHC. These terms are all connected to cell cycle control. It is the three same genes (*BTG2*, UniProt ID: P78543; *PRKDC,* UniProt ID: P78527; *RBM38*, UniProt ID: Q9H0Z9) that are associated to all terms involved in cell cycle control. The group colored in red is involved in apoptosis and response to DNA damage, and the warmer purple is types of cell-cell adhesions. In BMC, the groups are of other categories. Cell cycle response is still represented, but in to a lesser extent than in AHC. Other terms are presented, with groups connected to e.g. lymphocytes, interleukins, the TNF superfamily, interferons, and TLR signaling.

The BiNGO analysis output included an extensive list of GO terms (**Appendix 6**), in which many were comprehensive and general terms in the GO hierarchy, e.g. 'positive regulation of biological processes'. The analysis did not yield many significantly overrepresented GO terms of interest. Nevertheless, a few terms associated with inflammation were elevated (**Table 7**). The pattern of up- and downregulated genes is not as apparent here. The terms are mainly similar between the

diets, but the genes associated to them and the P values differ to some extent, as well as some terms being unique to the different diets. 'Platelet aggregation' is mentioned for AHC, which is common for the ClueGO and the BiNGO analyses. The top unique term assigned to BMC is 'regulation of ERK1 and ERK2 cascade'. The remaining terms are common between the diets. The most significant terms are connected to metabolism in both diets.

Table 6. The significantly most prominent GO terms from the ClueGO overrepresentation analysis. The table is an excerpt from the tables in appendix showing all overrepresented terms and attributing data. The remaining data can be viewed in Appendix 5.

	GO Term	**Associated Genes**
AHC	homotypic cell-cell adhesion	*ALOX12, FERMT3, ITGA2B, ITGB3, PLEK*
	platelet aggregation	*ALOX12, FERMT3, ITGA2B, ITGB3, PLEK*
	response to gamma radiation	*BCL2L1, HSF1, PRKDC*
	negative regulation of response to DNA damage stimulus	*BCL2L1, CD44, CD74, HSF1*
	regulation of intrinsic apoptotic signaling pathway in response to DNA damage	*BCL2L1, CD44, CD74*
	negative regulation of intrinsic apoptotic signaling pathway in response to DNA damage	*BCL2L1, CD44, CD74*
BMC	type I interferon production	*IRF1, POLR3H, PRKDC, RNF216, TICAM1*
	interleukin-12 production	*ARRB2, IRF1, LGALS9*
	regulation of type I interferon production	*IRF1, POLR3H, PRKDC, RNF216, TICAM1*
	regulation of interleukin-12 production	*ARRB2, IRF1, LGALS9*
	positive regulation of type I interferon production	*IRF1, POLR3H, PRKDC, TICAM1*
	interferon-beta production	*IRF1, RNF216, TICAM1*
	regulation of interferon-beta production	*IRF1, RNF216, TICAM1*
	positive regulation of cytokine biosynthetic process	*CD4, IRF1, TICAM1*
	regulation of toll-like receptor signaling pathway	*ARRB2, IRF1, TICAM1*

Table 7. An excerpt of GO terms from the BiNGO overrepresentation analysis. The remaining data can be viewed in Appendix 6.

	GO Term	Associated Genes
AHC	platelet aggregation	*PLEK, FERMT3*
	response to stress	*RBM38, TSC22D4, CD74, BTG2, PRKDC, ITGB3, DEFA4, PLEK, TPM1, F13A1, DEFA3, LSP1, MTF1, HSF1, MKNK2, CD44, TNRC6A, FERMT3, BCL2L1*
	positive regulation of cytokine-mediated signaling pathway	*CD74, AGPAT1*
	wound healing	*ITGB3, PLEK, TPM1, F13A1, CD44, FERMT3*
	negative regulation of DNA damage response, signal transduction by p53 class mediator	*CD74, CD44*
	homotypic cell-cell adhesion	*PLEK, FERMT3*
	hemostasis	*ITGB3, PLEK, F13A1, FERMT3*
	T cell activation	*FKBP1A, CD74, PRKDC, IRF1*
	T cell differentiation	*CD74, PRKDC, IRF1*
	leukocyte differentiation	*CD74, PRKDC, IRF1, JUNB*
BMC	regulation of ERK1 and ERK2 cascade	*CD74, VEGFB, ARRB2, DUSP6, CD44*
	leukocyte activation	*FKBP1A, CD74, CD4, WBP1, PRKDC, IMPDH1, IRF1, TICAM1*
	T cell activation	*FKBP1A, CD74, CD4, WBP1, PRKDC, IRF1*
	immune system process	*CD74, WBP1, PRKDC, IL1R2, NCF4, TCF7, PLEK, TICAM1, RASGRP4, FKBP1A, CD4, IMPDH1, IRF1, POLR3H, JUNB*
	positive regulation of cytokine-mediated signaling pathway	*CD74, AGPAT1*
	T cell differentiation	*CD74, CD4, PRKDC, IRF1*
	negative regulation of DNA damage response, signal transduction by p53 class mediator	*CD74, CD44*
	immune system development	*CD74, CD4, PRKDC, IRF1, PLEK, JUNB, RASGRP4*
	regulation of response to stress	*CD74, PLEK, VEGFB, ARRB2, TICAM1, RTN4, CD44*
	lymphocyte differentiation	*CD74, CD4, PRKDC, IRF1*

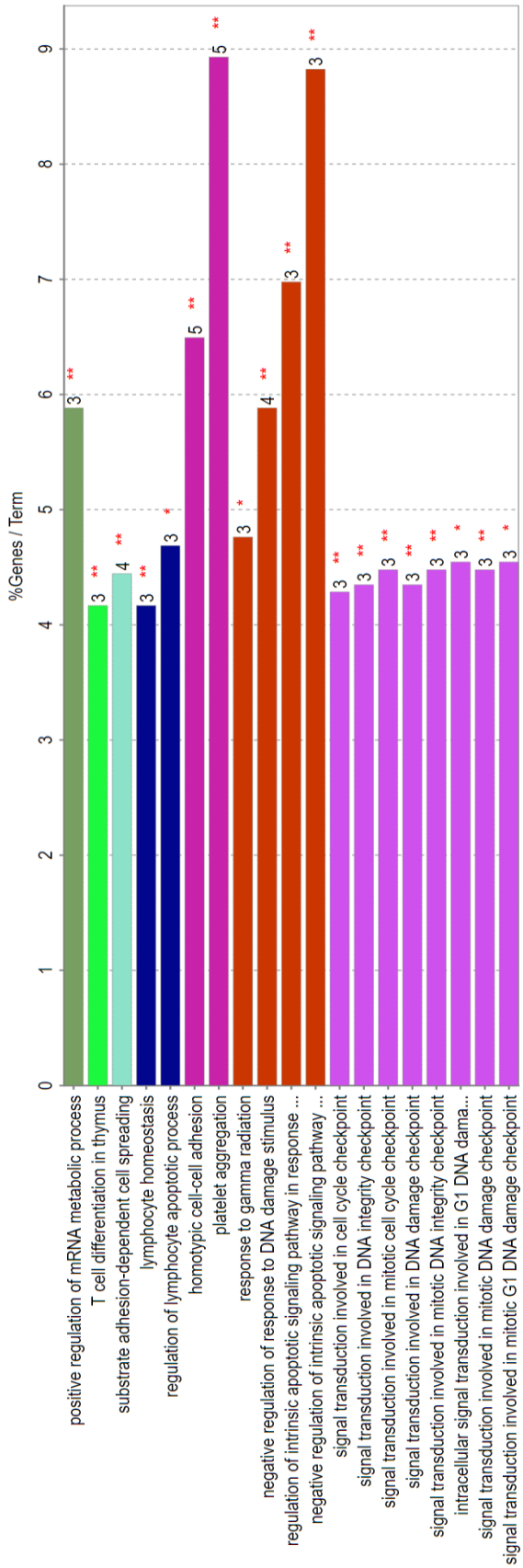


Figure 8. Results from ClueGO overrepresentation analysis for the AHC gene list. The AHC gene list is a merge of the gene lists based on KUL's analysis (Table 15) and NTNU's analysis (Table 14). The biological processes are divided into groups, each with a group color assigned to it. The processes in the same group are associated to each other based on kappa statistics. The bars represent the number of genes in the network associated with the GO/pathway terms specific for upregulated genes. The bar labels show the percentage of genes per term. The data and visualizations were produced using the ClueGO app in Cytoscape.

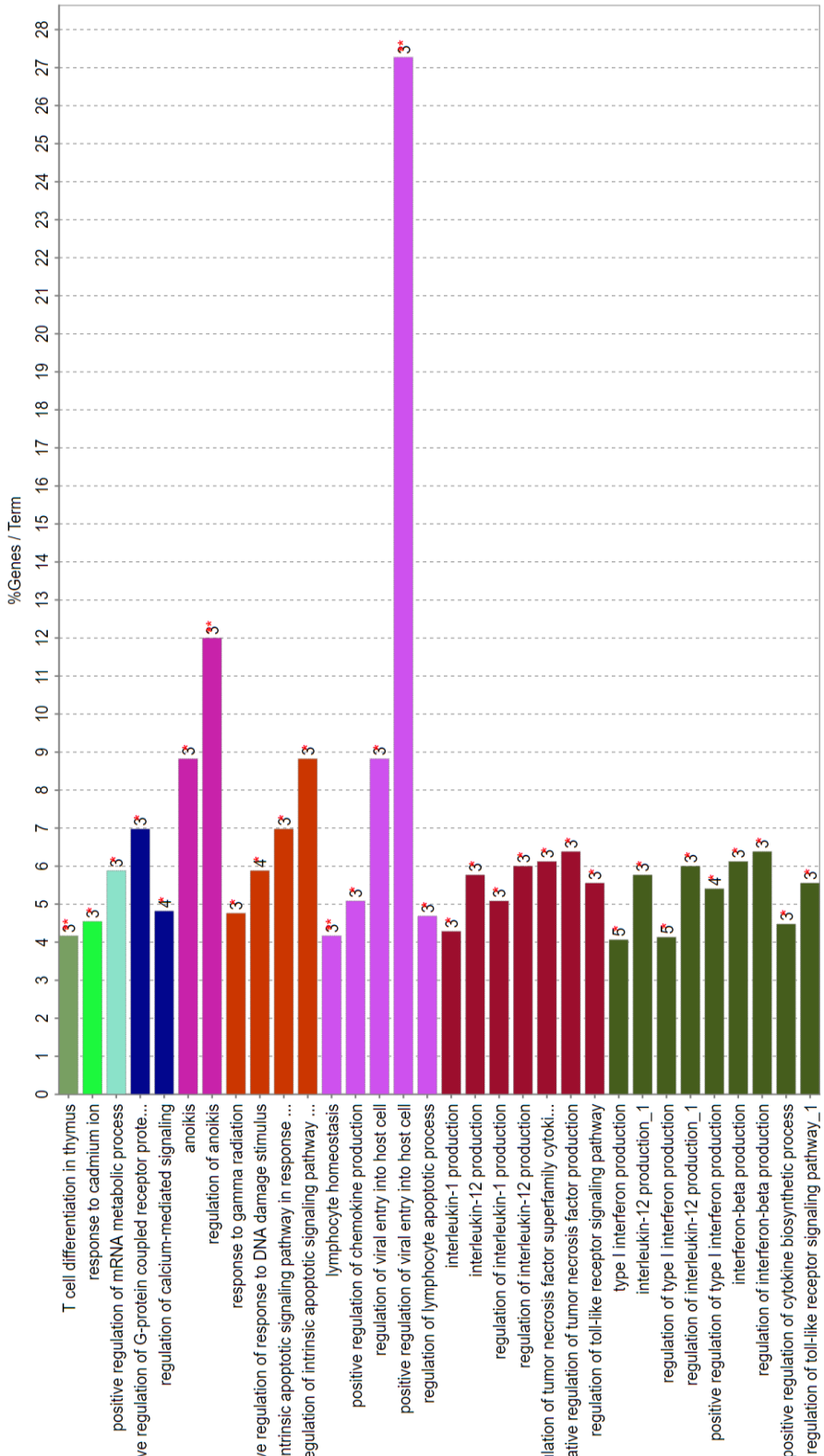


Figure 9. Results from ClueGO overrepresentation analysis for the BMC gene list. The BMC gene list is a merge of the gene lists based on KUL's analysis (Table 17Table 15) and NTNU's analysis (Table 16Table 14). The biological processes are divided into groups, each with a group color assigned to it. The processes in the same group are associated to each other based on kappa statistics. The bars represent the number of genes in the network associated with the GO/pathway terms specific for upregulated genes. The bar labels show the percentage of genes per term. The data and visualizations were produced using the ClueGO app in Cytoscape.

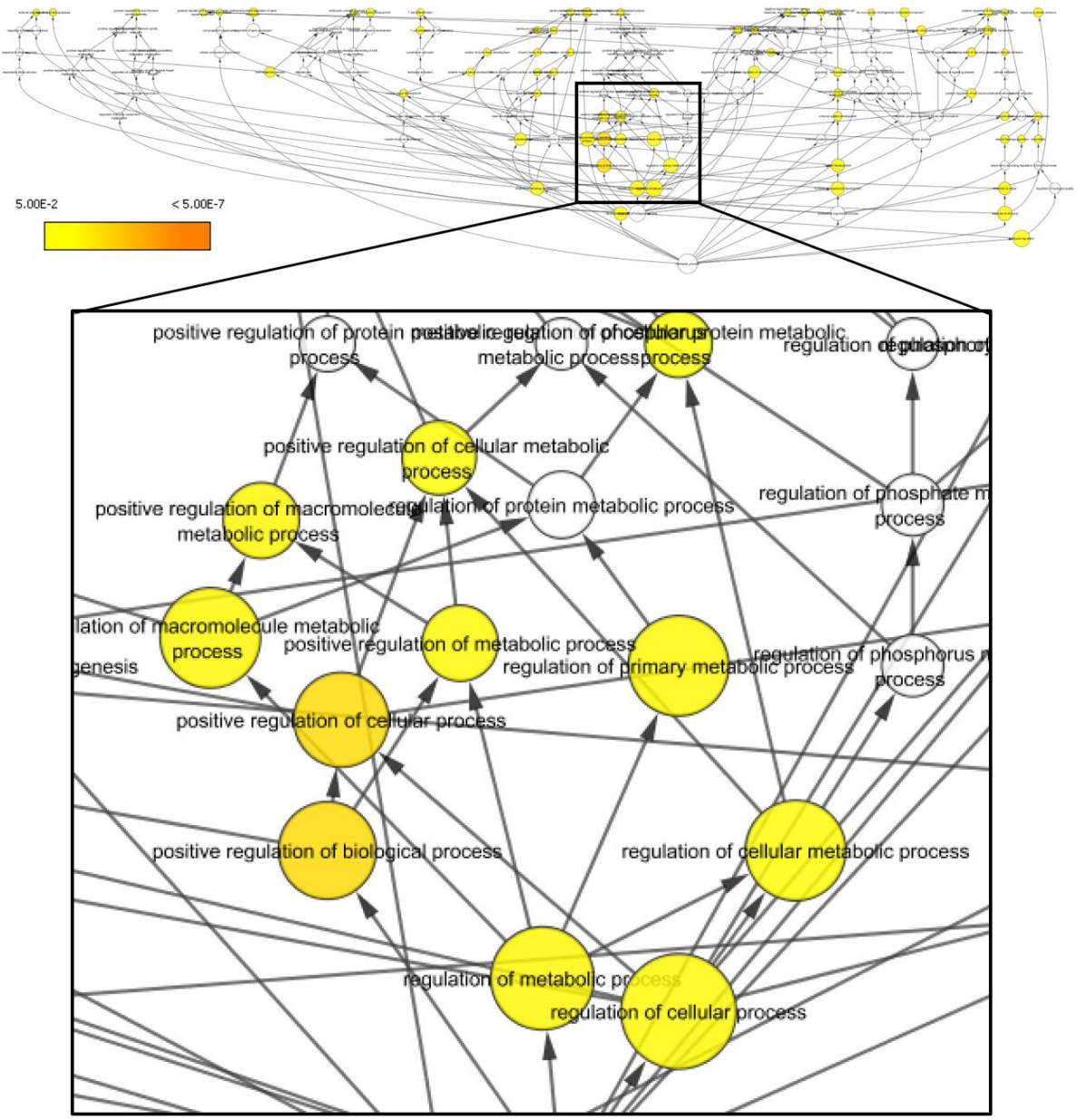


Figure 10. Visual representation of the results of the BiNGO analysis performed on the merged AHC gene list. The P values shown as a yellow-to-orange color gradient is based on a hypergeometric statistical test with Benjamini-Hochberg false discovery rate (FDR) correction. The data and visualizations were produced using the BiNGO app in Cytoscape. A blow-up of the area showing the most significant GO terms is shown.

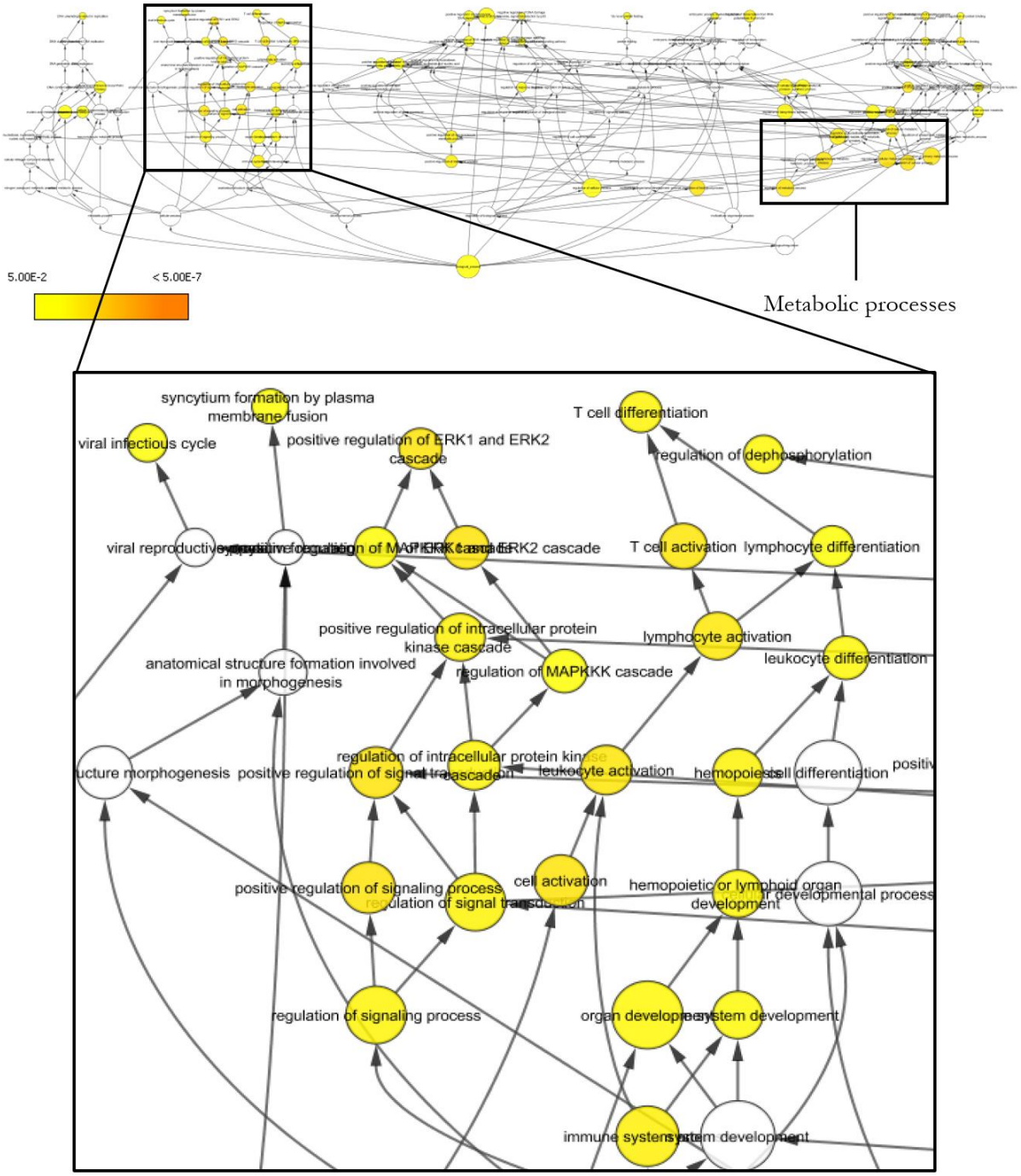


Figure 11. Visual representation of the results of the BiNGO analysis performed on the merged BMC gene list. The P values shown as a yellow-to-orange color gradient is based on a hypergeometric statistical test with Benjamini-Hochberg false discovery rate (FDR) correction. The data and visualizations were produced using the BiNGO app in Cytoscape.

3.3.2 Pathway-based analysis using Reactome

A pathway-based analysis was conducted, in which the gene sets were analyzed with respect to biological pathways. When analyzing the selected gene lists with the genes log2FC in the Reactome Pathway Database, few pathways (**Table 8**) were discovered to significantly change as a response to the diets. The results were, however, similar for both diets, regardless of statistical analysis. 'The Rho GTPase cycle' and 'Insulin-like Growth Factor-2 mRNA Binding Proteins (IGF2BPs/IMPs/VICKZs) bind RNA' are the top two pathways regardless of diet and statistical analysis.

The same analysis was performed on the full gene lists from the initial datasets to see if the results were different from that based on only the selected genes. The affected pathways for the full gene lists (**Table 9**) were indeed different from those for the selected ones. One pathway prominent compared to the others: 'Neutrophil degranulation', which has an evidently lower FDR compared to all other pathways, and which does appear in three out of four gene lists. The only exception is the BMC diet based on NTNU's statistical data, in which 'Neutrophil degranulation' does not appear.

3.4 Networks

To get an idea of whether there is a system level component suggesting coordinated function to the genes in the gene sets, a network was built for each of the diets: one for AHC (**Figure 12**) and one for BMC (**Figure 13**). The nodes are colored based on the genes' category after comparison of the cross-ranked gene lists (**Table 10**), and the edges are based on how GeneMANIA presented the interactions. Most of the genes that were in common between the diets (green nodes) are connected through interactions in the networks, at least for BMC. AHC did, on the other hand, not need as many 'filler' nodes to connect the genes in the gene list, and only 29 nodes are isolated from the network (**Table 13**). The BMC network has 57 isolated nodes, but most of them are from KUL's analysis only, and a few from NTNU's analysis only. 20 'filler' nodes have been introduced by GeneMANIA to connect the genes in the BMC gene list. The edges connecting the nodes are not equal nor directed.

Table 8. The pathways significantly affected by the diets according to the Reactome Pathway Database. The data used is the gene lists containing a selected genes in Appendix 3 for each diet and each statistical analysis, as well as the log2 FC for each respective gene.

Diet	Statistics	Pathways	FDR
AHC	NTNU	Rho GTPase cycle	1.00E-5
		Insulin-like Growth Factor-2 mRNA Binding Proteins (IGF2BPs/IMPs/VICKZs) bind RNA	2.09E-5
	KUL	Rho GTPase cycle	7.24E-6
		Insulin-like Growth Factor-2 mRNA Binding Proteins (IGF2BPs/IMPs/VICKZs) bind RNA	7.24E-6
		Interleukin-4 and 13 signaling	3.43E-1
		Platelet degranulation	3.43E-1
		Signaling by Rho GTPases	3.82E-1
		Alpha-defensins	3.82E-1
		Synthesis of 12-eicosatetraenoic acid derivatives	4.58E-1
BMC	NTNU	Rho GTPase cycle	1.31E-6
		Insulin-like Growth Factor-2 mRNA Binding Proteins (IGF2BPs/IMPs/VICKZs) bind RNA	7.25E-6
		Signaling by Rho GTPases	4.83E-2
	KUL	Rho GTPase cycle	5.48E-5
		Insulin-like Growth Factor-2 mRNA Binding Proteins (IGF2BPs/IMPs/VICKZs) bind RNA	5.86E-5

Table 9. The pathways significantly affected by the diets according to the Reactome Pathway Database. The data used is the full gene lists containing the all gene lists from the initial datasets for each diet and each statistical analysis, as well as the log2 FC for each respective gene.

Diet	Statistics	Pathways	FDR
AHC	NTNU	Neutrophil degranulation	9.27E-7
	KUL	Neutrophil degranulation	1.50E-8
		Antigen processing: Ubiquitination & Proteasome degradation	9.87E-1
		PD- 1 signaling	9.87E-1
		Abortive elongation of HIV- 1 transcript in the absence of Tat	9.87E-1
		Prostacyclin signaling through prostacyclin receptor	9.87E-1
		Translocation of ZAP-70 to immunological synapse	9.87E-1
		ERKs are inactivated	9.87E-1
		TCF7L2 mutant don't bind CTBP	9.87E-1
		HDACs deacetylate histones	9.87E-1
		Sema4D induced cell migration and growth-cone collapse	9.87E-1
		Insulin receptor recycling	9.87E-1
		Misspliced GSK3beta mutants stabilize beta-catenin	9.87E-1
		Tat-mediated HIC elongation arrest and recovery	9.87E-1
		Pausing and recovery of Tat-mediated HIV elongation	9.87E-1
		MAP2K and MAPK activation	9.87E-1
BMC	NTNU	Formation of the ternary complex, and subsequently, the 43S complex	2.71E-3
		Rho GTPase cycle	2.71E-3
		L13a-mediated translational silencing of Ceruloplasmin expression	2.71E-3
		Formation of a pool of free 40S subunits	8.64E-3
		GTP hydrolysis and joining of the 60S ribosomal subunit	9.09E-3
		Ribosomal scanning and start codon recognition	1.09E-2
		Peptide chain elongation	2.57E-2
		SRP-dependent cotranslational protein targeting to membrane	4.28E-2
		HSF1-dependent transactivation	4.28E-2
	KUL	Neutrophil degranulation	1.50E-8

Table 10. Color interpretation for the network presentations.

Color	Meaning
Dark green	Common to both gene lists of the particular diet, and do not appear in any of the gene lists for the other diet (**Table 22** for AHC, **Table 23** for BMC).
Light green	Common to both gene lists of the particular diet, but does also appear in at least one gene list for the other diet (see the 'Common for both lists' columns in **Table 18** and **Table 19**).
Dark blue	Appear in the particular diet based on NTNU' data only (see the 'Unique for NTNU's analysis' columns in **Table 18** and **Table 19**).
Light blue	Appear in the particular diet based on KUL's data only (see the 'Unique for KUL's analysis' columns in **Table 18** and **Table 19**).
Red	Selected from text-mining
Orange	Introduced by GeneMANIA

Table 11. Isolated nodes in the network presentations. The isolated nodes represent genes that did not have an apparent connection to any of the other genes, and thus did not get connected to the network. They were, however, included in the analysis.

AHC	BMC
CSDAP1, C6orf25, CHCHD10, DDX11L1, FAM100A, GPX1P1, GRINA, GPR162, INO80E, JUNB, LRG1, LY6E, LSP1, LOC100130751, LOC407835, MTF1, MKNK2, NPIPB12, NPIPB11, NPIPB15, PDKP1, PHF1, P2RY13, RASGRP4, SLA, SH3BP1, TSC22D4, TNRC6A, VNN2	*ABTB1, AGPAT1, C6orf136, CLIC3, CORO7, C7orf149, CNOT7, CHMP6, CD68, DEFA3, FKBP1A, FAM58A, IL1R2, INF2, JUNB, KIAA1267, LRG1, LOC339047, LSP1, LU6E, LOC440353, LOG100130751, LOC728888, MED16, MKNK2, MTF1, NPIPL2, NINJ1, NIPSNAP1, PLEK, PANX2, PNPLA2, RPL36AP49, RASGRP4, RPL13P12, RNF220, SFRS18, SLA, SH3BP1, TSPAN18, TGOLN2, TSC22D4, TOP3B, TP53I13, UCP2, VEGFB, VAMP2, WBP1, ZFP106, ZGPAT, ZNF746*

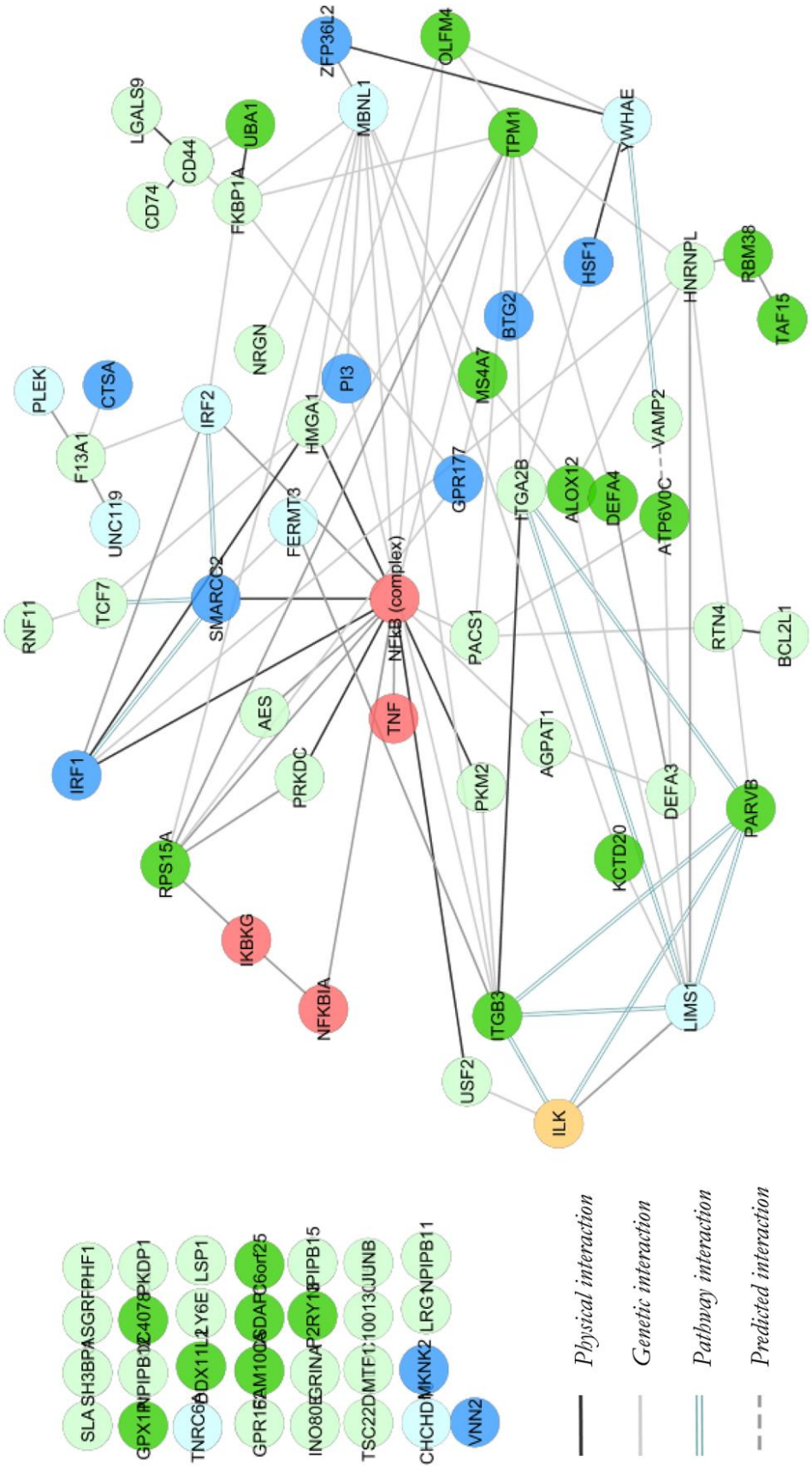


Figure 12. Final AHC network constructed in Cytoscape. The darker green nodes are common to both AHC gene lists, but do not appear in any BMC gene list. The lighter green nodes are common to both AHC gene lists, but do also appear in one or two BMC gene list(s). Both the darker and the lighter blue nodes are selected from AHC gene lists, whereas the darker nodes based on NTNU's statistical data, and the lighter blue on KUL's statistical data. The red node is selected from text-mining, whereas the orange ones are introduced by GeneMANIA. The edges are not directed. The nature of the interactions behind the edges can be viewed down to the left. Isolated nodes are located up to the left.

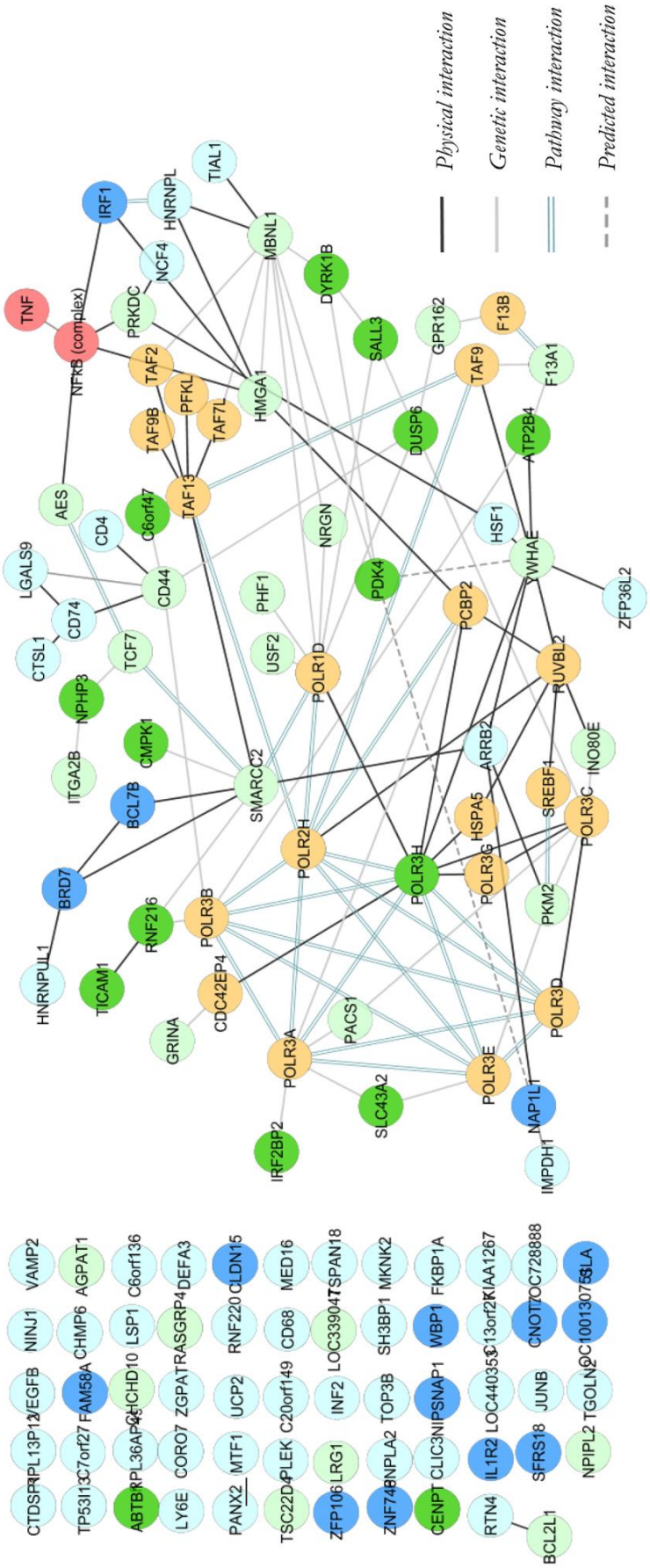


Figure 13. Final BMC network constructed in Cytoscape. The darker green nodes are common to both BMC gene lists, but do not appear in any AHC gene list. The lighter green nodes are common to both BMC gene lists, but do also appear in one or two AHC gene list(s). Both the darker and the lighter blue nodes are selected from BMC gene lists, whereas the darker nodes based on NTNU's statistical data, and the lighter blue on KUL's statistical data. The red node is selected from text-mining, whereas the orange ones are introduced by GeneMANIA. The edges are not directed. The nature of the interactions behind the edges can be viewed to the right. Isolated nodes are located to the left.

3.5 Network-based analysis

3.5.1 Graph-based analysis

The network was analyzed using NetworkAnalyzer, yielding different values and graphs providing information regarding the network's properties. A summary of the simple parameters provided can be viewed in **Table 13**. Among the complex parameters are betweenness centrality, which appears to be similar for AHC (**Figure 14 A**) and BMC (**Figure 14 B**). The most prominent difference is the node with 15 neighbors in AHC, with a relatively high betweenness centrality. This node is identified as the NF-κB complex. In BMC, the NF-κB complex have 5 neighbors, in which 2 of them are genes introduced to the network by GeneMANIA. NF-κB is also more connected in AHC compared to BMC. Several genes have multiple neighbors, and the genes with 8 or more neighbors are presented in (**Table 12**). The node degree distributions are similar for the two diets (**Figure 15**), which both show a decrease at the in number of nodes as the node degree increases.

Table 12. Nodes with eight or more neighbors. Eight neighbors was chosen as a cut-off because it is ~half of the number of neighbors for the most connected node, which have fifteen. Node names, associated UniProt IDs and number of neighbors are given in the table.

AHC			**BMC**		
NF-kB	-	15	*POLR3H*	(UniProt ID: Q9Y535)	10
MBNL1	(UniProt ID: Q9NR56)	10	*MBLN1*	(UniProt ID: Q9NR56)	8
TPM1	(UniProt ID: P09493)	9	*SMARCC2*	(UniProt ID: Q8TAQ2)	8
LIMS1	(UniProt ID: P48059)	8	*YWHAE*	(UniProt ID: P62258)	8
ITGB3	(UniProt ID: P05106)	8			

Table 13. Summary of the parameters provided by the graph-based analysis performed for both AHC and BMC in Cytoscape using the NetworkAnalyzer plug-in. The isolated nodes were included in all analyses.

Parameter	**AHC**	**BMC**
Clustering coefficient	0.094	0.092
Connected components	30	59
Network diameter	6	8
Network radius	4	1
Network centralization	0.159	0.080
Shortest paths	2863 (42%)	4832 (29%)
Characteristic path length	3.212	3.596
Average number of neighbors	2.265	1.860
Number of nodes	83	129
Network density	0.028	0.015
Network heterogeneity	1.213	1.387
Isolated nodes	29	57

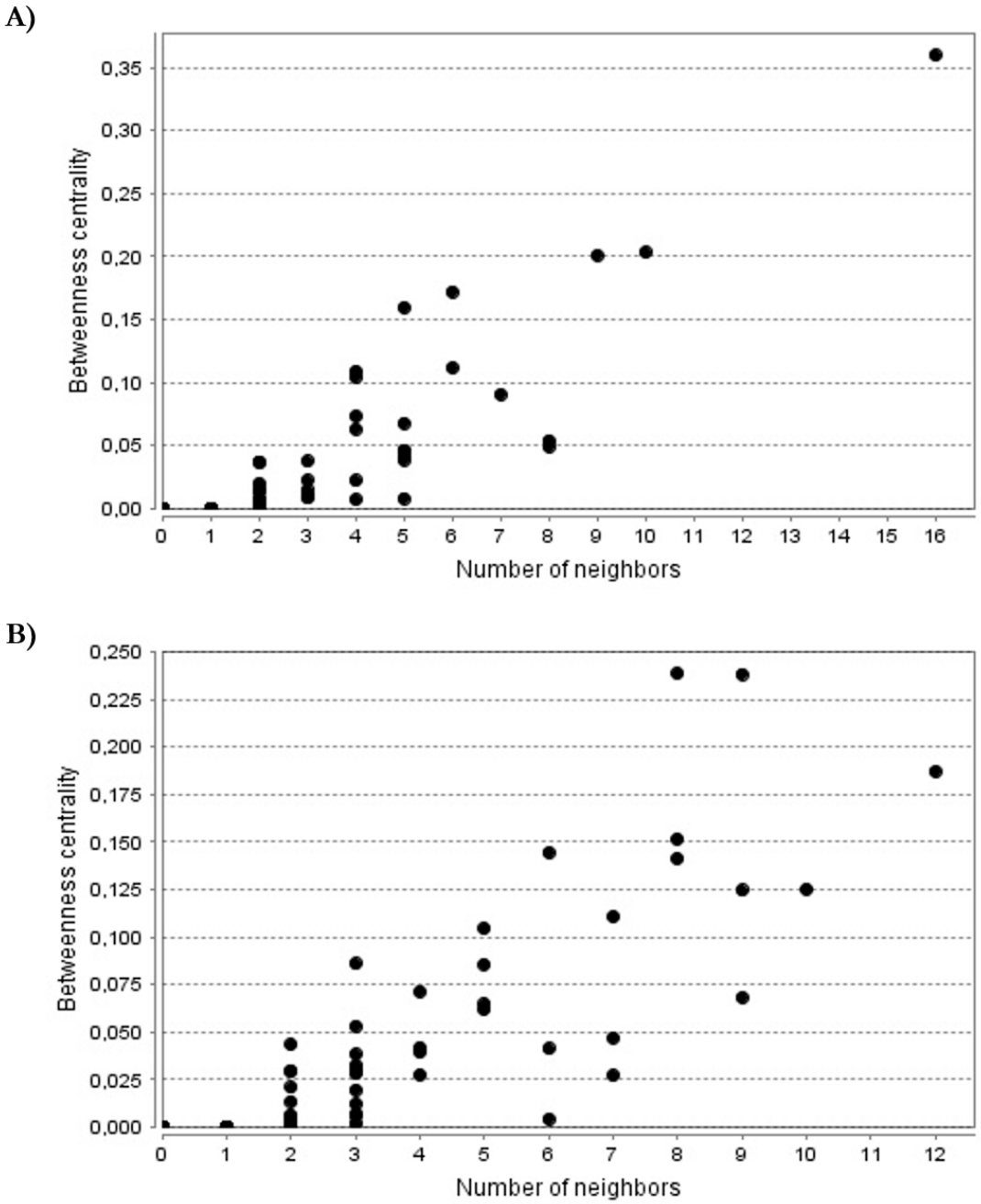
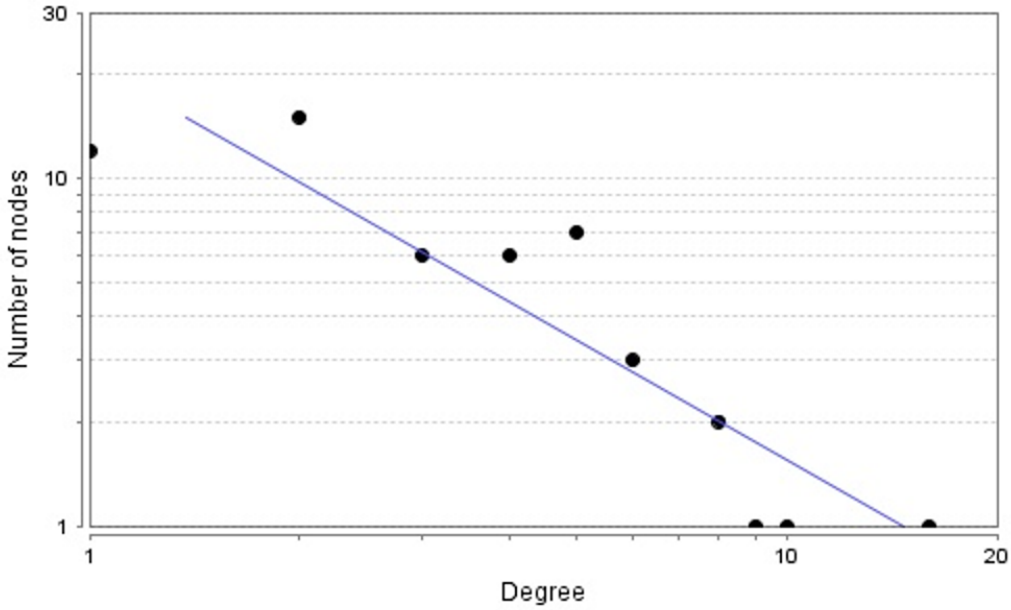


Figure 14. Graphical presentation of the betweenness centrality (In-Betweenness) for the nodes in the finished **A)** AHC and **B)** BMC networks. Each dot in the graph represents a node in the network. The horizontal axes show the number of neighbors, and thus give information regarding connectivity. The vertical axes show the betweenness centrality, which refers to the number of times a node acts as a bridge along the shortest path between two other nodes. The graphical results are produced using the NetworkAnalyzer plug-in in Cytoscape.

A)



B)

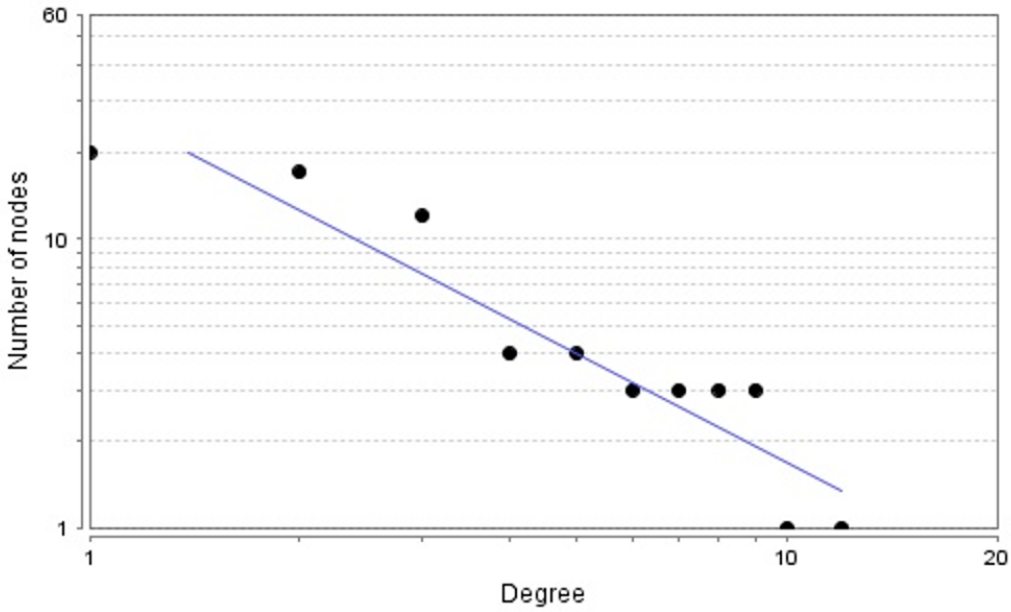


Figure 15. Graphical presentation of the node degree distribution for the nodes in the finished **A)** AHC and **B)** BMC networks. Each dot in the graph represents a node in the network. The horizontal axes show the node degree. The vertical axes show the number of nodes in the network possessing the given node degrees. The fitted line in blue is an $y = ax^b$ power law.

3.5.2 Superimposing of gene expression data from the microarray

To visualize the gene expression data in Cytoscape, the genes' respective log2 FC values were superimposed onto the two networks, together with the P values from both statistical analyses. The result was two visually different network presentations for each diet, four networks in total. The log2 FC values for AHC and BMC do not differ remarkably from each other. There are no genes that are downregulated in one diet and upregulated in the other, or the other way around. The only difference is *how* up- or downregulated the genes are. The P values from the KUL analysis and the NTNU analysis appear to be similar for AHC (**Figure 16** and **Figure 17**), whereas they differ distinctly for BMC (**Figure 18** and **Figure 19**). The difference in P value is most notable in the isolated genes. However, when reviewing the log2 FC calculations, the same trend is observed between the diets for the unique genes as for the common genes. The difference is not as extensive for genes connected to the network. The genes that are unique to each diet is not comparable in these results, but they trend the same.

The NFkB complex is a yellow node due to it being a protein complex. However, when looking at the calculated log2 FC values, the *NFKB1* gene has a log2 FC = -0.08187448 in AHC and log2 FC = -0.158690111 in BMC. No data were found for *RELA*. The NF-κB complex can thus be considered downregulated in both diets, but to a higher extent in BMC compared to AHC. No data were found for *IKBKG* either. However, a gene encoding another subunit of the IKK complex, *IKBKB* (UniProt ID: O14920), was found in the data sets. *IKBKB* has a log2 FC = -0.138284136 in AHC, and a log2 FC = -0.10641215 in BMC. The transcribed protein participates in phosphorylation of NF-kappaB inhibitors, such as IKK, and IKK-related kinases, e.g. *TBK1* (UniProt ID: Q9UHD2). *TBK1* plays an essential role in regulation of inflammatory responses to foreign agents, and have a log2 FC = -0.120653312 in AHC and -0.211656389 in BMC. The gene encoding the cytokine TNF, which is known to be involved in inflammatory responses by inducing NF-κB activation, is also downregulated in both diets (log2 FC = -0.164197696 in AHC, -0.37691091 in BMC). Even though not included in the network, the different TNF receptors in the data sets (*TNFRSF14*, *TNFRSF17*, *TNFRSF19*, *TNFRSF4*, and *TNFRSF9*) are downregulated in both diets as well.

The most downregulated gene in both AHC and BMC, but to a greater extent in BMC, is *IRF1* (UniProt ID: P10914). *IRF1* encodes the transcriptional regulator 'Interferon regulatory factor 1', which function as an activator for several genes involved in anti-viral response, anti-bacterial response, anti-proliferative response, apoptosis, immune response, DNA damage responses and DNA repair.

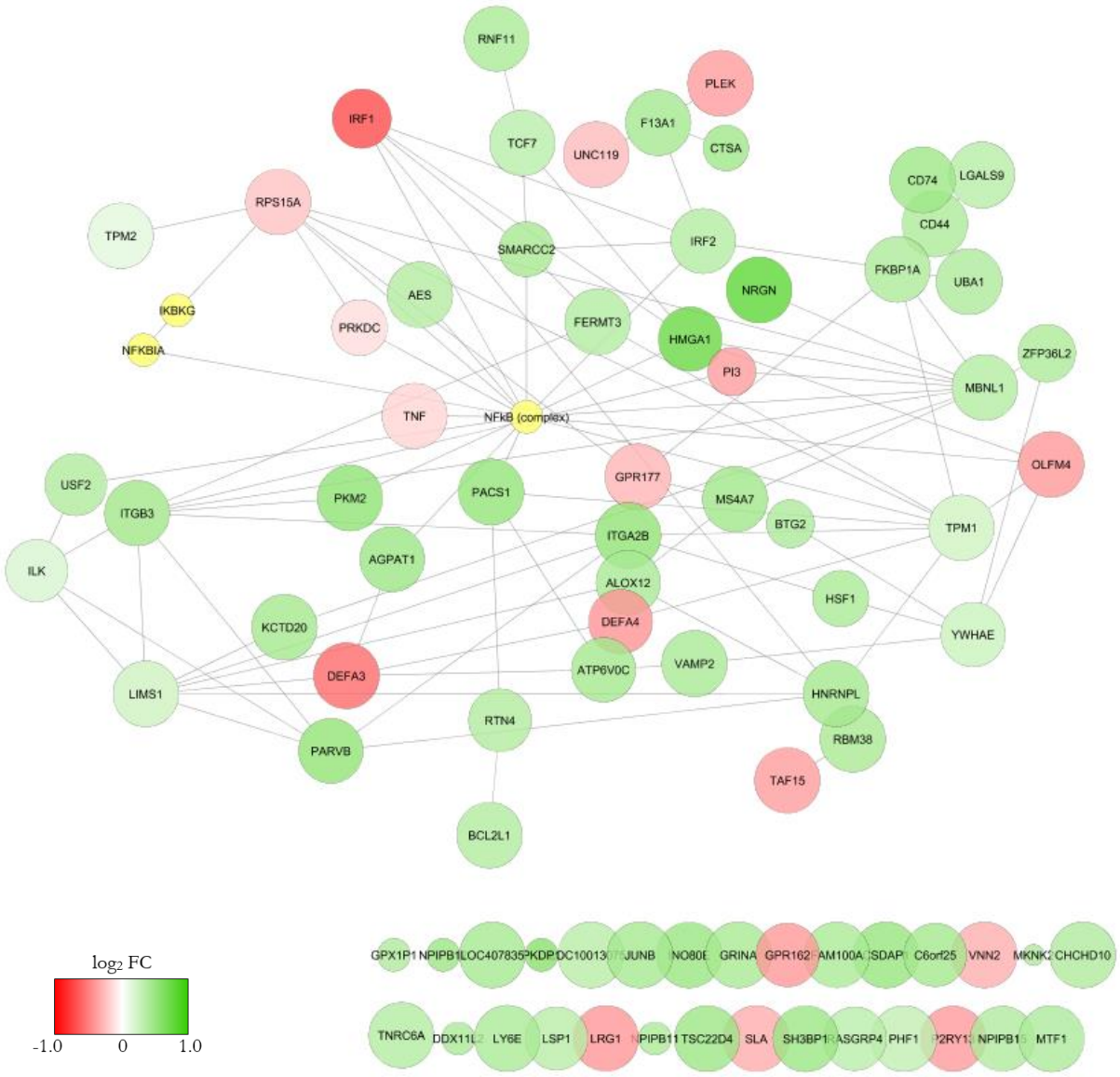


Figure 16. The finished AHC network and the isolated nodes with overlay of log2 FC and P value data. The log2 FC values used are the calculated mean values for each gene and are visualized according to the color gradient down to the left. The P values used are produced by KUL. The bigger node, the lower P value. The yellow nodes were not identified in the initial data set and does hence not have any values dedicated to them.

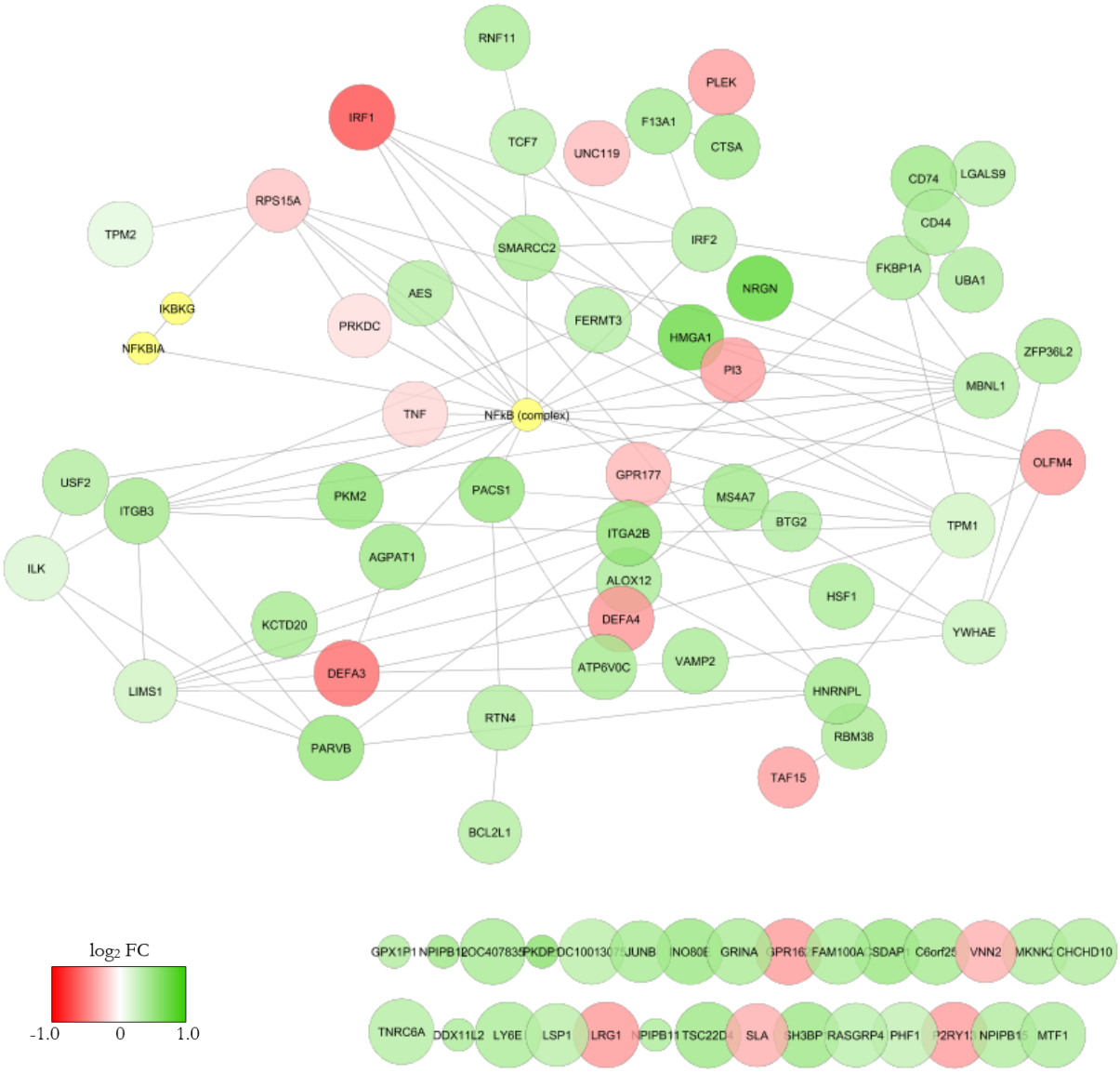


Figure 17. The finished AHC network and the isolated nodes with overlay of log2 FC and P value data. The log2 FC values used are the calculated mean values for each gene and are visualized according to the color gradient down to the left. The P values used are produced by NTNU. The bigger node, the lower P value. The yellow nodes were not identified in the initial data set and does hence not have any values dedicated to them.

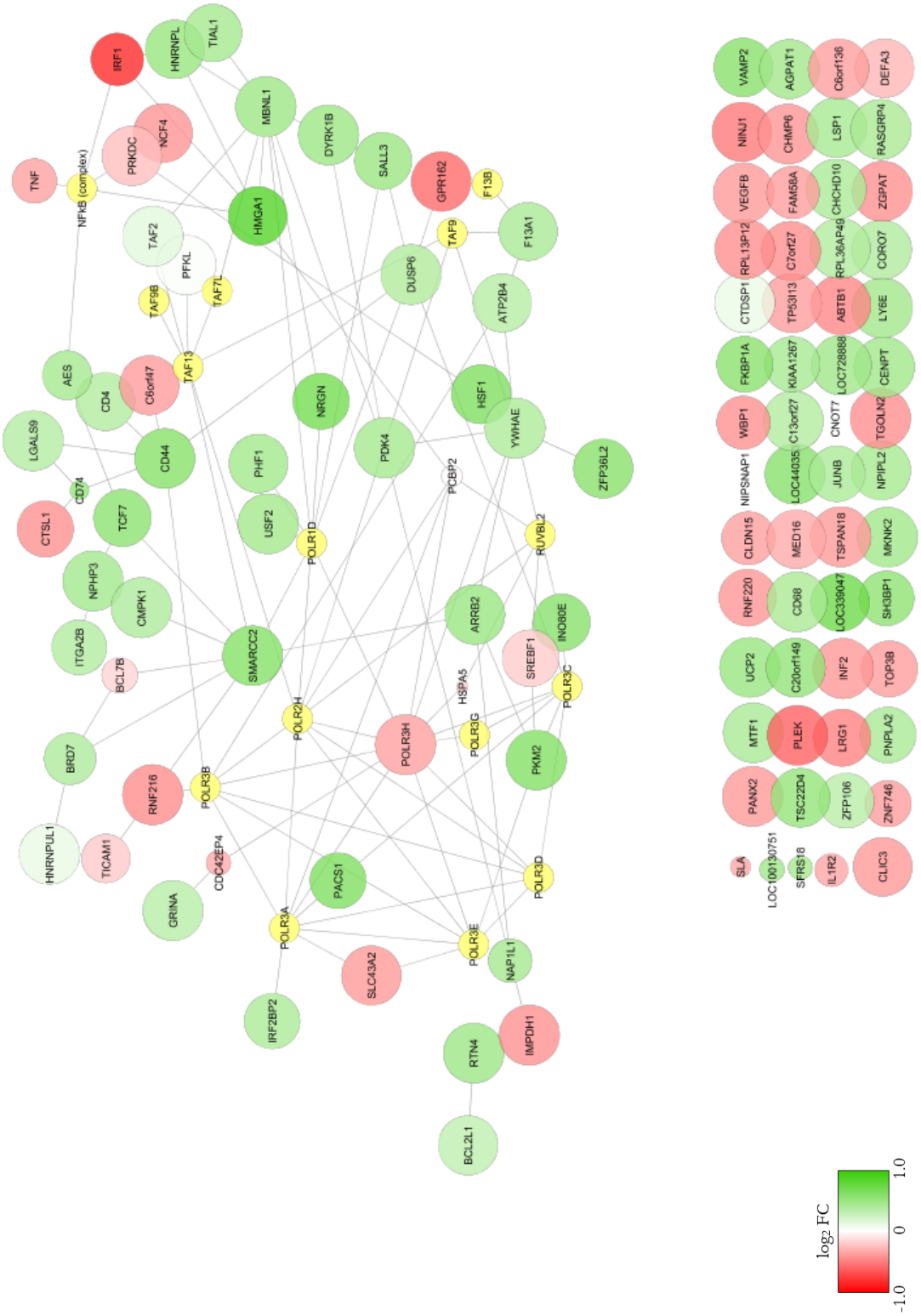


Figure 18. The finished BMC network and the isolated nodes with overlay of log2 FC and P value data. The log2 FC values used are the calculated mean values for each gene and are visualized according to the color gradient down to the left. The P values used are produced by KUL. The bigger node, the lower P value. The yellow nodes were not identified in the initial data set and does hence not have any values dedicated to them.

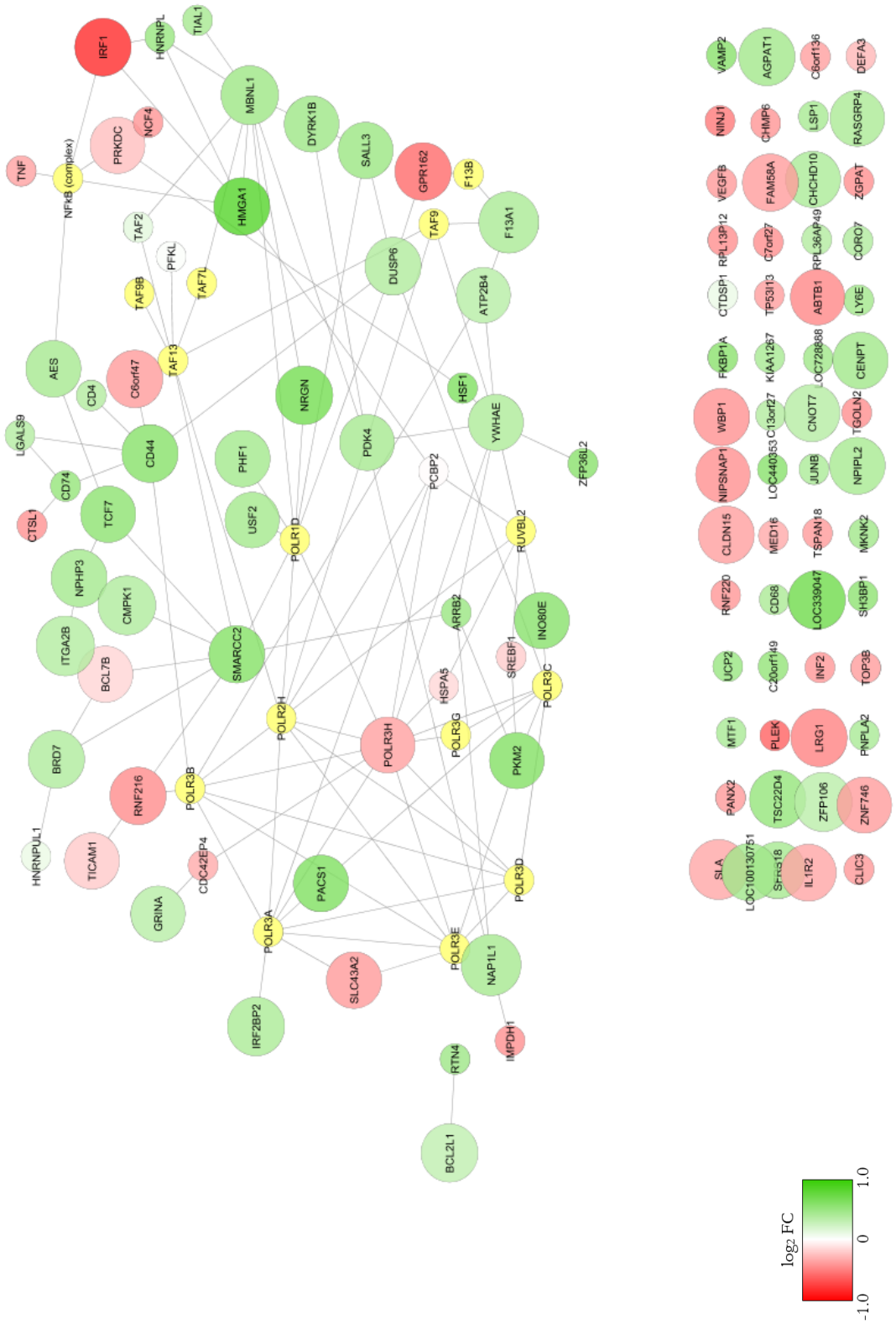


Figure 19. The finished BMC network and the isolated nodes with overlay of log2 FC and P value data. The log2 FC values used are the calculated mean values for each gene and are visualized according to the color gradient down to the left. The P values used are produced by NTNU. The bigger node, the lower P value. The yellow nodes were not identified in the initial data set and does hence not have any values dedicated to them.

4 Discussion

In the diet intervention study described by Arbo et al. (2010), results indicate that NF-κB was activated in response to AHC and inhibited in response to BMC, and that this suggests NF-κB having a key role in regulating the early diet specific changes in the current study. The pathways mentioned as most notably exhibiting gene expression changes due to the diets, are connected to processes such as apoptosis, proliferation/cell cycle regulation, and stress/immunity. The genes highlighted in the study is said to be part of these processes. However, it is also mentioned that very few genes showed differential regulation by the two diets: the overlap between AHC and BMC is described as extensive, and the majority of the genes changed in the same direction in both diets. In this thesis, both discrepancies and consensus to the initial conclusions have occurred. Throughout this discussion, the discoveries done will be assessed and compared to the discoveries Arbo et al. (2010) where appropriate.

The first observation that comes to mind is the difference in gene lists: just six of the genes (*IRF1*, *BCL2L1*, *BTG2*, *NAP1L1*, *F13A1*, and *CD44*) mentioned in Arbo et al. (2010) appear in either gene list produced in this thesis. The results from Arbo et al. (2010) are based on NTNU's analysis only, whereas the results in this thesis are a combination of NTNU's and KUL's analyses – a factor that possibly could contribute to the results. When addressing the difference between NTNU and KUL, the CAT analysis showed that the gene ranks differed greatly, at least for AHC, even though the statistical ('equalStat') CAT analysis could indicate some similarities. The substantial difference in 'equalRank' and 'equalStat' results is believed to be due to a bigger possibility of having the same P value compared to having the exact same rank in the gene list. The CAT analysis also show that the different diets share statistical data, which corresponds to the statement of Arbo et al. (2010), saying that there is indeed an extensive overlap between AHC and BMC. Regardless, the different FDR corrections used in the two analyses seem to have produced different results that affect the final significant genes. The statistical differences could be a contributing factor to the differences in selected genes.

The gene lists in this thesis are based on both P value and log2 FC. It is, however, important to keep in mind that a P=0.05 not necessarily should be an absolute cut-off, even though this is done

in this thesis. It could be interesting to experiment with different cut-offs for P value. The elimination of genes with a log2 FC<0.38 was a randomly chosen gene expression change, calculated to fit a 40% up- or downregulation, still a minor change. All values in the diet intervention study had a log2 FC<1, meaning no genes met the criteria of a two-fold up or downregulation. However, even minor changes in gene expression can contribute to major changes in cellular response, and thus result in bigger changes in an organism. In addition, it is likely that not all leukocytes in the analyzed blood samples exhibit a response to inflammation. If only a smaller fraction of the leukocytes displays relatively substantial changes, the response could possibly be 'diluted' due to numerous non-affected cells. Regardless, not many genes met the requirements, as can be viewed in the Volcano plots.

Most of the genes were eliminated due to log2 FC close to zero. In the log2 FC plot, the mean log2 FC for each gene was yet to be calculated. Using mean values would most likely yield even more results close to zero. A reason behind the low log2 FC values could be that AHC and BMC did not affect the participants considerably. However, it could be considered that the relatively short timeline of the diet intervention study may contribute. Even though the cells in our bodies respond quickly to environmental changes, a 7-days diet might not be considered extensive enough to cover potential long-term effects of a specific diet. A longer intervention study, as well more participants in the study, among them normal-weight people, could be considered. Due to the already existing overweight of the participants in the diet intervention study, it could be reasonable to believe that inflammatory processes were already in action as the study began, thus resulting in small changes. When reviewing the initial gene expression data, e.g. *NFKB1* was found to be among the top 200 most expressed genes on day 0 for AHC.

Regardless, the genes exceeding the limits using non-average log2 FC values are more numerous in the BMCs. This could indicate that the system as a whole was more affected by BMC than it was to AHC, which suggests a possibility of the individuals having a diet relatively high in carbohydrates to begin with. On average, Norwegians have a diet consisting of 47% carbohydrates (Helsedirektoratet, 2016), which is relatively high compared to the BMC diet. It can thus be assumed that the reduction in dietary carbohydrates could results in a bigger transcriptome change than continuing a higher-carb diet. Gene expression was nevertheless changed in both diets, which might indicate that the participants in the diet intervention study reacted to being on a diet, regardless of which one. In addition to log2 FC, the Volcano plots provided information regarding P values. An interesting observation with respect to P value is the obvious cut-off in both NTNU analyses, in which the data seem to have been modified by removing almost every gene with a P>0.05, which could affect the basis of comparison for the CAT analysis performed prior to the

Volcano plots. The criteria behind this removal is not known. Regardless of the cut-off, only genes with P<0.05 were included in the final gene lists for further analysis.

When addressing the genes that met the log2 FC>0.68 criteria, there number of genes are similar for AHC and BMC, and the genes are partially the same as well. However, in the upregulated category, *HMGA1* (UniProt ID: P17096) is more prominent in AHC. *HMGA1* is involved in regulation of mRNA transcription and processing. The kinase *PKM2* (UniProt ID: P14618), and the pseudogene *PKD1P1* (no UniProt ID) are more prominent in BMC. *DEFA3* (UniProt ID: P59666), which encodes a neutrophil defensin, is downregulated in AHC. *DEFA3* is indirectly connected to NF-κB in the final AHC network, whereas it is isolated in the final BMC network.

The genes in the final gene lists were built into two networks, one for each diet. Here, both statistical analyses were included in the same networks. The gene lists contain few genes in common with Arbo et al. (2010), even though the addition of *RELA*, *NFKB1* and *TNF* due to their proinflammatory properties made them easier to compare to Arbo et al.'s (2010) results. In Arbo et al. (2010), an upregulation of *RELA* for AHC, and a downregulation for *NFKB1* in BMC is described. There is also described a downregulation of *TNF* in AHC, but no change in *TNF* expression in BMC. In this thesis, *TNF* is downregulated in both diets, even though it is slightly less downregulated in AHC. NF-κB is described by Arbo et al. (2010) as downregulated in BMC, probably due to the downregulation in *NFKB1*, which supposedly lay the foundation of the conclusion saying that BMC alleviates proinflammatory symptoms. What is not mentioned is that *NFKB1* is downregulated also in AHC. No data were found for *RELA*, which made double-checking of upregulation in AHC difficult. In this thesis, the difference in NF-κB expression is thus not perceived as remarkable, even though it is indeed downregulated in BMC. *RELA* and *NFKB1* does not necessarily need to be upregulated for NF-κB activity to increase, it is possible that genes encoding IKKs are upregulated instead, and thus activating already existing NF-κB. However, only one such gene was identified during the analysis in this thesis: *IKBKB*, which was downregulated in both diets, and even slightly more downregulated in AHC compared to BMC. Consequently, the alleviating effects of BMC on proinflammatory responses are not striking in this thesis, even though the possibility is not disregarded.

Nevertheless, NF-κB does stand out in AHC in the graph-based analysis conducted on the network. NF-κB are more connected in AHC compared to BMC, and the betweenness centrality of NF-κB in AHC, meaning the influence it has on the other interactions in the network, is also relatively high compared to every other node in either of the networks. NF-κB can thus be said to inhibit hub properties, meaning that the genes in the AHC gene sets are more connected to NF-

κB than the genes in the BMC gene sets. This might indicate a correlation between the AHC gene sets and proinflammatory response. The node degree distribution in both AHC and BMC leads toward a scale-free network, not a random one, which makes it believable to think that both networks are representable for a biological system containing hubs (Albert, 2005). Several genes in the network did actually inhibit hub properties. In AHC, these genes were mainly connected to integrins and actin-binding, while the genes in BMC mainly were connected to transcriptional regulation. *POLR3H* (UniProt ID: Q9Y535), the node with the most neighbors in BMC, is an RNA polymerase III subunit involved in recognition of bacteria and DNA viruses, could be considered as involved in inflammatory response. So cans the integrin-related genes in AHC, which can contribute to leukocytes' adhesion to the blood vein walls during inflammation (Gahmberg et al., 1998). It is, however, important to remember that the networks mainly were annotated using GeneMANIA, and that the text-mining approach did not contribute much to ensure the connections between the results.

To get an idea of the connections without analyzing each gene, the processes they are involved in were analyzed. Arbo et al. (2010) did the same, and highlighted apoptosis, proliferation/cell cycle regulation, and stress/immunity as prominent processes. Using ClueGO, BiNGO and Reactome, to some extent related results were observed in this thesis. The most prominent processes associated with the AHC gene list are connected to cell-cell adhesions, cell cycle control, apoptosis, and response to DNA damage. For BMC, processes involving lymphocytes, interleukins, the TNF superfamily, interferons, and TLR signaling are prominent, but the genes connected to them are mainly downregulated. Common for both diets are 'The Rho GTPase cycle', 'Insulin-like Growth Factor-2 mRNA Binding Proteins (IGF2BPs/IMPs/VICKZs) bind RNA', and 'Neutrophil degranulation'. Rho GTPases act as molecular switches in response to extracellular signals. By acting together with the actin cytoskeleton, it can induce change in cell morphology, chemotaxis and cell cycle progression (Hall, 1998). The expression of IGF2BP family members has been implicated in various cancers (Bell et al., 2013). Four of the genes that are associated with control IGF2BPs (*CTNNB1*, *MYC*, *TCF4*, *NFKB1*) are mentioned in Arbo et al. (2010). Neutrophils are critical inflammatory cells that cause tissue damage in a range of diseases and disorders (Lacy, 2006). They mature as a response to the appropriate cytokines and release a variety of substances through degranulation, including antimicrobial proteins and enzymes, reactive oxygen espies and cytokines, and in this way, kill extracellular bacteria and recruit additional leukocytes to the region of infection/inflammation. An interesting observation here, is that Rho GTPases are involved in signaling pathways which can lead to Ca^{2+}-dependent neutrophil degranulation. These processes can thus be

said to be connected to apoptosis, proliferation/cell cycle regulation, and stress/immunity, which makes Arbo et al. (2010) and this thesis concur to some extent.

The approaches for interpretation of data in this thesis and in Arbo et al. (2010) differ, which further might have contributed to the differences in results. However, based on the differences identified between the NTNU and KUL gene lists, the new analysis conducted can be said to affect the results. After taking KUL's data into consideration and interpreting the results in a new manner, the connection to proinflammatory response is weakened compared to presented results in Arbo et al. (2010).

5 Conclusion

The genes that were affected in response to the diets, and the processes they influence, can be related to proinflammatory processes. However, the connection is not striking. There have been induced some changes on a transcriptional level in the participants of the diet intervention study, but the changes are barely perceived as considerable. Every gene that has been studied have changed in the same manner in both diets. After taking KUL's data into consideration and interpreting the results in a new manner, the connection to proinflammatory response is weakened compared to what was presented in Arbo et al. (2010). The tendencies are, however, existing, making this an interesting subject for further studies. Dietary diseases are still a rising problem, and addressing them properly could be a crucial step for avoiding them in the future. Further research and a more extensive diet intervention study could possibly lead to new knowledge of the subject.

Bibliography

Albert, R. (2005). Scale-free networks in cell biology. *Journal of cell science, 118*(21), 4947-4957.

Arbo, I., Brattbakk, H.-R., Langaas, M., Kuiper, M., Kulseng, B., Lindberg, F., & Johansen, B. (2010). *A balanced macronutrient diet induces changes in a host of pro-inflammatory biomarkers, rendering a more healthy phenotype; a randomized cross-over trial.* NTNU, Trondheim.

Barabasi, A.-L., & Oltvai, Z. N. (2004). Network biology: understanding the cell's functional organization. *Nature Reviews Genetics, 5*(2), 101-113.

Barnes , P. J., & Karin , M. (1997). Nuclear Factor-κB — A Pivotal Transcription Factor in Chronic Inflammatory Diseases. *New England Journal of Medicine, 336*(15), 1066-1071. doi:10.1056/nejm199704103361506

Bell, J. L., Wächter, K., Mühleck, B., Pazaitis, N., Köhn, M., Lederer, M., & Hüttelmaier, S. (2013). Insulin-like growth factor 2 mRNA-binding proteins (IGF2BPs): post-transcriptional drivers of cancer progression? *Cellular and Molecular Life Sciences, 70*(15), 2657-2675. doi:10.1007/s00018-012-1186-z

Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., . . . Galon, J. (2009). ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics, 25*(8), 1091-1093.

Brasch, M. A., Hartley, J. L., & Vidal, M. (2004). ORFeome cloning and systems biology: standardized mass production of the parts from the parts-list. *Genome research, 14*(10b), 2001-2009.

Chen, F., Castranova, V., Shi, X., & Demers, L. M. (1999). New insights into the role of nuclear factor-κB, a ubiquitous transcription factor in the initiation of diseases. *Clinical chemistry, 45*(1), 7-17.

Cline, M. S., Smoot, M., Cerami, E., Kuchinsky, A., Landys, N., Workman, C., . . . Gross, B. (2007). Integration of biological networks and gene expression data using Cytoscape. *Nature protocols, 2*(10), 2366-2382.

Cui, X., & Churchill, G. A. (2003). Statistical tests for differential expression in cDNA microarray experiments. *Genome biology, 4*(4), 210. doi:10.1186/gb-2003-4-4-210

Dong, J., & Horvath, S. (2007). Understanding network concepts in modules. *BMC systems biology, 1*(1), 24.

Frisch, M., Klocke, B., Haltmeier, M., & Frech, K. (2009). LitInspector: literature and signal transduction pathway mining in PubMed abstracts. *Nucleic acids research, 37*(suppl 2), W135-W140.

Gahmberg, C. G., Valmu, L., Fagerholm, S., Kotovuori, P., Ihanus, E., Tian, L., & Pessa-Morikawa, T. (1998). Leukocyte integrins and inflammation. *Cellular and Molecular Life Sciences CMLS, 54*(6), 549-555. doi:10.1007/s000180050183

Ge, H., Walhout, A. J., & Vidal, M. (2003). Integrating 'omic'information: a bridge between genomics and systems biology. *TRENDS in Genetics, 19*(10), 551-560.

Gupta, A., & Marchionni, L. (2012). Computing and plotting agreement among ranked vectors of features with matchBox. In J. H. U. S. o. M. The Sidney Kimmel Comprehensive Cancer Center (Ed.).

Hall, A. (1998). Rho GTPases and the Actin Cytoskeleton. *Science, 279*(5350), 509.

Helsedirektoratet. (2016). *Utviklingen i norsk kosthold 2016*. Retrieved from https://helsedirektoratet.no/Lists/Publikasjoner/Attachments/1257/Utviklingen-i-norsk-kosthold-2016-IS-2558.pdf

Hoffmann, R., & Valencia, A. (2005). Implementing the iHOP concept for navigation of biomedical literature. *Bioinformatics, 21*(suppl 2), ii252-ii258.

Irizarry, R. A., Warren, D., Spencer, F., Kim, I. F., Biswal, S., Frank, B. C., . . . Yu, W. (2005). Multiple-laboratory comparison of microarray platforms. *Nat Meth, 2*(5), 345-350. doi:http://www.nature.com/nmeth/journal/v2/n5/suppinfo/nmeth756_S1.html

Kim, T. H., & Ren, B. (2006). Genome-wide analysis of protein-DNA interactions. *Annu. Rev. Genomics Hum. Genet., 7*, 81-102.

Kitano, H. (2002). Systems biology: a brief overview. *Science, 295*(5560), 1662-1664.

Lacy, P. (2006). Mechanisms of Degranulation in Neutrophils. *Allergy, Asthma, and Clinical Immunology : Official Journal of the Canadian Society of Allergy and Clinical Immunology, 2*(3), 98-108. doi:10.1186/1710-1492-2-3-98

Luo, J.-L., Kamata, H., & Karin, M. (2005). IKK/NF-κB signaling: balancing life and death – a new approach to cancer therapy. *Journal of Clinical Investigation, 115*(10), 2625-2632. doi:10.1172/JCI26322

Maere, S., Heymans, K., & Kuiper, M. (2005). BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics, 21*(16), 3448-3449.

Marchionni, L., & Gupta, A. (2013). Package 'matchBox'.

Max-Planck-Institut für Informatik. NetworkAnalyzer Online Help. Retrieved from http://med.bioinf.mpi-inf.mpg.de/netanalyzer/help/2.7/

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., . . . Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research, 13*(11), 2498-2504.

Smoot, M., Albrecht, M., & Assenov, Y. (2016). NetworkAnalyzer. Retrieved from http://apps.cytoscape.org/apps/networkanalyzer

Tak, P. P., & Firestein, G. S. (2001). NF-κB: a key role in inflammatory diseases. *The Journal of Clinical Investigation, 107*(1), 7-11. doi:10.1172/JCI11830

UniProt Consortium. (2009). The Universal Protein Resource (UniProt) 2009. *Nucleic Acids Res, 37*, D169-D174.

Vastrik, I., D'Eustachio, P., Schmidt, E., Joshi-Tope, G., Gopinath, G., Croft, D., . . . Lewis, S. (2007). Reactome: a knowledge base of biologic pathways and processes. *Genome biology, 8*(3), 1.

Viswanathan, G. A., Seto, J., Patil, S., Nudelman, G., & Sealfon, S. C. (2008). Getting started in biological pathway construction and analysis. *PLoS Comput Biol, 4*(2), e16.

Warde-Farley, D., Donaldson, S. L., Comes, O., Zuberi, K., Badrawi, R., Chao, P., . . . Lopes, C. T. (2010). The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic acids research, 38*(suppl 2), W214-W220.

Whitehead Institute for Biomedical Research. (2013). Bioinformatics & Research Computing: Compare two lists. Retrieved from http://jura.wi.mit.edu/bioc/tools/compare.php

Wierling, C., Herwig, R., & Lehrach, H. (2007). Resources, standards and tools for systems biology. *Briefings in functional genomics & proteomics, 6*(3), 240-251.

Yoon, J., Blumer, A., & Lee, K. (2006). An algorithm for modularity analysis of directed and weighted biological networks based on edge-betweenness centrality. *Bioinformatics, 22*(24), 3106-3108.

Appendices

A.1 RStudio codes

CAT analysis

This code is written for analysis using the 'equalStat' parameter. To run the code using 'equalRank', change `method = "equalStat"` to `method = "equalRank"` in every computeCat function.

```
library(matchBox)

################### Diet A NTNU vs Diet A Wim ###################
data1 <- read.csv(file = "Diet_A_NTNU.csv", header = T, sep = ";",
stringsAsFactors = F)
data1 <- na.omit(data1)
data1$Padj.dietA <- as.numeric(data1$Padj.dietA)
data1$log2FC_mean_dietA <- as.numeric(data1$log2FC_mean_dietA)
data2 <- read.csv(file = "Diet_A_Wim.csv", header = T, sep = ";",
stringsAsFactors = F)
data2 <- na.omit(data2)
data2$Padj.dietA <- as.numeric(data2$Padj.dietA)
data2$log2FC_mean_dietA <- as.numeric(data2$log2FC_mean_dietA)

#Data merge
data1 <- filterRedundant(data1, idCol = "Name", byCol = "Padj.dietA",
decreasing = F)
data2 <- filterRedundant(data2, idCol = "Name", byCol = "Padj.dietA",
decreasing = F)
merge1 <- merge(data1, data2, by = "Name", all.x = F)
merge1$log2FC_mean_dietA.x <- NULL
merge1$log2FC_mean_dietA.y<- NULL
View(merge1)
CAT <- computeCat(merge1, size = nrow(merge1), idCol = "Name", de-
creasing = F, method = "equalStat")
plotCat(CAT, whichToPlot = 1:length(CAT))
View(CAT)
write.table(CAT, file = "AvsA_equalStat.csv", sep = ";", col.names =
T, row.names = F, quote = F)


################### Diet B NTNU vs Diet B Wim ###################
data1 <- read.csv(file = "Diet_B_NTNU.csv", header = T, sep = ";",
stringsAsFactors = F)
data1 <- na.omit(data1)
data1$Padj.dietB <- as.numeric(data1$Padj.dietB)
data1$log2FC_mean_dietB <- as.numeric(data1$log2FC_mean_dietB)
data2 <- read.csv(file = "Diet_B_Wim.csv", header = T, sep = ";",
stringsAsFactors = F)
data2 <- na.omit(data2)
data2$Padj.dietB <- as.numeric(data2$Padj.dietB)
data2$log2FC_mean_dietB <- as.numeric(data2$log2FC_mean_dietB)

#Data merge
data1 <- filterRedundant(data1, idCol = "Name", byCol = "Padj.dietB",
decreasing = F)
```

```
data2 <- filterRedundant(data2, idCol = "Name", byCol = "Padj.dietB",
decreasing = F)
merge1 <- merge(data1, data2, by = "Name", all.x = F)
View(merge1)
merge1$log2FC_mean_dietB.x <- NULL
merge1$log2FC_mean_dietB.y<- NULL
CAT <- computeCat(merge1, size = nrow(merge1), idCol = "Name", de-
creasing = F, method = "equalStat")
plotCat(CAT, whichToPlot = 1:length(CAT))
View(CAT)
write.table(CAT, file = "BvsB_equalStat.csv", sep = ";", col.names =
T, row.names = F, quote = F)

################### Diet A NTNU vs diet B NTNU ###################
data1 <- read.csv(file = "Diet_A_NTNU.csv", header = T, sep = ";",
stringsAsFactors = F)
data1 <- na.omit(data1)
data1$Padj.dietA <- as.numeric(data1$Padj.dietA)
data1$log2FC_mean_dietA <- as.numeric(data1$log2FC_mean_dietA)
data2 <- read.csv(file = "Diet_B_NTNU.csv", header = T, sep = ";",
stringsAsFactors = F)
data2 <- na.omit(data2)
data2$Padj.dietB <- as.numeric(data2$Padj.dietB)
data2$log2FC_mean_dietB <- as.numeric(data2$log2FC_mean_dietB)

#Data merge
data1 <- filterRedundant(data1, idCol = "Name", byCol = "Padj.dietA",
decreasing = F)
data2 <- filterRedundant(data2, idCol = "Name", byCol = "Padj.dietB",
decreasing = F)
merge1 <- merge(data1, data2, by = "Name", all.x = F)
View(merge1)
merge1$log2FC_mean_dietA <- NULL
merge1$log2FC_mean_dietB<- NULL
CAT <- computeCat(merge1, size = nrow(merge1), idCol = "Name", de-
creasing = F, method = "equalStat")
plotCat(CAT, whichToPlot = 1:length(CAT))
View(CAT)
write.table(CAT, file = "NTNU_AvsB_equalStat.csv", sep = ";",
col.names = T, row.names = F, quote = F)

################### Diet A Wim vs diet B Wim ###################
data1 <- read.csv(file = "Diet_A_Wim.csv", header = T, sep = ";",
stringsAsFactors = F)
data1 <- na.omit(data1)
data1$Padj.dietA <- as.numeric(data1$Padj.dietA)
data1$log2FC_mean_dietA <- as.numeric(data1$log2FC_mean_dietA)
data2 <- read.csv(file = "Diet_B_Wim.csv", header = T, sep = ";",
stringsAsFactors = F)
data2 <- na.omit(data2)
data2$Padj.dietB <- as.numeric(data2$Padj.dietB)
data2$log2FC_mean_dietB <- as.numeric(data2$log2FC_mean_dietB)
```

```
#Data merge
data1 <- filterRedundant(data1, idCol = "Name", byCol = "Padj.dietA",
decreasing = F)
data2 <- filterRedundant(data2, idCol = "Name", byCol = "Padj.dietB",
decreasing = F)
merge1 <- merge(data1, data2, by = "Name", all.x = F)
View(merge1)
merge1$log2FC_mean_dietA <- NULL
merge1$log2FC_mean_dietB<- NULL
CAT <- computeCat(merge1, size = nrow(merge1), idCol = "Name", de-
creasing = F, method = "equalStat")
plotCat(CAT, whichToPlot = 1:length(CAT))
View(CAT)
write.table(CAT, file = "Wim_AvsB_equalStat.csv", sep = ";", col.names
= T, row.names = F, quote = F)
```

Volcano plot code

The code is here presented using AHC NTNU as an example. Smaller changes were made with respect to data read, x axis limits (`xlim`) in the plot function, and to output file data to fit the different data sets.

```
data <- read.csv("Up-Down_regulated genes/Diet A_NTNU.csv", header =
T, sep = ";", stringsAsFactors = F)
data$Padj.dietA <- as.numeric(data$Padj.dietA)
data$log2FC_mean_dietA <- as.numeric(data$log2FC_mean_dietA)
View(data)

#Volcano plot
with(data, plot(log2FC_mean_dietA, -log10(Padj.dietA), pch=20,
main="Volcano plot", xlim=c(-1,1)))

#pval<.05 in red
with(subset(data, Padj.dietA<.05 ), points(log2FC_mean_dietA, -
log10(Padj.dietA), pch=20, col="red"))

#log2FC > 0.38 in orange
with(subset(data, abs(log2FC_mean_dietA)>0.38), points(log2FC_mean_di-
etA, -log10(Padj.dietA), pch=20, col="orange"))

#log2FC > 0.38 & pval < 0.05 in green
with(subset(data, Padj.dietA<.05 & abs(log2FC_mean_dietA)>0.38),
points(log2FC_mean_dietA, -log10(Padj.dietA), pch=20, col="green"))

#log2FC > 0.5 & pval < 0.05 in blue
with(subset(data, Padj.dietA<.05 & abs(log2FC_mean_dietA)>0.5),
points(log2FC_mean_dietA, -log10(Padj.dietA), pch=20, col="blue"))

#log2FC > 0.68 & pval < 0.05 in turquoise
with(subset(data, Padj.dietA<.05 & abs(log2FC_mean_dietA)>0.68),
points(log2FC_mean_dietA, -log10(Padj.dietA), pch=20, col="tur-
quoise"))

#Get values with pval < 0.05 & log2FC > 0.38 and save them in a csv
file
outputdata <- data.frame(subset(data, data$Padj.dietA<.05 &
abs(log2FC_mean_dietA)>0.38))
write.table(outputdata, file = "SumDietANTNU.csv", sep = ";", quote =
F, col.names = T, row.names = F)

#Labels on blue and turquoise points
library(calibrate)
with(subset(data, Padj.dietA<.05 & abs(log2FC_mean_dietA)>0.5),
textxy(log2FC_mean_dietA, -log10(Padj.dietA), labs=Name, cex=.5))
```

A.2 CAT plots

Parameter: 'equalRank'

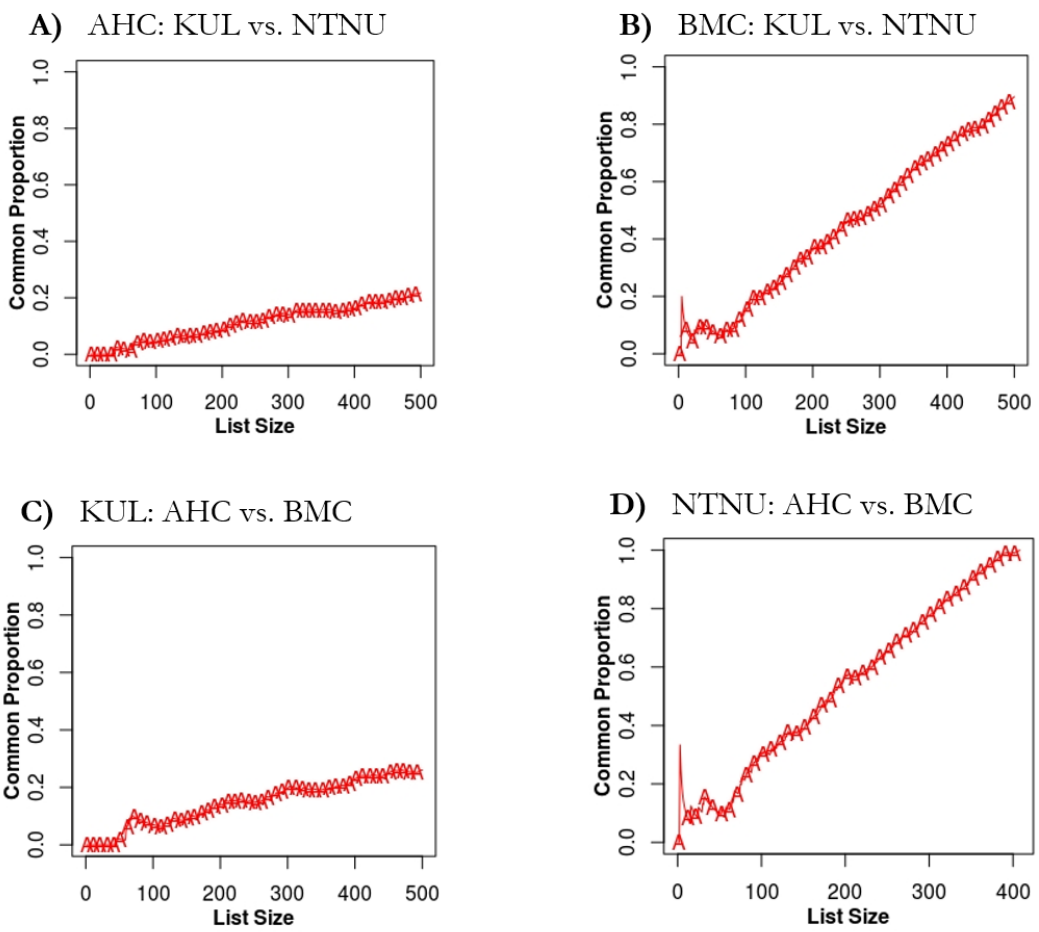


Figure 20. CAT plots produced using the 'equalRank' parameter. The initial gene lists for AHC and BMC were ranked from lowest to highest P value based on both KUL's and NTNU's analysis. The statistical analyses were compared for both AHC **(A)** and BMC **(B)**, and the diets were compared with respect to both KUL's **(C)** and NTNU's **(D)** analysis.

Parameter: 'equalStat'

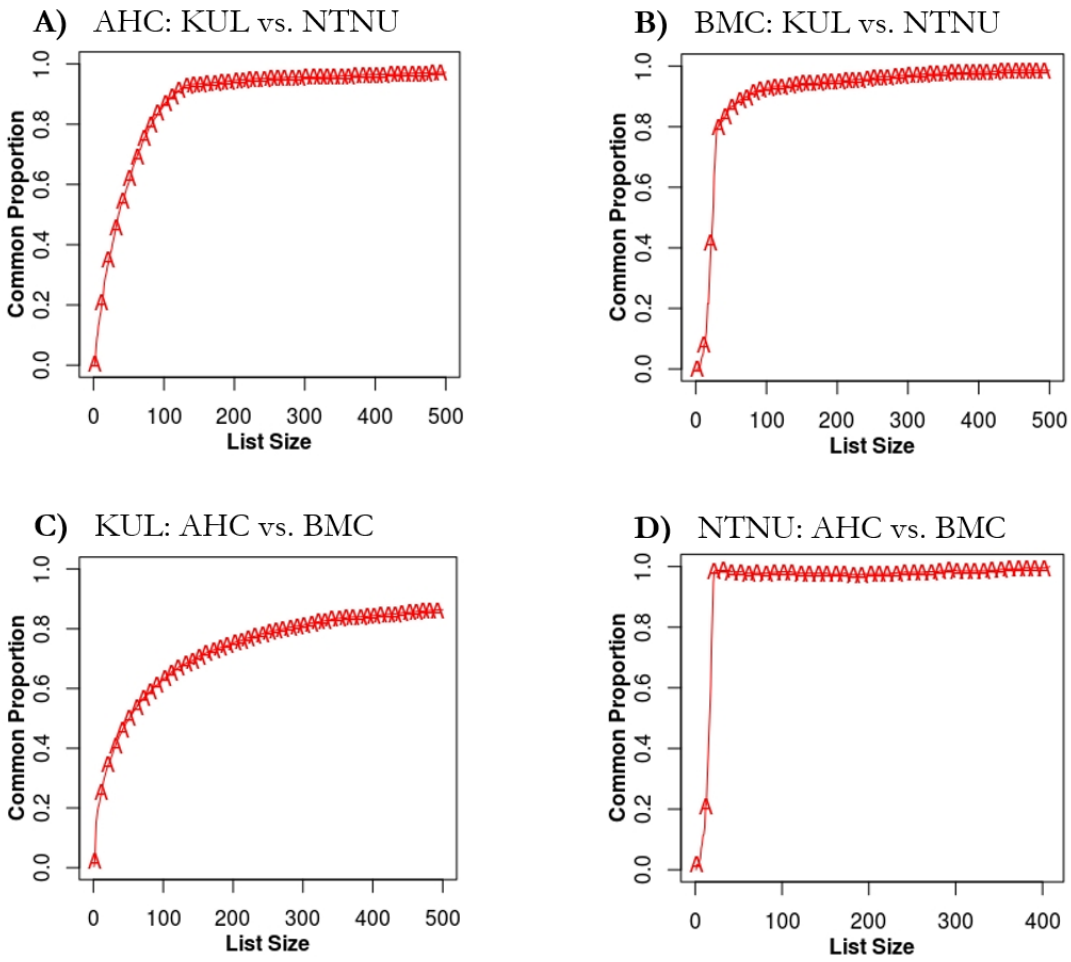


Figure 21. CAT plots produced using the 'equalStat' parameter. The initial gene lists for AHC and BMC were assigned P values based on both KUL's and NTNU's analysis. The statistical analyses were compared for both AHC **(A)** and BMC **(B)**, and the diets were compared with respect to both KUL's **(C)** and NTNU's **(D)** analysis.

A.3 Gene lists after cross-ranking

Table 14. Gene list for AHC based on the statistical data produced by NTNU. The genes are cross-ranked based both adjusted P value and log2FC.

Gene name	Full name	Log2 FC	P adj
NRGN	Neurogranin	0,785874310	0,000416813
DEFA3	Neutrophil defensin 3	-0,610251398	0,000440140
DEFA3	Neutrophil defensin 3	-0,605390274	0,000544368
IRF1	Interferon regulatory factor 1	-0,675673049	0,000739848
HMGA1	High mobility group protein HMG-I/HMG-Y	0,720552454	0,000952637
USF2	Upstream stimulatory factor 2	0,592587182	0,000846877
LOC339047	PKD1P1, polycystin 1, transient receptor potential channel interacting pseudogene 1	0,585702484	0,001133337
MS4A7	Membrane-spanning 4-domains subfamily A member 7	0,453569231	0,000440140
INO80E	INO80 complex subunit E	0,494609234	0,000875831
CD74	HLA class II histocompatibility antigen gamma chain	0,610058629	0,001358704
DEFA3	Neutrophil defensin 3	-0,577178692	0,001210528
CTSA	Lysosomal protective protein	0,454658501	0,000670523
PKM2	Pyruvate kinase PKM	0,432231814	0,000416813
GRINA	Protein lifeguard 1	0,658108966	0,002061457
LGALS9	Galectin-9	0,472252278	0,000799075
PKM2	Pyruvate kinase PKM	0,612912508	0,002223474
SLA	Src-like-adapter	-0,426170797	0,000161766
PKM2	Pyruvate kinase PKM	0,576699093	0,001367842
F13A1	Coagulation factor XIII A chain	0,437756665	0,000614478
PARVB	Beta-parvin	0,534067931	0,001676121
ITGB3	Integrin beta-3	0,454658488	0,001071993
CD74	HLA class II histocompatibility antigen gamma chain	0,473993051	0,001199907
AGPAT1	1-acyl-sn-glycerol-3-phosphate acyltransferase alpha	0,479608446	0,001425995
LRG1	Leucine-rich alpha-2-glycoprotein	-0,418436100	0,000544368
LY6E	Lymphocyte antigen 6E	0,410461033	0,000440140
TSC22D4	TSC22 domain family protein 4	0,474919175	0,001869523
KCTD20	BTB/POZ domain-containing protein KCTD20	0,438229436	0,001236490
PACS1	Phosphofurin acidic cluster sorting protein 1	0,524744470	0,003366308
ITGA2B	Integrin alpha-IIb	0,538815645	0,003906013
AES	Amino-terminal enhancer of split	0,423910891	0,000972378
OLFM4	Olfactomedin-4	-0,404168686	0,000508319
HNRNPL	Heterogeneous nuclear ribonucleoprotein L	0,445439227	0,001482677
VAMP2	Vesicle-associated membrane protein 2	0,423697080	0,001156693
SH3BP1	SH3 domain-binding protein 1	0,477171837	0,003386707

GPR162	Probable G-protein coupled receptor 162	-0,401875699	0,000614478
DEFA4	Neutrophil defensin 4	-0,400345191	0,000608033
CSDAP1	Y-box binding protein 3 pseudogene 1	0,503435985	0,006040751
C6orf25	Protein G6b	0,442302300	0,003295502
LOC407835	Mitogen-activated protein kinase kinase 2 pseudogene	0,429529180	0,002299286
LOC100130751	?	0,499233130	0,006277245
SMARCC2	SWI/SNF complex subunit SMARCC2	0,432116237	0,002840447
LOC441481	Glutathione peroxidase pseudogene 1	0,388188831	0,000750171
CD44	CD44 antigen	0,391022569	0,000923949
LSP1	Lymphocyte-specific protein 1	0,492140242	0,007030275
RNF11	RING finger protein 11	0,404962157	0,001364963
P2RY13	P2Y purinoceptor 13	-0,395416151	0,001156693
NPIPL2	Nuclear pore complex-interacting protein family member B15	0,445598057	0,005280867
ATP6V0C	V-type proton ATPase 16 kDa proteolipid subunit	0,441933012	0,005193471
LOC440353	Nuclear pore complex-interacting protein family, member B12	0,491915598	0,019692591
DEFA3	Neutrophil defensin 3	-0,415970243	0,003091785
ALOX12	Arachidonate 12-lipoxygenase, 12S-type	0,421899028	0,003371446
NA	*NA*	0,429270723	0,004627177
HSF1	Heat shock factor protein 1	0,405140744	0,002703660
FKBP1A	Peptidyl-prolyl cis-trans isomerase FKBP1A	0,454590780	0,010237587
RBM38	RNA-binding protein 38	0,431626935	0,007810615
GPR177	Protein wntless homolog	-0,380182909	0,001355033
JUNB	Transcription factor jun-B	0,433985735	0,034682590
RBM38	RNA-binding protein 38	0,403290280	0,004501506
PHF1	PHD finger protein 1	0,394063652	0,003139440
UBA1	Ubiquitin-like modifier-activating enzyme 1	0,386653768	0,002239957
MGC13005	DEAD/H box polypeptide 11 like 2	0,415587209	0,006353142
TCF7	Transcription factor 7	0,383192704	0,002407958
FAM100A	UBA-like domain-containing protein 1	0,406552150	0,005978949
PI3	Elafin	-0,412886761	0,006540036
TPM1	Tropomyosin alpha-1 chain	0,397098167	0,005062518
PRKDC	DNA-dependent protein kinase catalytic subunit	-0,389142526	0,003991718
NA	*NA*	0,380750645	0,002881443
RTN4	Reticulon-4	0,389556194	0,004731738
ZFP36L2	mRNA decay activator protein ZFP36L2	0,401825338	0,005999468
VNN2	Vascular non-inflammatory molecule 2	-0,418381380	0,049977201
RASGRP4	RAS guanyl-releasing protein 4	0,407536012	0,018248677

LOC728888	Nuclear pore complex-interacting protein family member B11	0,409530835	0,021993817
BCL2L1	Bcl-2-like protein 1	0,402780979	0,014786532
MKNK2	MAP kinase-interacting serine/threonine-protein kinase 2	0,380763132	0,005221225
MTF1	Metal regulatory transcription factor 1	0,380650086	0,005908555
RPS15A	40S ribosomal protein S15a	-0,387166063	0,007810615
BTG2	Protein BTG2	0,386849900	0,044598663
TAF15	TATA-binding protein-associated factor 2N	-0,382570254	0,040225811

Table 15. Gene list for AHC based on the statistical data produced by KUL. The genes are cross-ranked based on both adjusted P value and log2FC.

Gene name	Full name	Log2 FC	P adj
GRINA	Protein lifeguard 1	0,693714416	9,00E-05
DEFA3	Neutrophil defensin 3	-0,688533497	9,00E-05
LGALS9	Galectin-9	0,504723422	9,00E-05
LOC100130751	?	0,509230068	9,00E-05
PACS1	Phosphofurin acidic cluster sorting protein 1	0,547518324	2,00E-04
PKM2	Pyruvate kinase PKM	0,638477671	4,00E-04
NRGN	Neurogranin	0,813561724	0,0013
DEFA3	Neutrophil defensin 3	-0,683434566	8,00E-04
PKM2	Pyruvate kinase PKM	0,602777934	5,00E-04
LOC339047	PKD1P1, polycystin 1, transient receptor potential channel interacting pseudogene 1	0,600776281	5,00E-04
DEFA3	Neutrophil defensin 3	-0,656440818	0,0011
USF2	Upstream stimulatory factor 2	0,610103656	5,00E-04
CSDAP1	Y-box binding protein 3 pseudogene 1	0,523860651	4,00E-04
CD74	HLA class II histocompatibility antigen gamma chain	0,644901231	0,0014
DEFA3	Neutrophil defensin 3	-0,491143066	3,00E-04
SH3BP1	SH3 domain-binding protein 1	0,491773789	3,00E-04
CD74	HLA class II histocompatibility antigen gamma chain	0,499640611	5,00E-04
RBM38	RNA-binding protein 38	0,443326181	1,00E-04
NA	*NA*	0,420055129	9,00E-05
JUNB	Transcription factor jun-B	0,446964244	3,00E-04
ITGA2B	Integrin alpha-IIb	0,548291436	0,0016
VAMP2	Vesicle-associated membrane protein 2	0,429148495	9,00E-05
LY6E	Lymphocyte antigen 6E	0,41983234	9,00E-05
INO80E	INO80 complex subunit E	0,514833778	0,0019
FKBP1A	Peptidyl-prolyl cis-trans isomerase FKBP1A	0,48287916	0,0015
CHCHD10	Coiled-coil-helix-coiled-coil-helix domain-containing protein 10, mitochondrial	0,400503305	9,00E-05
MTF1	Metal regulatory transcription factor 1	0,409039501	9,00E-05
TSC22D4	TSC22 domain family protein 4	0,500044547	0,002
HMGA1	High mobility group protein HMG-I/HMG-Y	0,753453878	0,014
UBA1	Ubiquitin-like modifier-activating enzyme 1	0,405837951	9,00E-05
PLEK	Pleckstrin	-0,412199074	3,00E-04
CD44	CD44 antigen	0,4091406	2,00E-04
HNRNPL	Heterogeneous nuclear ribonucleoprotein L	0,453977066	0,0016
SLA	Src-like-adapter	-0,43439385	0,0013
TCF7	Transcription factor 7	0,394141612	9,00E-05
PARVB	Beta-parvin	0,552513807	0,0103

LOC407835	Mitogen-activated protein kinase kinase 2 pseudo-gene	0,44216574	0,0015
LSP1	Lymphocyte-specific protein 1	0,387701197	9,00E-05
AES	Amino-terminal enhancer of split	0,418898738	6,00E-04
KCTD20	BTB/POZ domain-containing protein KCTD20	0,454878769	0,0028
BCL2L1	Bcl-2-like protein 1	0,39073649	1,00E-04
AGPAT1	1-acyl-sn-glycerol-3-phosphate acyltransferase al-pha	0,50148461	0,0076
MS4A7	Membrane-spanning 4-domains subfamily A mem-ber 7	0,488745772	0,0039
F13A1	Coagulation factor XIII A chain	0,459969266	0,0036
MGC13005	DEAD/H box polypeptide 11 like 2	0,453141108	0,0031
OLFM4	Olfactomedin-4	-0,421921671	0,0015
ITGB3	Integrin beta-3	0,469310517	0,0059
TPM1	Tropomyosin alpha-1 chain	0,39725809	4,00E-04
LSP1	Lymphocyte-specific protein 1	0,527715236	0,0241
LOC440353	Nuclear pore complex-interacting protein family, member B12	0,510410873	0,0406
FAM100A	UBA-like domain-containing protein 1	0,425056575	0,0021
FERMT3	Fermitin family homolog 3	0,381980726	3,00E-04
ATP6V0C	V-type proton ATPase 16 kDa proteolipid subunit	0,457511459	0,0127
RASGRP4	RAS guanyl-releasing protein 4	0,430568816	0,0036
PRKDC	DNA-dependent protein kinase catalytic subunit	-0,388675954	5,00E-04
PKM2	Pyruvate kinase PKM	0,45165739	0,0086
NA	*NA*	0,454916319	0,0176
DEFA4	Neutrophil defensin 4	-0,451712509	0,0128
LOC728888	Nuclear pore complex-interacting protein family member B11	0,418731158	0,0034
NA	*NA*	-0,39033558	0,0014
PHF1	PHD finger protein 1	0,417839553	0,0035
LRG1	Leucine-rich alpha-2-glycoprotein	-0,410246032	0,0032
RNF11	RING finger protein 11	0,426046797	0,0076
RPS15A	40S ribosomal protein S15a	-0,407962228	0,0024
TAF15	TATA-binding protein-associated factor 2N	-0,403342564	0,002
LOC441481	Glutathione peroxidase pseudogene 1	0,392583677	0,0017
MBNL1	Muscleblind-like protein 1	0,38759471	0,0014
C6orf25	Protein G6b	0,443570248	0,0214
NPIPL2	Nuclear pore complex-interacting protein family member B15	0,438134884	0,0148
ALOX12	Arachidonate 12-lipoxygenase, 12S-type	0,428000375	0,011
UNC119	Protein unc-119 homolog A	-0,389845145	0,0015
RBM38	RNA-binding protein 38	0,411545073	0,0048
P2RY13	P2Y purinoceptor 13	-0,405606709	0,0058

LIMS1	LIM and senescent cell antigen-like-containing domain protein 1	0,384457939	0,0019
GPR162	Probable G-protein coupled receptor 162	-0,409735238	0,019
IRF2	Interferon regulatory factor 2	0,391549298	0,008
TNRC6A	Trinucleotide repeat-containing gene 6A protein	0,385622248	0,0039
RTN4	Reticulon-4	0,409769226	0,0474
YWHAE	14-3-3 protein epsilon	0,385477237	0,0239

Table 16. Gene list for BMC based on the statistical data produced by NTNU. The genes are cross-ranked based on both adjusted P value and log2FC.

Gene name	Full name	log2FC	P adj
LOC339047	PKD1P1, polycystin 1, transient receptor potential channel interacting pseudogene 1	0,719047197	0,023662603
IRF1	Interferon regulatory factor 1	-0,851221871	0,030740429
NRGN	Neurogranin	0,693988488	0,023662603
GPR162	Probable G-protein coupled receptor 162	-0,594660328	0,030740429
CD44	CD44 antigen	0,601589625	0,030868901
HMGA1	High mobility group protein HMG-I/HMG-Y	0,849385448	0,034174367
MBNL1	Muscleblind-like protein 1	0,491176575	0,021168067
LRG1	Leucine-rich alpha-2-glycoprotein	-0,502168085	0,030740429
TSC22D4	TSC22 domain family protein 4	0,530444619	0,031574297
AES	Amino-terminal enhancer of split	0,513955387	0,031574297
NA	*NA*	0,466875965	0,021168067
SLA	Src-like-adapter	-0,449542273	0,012249425
LOC100130751	?	0,577752335	0,033521227
NAP1L1	Nucleosome assembly protein 1-like 1	0,447204759	0,001834412
SMARCC2	SWI/SNF complex subunit SMARCC2	0,613911398	0,034766813
NPIPL2	Nuclear pore complex-interacting protein family member B15	0,484031392	0,030740429
PACS1	Phosphofurin acidic cluster sorting protein 1	0,635109786	0,039777016
F13A1	Coagulation factor XIII A chain	0,425507141	0,000875432
PKM2	Pyruvate kinase PKM	0,696306324	0,041855689
AGPAT1	1-acyl-sn-glycerol-3-phosphate acyltransferase alpha	0,462504414	0,031574297
TCF7	Transcription factor 7	0,570017223	0,041585990
CNOT7	CCR4-NOT transcription complex subunit 7	0,484389023	0,034766813
IRF2BP2	Interferon regulatory factor 2-binding protein 2	0,419109830	0,022588768
ZFP106	Zinc finger protein 106	0,403313303	0,018724670
RNF216	E3 ubiquitin-protein ligase RNF216	-0,460638149	0,033521227
NPHP3	Nephrocystin-3	0,456527638	0,033513026
PHF1	PHD finger protein 1	0,532610196	0,042016778
CMPK1	UMP-CMP kinase	0,395463361	0,018331504
GRINA	Protein lifeguard 1	0,683058276	0,046578562
INO80E	INO80 complex subunit E	0,593056361	0,042936300
ITGA2B	Integrin alpha-IIb	0,380515578	0,012917413
YWHAE	14-3-3 protein epsilon	0,518190842	0,042792212
BCL2L1	Bcl-2-like protein 1	0,445082955	0,033585124
USF2	Upstream stimulatory factor 2	0,620980601	0,048240776
SALL3	Sal-like protein 3	0,499933159	0,042196521
CENPT	Centromere protein T	0,471299008	0,040314785

BRD7	Bromodomain-containing protein 7	0,397385011	0,031574297
BCL7B	B-cell CLL/lymphoma 7 protein family member B	-0,521813957	0,046686351
C6orf47	Uncharacterized protein C6orf47	-0,400437718	0,032929123
DYRK1B	Dual specificity tyrosine-phosphorylation-regulated kinase 1B	0,487147376	0,043999479
WBP1	WW domain-binding protein 1	-0,438271783	0,034766813
PRKDC	DNA-dependent protein kinase catalytic subunit	-0,467379753	0,042792212
DUSP6	Dual specificity protein phosphatase 6	0,394407562	0,031574297
SFRS18	Arginine/serine-rich protein PNISR	0,442511784	0,040219449
IL1R2	Interleukin-1 receptor type 2	-0,432111307	0,037447629
PKM2	Pyruvate kinase PKM	0,517749706	0,048373968
FAM58A	Cyclin-related protein FAM58A	-0,397985023	0,034766813
ABTB1	Ankyrin repeat and BTB/POZ domain-containing protein 1	-0,477831942	0,048373968
SLC43A2	Large neutral amino acids transporter small subunit 4	-0,404969761	0,040587624
PDK4	[Pyruvate dehydrogenase (acetyl-transferring)] kinase isozyme 4, mitochondrial	0,448341027	0,044476083
CLDN15	Claudin-15	-0,383796373	0,034766813
TICAM1	TIR domain-containing adapter molecule 1	-0,440487334	0,042936300
CHCHD10	Coiled-coil-helix-coiled-coil-helix domain-containing protein 10, mitochondrial	0,41723006	0,042792212
NIPSNAP1	Protein NipSnap homolog 1	-0,448288528	0,046814842
PHF1	PHD finger protein 1	0,384062734	0,039338506
NPIPL2	Nuclear pore complex-interacting protein family member B15	0,391624837	0,041772349
RASGRP4	RAS guanyl-releasing protein 4	0,424935155	0,047121020
ZNF746	Zinc finger protein 746	-0,394959174	0,044054224
POLR3H	DNA-directed RNA polymerase III subunit RPC8	-0,380141385	0,045763314
ATP2B4	Plasma membrane calcium-transporting ATPase 4	0,385870842	0,048918385

Table 17. Gene list for BMC based on the statistical data produced by KUL. The genes are cross-ranked based on both adjusted P value and log2FC.

Gene name	Full name	Log2 FC	P adj
PKM2	Pyruvate kinase PKM	0,696306324	0,0012
USF2	Upstream stimulatory factor 2	0,620980601	0,0011
YWHAE	14-3-3 protein epsilon	0,518190842	5,00E-04
PKM2	Pyruvate kinase PKM	0,517749706	5,00E-04
FKBP1A	Peptidyl-prolyl cis-trans isomerase FKBP1A	0,671873555	0,0017
GRINA	Protein lifeguard 1	0,683058276	0,0023
LGALS9	Galectin-9	0,513059354	8,00E-04
PHF1	PHD finger protein 1	0,532610196	9,00E-04
PLEK	Pleckstrin	-0,633049378	0,0026
LSP1	Lymphocyte-specific protein 1	0,603393027	0,0021
CENPT	Centromere protein T	0,471299008	6,00E-04
LY6E	Lymphocyte antigen 6E	0,462036682	6,00E-04
TGOLN2	Trans-Golgi network integral membrane protein 2	-0,449519043	3,00E-04
SMARCC2	SWI/SNF complex subunit SMARCC2	0,613911398	0,0044
CD44	CD44 antigen	0,601589625	0,0039
LOC440353	Nuclear pore complex-interacting protein family, member B12	0,590926133	0,0033
RTN4	Reticulon-4	0,566005247	0,0032
IMPDH1	Inosine-5'-monophosphate dehydrogenase 1	-0,444833702	5,00E-04
MBNL1	Muscleblind-like protein 1	0,491176575	0,0013
NPHP3	Nephrocystin-3	0,456527638	8,00E-04
ZGPAT	Zinc finger CCCH-type with G patch domain-containing protein	-0,446692663	7,00E-04
GPR162	Probable G-protein coupled receptor 162	-0,594660328	0,0063
DYRK1B	Dual specificity tyrosine-phosphorylation-regulated kinase 1B	0,487147376	0,0017
CTDSP1	Carboxy-terminal domain RNA polymerase II polypeptide A small phosphatase 1	0,505736138	0,0019
HSF1	Heat shock factor protein 1	0,636529652	0,0103
PDK4	[Pyruvate dehydrogenase (acetyl-transferring)] kinase isozyme 4, mitochondrial	0,448341027	9,00E-04
PKM2	Pyruvate kinase PKM	0,672293787	0,0116
TCF7	Transcription factor 7	0,570017223	0,0066
UCP2	Mitochondrial uncoupling protein 2	0,507509426	0,0027
LSP1	Lymphocyte-specific protein 1	0,537230057	0,0046
HNRNPUL1	Heterogeneous nuclear ribonucleoprotein U-like protein 1	0,438970522	9,00E-04
HMGA1	High mobility group protein HMG-I/HMG-Y	0,849385448	0,014
ZFP36L2	mRNA decay activator protein ZFP36L2	0,587802035	0,0092
AES	Amino-terminal enhancer of split	0,513955387	0,0042

NA	*NA*	0,466875965	0,0018
RNF216	E3 ubiquitin-protein ligase RNF216	-0,460638149	0,0016
LOC339047	Polycystin 1, transient receptor potential channel interacting pseudogene 1	0,719047197	0,0180
NA	*NA*	0,553203808	0,0090
MKNK2	MAP kinase-interacting serine/threonine-protein kinase 2	0,469126564	0,0025
NINJ1	Ninjurin-1	-0,532026589	0,0073
ABTB1	Ankyrin repeat and BTB/POZ domain-containing protein 1	-0,477831942	0,0034
ARRB2	Beta-arrestin-2	0,505803769	0,0052
RTN4	Reticulon-4	0,405850611	6,00E-04
INF2	Inverted formin-2	-0,417067674	9,00E-04
LOC728888	Nuclear pore complex interacting protein family member B11	0,474417327	0,0045
NRGN	Neurogranin	0,693988488	0,0351
C20orf149	Pancreatic progenitor cell differentiation and proliferation factor	0,541121232	0,0112
VAMP2	Vesicle-associated membrane protein 2	0,587794541	0,0141
TSPAN18	Tetraspanin-18	-0,402539211	8,00E-04
CHCHD10	Coiled-coil-helix-coiled-coil-helix domain-containing protein 10, mitochondrial	0,417230060	0,0013
PANX2	Pannexin-2	-0,403778469	9,00E-04
TSC22D4	TSC22 domain family protein 4	0,530444619	0,0118
PACS1	Phosphofurin acidic cluster sorting protein 1	0,635109786	0,0375
INO80E	INO80 complex subunit E	0,593056361	0,0290
MED16	Mediator of RNA polymerase II transcription subunit 16	-0,411724432	0,0013
VEGFB	Vascular endothelial growth factor B	-0,405185547	0,0012
FKBP1A	Peptidyl-prolyl cis-trans isomerase FKBP1A	0,592831615	0,0317
RPL36AP49	Ribosomal protein L36a pseudogene 49	0,383858855	7,00E-04
CMPK1	UMP-CMP kinase	0,395463361	0,0012
DUSP6	Dual specificity protein phosphatase 6	0,394407562	0,0011
RPL13P12	Ribosomal protein L13 pseudogene 12	-0,449870438	0,0057
RASGRP4	RAS guanyl-releasing protein 4	0,416862274	0,0018
HNRNPL	Heterogeneous nuclear ribonucleoprotein L	0,508626578	0,0138
AGPAT1	1-acyl-sn-glycerol-3-phosphate acyltransferase alpha	0,462504414	0,0095
CHMP6	Charged multivesicular body protein 6	-0,428867696	0,0037
CD68	Macrosialin	0,423481253	0,0031
PRKDC	DNA-dependent protein kinase catalytic subunit	-0,467379753	0,0111
NPIPL2	Nuclear pore complex-interacting protein family member B15	0,484031392	0,0129
MTF1	Metal regulatory transcription factor 1	0,429124710	0,0056
ATP2B4	Plasma membrane calcium-transporting ATPase 4	0,385870842	0,0016

C7orf27	BRCA1-associated ATM activator 1	-0,452368688	0,0106
RASGRP4	RAS guanyl-releasing protein 4	0,424935155	0,0045
TCF7	Transcription factor 7	0,565927333	0,0456
LRG1	Leucine-rich alpha-2-glycoprotein	-0,502168085	0,0258
SH3BP1	SH3 domain-binding protein 1	0,563846541	0,0467
NA	*NA*	0,537495435	0,0439
SALL3	Sal-like protein 3	0,499933159	0,0281
C6orf136	Uncharacterized protein C6orf136	-0,383182004	0,0016
TIAL1	Nucleolysin TIAR	0,433213340	0,0099
C13orf27	Testis-expressed protein 30	0,383802088	0,0019
BCL2L1	Bcl-2-like protein 1	0,445082955	0,0131
CD74	HLA class II histocompatibility antigen gamma chain	0,438487679	0,0126
POLR3H	DNA-directed RNA polymerase III subunit RPC8	-0,380141385	0,0019
TIAL1	Nucleolysin TIAR	0,463591856	0,0368
SLC43A2	Large neutral amino acids transporter small subunit 4	-0,404969761	0,0078
NCF4	Neutrophil cytosol factor 4	-0,428062793	0,0128
CORO7	Coronin-7	0,390991772	0,0073
TICAM1	TIR domain-containing adapter molecule 1	-0,440487334	0,0290
C6orf47	Uncharacterized protein C6orf47	-0,400437718	0,0096
NPIPL2	Nuclear pore complex-interacting protein family member B15	0,391624837	0,0082
F13A1	Coagulation factor XIII A chain	0,425507141	0,0192
CTSL1	Cathepsin L	-0,439978410	0,0352
TP53I13	Tumor protein p53-inducible protein 13	-0,393590650	0,0098
CLIC3	Chloride intracellular channel protein 3	-0,407891983	0,0134
RNF220	E3 ubiquitin-protein ligase RNF220	-0,419772074	0,0283
PNPLA2	Patatin-like phospholipase domain-containing protein 2	0,428958369	0,0387
KIAA1267	KAT8 regulatory NSL complex subunit 1	0,437244256	0,0440
IRF2BP2	Interferon regulatory factor 2-binding protein 2	0,419109830	0,0410
JUNB	Transcription factor jun-B	0,414445837	0,0396
CD4	T-cell surface glycoprotein CD4	0,384500142	0,0183
DEFA3	Neutrophil defensin 3	-0,385115300	0,0250
PHF1	PHD finger protein 1	0,384062734	0,0260
TOP3B	DNA topoisomerase 3-beta-1	-0,395243472	0,0373
ITGA2B	Integrin alpha-IIb	0,380515578	0,0483

A.4 Results of the comparison of gene lists after cross-ranking

Table 18. Unique and common genes after comparing the two gene lists for AHC produced by the Volcano plot code, based on both KUL's analysis (Table 15) and NTNU's analysis (Table 14). The lists are ranked alphabetically.

Unique for KUL's analysis	Unique for NTNU's analysis	Common for both lists	
CHCHD10	*BTG2*	*AES*	*LRG1*
FERMT3	*CTSA*	*AGPAT1*	*LSP1*
IRF2	*GPR177*	*ALOX12*	*LY6E*
LIMS1	*HSF1*	*ATP6V0C*	*MGC13005*
MBNL1	*IRF1*	*BCL2L1*	*MS4A7*
PLEK	*MKNK2*	*C6orf25*	*MTF1*
TNRC6A	*PI3*	*CD44*	*NA*
UNC119	*SMARCC2*	*CD74*	*NPIPL2*
YWHAE	*VNN2*	*CSDAP1*	*NRGN*
	ZFP36L2	*DEFA3*	*OLFM4*
		DEFA4	*P2RY13*
		F13A1	*PACS1*
		FAM100A	*PARVB*
		FKBP1A	*PHF1*
		GPR162	*PKM2*
		GRINA	*PRKDC*
		HMGA1	*RASGRP4*
		HNRNPL	*RBM38*
		INO80E	*RNF11*
		ITGA2B	*RPS15A*
		ITGB3	*RTN4*
		JUNB	*SH3BP1*
		KCTD20	*SLA*
		LGALS9	*TAF15*
		LOC100130751	*TCF7*
		LOC339047	*TPM1*
		LOC407835	*TSC22D4*
		LOC440353	*UBA1*
		LOC441481	*USF2*
		LOC728888	*VAMP2*

Table 19. Unique and common genes after comparing the two gene lists for BMC produced by the Volcano plot code, based on both KUL's analysis (Table 17) and NTNU's analysis (Table 16). The lists are alphabetically ordered.

Unique for KUL's analysis		Unique for NTNU's analysis	Common for both lists	
ARRB2	*LSP1*	*BCL7B*	*ABTB1*	*MBNL1*
C13orf27	*LY6E*	*BRD7*	*AES*	*NA*
C20orf149	*MED16*	*CLDN15*	*AGPAT1*	*NPHP3*
C6orf136	*MKNK2*	*CNOT7*	*ATP2B4*	*NPIPL2*
C7orf27	*MTF1*	*FAM58A*	*BCL2L1*	*NRGN*
CD4	*NCF4*	*IL1R2*	*C6orf47*	*PACS1*
CD68	*NINJ1*	*IRF1*	*CD44*	*PDK4*
CD74	*PANX2*	*LOC100130751*	*CENPT*	*PHF1*
CHMP6	*PLEK*	*NAP1L1*	*CHCHD10*	*PKM2*
CLIC3	*PNPLA2*	*NIPSNAP1*	*CMPK1*	*POLR3H*
CORO7	*RNF220*	*SFRS18*	*DUSP6*	*PRKDC*
CTDSP1	*RPL13P12*	*SLA*	*DYRK1B*	*RASGRP4*
CTSL1	*RPL36AP49*	*WBP1*	*F13A1*	*RNF216*
DEFA3	*RTN4*	*ZFP106*	*GPR162*	*SALL3*
FKBP1A	*SH3BP1*	*ZNF746*	*GRINA*	*SLC43A2*
HNRNPL	*TGOLN2*		*HMGA1*	*SMARCC2*
HNRNPUL1	*TIAL1*		*INO80E*	*TCF7*
HSF1	*TOP3B*		*IRF2BP2*	*TICAM1*
IMPDH1	*TP53I13*		*ITGA2B*	*TSC22D4*
INF2	*TSPAN18*		*LOC339047*	*USF2*
JUNB	*UCP2*		*LRG1*	*YWHAE*
KIAA1267	*VAMP2*			
LGALS9	*VEGFB*			
LOC440353	*ZFP36L2*			
LOC728888	*ZGPAT*			

Table 20. Unique and common genes after comparing the two gene lists for AHC (Table 15) and BMC (Table 17) based on KUL's analysis only. The lists are alphabetically ordered.

Unique for AHC	Unique for BMC		Common for both lists	
ALOX12	*ABTB1*	*MKNK2*	*AES*	*LSP1*
ATP6V0C	*ARRB2*	*NCF4*	*AGPAT1*	*LY6E*
C6orf25	*ATP2B4*	*NINJ1*	*BCL2L1*	*MBNL1*
CSDAP1	*C13orf27*	*NPHP3*	*CD44*	*MTF1*
DEFA4	*C20orf149*	*PANX2*	*CD74*	*NA*
FAM100A	*C6orf136*	*PDK4*	*CHCHD10*	*NPIPL2*
FERMT3	*C6orf47*	*PNPLA2*	*DEFA3*	*NRGN*
IRF2	*C7orf27*	*POLR3H*	*F13A1*	*PACS1*
ITGB3	*CD4*	*RNF216*	*FKBP1A*	*PHF1*
KCTD20	*CD68*	*RNF220*	*GPR162*	*PKM2*
LIMS1	*CENPT*	*RPL13P12*	*GRINA*	*PLEK*
LOC100130751	*CHMP6*	*RPL36AP49*	*HMGA1*	*PRKDC*
LOC407835	*CLIC3*	*SALL3*	*HNRNPL*	*RASGRP4*
LOC441481	*CMPK1*	*SLC43A2*	*INO80E*	*RTN4*
MGC13005	*CORO7*	*SMARCC2*	*ITGA2B*	*SH3BP1*
MS4A7	*CTDSP1*	*TGOLN2*	*JUNB*	*TCF7*
OLFM4	*CTSL1*	*TIAL1*	*LGALS9*	*TSC22D4*
P2RY13	*DUSP6*	*TICAM1*	*LOC339047*	*USF2*
PARVB	*DYRK1B*	*TOP3B*	*LOC440353*	*VAMP2*
RBM38	*HNRNPUL1*	*TP53I13*	*LOC728888*	*YWHAE*
RNF11	*HSF1*	*TSPAN18*	*LRG1*	
RPS15A	*IMPDH1*	*UCP2*		
SLA	*INF2*	*VEGFB*		
TAF15	*IRF2BP2*	*ZFP36L2*		
TNRC6A	*KIAA1267*	*ZGPAT*		
TPM1	*MED16*			
UBA1				
UNC119				

Table 21. Unique and common genes after comparing the two gene lists for AHC (Table 14) and BMC (Table 16) based on NTNU's analysis only. The lists are alphabetically ordered.

Unique for AHC		Unique for BMC	Common for both lists
ALOX12	*LSP1*	*ABTB1*	*AES*
ATP6V0C	*LY6E*	*ATP2B4*	*AGPAT1*
BTG2	*MGC13005*	*BCL7B*	*BCL2L1*
C6orf25	*MKNK2*	*BRD7*	*CD44*
CD74	*MS4A7*	*C6orf47*	*F13A1*
CSDAP1	*MTF1*	*CENPT*	*GPR162*
CTSA	*OLFM4*	*CHCHD10*	*GRINA*
DEFA3	*P2RY13*	*CLDN15*	*HMGA1*
DEFA4	*PARVB*	*CMPK1*	*INO80E*
FAM100A	*PI3*	*CNOT7*	*IRF1*
FKBP1A	*RBM38*	*DUSP6*	*ITGA2B*
GPR177	*RNF11*	*DYRK1B*	*LOC100130751*
HNRNPL	*RPS15A*	*FAM58A*	*LOC339047*
HSF1	*RTN4*	*IL1R2*	*LRG1*
ITGB3	*SH3BP1*	*IRF2BP2*	*NA*
JUNB	*TAF15*	*MBNL1*	*NPIPL2*
KCTD20	*TPM1*	*NAP1L1*	*NRGN*
LGALS9	*UBA1*	*NIPSNAP1*	*PACS1*
LOC407835	*VAMP2*	*NPHP3*	*PHF1*
LOC440353	*VNN2*	*PDK4*	*PKM2*
LOC441481	*ZFP36L2*	*POLR3H*	*PRKDC*
LOC728888		*RNF216*	*RASGRP4*
		SALL3	*SLA*
		SFRS18	*SMARCC2*
		SLC43A2	*TCF7*
		TICAM1	*TSC22D4*
		WBP1	*USF2*
		YWHAE	
		ZFP106	
		ZNF746	

Table 22. The genes common in AHC for both statistical analyses, and which does not appear in any BMC. The list is alphabetically ordered.

Genes
ALOX12
ATP6V0C
C6orf25
CSDAP1
DEFA4
FAM100A
ITGB3
KCTD20
LOC407835
LOC441481
MGC13005
MS4A7
OLFM4
P2RY13
PARVB
RBM38
RNF11
RPS15A
TAF15
TPM1
UBA1

Table 23. The genes common in BMC for both statistical analyses, and which does not appear in any AHC. The list is alphabetically ordered.

Genes
ABTB1
ATP2B4
C6orf47
CENPT
CMPK1
DUSP6
DYRK1B
IRF2BP2
NPHP3
PDK4
POLR3H
RNF216
SALL3
SLC43A2
TICAM1

A.5 ClueGO results

Table 24. Results from the ClueGO overrepresentation analysis for a merged gene list for AHC. The analysis is conducted on a gene list merge of the AHC based on KUL's analysis and of the AHC based on NTNU's analysis. The given P values are corrected values, and both are corrected using Bonferroni step down. The terms are sorted with respect to ontology groups. Ontology source: GO_BiologicalProcess-GOA_23.02.2017_10h01. The analysis was performed in Cytoscape using the ClueGO plug-in.

GO ID	GO Term	Term P value	Group P value	Associated genes
1903313	positive regulation of mRNA metabolic process	7,9E-3	3,9E-3	*[BTG2, HSF1, ZFP36L2]*
33077	T cell differentiation in thymus	2,7E-3	5,4E-3	*[CD74, PRKDC, ZFP36L2]*
34446	substrate adhesion-dependent cell spreading	4,0E-3	2,0E-3	*[FERMT3, ITGB3, LIMS1, OLFM4]*
2260	lymphocyte homeostasis	2,7E-3	9,0E-3	*[CD74, LGALS9, TSC22D4]*
70228	regulation of lymphocyte apoptotic process	11,0E-3	9,0E-3	*[CD74, LGALS9, TSC22D4]*
34109	homotypic cell-cell adhesion	150,0E-6	81,0E-6	*[ALOX12, FERMT3, ITGA2B, ITGB3, PLEK]*
70527	platelet aggregation	33,0E-6	81,0E-6	*[ALOX12, FERMT3, ITGA2B, ITGB3, PLEK]*
10332	response to gamma radiation	12,0E-3	930,0E-6	*[BCL2L1, HSF1, PRKDC]*
2001021	negative regulation of response to DNA damage stimulus	1,6E-3	930,0E-6	*[BCL2L1, CD44, CD74, HSF1]*
1902229	regulation of intrinsic apoptotic signaling pathway in response to DNA damage	5,4E-3	930,0E-6	*[BCL2L1, CD44, CD74]*
1902230	negative regulation of intrinsic apoptotic signaling pathway in response to DNA damage	3,3E-3	930,0E-6	*[BCL2L1, CD44, CD74]*
72395	signal transduction involved in cell cycle checkpoint	4,9E-3	7,4E-3	*[BTG2, PRKDC, RBM38]*
72401	signal transduction involved in DNA integrity checkpoint	7,1E-3	7,4E-3	*[BTG2, PRKDC, RBM38]*
72413	signal transduction involved in mitotic cell cycle checkpoint	8,8E-3	7,4E-3	*[BTG2, PRKDC, RBM38]*
72422	signal transduction involved in DNA damage checkpoint	7,1E-3	7,4E-3	*[BTG2, PRKDC, RBM38]*
1902403	signal transduction involved in mitotic DNA integrity checkpoint	8,8E-3	7,4E-3	*[BTG2, PRKDC, RBM38]*

1902400	intracellular signal transduction involved in G1 DNA damage checkpoint	10,0E-3	7,4E-3	*[BTG2, PRKDC, RBM38]*
1902402	signal transduction involved in mitotic DNA damage checkpoint	8,8E-3	7,4E-3	*[BTG2, PRKDC, RBM38]*
72431	signal transduction involved in mitotic G1 DNA damage checkpoint	10,0E-3	7,4E-3	*[BTG2, PRKDC, RBM38]*

Table 25. Results from the ClueGO overrepresentation analysis for a merged gene list for BMC. The analysis is conducted on a gene list merge of the BMC based on KUL's analysis and of the BMC based on NTNU's analysis The given P values are corrected values, and both are corrected using Bonferroni step down. The terms are sorted with respect to ontology groups. Ontology source: GO_BiologicalProcess-GOA_23.02.2017_10h01. The analysis was performed in Cytoscape using the ClueGO plug-in.

GO ID	GO Term	Term P value	Group P value	Associated genes
33077	T cell differentiation in thymus	7,4E-3	7,4E-3	*[CD74, PRKDC, ZFP36L2]*
46686	response to cadmium ion	23,0E-3	11,0E-3	*[FAM58A, HSF1, MTF1]*
1903313	positive regulation of mRNA metabolic process	28,0E-3	11,0E-3	*[CNOT7, HSF1, ZFP36L2]*
45744	negative regulation of G-protein coupled receptor protein signaling pathway	24,0E-3	3,8E-3	*[ARRB2, ATP2B4, PLEK]*
50848	regulation of calcium-mediated signaling	17,0E-3	3,8E-3	*[ATP2B4, CD4, FKBP1A, PLEK]*
43276	anoikis	13,0E-3	4,3E-3	*[AES, BCL2L1, PDK4]*
2000209	regulation of anoikis	7,2E-3	4,3E-3	*[AES, BCL2L1, PDK4]*
10332	response to gamma radiation	30,0E-3	4,9E-3	*[BCL2L1, HSF1, PRKDC]*
2001021	negative regulation of response to DNA damage stimulus	10,0E-3	4,9E-3	*[BCL2L1, CD44, CD74, HSF1]*
1902229	regulation of intrinsic apoptotic signaling pathway in response to DNA damage	24,0E-3	4,9E-3	*[BCL2L1, CD44, CD74]*
1902230	negative regulation of intrinsic apoptotic signaling pathway in response to DNA damage	13,0E-3	4,9E-3	*[BCL2L1, CD44, CD74]*
2260	lymphocyte homeostasis	7,4E-3	13,0E-3	*[CD74, LGALS9, TSC22D4]*
32722	positive regulation of chemokine production	29,0E-3	13,0E-3	*[CD74, LGALS9, TICAM1]*
46596	regulation of viral entry into host cell	13,0E-3	13,0E-3	*[CD4, CD74, LGALS9]*
46598	positive regulation of viral entry into host cell	570,0E-6	13,0E-3	*[CD4, CD74, LGALS9]*
70228	regulation of lymphocyte apoptotic process	26,0E-3	13,0E-3	*[CD74, LGALS9, TSC22D4]*
32612	interleukin-1 production	13,0E-3	3,6E-3	*[ARRB2, IL1R2, LGALS9]*
32615	interleukin-12 production	26,0E-3	3,6E-3	*[ARRB2, IRF1, LGALS9]*
32652	regulation of interleukin-1 production	29,0E-3	3,6E-3	*[ARRB2, IL1R2, LGALS9]*
32655	regulation of interleukin-12 production	29,0E-3	3,6E-3	*[ARRB2, IRF1, LGALS9]*
1903556	negative regulation of tumor necrosis factor superfamily cytokine production	30,0E-3	3,6E-3	*[ARRB2, HSF1, LGALS9]*

32720	negative regulation of tumor necrosis factor production	28,0E-3	3,6E-3	*[ARRB2, HSF1, LGALS9]*
34121	regulation of toll-like receptor signaling pathway	26,0E-3	3,6E-3	*[ARRB2, IRF1, TICAM1]*
32606	type I interferon production	10,0E-3	520,0E-6	*[IRF1, POLR3H, PRKDC, RNF216, TICAM1]*
32615	interleukin-12 production	26,0E-3	520,0E-6	*[ARRB2, IRF1, LGALS9]*
32479	regulation of type I interferon production	10,0E-3	520,0E-6	*[IRF1, POLR3H, PRKDC, RNF216, TICAM1]*
32655	regulation of interleukin-12 production	29,0E-3	520,0E-6	*[ARRB2, IRF1, LGALS9]*
32481	positive regulation of type I interferon production	12,0E-3	520,0E-6	*[IRF1, POLR3H, PRKDC, TICAM1]*
32608	interferon-beta production	30,0E-3	520,0E-6	*[IRF1, RNF216, TICAM1]*
32648	regulation of interferon-beta production	28,0E-3	520,0E-6	*[IRF1, RNF216, TICAM1]*
42108	positive regulation of cytokine biosynthetic process	18,0E-3	520,0E-6	*[CD4, IRF1, TICAM1]*
34121	regulation of toll-like receptor signaling pathway	26,0E-3	520,0E-6	*[ARRB2, IRF1, TICAM1]*

A.6 BiNGO results

Table 26. BiNGO overrepresentation analysis results for AHC. The analysis is conducted on a gene list merge of the AHC based on KUL's analysis and of the AHC based on NTNU's analysis. The P values are corrected using Benjamini-Hochberg FDR.

GO ID	GO Term	P value	Associated genes
48518	positive regulation of biological process	1.6445E-3	*YWHAE BTG2 PRKDC ITGB3 PLEK ALOX12 AGPAT1 RASGRP4 RTN4 RPS15A HSF1 LGALS9 GPR177 JUNB RBM38 CD74 SMARCC2 TPM1 HMGA1 USF2 FKBP1A IRF1 MTF1 CD44 BCL2L1*
48522	positive regulation of cellular process	2.1619E-3	*YWHAE RBM38 CD74 SMARCC2 PRKDC ITGB3 PLEK TPM1 HMGA1 ALOX12 AGPAT1 USF2 RASGRP4 RTN4 FKBP1A RPS15A IRF1 MTF1 LGALS9 GPR177 JUNB CD44 BCL2L1*
6950	response to stress	2.0238E-2	*RBM38 TSC22D4 CD74 BTG2 PRKDC ITGB3 DEFA4 PLEK TPM1 F13A1 DEFA3 LSP1 MTF1 HSF1 MKNK2 CD44 TNRC6A FERMT3 BCL2L1*
70527	platelet aggregation	2.0238E-2	*PLEK FERMT3*
3229	ventricular cardiac muscle tissue development	2.0238E-2	*FKBP1A TPM1 LY6E*
55010	ventricular cardiac muscle tissue morphogenesis	2.0238E-2	*FKBP1A TPM1 LY6E*
3208	cardiac ventricle morphogenesis	2.0238E-2	*FKBP1A TPM1 LY6E*
1961	positive regulation of cytokine-mediated signaling pathway	2.0238E-2	*CD74 AGPAT1*
42060	wound healing	2.0238E-2	*ITGB3 PLEK TPM1 F13A1 CD44 FERMT3*
35468	positive regulation of signaling pathway	2.0238E-2	*FKBP1A CD74 ITGB3 LGALS9 AGPAT1 GPR177 CD44 RASGRP4*
60415	muscle tissue morphogenesis	2.0238E-2	*FKBP1A TPM1 LY6E*
55008	cardiac muscle tissue morphogenesis	2.0238E-2	*FKBP1A TPM1 LY6E*
48731	system development	2.2173E-2	*YWHAE CD74 BTG2 MBNL1 SMARCC2 PRKDC ITGB3 PLEK TPM1 USF2 RASGRP4 RTN4 NRGN AES FKBP1A IRF1 MTF1 HSF1 JUNB CD44 LY6E BCL2L1*
43518	negative regulation of DNA damage response, signal transduction by p53 class mediator	2.2173E-2	*CD74 CD44*
60255	regulation of macromolecule metabolic process	2.2173E-2	*YWHAE BTG2 PRKDC ITGB3 PHF1 TCF7 SLA ALOX12 ZFP36L2 HSF1 MKNK2 JUNB RBM38 TSC22D4 CD74 MBNL1 SMARCC2 HMGA1 USF2 AES FKBP1A IRF1 MTF1 IRF2 CD44 TNRC6A BCL2L1*
3231	cardiac ventricle development	2.2173E-2	*FKBP1A TPM1 LY6E*
35303	regulation of dephosphorylation	2.2173E-2	*YWHAE FKBP1A PLEK*

48513	organ development	2.2173E-2	*YWHAE CD74 MBNL1 PRKDC ITGB3 PLEK TPM1 USF2 RASGRP4 RTN4 AES FKBP1A IRF1 HSF1 JUNB CD44 LY6E BCL2L1*
10647	positive regulation of cell communication	2.2173E-2	*FKBP1A CD74 ITGB3 LGALS9 AGPAT1 GPR177 CD44 RASGRP4*
60136	embryonic process involved in female pregnancy	2.2385E-2	*HSF1 JUNB*
3206	cardiac chamber morphogenesis	2.2385E-2	*FKBP1A TPM1 LY6E*
50896	response to stimulus	2.2385E-2	*BTG2 PRKDC ITGB3 PLEK TCF7 F13A1 ALOX12 LSP1 RASGRP4 RTN4 RPS15A HSF1 MKNK2 JUNB RBM38 TSC22D4 CD74 DEFA4 TPM1 DEFA3 USF2 AES MTF1 UNC119 CD44 TNRC6A FERMT3 BCL2L1*
51704	multi-organism process	2.2453E-2	*YWHAE RPS15A PACS1 ITGB3 DEFA4 HSF1 HMGA1 DEFA3 PI3 JUNB ATP6V0C*
30097	hemopoiesis	2.3214E-2	*CD74 PRKDC IRF1 PLEK JUNB RASGRP4*
34109	homotypic cell-cell adhesion	2.3850E-2	*PLEK FERMT3*
10604	positive regulation of macromolecule metabolic process	2.4665E-2	*FKBP1A CD74 SMARCC2 PRKDC ITGB3 IRF1 MTF1 HMGA1 ALOX12 JUNB USF2 CD44*
10608	posttranscriptional regulation of gene expression	2.4665E-2	*RBM38 PRKDC MKNK2 SLA ZFP36L2 TNRC6A*
3205	cardiac chamber development	2.5333E-2	*FKBP1A TPM1 LY6E*
19222	regulation of metabolic process	2.5333E-2	*YWHAE BTG2 PRKDC ITGB3 PHF1 PLEK TCF7 SLA ALOX12 ZFP36L2 HSF1 MKNK2 JUNB RBM38 TSC22D4 CD74 MBNL1 SMARCC2 TPM1 HMGA1 USF2 AES FKBP1A IRF1 MTF1 IRF2 CD44 TNRC6A BCL2L1*
31529	ruffle organization	2.5439E-2	*PLEK TPM1*
45767	regulation of anti-apoptosis	2.5439E-2	*BTG2 RTN4 BCL2L1*
80090	regulation of primary metabolic process	2.6914E-2	*YWHAE BTG2 PRKDC ITGB3 PHF1 PLEK TCF7 SLA ZFP36L2 HSF1 MKNK2 JUNB RBM38 TSC22D4 CD74 MBNL1 SMARCC2 TPM1 HMGA1 USF2 AES FKBP1A IRF1 MTF1 IRF2 CD44 TNRC6A*
48534	hemopoietic or lymphoid organ development	2.9431E-2	*CD74 PRKDC IRF1 PLEK JUNB RASGRP4*
6928	cellular component movement	2.9995E-2	*YWHAE VNN2 PRKDC ITGB3 TPM1 ALOX12 LSP1 CD44*
48856	anatomical structure development	2.9995E-2	*YWHAE CD74 BTG2 MBNL1 SMARCC2 PRKDC ITGB3 PLEK TPM1 USF2 RASGRP4 RTN4 NRGN AES FKBP1A IRF1 MTF1 HSF1 JUNB CD44 LY6E BCL2L1*
43516	regulation of DNA damage response, signal transduction by p53 class mediator	2.9995E-2	*CD74 CD44*

7596	blood coagulation	2.9995E-2	*ITGB3 PLEK F13A1 FERMT3*
50817	coagulation	2.9995E-2	*ITGB3 PLEK F13A1 FERMT3*
32502	developmental process	2.9995E-2	*YWHAE BTG2 PRKDC ITGB3 PLEK RASGRP4 RTN4 NRGN HSF1 GPR177 JUNB CD74 MBNL1 SMARCC2 TPM1 USF2 AES FKBP1A LRG1 IRF1 MTF1 CD44 LY6E LIMS1 BCL2L1*
50794	regulation of cellular process	2.9995E-2	*YWHAE BTG2 PRKDC ITGB3 PHF1 PLEK TCF7 SLA ALOX12 AGPAT1 ZFP36L2 RASGRP4 RTN4 NRGN RPS15A SH3BP1 HSF1 MKNK2 LGALS9 GPR177 JUNB RBM38 TSC22D4 CD74 MBNL1 SMARCC2 TPM1 HMGA1 USF2 AES FKBP1A IRF1 MTF1 IRF2 UNC119 CD44 TNRC6A FERMT3 BCL2L1*
48646	anatomical structure formation involved in morphogenesis	2.9995E-2	*FKBP1A PRKDC ITGB3 TPM1 JUNB CD44 RTN4*
9893	positive regulation of metabolic process	2.9995E-2	*FKBP1A CD74 SMARCC2 PRKDC ITGB3 IRF1 MTF1 HMGA1 ALOX12 JUNB USF2 CD44*
2520	immune system development	2.9995E-2	*CD74 PRKDC IRF1 PLEK JUNB RASGRP4*
1775	cell activation	2.9995E-2	*FKBP1A CD74 PRKDC IRF1 PLEK FERMT3*
9967	positive regulation of signal transduction	2.9995E-2	*FKBP1A CD74 LGALS9 GPR177 CD44 RASGRP4*
7599	hemostasis	3.1372E-2	*ITGB3 PLEK F13A1 FERMT3*
23056	positive regulation of signaling process	3.1494E-2	*FKBP1A CD74 LGALS9 GPR177 CD44 RASGRP4*
6458	'de novo' protein folding	3.4132E-2	*FKBP1A CD74*
48523	negative regulation of cellular process	3.4352E-2	*YWHAE RBM38 CD74 BTG2 SMARCC2 ITGB3 PLEK TPM1 HMGA1 ALOX12 RTN4 AES IRF2 HSF1 CD44 TNRC6A BCL2L1*
9628	response to abiotic stimulus	3.8122E-2	*TSC22D4 BTG2 PRKDC HSF1 UNC119 JUNB BCL2L1*
31323	regulation of cellular metabolic process	3.8122E-2	*YWHAE BTG2 PRKDC ITGB3 PHF1 PLEK TCF7 SLA ZFP36L2 HSF1 MKNK2 JUNB RBM38 TSC22D4 CD74 MBNL1 SMARCC2 TPM1 HMGA1 USF2 AES FKBP1A IRF1 MTF1 IRF2 CD44 TNRC6A*
42110	T cell activation	3.8378E-2	*FKBP1A CD74 PRKDC IRF1*
7275	multicellular organismal development	3.9614E-2	*YWHAE CD74 BTG2 MBNL1 SMARCC2 PRKDC ITGB3 PLEK TPM1 USF2 RASGRP4 RTN4 NRGN AES FKBP1A IRF1 MTF1 HSF1 GPR177 JUNB CD44 LY6E BCL2L1*
14706	striated muscle tissue development	3.9614E-2	*FKBP1A MBNL1 TPM1 LY6E*
50832	defense response to fungus	3.9614E-2	*DEFA4 DEFA3*
7229	integrin-mediated signaling pathway	3.9614E-2	*ITGB3 PLEK ITGA2B*
50731	positive regulation of peptidyl-tyrosine phosphorylation	4.0213E-2	*CD74 ITGB3 CD44*

48738	cardiac muscle tissue development	4.0213E-2	*FKBP1A TPM1 LY6E*
48585	negative regulation of response to stimulus	4.1514E-2	*CD74 CD44 RTN4 AES*
2244	hemopoietic progenitor cell differentiation	4.1514E-2	*PRKDC PLEK*
2521	leukocyte differentiation	4.1514E-2	*CD74 PRKDC IRF1 JUNB*
1701	in utero embryonic development	4.2954E-2	*MBNL1 HSF1 JUNB LY6E BCL2L1*
31325	positive regulation of cellular metabolic process	4.2954E-2	*FKBP1A CD74 SMARCC2 PRKDC ITGB3 IRF1 MTF1 HMGA1 JUNB USF2 CD44*
30217	T cell differentiation	4.2954E-2	*CD74 PRKDC IRF1*
32233	positive regulation of actin filament bundle assembly	4.2954E-2	*PLEK TPM1*
60537	muscle tissue development	4.2954E-2	*FKBP1A MBNL1 TPM1 LY6E*
65007	biological regulation	4.3739E-2	*YWHAE BTG2 PRKDC ITGB3 PHF1 PLEK TCF7 F13A1 SLA ALOX12 AGPAT1 ZFP36L2 RASGRP4 RTN4 NRGN RPS15A SH3BP1 HSF1 MKNK2 LGALS9 GPR177 JUNB RBM38 TSC22D4 CD74 MBNL1 SMARCC2 TPM1 HMGA1 USF2 AES FKBP1A IRF1 MTF1 IRF2 UNC119 CD44 TNRC6A LY6E FERMT3 BCL2L1*
32268	regulation of cellular protein metabolic process	4.3739E-2	*YWHAE FKBP1A CD74 ITGB3 MKNK2 SLA CD44 TNRC6A*
10740	positive regulation of intracellular protein kinase cascade	4.4347E-2	*FKBP1A CD74 LGALS9 GPR177 CD44*

Table 27. BiNGO overrepresentation analysis results for BMC. The analysis is conducted on a gene list merge of the BMC based on KUL's analysis and of the BMC based on NTNU's analysis. The P values are corrected using Benjamini-Hochberg FDR.

GO ID	GO Term	P value	Associated genes
70372	regulation of ERK1 and ERK2 cascade	4.1199E-3	*CD74 VEGFB ARRB2 DUSP6 CD44*
48522	positive regulation of cellular process	4.1199E-3	*YWHAE PRKDC PLEK ARRB2 AGPAT1 RTN4 RASGRP4 TIAL1 LGALS9 JUNB CD74 SMARCC2 HMGA1 VEGFB DYRK1B NAP1L1 TICAM1 USF2 DUSP6 FKBP1A CD4 CNOT7 MTF1 IRF1 CD44 BCL2L1 PNPLA2*
48518	positive regulation of biological process	4.1199E-3	*YWHAE PRKDC PLEK ARRB2 AGPAT1 RTN4 RASGRP4 TIAL1 HSF1 LGALS9 JUNB CD74 SMARCC2 HMGA1 VEGFB DYRK1B NAP1L1 TICAM1 USF2 DUSP6 FKBP1A CD4 CNOT7 MTF1 IRF1 CD44 BCL2L1 PNPLA2*
31323	regulation of cellular metabolic process	4.1199E-3	*YWHAE PRKDC PHF1 TCF7 PLEK SLA ARRB2 MED16 ZFP36L2 TIAL1 SALL3 HSF1 MKNK2 PDK4 ZNF746 BRD7 JUNB TSC22D4 CD74 MBNL1 ZGPAT SMARCC2 HMGA1 VEGFB DYRK1B IRF2BP2 TICAM1 USF2 DUSP6 AES FKBP1A CD4 CNOT7 HNRNPUL1 CTDSP1 MTF1 IRF1 CD44 PNPLA2*
19222	regulation of metabolic process	4.1199E-3	*YWHAE PRKDC PHF1 TCF7 PLEK SLA ARRB2 MED16 ZFP36L2 TIAL1 SALL3 HSF1 MKNK2 PDK4 ZNF746 BRD7 JUNB TSC22D4 CD74 MBNL1 ZGPAT SMARCC2 HMGA1 VEGFB DYRK1B IRF2BP2 TICAM1 USF2 DUSP6 AES FKBP1A CD4 CNOT7 HNRNPUL1 CTDSP1 MTF1 IRF1 CD44 BCL2L1 PNPLA2*
1775	cell activation	4.1199E-3	*FKBP1A CD74 CD4 WBP1 PRKDC IMPDH1 IRF1 PLEK TICAM1*
9967	positive regulation of signal transduction	4.1199E-3	*FKBP1A CD74 CD4 VEGFB LGALS9 ARRB2 TICAM1 RASGRP4 CD44*
60255	regulation of macromolecule metabolic process	4.1199E-3	*YWHAE PRKDC PHF1 TCF7 SLA ARRB2 MED16 ZFP36L2 TIAL1 SALL3 HSF1 MKNK2 ZNF746 BRD7 JUNB TSC22D4 CD74 MBNL1 ZGPAT SMARCC2 HMGA1 VEGFB DYRK1B IRF2BP2 TICAM1 USF2 AES FKBP1A CD4 CNOT7 HNRNPUL1 CTDSP1 MTF1 IRF1 CD44 BCL2L1*
23056	positive regulation of signaling process	4.1199E-3	*FKBP1A CD74 CD4 VEGFB LGALS9 ARRB2 TICAM1 RASGRP4 CD44*
70374	positive regulation of ERK1 and ERK2 cascade	4.1199E-3	*CD74 VEGFB ARRB2 CD44*

80090	regulation of primary metabolic process	4.1199E-3	*YWHAE PRKDC PHF1 TCF7 PLEK SLA ARRB2 MED16 ZFP36L2 TIAL1 SALL3 HSF1 MKNK2 ZNF746 BRD7 JUNB TSC22D4 CD74 MBNL1 ZGPAT SMARCC2 HMGA1 VEGFB DYRK1B IRF2BP2 TICAM1 USF2 AES FKBP1A CD4 CNOT7 HNRNPUL1 CTDSP1 MTF1 IRF1 CD44 PNPLA2*
45321	leukocyte activation	5.2736E-3	*FKBP1A CD74 CD4 WBP1 PRKDC IMPDH1 IRF1 TICAM1*
35468	positive regulation of signaling pathway	5.2736E-3	*FKBP1A CD74 CD4 VEGFB LGALS9 ARRB2 TICAM1 AGPAT1 RASGRP4 CD44*
42110	T cell activation	5.2736E-3	*FKBP1A CD74 CD4 WBP1 PRKDC IRF1*
43393	regulation of protein binding	6.6083E-3	*FKBP1A ARRB2 TICAM1 AES*
31325	positive regulation of cellular metabolic process	6.6083E-3	*CD74 SMARCC2 PRKDC HMGA1 VEGFB DYRK1B TICAM1 USF2 FKBP1A CD4 CNOT7 MTF1 IRF1 JUNB CD44 PNPLA2*
46649	lymphocyte activation	6.6083E-3	*FKBP1A CD74 CD4 WBP1 PRKDC IMPDH1 IRF1*
10647	positive regulation of cell communication	7.8333E-3	*FKBP1A CD74 CD4 VEGFB LGALS9 ARRB2 TICAM1 AGPAT1 RASGRP4 CD44*
9893	positive regulation of metabolic process	1.0856E-2	*CD74 SMARCC2 PRKDC HMGA1 VEGFB DYRK1B TICAM1 USF2 FKBP1A CD4 CNOT7 MTF1 IRF1 JUNB CD44 PNPLA2*
10468	regulation of gene expression	1.0994E-2	*PRKDC PHF1 TCF7 SLA ARRB2 MED16 ZFP36L2 TIAL1 SALL3 HSF1 MKNK2 ZNF746 BRD7 JUNB TSC22D4 MBNL1 ZGPAT SMARCC2 HMGA1 VEGFB DYRK1B IRF2BP2 TICAM1 USF2 AES CNOT7 HNRNPUL1 CTDSP1 MTF1 IRF1 BCL2L1*
10604	positive regulation of macromolecule metabolic process	1.3966E-2	*CD74 SMARCC2 PRKDC HMGA1 VEGFB DYRK1B TICAM1 USF2 FKBP1A CD4 CNOT7 MTF1 IRF1 JUNB CD44*
2376	immune system process	1.4447E-2	*CD74 WBP1 PRKDC IL1R2 NCF4 TCF7 PLEK TICAM1 RASGRP4 FKBP1A CD4 IMPDH1 IRF1 POLR3H JUNB*
10740	positive regulation of intracellular protein kinase cascade	1.4590E-2	*FKBP1A CD74 VEGFB LGALS9 ARRB2 TICAM1 CD44*
6357	regulation of transcription from RNA polymerase II promoter	1.4590E-2	*SMARCC2 PRKDC TCF7 MED16 USF2 AES TIAL1 CNOT7 CTDSP1 MTF1 IRF1 BRD7 JUNB*
1961	positive regulation of cytokine-mediated signaling pathway	1.5980E-2	*CD74 AGPAT1*

50731	positive regulation of peptidyl-tyrosine phosphorylation	1.5980E-2	*CD74 CD4 VEGFB CD44*
30097	hemopoiesis	1.6667E-2	*CD74 CD4 PRKDC IRF1 PLEK JUNB RASGRP4*
9966	regulation of signal transduction	1.7396E-2	*CD74 ZGPAT PLEK VEGFB ARRB2 TICAM1 AGPAT1 RASGRP4 DUSP6 FKBP1A CD4 LGALS9 CD44*
23051	regulation of signaling process	1.7866E-2	*CD74 ZGPAT PLEK VEGFB ARRB2 TICAM1 AGPAT1 RASGRP4 DUSP6 FKBP1A CD4 LGALS9 CD44*
30217	T cell differentiation	1.7900E-2	*CD74 CD4 PRKDC IRF1*
43518	negative regulation of DNA damage response, signal transduction by p53 class mediator	1.9279E-2	*CD74 CD44*
10627	regulation of intracellular protein kinase cascade	1.9279E-2	*FKBP1A CD74 VEGFB LGALS9 ARRB2 TICAM1 DUSP6 CD44*
60136	embryonic process involved in female pregnancy	2.4907E-2	*HSF1 JUNB*
48534	hemopoietic or lymphoid organ development	2.4907E-2	*CD74 CD4 PRKDC IRF1 PLEK JUNB RASGRP4*
31326	regulation of cellular biosynthetic process	2.5949E-2	*PRKDC PHF1 TCF7 PLEK SLA ARRB2 MED16 TIAL1 SALL3 HSF1 MKNK2 PDK4 ZNF746 BRD7 JUNB TSC22D4 ZGPAT SMARCC2 HMGA1 DYRK1B IRF2BP2 TICAM1 USF2 AES CD4 CNOT7 HNRNPUL1 CTDSP1 MTF1 IRF1*
48585	negative regulation of response to stimulus	2.5949E-2	*CD74 ARRB2 RTN4 CD44 AES*
2521	leukocyte differentiation	2.5949E-2	*CD74 CD4 PRKDC IRF1 JUNB*
35303	regulation of dephosphorylation	2.5949E-2	*YWHAE FKBP1A PLEK*
48513	organ development	2.5949E-2	*YWHAE CD74 MBNL1 PRKDC PLEK VEGFB USF2 RTN4 RASGRP4 AES FKBP1A CD4 SALL3 IRF1 NINJ1 HSF1 NPHP3 JUNB CD44 LY6E BCL2L1*
9889	regulation of biosynthetic process	2.6824E-2	*PRKDC PHF1 TCF7 PLEK SLA ARRB2 MED16 TIAL1 SALL3 HSF1 MKNK2 PDK4 ZNF746 BRD7 JUNB TSC22D4 ZGPAT SMARCC2 HMGA1 DYRK1B IRF2BP2 TICAM1 USF2 AES CD4 CNOT7 HNRNPUL1 CTDSP1 MTF1 IRF1*
2520	immune system development	2.9088E-2	*CD74 CD4 PRKDC IRF1 PLEK JUNB RASGRP4*

50730	regulation of peptidyl-tyrosine phosphorylation	2.9088E-2	*CD74 CD4 VEGFB CD44*
31399	regulation of protein modification process	3.2407E-2	*YWHAE FKBP1A CD74 CD4 VEGFB ARRB2 TICAM1 CD44*
32268	regulation of cellular protein metabolic process	3.4952E-2	*YWHAE FKBP1A CD74 CD4 MKNK2 VEGFB SLA ARRB2 TICAM1 CD44*
51252	regulation of RNA metabolic process	3.5150E-2	*TSC22D4 MBNL1 ZGPAT SMARCC2 PRKDC TCF7 HMGA1 DYRK1B MED16 USF2 ZFP36L2 AES TIAL1 CNOT7 CTDSP1 MTF1 IRF1 HSF1 ZNF746 BRD7 JUNB*
43410	positive regulation of MAPKKK cascade	3.5150E-2	*CD74 VEGFB ARRB2 CD44*
10557	positive regulation of macromolecule biosynthetic process	3.9190E-2	*CD4 SMARCC2 CNOT7 PRKDC MTF1 IRF1 HMGA1 DYRK1B TICAM1 JUNB USF2*
19058	viral infectious cycle	3.9907E-2	*CD4 HMGA1 USF2*
50794	regulation of cellular process	3.9907E-2	*YWHAE CLIC3 PHF1 PLEK SLA ARRB2 MED16 NRGN TIAL1 SALL3 PDK4 LGALS9 JUNB TSC22D4 MBNL1 ZGPAT SMARCC2 DYRK1B TICAM1 DUSP6 AES HNRNPUL1 CTDSP1 MTF1 IRF1 CD44 PRKDC TCF7 AGPAT1 RTN4 ZFP36L2 RASGRP4 SH3BP1 HSF1 MKNK2 ZNF746 NPHP3 BRD7 CD74 HMGA1 VEGFB IRF2BP2 NAP1L1 USF2 FKBP1A CD4 CNOT7 BCL2L1 PNPLA2*
10556	regulation of macromolecule biosynthetic process	4.2622E-2	*PRKDC PHF1 TCF7 SLA ARRB2 MED16 TIAL1 SALL3 HSF1 MKNK2 ZNF746 BRD7 JUNB TSC22D4 ZGPAT SMARCC2 HMGA1 DYRK1B IRF2BP2 TICAM1 USF2 AES CD4 CNOT7 HNRNPUL1 CTDSP1 MTF1 IRF1*
43516	regulation of DNA damage response, signal transduction by p53 class mediator	4.2770E-2	*CD74 CD44*
50896	response to stimulus	4.3178E-2	*ZFP106 WBP1 PRKDC NCF4 TCF7 PLEK F13A1 ARRB2 LSP1 RTN4 RASGRP4 TIAL1 HSF1 MKNK2 UCP2 JUNB TSC22D4 CD74 IL1R2 DEFA3 TICAM1 USF2 DUSP6 AES C7ORF27 CD4 HNRNPUL1 IMPDH1 MTF1 NINJ1 POLR3H CD44 BCL2L1*
34645	cellular macromolecule biosynthetic process	4.4898E-2	*TOP3B WBP1 HMGA1 VEGFB SLA NAP1L1 ARRB2 MED16 USF2 ABTB1 CD4 IRF1 POLR3H JUNB*

31401	positive regulation of protein modification process	4.5884E-2	*FKBP1A CD74 CD4 VEGFB TICAM1 CD44*
6268	DNA unwinding involved in replication	4.5884E-2	*TOP3B HMGA1*
768	syncytium formation by plasma membrane fusion	4.5884E-2	*DYRK1B CD44*
45893	positive regulation of transcription, DNA-dependent	4.5884E-2	*SMARCC2 CNOT7 PRKDC MTF1 IRF1 HMGA1 DYRK1B JUNB USF2*
30098	lymphocyte differentiation	4.5884E-2	*CD74 CD4 PRKDC IRF1*
80134	regulation of response to stress	4.6308E-2	*CD74 PLEK VEGFB ARRB2 TICAM1 RTN4 CD44*
43408	regulation of MAPKKK cascade	4.6308E-2	*CD74 VEGFB ARRB2 DUSP6 CD44*
51254	positive regulation of RNA metabolic process	4.6308E-2	*SMARCC2 CNOT7 PRKDC MTF1 IRF1 HMGA1 DYRK1B JUNB USF2*
9059	macromolecule biosynthetic process	4.6308E-2	*TOP3B WBP1 HMGA1 VEGFB SLA NAP1L1 ARRB2 MED16 USF2 ABTB1 CD4 IRF1 POLR3H JUNB*
44249	cellular biosynthetic process	4.6527E-2	*TOP3B CD74 WBP1 HMGA1 VEGFB ATP2B4 SLA NAP1L1 ARRB2 MED16 AGPAT1 USF2 ABTB1 CD4 IMPDH1 IRF1 CMPK1 POLR3H JUNB*
31328	positive regulation of cellular biosynthetic process	4.6527E-2	*CD4 SMARCC2 CNOT7 PRKDC MTF1 IRF1 HMGA1 DYRK1B TICAM1 JUNB USF2*
6458	'de novo' protein folding	4.6527E-2	*FKBP1A CD74*
32091	negative regulation of protein binding	4.6527E-2	*ARRB2 AES*