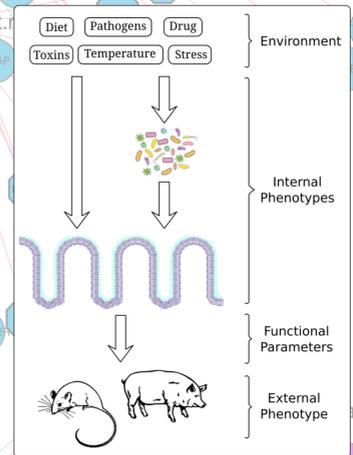
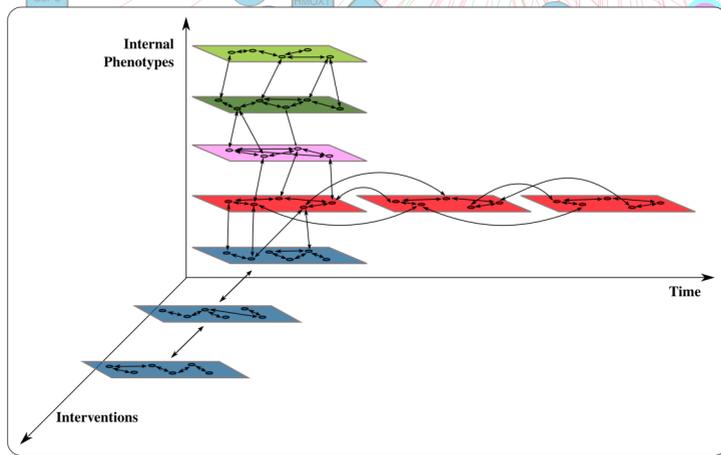
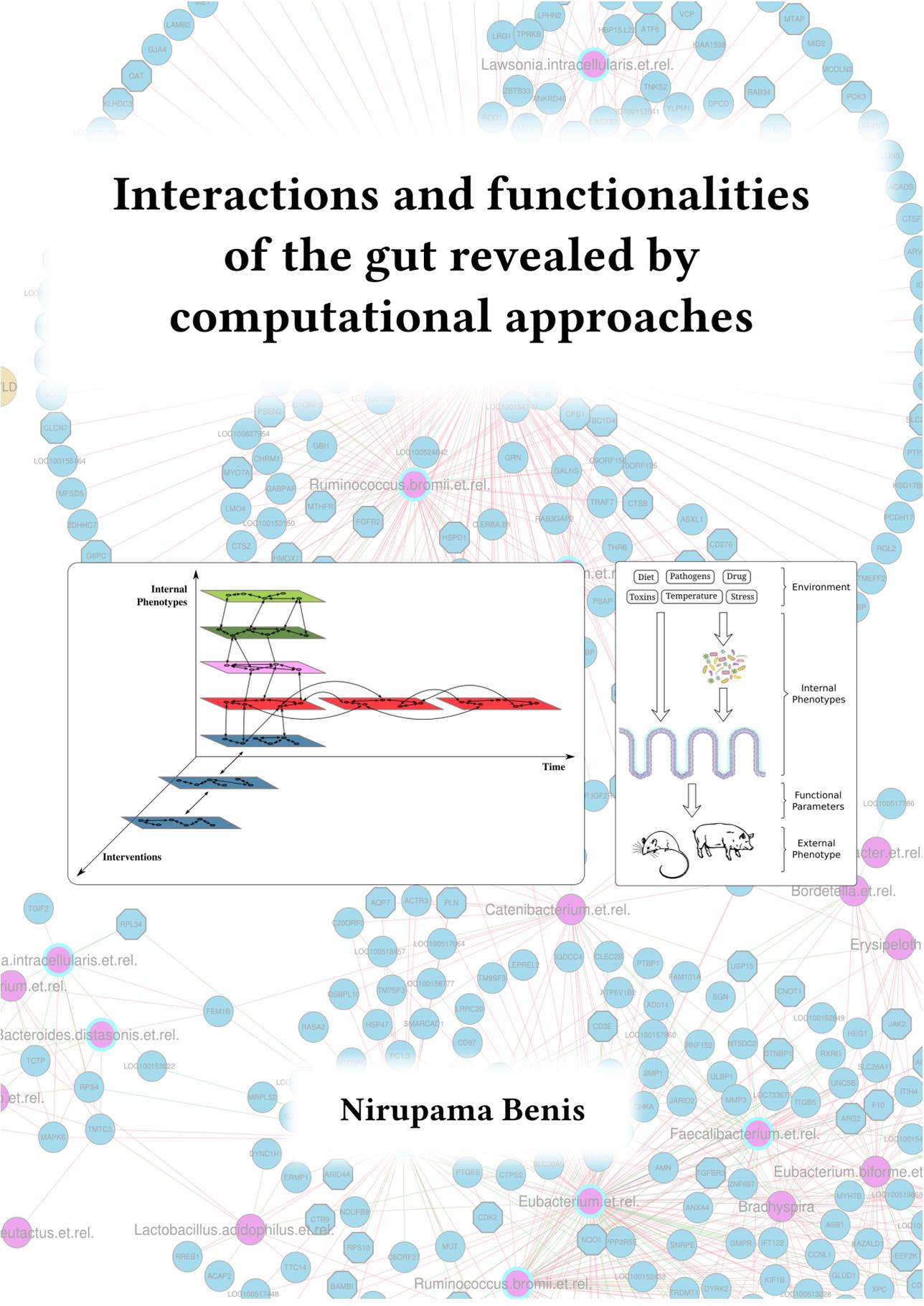


# Interactions and functionalities of the gut revealed by computational approaches



## Nirupama Benis





**Interactions and functionalities of the  
gut revealed by  
computational approaches**

**Nirupama Benis**

**Thesis committee****Promotors**

Prof. Dr M.A. Smits

Special professor, Intestinal Health of Animals  
Wageningen University & Research

Prof. Dr V.A.P. Martins dos Santos

Professor of Systems and Synthetic Biology  
Wageningen University & Research

**Co-promotors**

Dr D. Schokker

Scientist, Animal Breeding & Genomics  
Wageningen University & Research

Dr M Suarez-Diez

Assistant Professor, Systems and Synthetic Biology  
Wageningen University & Research

**Other members**

Dr G. Hooiveld, Wageningen University & Research

Prof. Dr T.A. Niewold, University of Leuven, Belgium

Prof. Ir D. de Ridder, Wageningen University & Research

Dr M. Roos, Leiden University Medical Center

This research was conducted under the auspices of the Graduate School Wageningen Institute of Animal Sciences.

# **Interactions and functionalities of the gut revealed by computational approaches**

**Nirupama Benis**

## **Thesis**

submitted in fulfilment of the requirements for the degree of doctor at Wageningen

University by the authority of the Rector Magnificus,

Prof. Dr A.P.J. Mol,

in the presence of the

Thesis Committee appointed by the Academic Board

to be defended in public

on Tuesday 4 July 2017

at 11 a.m. in the Aula.

Nirupama Benis

Interactions and functionalities of the gut revealed by computational approaches

A4, 194 pages.

PhD thesis, Wageningen University, Wageningen, the Netherlands (2017)

With references, with summary in English

ISBN 978-94-6343-454-6

DOI <http://dx.doi.org/10.18174/416470>

# TABLE OF CONTENTS

<b>CHAPTER 1</b> .....	<b>7</b>
GENERAL INTRODUCTION	
<b>CHAPTER 2</b> .....	<b>27</b>
NETWORK ANALYSIS OF TEMPORAL FUNCTIONALITIES OF THE GUT INDUCED BY PERTURBATIONS IN NEW-BORN PIGLETS.	
<b>CHAPTER 3</b> .....	<b>59</b>
MULTI-LEVEL INTEGRATION OF ENVIRONMENTALLY PERTURBED INTERNAL PHENOTYPES REVEALS KEY POINTS OF CONNECTIVITY BETWEEN THEM	
<b>CHAPTER 4</b> .....	<b>89</b>
BUILDING PATHWAY GRAPHS FROM BIOPAX DATA IN R	
<b>CHAPTER 5</b> .....	<b>101</b>
FUNCTIONAL PLASTICITY OF MURINE INTESTINAL MUCOSAL TISSUE.	
<b>CHAPTER 6</b> .....	<b>149</b>
MODELLING THE IMPACT OF DIET ON IMMUNITY PARAMETERS USING HIGH-THROUGHPUT DATA	
<b>CHAPTER 7</b> .....	<b>177</b>
GENERAL DISCUSSION	
<b>SUMMARY</b> .....	<b>197</b>
<b>REFERENCES</b> .....	<b>205</b>
<b>ACKNOWLEDGEMENTS</b> .....	<b>235</b>
<b>TRAINING AND SUPERVISION PLAN</b> .....	<b>241</b>
<b>LIST OF PUBLICATIONS</b> .....	<b>245</b>



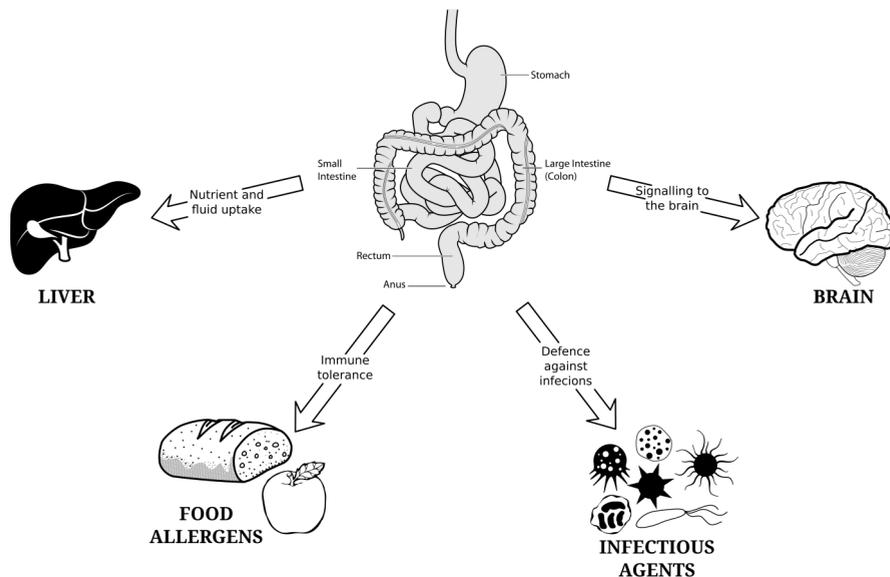
# **Chapter 1**

## **General Introduction**



## The pleiotropic functions of the gastrointestinal tract

Health of an organism, whether human or animal, is difficult to define, although it is the subject of extensive research activities all over the world. Many of these studies focus on restoring health, including cures for diseases. Other research is directed towards the development of products and methods to prevent diseases, e.g. the development of diagnostics and vaccines. More recent studies are focussing on promoting and maintaining health e.g. with nutrition and change of lifestyle.



**Figure 1. Functionalities of the gastrointestinal tract (GIT).** The GIT is depicted with its various functionalities. The gut is involved in digestion and absorption of food or feed nutrients. The gut is also involved in immune functions like tolerance of antigens in food and defending the organism against possible pathogens. There is also extensive signalling from the gut to other organs like the liver which receives absorbed nutrients and the brain which receives signals from the nervous system. (Image adapted from Bischoff (Bischoff 2011))

Intestinal health significantly contributes to the overall health of an organism whether human (Bischoff 2011; J H Cummings et al. 2004) or animal (Kogut and Arsenault 2016). The gastro-intestinal tract (GIT) is considered to be healthy when it is able to efficiently digest and absorb components of feed or food, when the intestinal microbiota composition

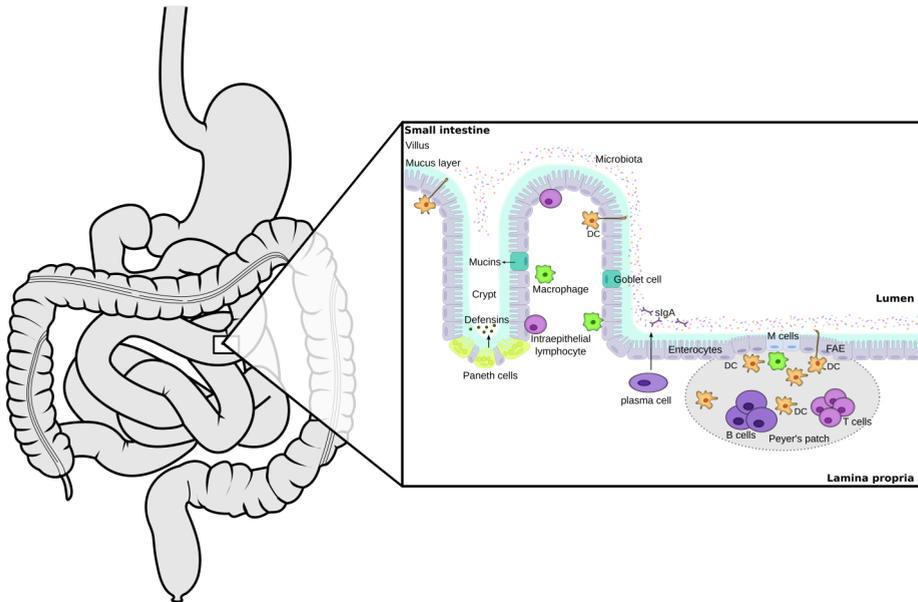
and diversity is “in balance”, when pathogenic micro-organisms are absent or cannot colonize and/or expand, and when the gut mucosal immune system is capable of efficiently neutralizing potential disease causing stressors (see Figure 1). In addition to this, the GIT has several signalling functions towards other parts of the body, which also contribute to the health of the organism.

The GIT, which starts from the mouth, continues into the stomach and intestine and ends at the anus, transports, digests and absorbs food/feed components before the undigested waste is expelled through the anus. During transport, food/feed components (carbohydrates, fat, proteins) are digested by enzymes, bile and by the fermenting action of microbes present in the lumen of the gut. Dietary components are broken up into nutrients that are absorbed by the epithelial cells of the intestine and subsequently transported to different parts of the body via the bloodstream. The presence of specific nutritional components is signalled by the gut towards the brain via the gut-brain axis, and the brain then regulates hormonal levels for digestion in the gut, for example insulin levels in the blood are modulated when glucose is absorbed by the gut (Tsurugizawa et al. 2009; Berthoud 2008).

The epithelial layer of the GIT is a single layer of cells that forms the luminal surface. Besides its role in digestion, the epithelial layer of the GIT has an important barrier function to defend the body against the outside world. It protects the body from the contents of the lumen. The gut-associated immune system, which is most abundant in the small intestine, constantly monitors the content of the gut for toxic components, food/feed based antigens and antigens produced by commensal and (opportunistic) pathogenic organisms. The immune system is able to recognize several antigens that could be part of the food/feed and to induce immunological tolerance (Pabst and Mowat 2012; Castro-Sánchez and Martín-Villa 2013). Immune tolerance also plays a role in dampening the immune response to commensal bacteria that reside in the gut and perform several vital functions. The mucosal immune system is able to discriminate between commensal microbes and pathogens. Discrimination occurs by pattern recognition receptors on and within epithelial cells that can recognize specific pathogen-associated molecular patterns. In case antigens of (potential) pathogens are recognized, local and systemic immune responses are induced to defend the organism from disease development (Artis 2008).

Thus, the GIT exhibits a large variety of functions that are essential for the performance and health of the organism. Improved understanding of the mechanisms underlying the pleiotropic functionalities of the gut will allow the development of (nutritional) interventions to modulate intestinal functionalities, thereby contributing to human and animal health.

### Morphology of the GIT tract



**Figure 2: Gut interface with the environment.** The image is an illustration of a section of the small intestine showing the different cell types at the epithelial surface of the tissue. The epithelial layer is covered by a mucus layer that separates the cells from microbes in the lumen. A Peyer's patch is also shown with several immune cells. Adapted from (Rossi, van Baarlen, and Wells 2011)

There are several distinct sections of the GIT: the stomach; the small intestine; the large intestine; and the caecum. Each section has a different function for digestion and immune actions. In all the sections the luminal surface area is increased with projections of the epithelial layer that forms villi (projections) and crypts (invaginations). The small intestine can be further divided into the duodenum, jejunum and the ileum. These sub sections have different morphologies but the small intestine in general has a high villus to crypt ratio and a higher concentration of mucosal immune cells than other sections of the gut.

The various sections of the intestine have common cellular structures that play a role in the functioning of the GIT as a whole. The different cell types present in the gut mucosa allow it to perform a multitude of processes. The different cell types are represented in a cartoon of the mucosal layer in Figure 2. Enterocytes are the major cell type of the intestinal epithelium and contribute significantly to nutrient absorption. Goblet cells also form part of the single epithelial layer and secrete mucus, a thick layer of glycosylated proteins that acts as the first line of defence against the content of the gut lumen (Peterson and Artis 2014). Paneth cells and M-cells are epithelial cells specialized in immune functions. The mucosal layer of the intestinal tract contains a large number of different types of immune cells that are part of the innate and/or the adaptive immune system (Mowat and Agace 2014). Scattered along the intestine are numerous lymphoid tissue structures, called gut associated lymphoid tissues with specialized immune structures; the Peyer's patch(es) in the small intestine and the lymphoid follicles all over the intestine. These structures are involved in the regulation of the immune response of the gut.

### **Microbiota**

A large number of microbial communities reside in the lumen of the gut. The number of microbial cells is comparable to the number of cells in an organism (Sender, Fuchs, and Milo 2016). These microbes have evolved with the host to form a mutually beneficial relationship (O'Hara and Shanahan 2006; Scholtens et al. 2012). Bacteria are most abundant micro-organisms of the mammalian microbiota, 2-3 times the number of eukaryotes and archaea (Sender, Fuchs, and Milo 2016) and are the subject of most microbiota research. There is a large diversity (number and distribution of unique bacterial taxonomic groups) in the species of bacteria found in the gut. Yet a majority of the gut bacterial groups belong to a few dominant phyla, Bacteroidetes and Firmicutes are the largest contributors to gut microbes, followed by Actinobacteria, Proteobacteria and Verrucomicrobia (Lozupone et al. 2012).

The number and concentration of the bacteria increases along the GIT with lower numbers in the small intestine than in the colon. The duodenum and jejunum are estimated to have about  $10^7$  bacterial cells. The ileum is estimated to have about  $10^{11}$  bacteria, whereas the largest number of bacteria is observed in the colon,  $10^{14}$  (Sender, Fuchs, and Milo 2016).

The composition of gut microbiota is dynamic and changes with age as the morphology and physiology of the gut undergoes changes with the growth of the organism (Lozupone et al. 2012; Arrieta et al. 2014; Odamaki et al. 2016). Several studies have shown that the microbial populations are less steady and less diverse in early life and that they stabilize in adulthood (Scholtens et al. 2012; Ottman et al. 2012). Although it is known that there is an active cross-talk between microbes in the lumen of the gut and cells in the intestinal mucosa, not much is currently known about the microbial factors involved and the biological processes they modulate in the gut system.

Microbiota contribute to the breakdown of food/feed components such that they can be absorbed by the gut tissue into the bloodstream. The colon is a major site of fermentation of enzymatically indigestible feed/food components due to presence of the large bacterial population in this anaerobic environment. Several bacterial groups are known to ferment polysaccharides that the mammalian digestive system is unable to break down (Tremaroli and Backhed 2012; Macfarlane and Macfarlane 1997) and ferment undigested protein. The main products of this fermentation are Short Chain Fatty Acids (SCFA) (Haenen, Silva, et al. 2013; den Besten et al. 2013; Topping and Clifton 2001; Macfarlane and Macfarlane 1997). SCFA is readily absorbed by intestinal mucosal tissues and forms the major source of energy for this tissue, they are also known to contribute to intestinal health (den Besten et al. 2013). Microbiota also contributes to vitamin metabolism as bacterial enzymes are essential for the process, and bacteria have been observed to produce vitamin K (Sherwood L 1996). Additionally, immune functions of the gut mucosal tissues are modulated by commensal microbes in the lumen of the gut, for instance, by stimulating the production of antimicrobial proteins and increasing the production of mucosal immunoglobulins and intra-epithelial lymphocytes (Kabat, Srinivasan, and Maloy 2014; Arrieta and Finlay 2012; Ottman et al. 2012).

### **Perturbation of gut homeostasis**

With the aid of specialized cells and by communicating with luminal microbiota and with other organs of the body, the gut maintains homeostasis, an equilibrium state essential for proper functioning of the gut (J H Cummings et al. 2004). This state of homeostasis is predicated on a balance between the host cells in the mucosa of the GIT, the microbiota

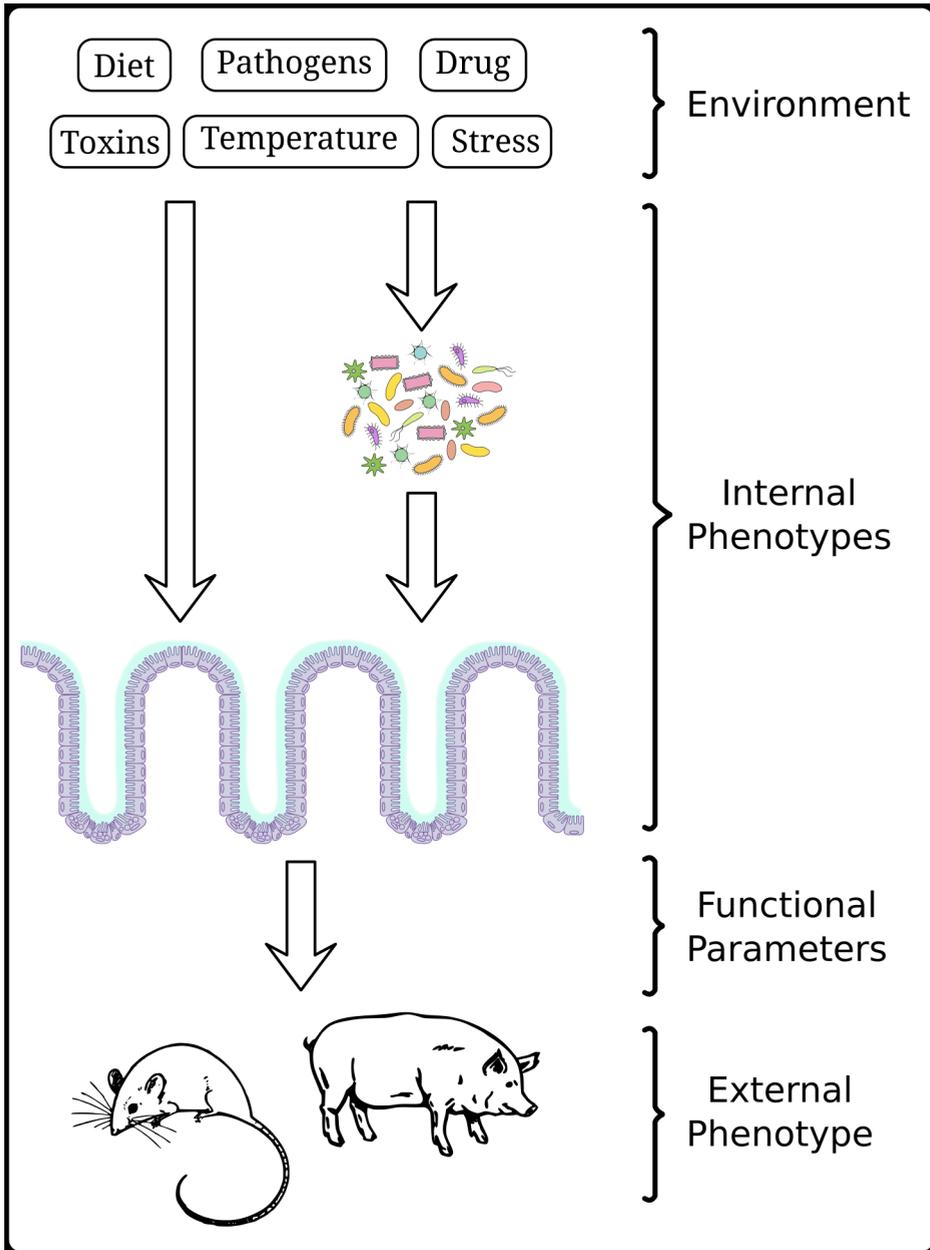
1

and the external environment. There are several different ways this homeostasis can be disturbed, for example by changes in the microbial composition of the gut (through the use of antibiotics) or stimulation of the immune system by dietary components, drugs, disease causing organisms or other changes in the environment. Many different studies have been published on changes in gut tissues and in resident microbiota after dietary interventions. In many cases, the effect of changes in the diet is not always restricted to local effects in the intestine but also induces changes at the systemic level e.g. changes in metabolite and cytokine levels in serum (Kar et al. 2016). Extreme changes in diets may lead to microbial dysbiosis, which is an acute change in the composition and diversity of the microbiota causing a change in microbial functionality that eventually harms the host. Severe changes in the microbiota due to administration of antibiotics, especially early in life, can have impacts on the host immune system (Olszak et al. 2012; Nylund et al. 2014; Schokker et al. 2014; Schokker et al. 2015). An infection with microbial pathogens is another extreme change that activates the immune system while also reshaping the structure of the microbial community. Other environmental changes like a stressful environment, e.g. maternal separation, and changes in temperature are also known to cause changes in both intestinal cells and microbiota. Perturbations of homeostasis in the gut are manifested in changes in the microbial population, changes in signalling to other organs, changes in the expression of immunity parameters, and modified metabolic capacity.

Several different environmental factors can affect the functionalities of gut mucosal tissues either directly or through the microbiota, as depicted in Figure 3. For example, the administration of antibiotics is expected to affect functions of the gut by changing the microbial composition and diversity that results in a change in microbiome-host crosstalk (Willing, Russell, and Finlay 2011). As shown in Figure 3, changes in environmental factors may initially lead to changes in so-called internal phenotypes and ultimately in a change in the expression of external phenotypes, for example weight or production traits of livestock species. Internal phenotypes encompass profiles of several measurable molecular/cellular components at different biological levels (Fontanesi 2016; Houle, Govindaraju, and Omholt 2010). A few examples of internal phenotypes are the transcriptome: the transcripts (mRNA) that are expressed in a specific cell or tissue that are encoded by the host genome, the proteome: all the host encoded proteins that are expressed in a cell/tissue or present in a particular sample (blood), the metabolome: all the metabolites

present in a specific cell or tissue or body fluid sample produced either by the organism or by its associated microbiome.

While a disruption in the normal functioning of the gut can be observed and measured (especially in a pathological state), it is not yet known how to measure the homeostatic status of the gut. There are several measurable parameters that are commonly used in clinical diagnosis as an indicator of a pathological state. In principle, this concept could also be used to measure the competence of an organisms to defend itself against developing disease (immune competence), which is only one aspect of health. Although these parameters have not been used in this context, it is likely that measurements of immune competence would include immunity parameters that are functional related to immunity-based external phenotypes. Normal levels of immunity parameters like Immunoglobulin A levels and lymphocyte counts have been determined in healthy adults (J H Cummings et al. 2004; Moss and Drayson 2015) and their role in the gut has been extensively studied. It is logical to assume that the levels of such immunity parameters result from biological processes (internal phenotypes) of the mucosal tissue of the gut in interaction with several environmental factors, including dietary components. However, relationships between internal phenotypes of the gut mucosal tissue with local and/or systemic immunity parameters, have not been defined yet.



**Figure 3: Effects of environmental factors on an organism mediated by the GIT.** This schema shows how environmental factors can affect the gut system, either directly or through the microbiota. These effects can be measured by studying changes in the internal phenotypes or in biomarkers that are indicative of changes in external phenotypes. (Depiction of the intestine taken from (Rossi, van Baarlen, and Wells 2011))

## **Functional plasticity of gut mucosal tissue**

Due to its various functionalities, the tissue of the GIT is able to respond to many different stressors in an appropriate way. For example, the gut responds to damage to the epithelial layer with cell proliferation (IJssennagger et al. 2012); when pathogens are recognised, immunity parameters are highly expressed (Rossi, van Baarlen, and Wells 2011); to prevent the invasion of foreign substances the barrier function is improved (Wells et al. 2017); when microbes are recognised too close to the epithelial layer the production of antimicrobials is increased (Donaldson, Lee, and Mazmanian 2016). All these changes in functionalities requires a high level of functional plasticity. This plasticity of the gut is a phenomenon that is poorly understood. Simultaneous evaluation of internal phenotypes of GIT mucosal tissues, exposed to different stressors that evoke different types of responses, could provide first insights into the way the gut displays its functional plasticity. Unfortunately, most studies focus on the effect of single stressors and there is limited information available for multiple stressors in the same laboratory conditions.

## **Measurement of internal phenotypes**

Several studies have measured the response of one or more internal phenotypes of the gut in response to an external challenge. There is a vast amount of data available from *in vivo* experiments on model organisms, like mice and pigs. In most *in vivo* experiments that study the gut, the organism is given a stimulus; a change in the environment that is expected to alter the normal functions of the intestine; and parts of the intestine are sectioned and analysed with different technologies to obtain either quantitative or qualitative data.

With transcriptomic approaches, the transcriptome of the organism (collective RNA), in a particular tissue, at a specific time point. With this data, it is possible to define which genes are expressed in that tissue and which proteins are most likely to be found in that tissue. In some studies, the expression of only a few key genes are analysed using variations of polymerase chain reaction (a technology that can amplify a certain sequence of nucleotides) to understand processes known to be modulated by these genes. Instead of concentrating only on a few genes, genome wide transcriptomics gives an overview of the response of all the genes to the challenge. An efficient and cheap method of genome-wide gene expression measurements is the use of microarrays that provides a profile of the level

of expression of all the genes encoded by the genome of the host (Sanden, Lin, and Burzykowski 2008).

1 The next level of biological measurement is to measure the translation of transcripts to proteins by methods, collectively called, proteomics. Even though the transcriptome gives an indication of the proteins that can be expected to be present in a tissue, measuring the actual protein contents is more accurate for functional characterization, as protein levels are more closely related to cellular functions than transcript levels. There are different ways of profiling the amount of all the proteins in a sample; several techniques utilizing mass spectrometry methods like liquid chromatography- mass spectrometry (LC-MS). Such MS based methods can identify a variety of proteins by comparing the mass spectrometry results to a database of known proteins. Protein microarrays use immunoassay techniques or functional proteins to bind to different proteins in the sample. These arrays are useful for identifying a specific group of biomarkers in the samples.

Measuring metabolomics profiles of tissues and samples provides information on the metabolites that the organism is able to produce by itself or with the help of its associated microbiome. Technologies used for this measurement (e.g. mass spectrometry, high performance liquid-phase chromatography) can give quantitative values of known metabolites when a targeted search is performed, by comparing the obtained biochemical signals to signals of known substances. The quantifiable metabolites are fewer than the metabolites that can be produced by the host or the microbes. Hence it is sometimes preferable to perform an untargeted search where unknown metabolites are also recorded.

The composition and diversity of intestinal microbiota can be measured either from the faeces, which reflects the microbial composition in the colon, or from the luminal content in sections of the intestine. Microarrays for bacteria are designed to detect bacteria with known 16S rDNA sequences and are tailored to the known composition in an organism, e.g. HITChip tailored for the human intestinal tract (Rajilić-Stojanović et al. 2009) and PITChip tailored for the porcine intestinal tract (Pérez Gutiérrez et al. 2010). Another popular measurement is to sequence certain regions of the 16S rDNA of the all the bacterial content and then identify the present bacteria by comparing measurements with sequences in databases (Ju and Zhang 2015).

**Set-up of intervention experiments**

Animal experiments generally have a group of animals that are subject to the intervention or challenge. Usually, there are a comparable number of animals that are under the same conditions as the experimental animals except for the challenge under consideration. This group of animals is called the control or reference group. Data from the experimental animals is compared to the data from the control animals to obtain a differential value for that biological measurement. All the data described in the previous section can be subjected to this differential analysis to understand the effects of the challenges at the corresponding biological levels. This approach enables quantification of how an experimental condition affects the system or organism. For example, differential gene expression of certain genes indicates that the intervention affects the organism by regulation of the expression of those particular genes. Differential analysis requires the appropriate statistical analysis methods to take into account variability associated to small sample size, biological noise, intrinsic variability and additional external factors that could potentially bias or influence the comparison, the so-called blocking factors, such as age or sex.

**Analysis of individual data**

The application of high throughput –omics measurements provide an unprecedented wealth of data. The interpretation and analysis of these large datasets requires the use of computational tools to extract relevant biological knowledge. The majority of “-omics” data are/were generated to provide answers to research questions specific to the measured data type, for example, certain proteomic measurements are aimed at finding post-translational modifications in proteins. But these different types of data can also be used to gauge the effects of the challenge on the system. However, effects of challenges do not only display changes at a single biological level, but could affect several, hence multiple levels need to be measured to understand the effects of the challenges.

The most common data analysis method relies on fitting linear models to each data-point to account for the contribution of the blocking factors. This approach is widely used in gene expression analysis as it is implemented in the R package linear models for microarray data (LIMMA) (Smyth 2005; Phipson et al. 2016), which has become a *de facto* standard. In the case of gene expression, linear modelling assumes that the changes in gene

1

expression, e.g. due to a drug, are linear to the gene expression in the control samples. Genes that fit this pattern of change are deemed to be relevant to the given challenge before an estimation of significance. Linear model fitting a technique that is quite flexible and can be extended to other types of biological data besides gene expression. However, since the responses of biological systems are rarely linear, several statistical methods are often used to augment the results of differential expression or abundance and to account for possible non linearities, the methods use generalized linear modelling, such as the one implemented in edgeR (McCarthy, Chen, and Smyth 2012).

Most of these analyses only concentrate on measurements performed at one time point after the intervention. Studying more time points will give an improved understanding of the dynamics of the response. However, given the disruptive characteristics of most of the measurements, that often require the sacrifice of the host to collect samples, such time course analysis requires a large number of subjects and presents challenges of its own regarding the experimental design. Similarly, spatial location plays a critical role, since at each section of the intestine the stages of digestion differ.

Although individual internal phenotypes of the gut have been studied extensively, it is not known if or how they contribute to the establishment of external phenotypes. There is theoretical knowledge on how these internal phenotypes are connected to each other, for example, the amount of a metabolite present in the tissue is dependent on the level of expression of the enzyme that catalyses the biochemical reaction that produces the metabolite, and this protein level depends on the number of transcripts from the gene encoding this protein. So far there has not been data-driven confirmation of the existence of links between internal phenotypes.

### **Databases and knowledgebases**

There are several databases where biological measurements can be/are stored for independent confirmation of the results and further reuse. NCBI and EBI have several databases for specific types of biological data, genomic sequences, protein sequences, genome wide association studies (GWAS) results etc. The most structured databases are those that store transcriptomics data, Gene Expression Omnibus (GEO) (Barrett et al. 2013) from NCBI and Array Express (Brazma et al. 2003) from EBI. But there are also repositories

that can handle more than one type of data like Zenodo (<http://zenodo.org/>) and Dataverse (Crosas 2011). All these databases have basic search functions to find a dataset based on a unique identifier, keywords in the experimental description of the data, specific technological platforms etc. Some of these databases also have integrated tools to perform basic data analysis functions, for example, GEO has an R user interface that allows simple differential analysis. There are a few databases that integrate all the submitted data to allow users to query the results of basic analysis on the datasets rather than the raw data itself. One such database is the database of Genotypes and Phenotypes (dbGaP) (Tryka et al. 2014) from NCBI where data from GWAS studies can be submitted. A search tool in this database, the Phenotype Genotype Integrator (Ramos et al. 2014), retrieves gene to phenotype associations based on a user-given threshold.

A knowledge base is a specific type of database that contains information that can be used in analysis of different types of data. In gene set knowledge-bases genes are grouped together based on annotation of the genes. Gene Ontology (Ashburner et al. 2000) has several genes classified under functional terms that could be biological processes, molecular functions or cellular locations. Other knowledge bases not only group genes but also add information on the interactions between these genes to form pathways that perform a particular function. Such knowledge-bases, like Reactome (Fabregat et al. 2016) and Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa and Goto 2000), use literature findings to build the pathways. This information is very useful in understanding the possible functionalities of significant genes obtained from a differential analysis. The GreenGenes (DeSantis et al. 2006) knowledge-base has information on mapping 16S rRNA sequences of bacteria to taxonomic units. As previously mentioned, this type of information is crucial in identifying the bacterial groups present in gut microbiota samples.

### **Data integration methods**

In most studies, differential analysis of multiple measurements is performed separately, e.g. on several time points. For differential expression or abundance of the internal phenotypes and the results of each of the time points are interpreted in the context of the results of the other measurements. In some cases, the time points themselves are used for differ-

1

ential analysis, for example the gene expression at time point one is compared to the expression at time point two. These methods analyse changes caused by either the challenge or by the passage of time and do not give information on consistent changes in expression or abundance profiles. However, this type of information can be obtained by integrating different datasets (from different time points or different challenges) and analysing them together. Such approaches are frequently called horizontal or lateral integration to distinguish it from vertical integration (Tseng, Ghosh, and Feingold 2012; Xia et al. 2013). Vertical integration is the case where different data types corresponding to the same time point and challenge are analysed together. So far there are not many methods available to perform this integration over time or challenges.

When different types of data need to be analysed, several studies use individual analysis of each data type. For example, the effects of the use of novel protein sources in animal diets on the small intestinal tissue, on the microbiota of the small intestine and on cytokines in the blood are analysed separately. The results of these individual analyses are then interpreted in the context of each other (Kar et al. 2016). But this method does not reveal interactions between the different internal phenotypes that are measured. Integration of the measurements of different biological levels is a preferred method to gain a holistic view of the GIT system. Most computational methods currently available, can only handle one type of data as measured on one time point. Thus, there is a need for methods that can use existing single data analysis tools for integrative analysis. Many integrative tools are built to integrate specific types of data, e.g. the tool Conexic (Akavia et al. 2010) has been developed to find driver mutations in cancer with transcriptomic and genomic information. Other tools can work on any type of data, e.g. MixOmics (Dejean et al. 2011; González et al. 2012), Kernel Fusion (De Bie et al. 2007) and Similarity Fusion Network (B. Wang et al. 2014). The flexibility that such general tools provide, allows to use the variety of biological data that is necessary to provide an integrated picture of complex systems. For example, nutritional data could be integrated with transcriptomic data to pinpoint the parts of the diet that cause transcriptomic changes. Such tools use different approaches to integrate input data, but the first step usually involves pre-processing of each data on its own. Kernel Fusion uses kernel matrices of the input omics data and Similarity Fusion Network requires a similarity network from each of the data. The tool MixOmics uses sparse partial least squares regression to pre-select variables which is a good solution

to the dimensionality problem in biological measurements, where the number of variables (e.g. genes) is larger than the number of samples.

### **Systems biology (of the gut)**

The dynamic functioning of the GIT and its plasticity results from interactions between several sub-systems like the digestive system, the immune system, endocrine system and the microbiome. In order to better understand the mechanisms behind the functional plasticity, a holistic approach is required, so that the different sub-systems and their relationships are analysed together.

This holistic approach is one of the hallmarks of Systems Biology, a field of study where a living system is studied as several sub-systems together rather than one sub-system at a time (Kitano 2002). Systems Biology aims to understand biological systems as a sum of all its sub-systems especially by including the interactions between them. Given the complexity of the GIT, its spatial heterogeneity and its functional plasticity, a Systems Biology approach is required to gain an improved understanding on how intestinal homeostasis is maintained and using which underlying mechanisms. Synergies and interactions between the components of this system determine the response of the system to external stressors. Thus, it is essential to gain a systems perspective of the functioning and cross-talk between these components across space and over time.

Another hallmark of Systems Biology is the construction of models that are able to account for the response of the system. The models can be built bottom-up or top-down (Bruggeman et al. 2007). Bottom-up modelling approaches involve mechanistic models explaining the observed behaviour from the knowledge of the system's components and their relationships. The large volume of available high-throughput data, enables top-down approaches, combining system-wide data with mathematical modelling to gain insights into the molecular networks under study. Some examples of existing models built using a top-down approach are a metabolic model of enterocytes (Sahoo and Thiele 2013), a model of metabolic interactions between gut microbes (Steinway et al. 2015) and a model of mucosal immune responses (Mei et al. 2015). These models use information from literature and experimental data.

Besides representing the system, models can be used to predict reactions of the system to environments it has not been exposed. The model only contains the parts of the system that are relevant to the question at hand. For example, statistical models can be built using a large amount of data of the varying parameters and the predictors, like different types of feed and the changes in weight of animals, and can make accurate predictions on weight without information on the mechanisms causing the change. Models in systems biology are built from different but complementary data. For example, genome scale metabolic models are based on biochemical interaction information, genomic data and can be further constrained by experimental results (Orth, Thiele, and Palsson 2010).

Although there are models describing or predicting functionalities in the gut, there is no single model that can represent the entire gut. This is largely due to the complexity of the gut with the different cell types performing different functions, the lack of a complete understanding of the role of the microbiota and the complex dynamic changes in the gut when exposed to a changing environment. Thus, before a multiscale (cellular level and molecular level) model that can represent or predict all the functionalities of the gut to a variety of challenges (a virtual gut) can be built, several gaps in knowledge have to be filled with data acquisition and data analysis. The response of the different sub-systems of the gut to challenges must be catalogued, interactions between these sub-systems must be studied and mechanistic information on all these processes must be obtained.

### **Research objectives and thesis outline**

The objective of the research, as described in this thesis, was to i) develop and apply computational methods to get improved insight into time-dependent dynamics of a gut system, ii) to obtain information on key links between internal phenotypic or different biological levels, iii) to understand the functional plasticity of the gut mucosal tissue by studying effects of dietary / drug / immune system targeted interventions, and finally, iv) to develop a concept of a model that allows to make links between internal phenotypes and immunity parameters functionally related to external immune-based phenotypes.

We aimed to uncover the effect of an early life perturbation and growth on animals and in **Chapter 2** we analysed data from an antibiotic intervention on pigs which was administered early in life and disrupted the homeostatic status of the microbiota. We perform a

dynamic analysis in order to find if there were long lasting changes in the animals due to this early life perturbation and to find the drivers of these changes. We used a two-dimensional method that can analyse data between experimental treatments and across time simultaneously. This analysis was applied on two internal phenotypes, transcriptomics and microbiota, which were analysed along two dimensions and the results were integrated to uncover the interactions between them.

The second objective was to get improved insights into the interactions between internal phenotypic layers when stimulated by external factors. In **Chapter 3**, we observe the response of another model animal, mice, to changes in the diet using five measurements of different internal phenotypes. This lateral data integration was done across different physiological levels to find interactions between internal phenotypes in the gut and those in systemic levels. We used a flexible multi-omics integration method to link all the five measurements to each other.

In **Chapter 5** we attempted to identify mechanisms behind the functional plasticity of the gut using the same type of data from several different experiments. Here we used a high-level data integration approach where we analysed each transcriptomic dataset separately and then combined the results. This horizontal data integration showcased the gut mucosal response towards different environmental challenges. The method developed in **Chapter 4** enables this analysis on a pathway level. By observing changes in the functionality of the gut mucosa under different experimental conditions involving dietary changes, drug administration and immune challenges, we aimed to identify mechanisms of functional plasticity of the gut.

The fourth objective was to find associations between dietary changes and parameters that are functionally related to external immune-based phenotypes. In **Chapter 6** we used the internal phenotype transcriptomics to link dietary changes to a few functional parameters that are related to the immune system. We used the high-level data integration technique similar to that used in **Chapter 5** to analyse and integrate the transcriptomic data. The results obtained were then linked to functional parameters with information from a knowledge-base.

**Chapter 7** discusses the result of **Chapters 2-6** in a broader context. It describes relevance of the work done in my thesis for improving scientific knowledge and its potential applications in the furthering insights into gut functionalities. I also discuss the limitations of this work and suggestions for general data management to improve the results of analyses as performed in my thesis. There are several indications for future work based on the results of my work.

# Chapter 2

## Network analysis of temporal functionalities of the gut induced by perturbations in new-born piglets.

**Nirupama Benis**; Host Microbe Interactomics, Wageningen University.

**Dirkjan Schokker**; Wageningen UR Livestock Research, Wageningen University.

**Maria Suarez-Diez**; Systems and Synthetic biology, Wageningen University.

**Vitor AP Martins dos Santos**; Systems and Synthetic biology, Wageningen University; Lifeglimmer GmbH, Berlin, Germany.

**Hauke Smidt**; Laboratory of Microbiology, Wageningen University.

**Mari A Smits**; Host Microbe Interactomics, Wageningen University; Wageningen UR Livestock Research, Wageningen University, Central Veterinary Institute, Wageningen University.

**Published in**

BMC Genomics 2015 16:556



## **Abstract**

### **Background**

Evidence is accumulating that perturbation of early life microbial colonization of the gut induces long-lasting adverse health effects in individuals. Understanding the mechanisms behind these effects will facilitate modulation of intestinal health. The objective of this study was to identify biological processes involved in these long-lasting effects and the (molecular) factors that regulate them. We used an antibiotic and the same antibiotic in combination with stress on piglets as an early life perturbation. Then we used host gene expression data from the gut (jejunum) tissue and community-scale analysis of gut microbiota from the same location of the gut, at three different time-points to gauge the reaction to the perturbation. We analysed the data by a new combination of existing tools. First, we analysed the data in two dimensions, treatment and time, with quadratic regression analysis. Then we applied network-based data integration approaches to find correlations between host gene expression and the resident microbial species.

### **Results**

The use of a new combination of data analysis tools allowed us to identify significant long-lasting differences in jejunal gene expression patterns resulting from the early life perturbations. In addition, we were able to identify potential key gene regulators (hubs) for these long-lasting effects. Furthermore, data integration also showed that there are a handful of bacterial groups that were associated with temporal changes in gene expression.

### **Conclusion**

The applied systems-biology approach allowed us to take the first steps in unravelling biological processes involved in long lasting effects in the gut due to early life perturbations. The observed data are consistent with the hypothesis that these long lasting effects are due to differences in the programming of the gut immune system as induced by the temporary early life changes in the composition and/or diversity of microbiota in the gut.

**Keywords:**

Gene expression, microbiota, data-integration, long-term effects, early life perturbations, antibiotic, stress, pig intestine

## 1. Background

Evidence is accumulating that perturbations of the early life colonization of the gastrointestinal (GI) tract by microbes induce long-lasting health effects in individuals (Foliaki et al. 2009; Hoskin-Parr et al. 2013). Though these effects have been studied and documented, the system components involved in the induction and maintenance of such long-lasting effects have not yet been studied in detail. This is because the GI tract itself is a complex and dynamic system with variable interactions between the host tissue, resident microbiota and nutritional factors. The host tissue is comprised of different cells with different functions, varying from the digestion and uptake of nutrients to providing resistance towards toxic components in the diet. The mucosal layer of the GI tract separates the lumen of the digestive tract from the rest of the body and contains the largest repertoire of immune cells that display a pleiotropy of immune signalling and response functions. In addition, the mucosal layer also generates endocrine responses to the lumen of the gut and to the rest of the body. The lumen of the GI tract harbours a complex ecosystem of a huge number and a large variety of micro-organisms, collectively called “microbiota” (Relman 2012).

Gut microbiota play an important role in modulating diverse gastrointestinal functions, ranging from enzymatic digestion to modulation of immune responses (Lora V. Hooper 2004; O’Hara and Shanahan 2006; Tremaroli and Backhed 2012; Arrieta and Finlay 2012; Kabat, Srinivasan, and Maloy 2014; Scholtens et al. 2012; Ottman et al. 2012). In turn, the host immune system has a regulatory effect on the composition and diversity of the microbiota (Willing, Gill, and Finlay 2010, by mechanisms known as immune exclusion (Mantis, Rol, and Corthesy 2011). Early life colonization of the gut with microbiota is an important driver for the development and ultimate functionality of the GI tract in mammals (Palmer et al. 2007; Inman et al. 2010; Kabat, Srinivasan, and Maloy 2014). Long-lasting effects on the host due to disruption of early life colonisation of the gut have been demonstrated on the level of disease susceptibility, immune parameters (Olszak et al. 2012) and the composition and diversity of microbiota (Nylund et al. 2014). Early life environmental factors, such as caesarean delivery (C. E. Cho and Norman 2013), breastfeeding (Weng and Walker 2013; D. Kelly and Coutts 2000; Newburg and Walker 2007), exposure to stress (Jarillo-Luna et al. 2007; O’Mahony et al. 2009), and the use of antibiotics

(Schumann et al. 2005; Tanaka et al. 2009 influence the microbial colonisation of the gut and affect the development and programming of the mucosal and systemic immune system (Nylund et al. 2014; Arrieta et al. 2014. Such early life factors may also result in variation of the abilities of the microbiota to ferment carbohydrates into short-chain fatty acids (I. Cho et al. 2012 and/or to ferment indigestible proteins in later stages of life.

The effect of the use of antibiotics on the physiology of the host is believed to be due to the primary effect of antibiotics on the loss/change in bacterial (sub)-populations, especially in the GI tract (Willing, Russell, and Finlay 2011. The spectrum of the antibiotic and its mode of action determines the effects it has on the gut microbial community (Willing, Russell, and Finlay 2011. Some reports in this area also suggests that antibiotics have a direct effect on the immune system of the host (Willing, Russell, and Finlay 2011; Niewold 2007. For young piglets it has recently been demonstrated that an antibiotic treatment at day 4 after birth causes changes in microbial populations and in tissue gene expression patterns in the small intestine (Schokker et al. 2014; Schokker et al. 2015. Stress is another factor that can influence the functionality of the GI tract by a temporary secretion of hormones. Such short term hormonal secretions may cause several long lasting effects on the GI immune system (Jacobson and Sapolsky 1991; E A Mayer 2000; Braun et al. 2013 and on microbiota composition (Collins and Bercik 2009. This has led to studies focusing on the brain-gut-enteric microbiota axis (Konturek, Brzozowski, and Konturek 2011; Montiel-Castro et al. 2013; Emeran A Mayer, Savidge, and Shulman 2014. The effect of stress on the GI tissue is most obvious in changes of morphology and functions of the gut (Söderholm and Perdue 2001; Bhatia and Tandon 2005; Taché et al. 2001.

The first objective of this study was to develop a workflow that can be used to analyse biological data in two dimensions simultaneously, over time and between treatments. The second objective was to apply this workflow on two types of gut-related data as measured in an experiment with pigs exposed to an early life perturbation, followed by effect measurements at three different time-points later in life. Using these methods we endeavour to identify gut system components that contribute to induction and maintenance of long lasting effects of early life perturbations. We aim to find major host- and microbe-related components that propagate or regulate these long term effects. Such components may

form potential targets to modulate early life events that affect later life immunologic performance. We used piglets as a model and the exposure to an antibiotic and/or stress on day four after birth as the perturbations. We measured whole genome gene expression profiles of intestinal tissues and provided community scale data on the composition and diversity of microbiota on three different time-points, taken during different stages of the life-span of pigs, neonate, adolescent and full grown adult. In order to take into account the extreme changes in the physiology and morphology of the animal, we simultaneously studied the effect of the treatment and the time. Once the gene expression and microbiota data were analysed separately, we integrated both to obtain information on their possible interactions.

## **2. Materials and methods**

### **2.1. Experimental design**

The animal experiment consisted of one control group and two treatment groups, each consisting of 48 piglets derived from 16 different sows (TOPIGS20 (GY x NL)). Each litter contained 4 piglets of each treatment and control group. Litter-mates stayed with their sow until weaning at day 25. After weaning, the same number of piglets of each treatment and control group were housed together in pens. All pens were located in the same compartment. The first treatment group (Tr1) was given a dose of Tulathromycin (0.1 ml, 2.5mg/kg body weight) on the 4<sup>th</sup> day after birth and then left undisturbed until the time of tissue sampling after sacrifice. The same dose of the antibiotic was given to the second treatment group (Tr2) on the 4<sup>th</sup> day after birth but these piglets were also subjected to stressful conditions on the same day. The stress was in the form of handling of the piglets, which is common practice in intensive pig husbandry systems (eg, weighing, nail clipping). The control group (Ctrl) was not disturbed for the entirety of the experiment until the time of sacrifice for sampling.

Sampling was done on three time-points, day 8 after they were born, day 55 and day 176. On each of these days, 16 piglets from each treatment group and derived from 16 different sows were sacrificed and samples were collected for microarray and microbiota analysis. Further details of the experiment have been described elsewhere (Schokker et al. 2014). The experiment was approved by the institutional animal experiment committee “Dier

Experimenten Commissie (DEC) Lelystad” (2011077.b), in accordance with the Dutch regulations on animal experiments. Supplementary Figure 1 gives an overview of the experimental design.

## 2.2. Sample preparation and data generation

For transcriptome analysis, jejunal tissues scrapings were taken and RNA was extracted from the samples for microarray analysis as described in (Schokker et al. 2014). For microbiota analysis luminal contents were taken from the same location of the jejunum as the tissue scrapings and microbial DNA was extracted. The microbial composition was detected by the Pig Intestinal Tract Chip (PITChip) (Pérez Gutiérrez et al. 2010 version 2.0). The PITChip is a phylogenetic microarray with 3,299 oligonucleotides based on 16S rRNA gene sequences of 781 porcine intestinal microbial phylotypes (Haenen, Zhang, et al. 2013; Pérez Gutiérrez 2010). The protocol for hybridization and generation of data was performed essentially as described before (Rajilić-Stojanović et al. 2009; Schokker et al. 2014). More details on sample preparation is given in previous descriptions of this experiment (Schokker et al. 2014 and the data is available in GEO (Barrett et al. 2013 with the accession number GSE53170 (Schokker 2014).

## 2.3. Gene expression analysis

### Microarray normalization and quality control:

Background correction and quantile normalisation was performed on the microarray data (GSE53170) using the R package LIMMA (Smyth 2005 from Bioconductor (Gentleman et al. 2005). After quality control, six microarray samples were removed from the original 72 samples and all further analysis were done on the remaining 66 microarray samples. Data points below the 5<sup>th</sup> percentile of intensities were removed, this resulted in 25,915 genes from 66 microarray samples. Details of the applied analysis pipeline are in Supplementary Figure 2.

## Identification of differences in gene expression profiles

To identify genes with significant expression profile differences over time between the experimental groups, we used maSigPro (Conesa et al. 2006 from Bioconductor). The time profile (expression over the three time-points) from each treatment is compared to that of the control. Quadratic regression is used to retrieve genes with time profiles significantly different for the treatments versus the control group. The calculations were done with the default settings of the function maSigPro. With three time-points there are three regression coefficients calculated for each gene's expression profile. At least one of these coefficients has to be significant (with FDR corrected p-values less than 0.05) for the gene to be included in the result. The first coefficient ( $\beta_1$ ) denotes the difference in the first time point (day 8) between one treatment group and the control. The second one ( $\beta_2$ ) indicates the difference in slope between the first two timepoints. The third coefficient ( $\beta_3$ ) shows difference in curvature of the expression patterns and can thus capture long term differences between treatment and control expression patterns of the same gene. Each treatment is compared against the control group and this gives two lists, Tr1vsCtrl (Antibiotic vs Control) and Tr2vsCtrl (Antibiotic+Stress vs Control). For the rest of the analysis three lists of genes were used; genes unique to Tr1vsCtrl (**OnlyTr1**); gene unique to Tr2vsCtrl (**OnlyTr2**) and the overlapping genes (**Tr1&Tr2**).

GO (Gene Ontology) enrichment analysis (Dopazo 2006, with focus on the sub-ontology Biological Process, was performed using the R package topGO (Alexa and Rahnenführer 2009). The Fisher test was applied to obtain significantly enriched GO terms, only the terms with p-values below 0.01 were included.

## Functional interaction networks

Functional interaction networks among groups of genes selected from the data, were built using the Cytoscape (Shannon et al. 2003 Reactome FI (Functional Interaction) application (Croft 2013; Milacic et al. 2012; G. Wu, Feng, and Stein 2010 and visualised with the organic layout. The network was built from information in the Reactome FI database such that nodes are genes and edges are interactions, either known or predicted. Modules in the networks were identified using community structure detection as described in (Newman 2006). Using inbuilt functions in the Reactome FI app, GO enrichment analysis was performed on

each of the identified modules. The final networks consisted of modules with more than 5 nodes and with significantly enriched GO terms. The topological features of these final networks were determined using the NetworkAnalyzer (Assenov et al. 2008 tool in Cytoscape). Hubs were defined to be the nodes in the top 40% of the degree distribution, where the degree is the number of connections of each node. These hubs were analysed using information from [www.genecards.org](http://www.genecards.org).

## 2.4. Analysis of Microbiota

Microbial populations from the jejunum were analysed using the PITChip.2 (Pérez Gutiérrez et al. 2010, which provides information on three levels of taxonomic resolution, level 1 is similar to the phylogenetic family, level 2 corresponds mostly to the genus level and level 3 is more the species level. Level 2 data provided population percentages for 151 microbial genus groups. Initial exploratory analysis was done on the microbial data with the R package microbiome (<http://microbiome.github.com>) and the MySQL database as described by Rajilic-Stojanovic (Rajiliæ-Stojanoviæ et al. 2009). Microbial groups that do not differ between control and treatment piglets were filtered out with a threshold of 0.01. A two-way ANOVA analysis of the temporal patterns with the package maSigPro was used to obtain only the groups that change with treatment or time, with a statistical significance threshold of 0.05.

## 2.5. Integration of Gene expression and Microbiota data

The mixOmics R package (Lê Cao, González, and Déjean 2009; Dejean et al. 2011) was used to integrate gene expression and microbiota abundance data and to perform regression analysis of these two data types. The genes used were those selected using the method described in section 2.3, which leads to the identification of genes whose temporal expression profiles were significantly different between the control vs. the treatment group. The bacterial groups used for integration were also different either across time or between treatments. We set the microbial information as the independent variable and the gene expression data as the dependent variable. Interchanging the two set of parameters did not affect the results significantly. We only used the significant genes from the gene expression analysis as input for the data integration. Subsequently three networks were built using sparse Partial Least Squares for the each of the gene lists **OnlyTr1**, **OnlyTr2**

and **Tr1&Tr2**. The first two networks were based on six variables (treatment and control in three time-points) and the third one on nine variables (two treatments and the control in three time-points).

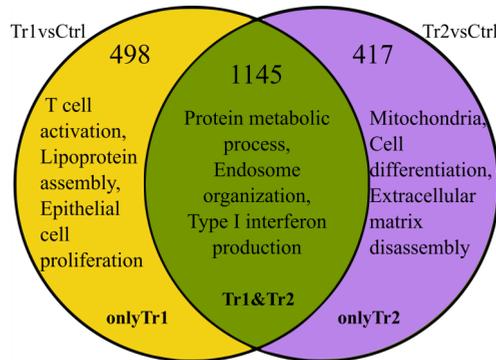
The resulting un-directed network connects the microbial groups and the genes that have absolute correlation values above 0.8. A positive high edge weight hints to a positive regulatory relationship while a low (negative) weight indicates a possible repression relationship. Networks were visualised using Cytoscape 3.1.0. Enrichment analysis was done on the gene neighbours of the bacterial groups via topGO. Fisher's test was applied on the results and only terms with p-values lesser than 0.01 were further analysed. The functionality of the bacterial groups was determined with the help of experts in the field.

### 3. Results

#### 3.1. Genes, gene networks and hubs

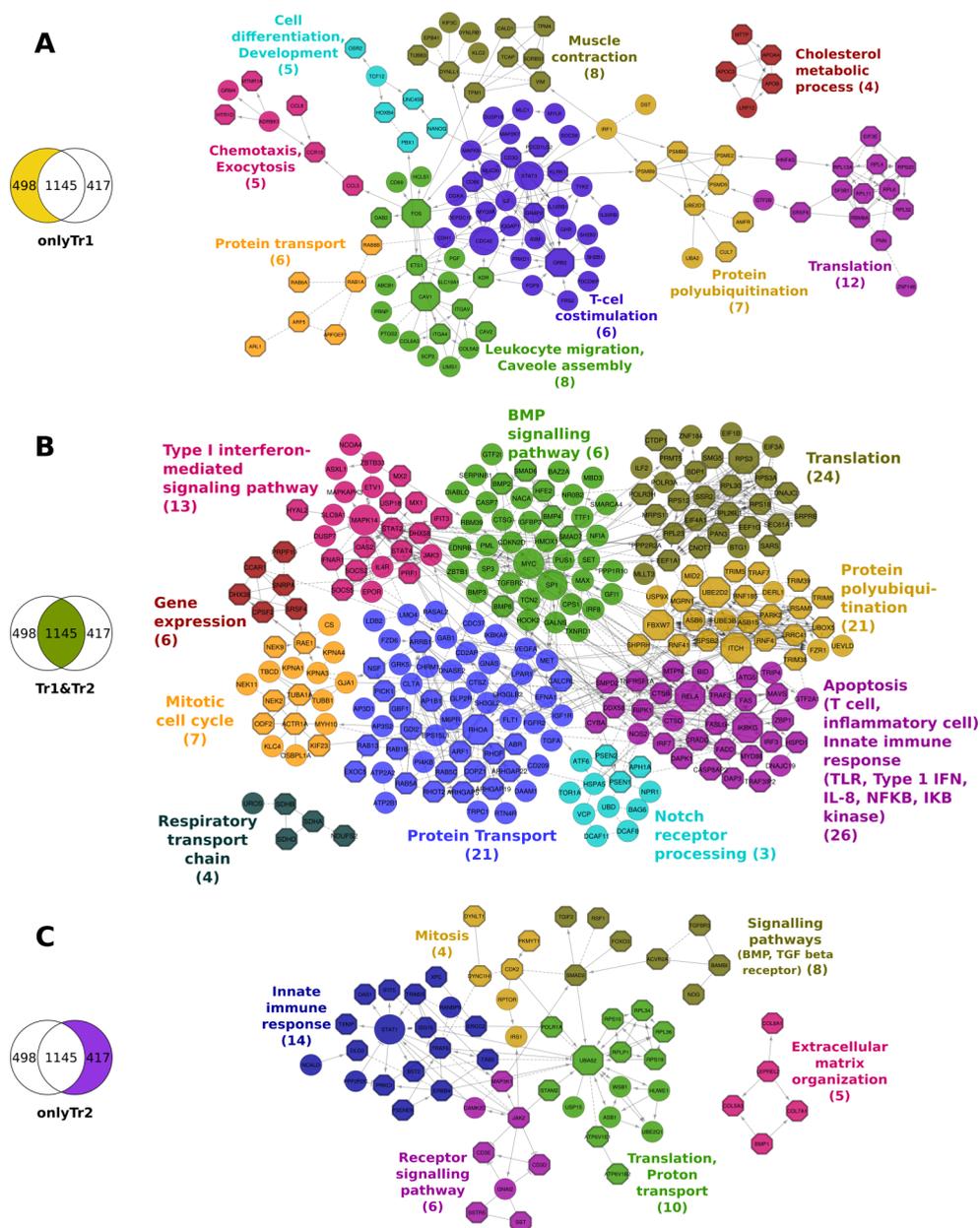
For gene expression analysis, the temporal profiles of the samples from each treatment group were compared with those of the control group. We used quadratic regression analysis to obtain two gene lists: one for Tr1vsCtrl (1,643 genes) and another one for Tr2vsCtrl (1,562 genes). These lists are in Supplementary File 1. The genes in these lists show significant differences in their gene expression profiles over time between the control and treatment groups. Supplementary Figure 3 shows the temporal expression profiles of some genes, showcasing different possible behaviours under different conditions. For example, the expression profiles over time of CHIT1 differs for all three conditions while the profiles for some other genes, such as GRB2 and MAPK14, show most difference in only one treatment compared to the Ctrl. In other cases, such as ERBB4, MX2 and RELA, significant differences in expression are restricted to a single time point. Based on the  $\hat{\alpha}$  coefficients (explained in the Materials and methods section 2.3), 60% of the **OnlyTr1** and **OnlyTr2** gene-lists have genes with long lasting differences between the treatment and control. The **Tr1&Tr2** list has 91% of the genes that have significant long term differences. The three gene lists; **OnlyTr1**, **Tr1&Tr2** and **OnlyTr2** consist of genes that have gene expression profiles over time that are significantly different in the treatment compared to the control group. These lists are used for the follow-up of the analyses. On these three lists,

functional analysis was done using GO enrichment analysis for biological processes with topGO. A summary of the results is presented in Figure 1.



**Figure 1: Summary of GO Enrichment analysis results from topGO.** Biological processes (Gene Ontology terms) are given based on manual interpretation of the most significant enriched terms obtained with topGO. The two circles represent the Tr1vsCtrl and Tr2vsCtrl comparisons. Numbers denote the number of input genes in topGO, these genes are have significantly different time profiles in the treatment vs the control groups. In the yellow, green and purple fields enriched processes are given for OnlyTr1, Tr1&Tr2, and OnlyTr2, respectively.

In order to verify the results of the topGO analysis and to get insight into networks of genes potentially involved in the induction and maintenance of the long-lasting effects, a network based analysis was used. The Reactome Functional Interaction (FI) database was used to build three functional interaction networks, one for each list of genes. In any of the explored gene sets, about 30% of the genes were represented in the FI networks. More than 50% of the nodes had significant long-term differences from the control, with the Tr1&Tr2 FI network having the biggest fraction (94%). Within each of the FI networks, we were able to identify several topological modules. We identified 9 modules with significant biological function in the **OnlyTr1** network, 10 modules in the **Tr1&Tr2** network, and 6 modules in the **OnlyTr2** network. The modules are depicted by different colours in Figure 2. Subsequently GO enrichment analysis for Biological Processes was performed on the genes in each of the identified modules. The GO terms with the highest enrichment score for each module are shown in Figure 2. The results of GO enrichment by both methods for the same gene lists are notably different especially in the **OnlyTr2** list. Information on the various topological parameters of the three networks can be found in the Supplementary File 2.



**Figure 2: Functional Interaction networks (A, B, C).** Genes are represented as nodes in the networks, all these genes have time profiles that are significantly different in the treatment than in the control. The edges represent interactions between genes as determined by Reactome. Arrows represent directed interactions, bar-headed arrows indicate inhibition reactions. Dotted lines indicate predicted relationships. Network A was built from **OnlyTr1** genes, network B from the common or overlapping genes (**Tr1&Tr2**) and network C from **OnlyTr2** genes. Colours in the network repre-

sent the network segmentation into modules. The text denotes the GO term that was most enriched for the genes in that module and the number in brackets denotes the number of genes associated with that particular GO term. Octagonal nodes are related to the GO term at the set p-value threshold. The nodes with a larger diameter are hubs in the networks. The nodes of the **Tr1&Tr2** network were rearranged for better visualisation of the modules; the network in the original structure is in **Supplementary Figure 4**.

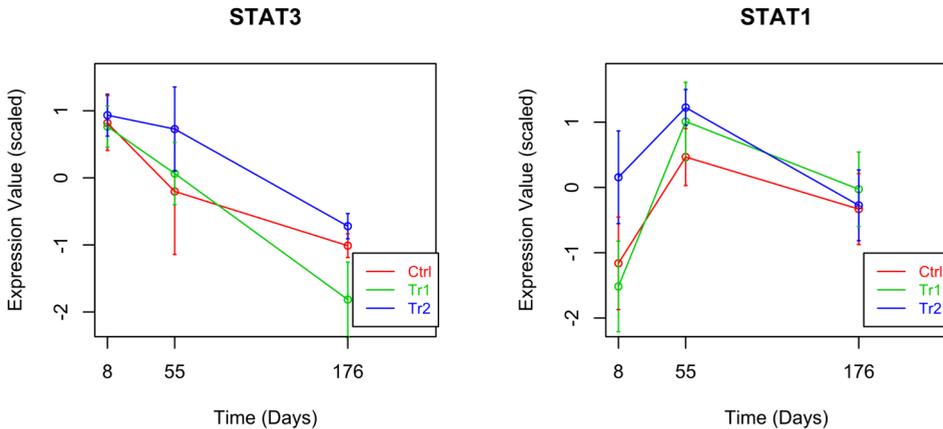
The results from topGO show that the **OnlyTr1** and **Tr1&Tr2** genes are mainly involved in innate and adaptive immune processes, whereas the **OnlyTr2** list is dominated by genes involved in developmental processes. Compared to the topGO analysis, the enrichment analysis from Reactome FI provided more detailed functional information. It showed that the genes in all three lists are involved in immune processes with dominance of adaptive immune processes in **OnlyTr1** and innate immune processes in **OnlyTr2**. The overlapping gene list **Tr1&Tr2** included immune signalling functions, interferon and interleukin genes, that are speculated to contribute to both types of immunity (Yu and Gaffen 2008; Langrish et al. 2004; Le Bon and Tough 2002).

**Table 1: Description of the 17 hubs in the three Functional Interaction networks.** The number of connections of each gene in their respective network is given along with the gene symbol. The information about the genes is adapted from [www.genecards.org](http://www.genecards.org). The hubs can be clustered into three broad groups, given in the last column, based on their functions. All these genes have significantly different time profiles in the treatment compared to the control.

	<b>Gene Symbol</b>	<b>Function</b>	<b>Summary</b>
Only- Tr1 hubs	GRB2 <i>17</i>	Important link between growth factor receptors on the cell surface and Ras signalling	<b>Cell cycle/ Proliferation</b>
	STAT3 <i>15</i>	In response to cytokines and growth factors, STAT family members are phosphorylated and translocate to the nucleus to function as transcription factors.	<b>Immune</b>
	CDC42 <i>12</i>	A GTP-ase involved in signalling for several processes, cell migration, morphology, endocytosis, cell cycle progression and cell proliferation.	<b>Cell cycle/ Proliferation</b>

	CAV1 <i>11</i>	Encodes a scaffolding protein that an essential part of caveolar membranes.	<b>Immune</b>
	FOS <i>11</i>	The FOS family encodes for leucine zipper proteins. Regulates cell proliferation, differentiation, transformation and apoptotic cell death.	<b>Cell cycle/ proliferation</b>
Tr1 & Tr2 hubs	MYC <i>31</i>	Transcription factor that activates growth related genes	<b>Cell cycle/ Proliferation</b>
	MAPK14 <i>24</i>	Important for the cascades of cellular responses evoked by extracellular stimuli leading to direct activation of transcription factors.	
	RELA <i>24</i>	Forms a complex with NFκB transcription factor and regulates the NFκB pathway.	<b>Immune</b>
	UBE2D2 <i>24</i>	Ubiquitin conjugating enzyme that catalyses covalent attachment of activated ubiquitin to other ubiquitin ligases	<b>Ubiquitination</b>
	ITCH <i>23</i>	Ubiquitin-protein ligase which accepts ubiquitin from an ubiquitin-conjugating enzyme and then directly transfers the ubiquitin to targeted substrates.	<b>Ubiquitination</b>
	IKBKG <i>22</i>	Regulatory subunit of the IKK core complex which phosphorylates inhibitors of NFκB and ultimately the degradation of the inhibitor	<b>Immune</b>
	SP1 <i>22</i>	Transcription factor (activator or repressor) that regulates many cellular processes	<b>Cell cycle/ Proliferation</b>
	RHOA <i>20</i>	Regulates signalling from plasma membrane receptors to the assembly of focal adhesions and actin stress fibers.	<b>Cell cycle/ Proliferation</b>
	RPS3 <i>20</i>	Component of the 40S ribosomal subunit, where translation is initiated	
	FBXW7 <i>19</i>	Part of the ubiquitin ligase complex (SCF) which recognizes and binds to phosphorylated targets.	<b>Ubiquitination</b>
Only- Tr2 hubs	UBA52 <i>20</i>	One of the four genes that code for ubiquitin, and ribosomal components which are part of the ribosome 60S subunit.	<b>Ubiquitination</b>
	STAT1 <i>12</i>	In response to cytokines and growth factors, STAT family members are phosphorylated and translocate to the nucleus to function as transcription factors.	<b>Immune</b>

The network analysis provided insight into genes with potential high level regulatory activity, i.e. genes in the network with high number of connections/edges. A list of these potential high level regulators or hubs and a gist of their known biological functions is given in Table 1. The temporal expression patterns of three of these 17 genes are shown in Figure 3 (all 17 of them can be found in the Supplementary Figure 4). Except for the two hubs of the OnlyTr1 FI network, all the hubs have long term differences in expression between the treatment and control based on the  $\beta_3$  regression coefficient. The hubs in all the three networks can be roughly assigned to three functional categories: immune, cell cycle or proliferation, and genes involved in ubiquitination. There are two genes that are not part of these three clusters, MAPK14 and RPS3, where the latter is an important component of the ribosome. MAP kinases act as an integration point for multiple biochemical signals, involved in a wide variety of cellular processes like proliferation, differentiation and development. They are activated by environmental stresses or cytokines.



**Figure 3: Gene expression patterns of important hubs.** In Figure 3 each graph depicts the temporal expression pattern of a single gene. These temporal changes are shown under three different conditions: Ctrl (red line), Tr1 (green line), and Tr2 (blue line). The x-axis indicates the time in days. The expression values (y-axis) are scaled such that the average expression of each gene is 0 and the standard deviation is 1.

All the hubs in the **OnlyTr1** are either immune or cell cycle/proliferation related genes. This is reflected in the positioning of these genes in the network. They are found in the two modules related to immunity; T-cell co-stimulation and leukocyte migration (Figure

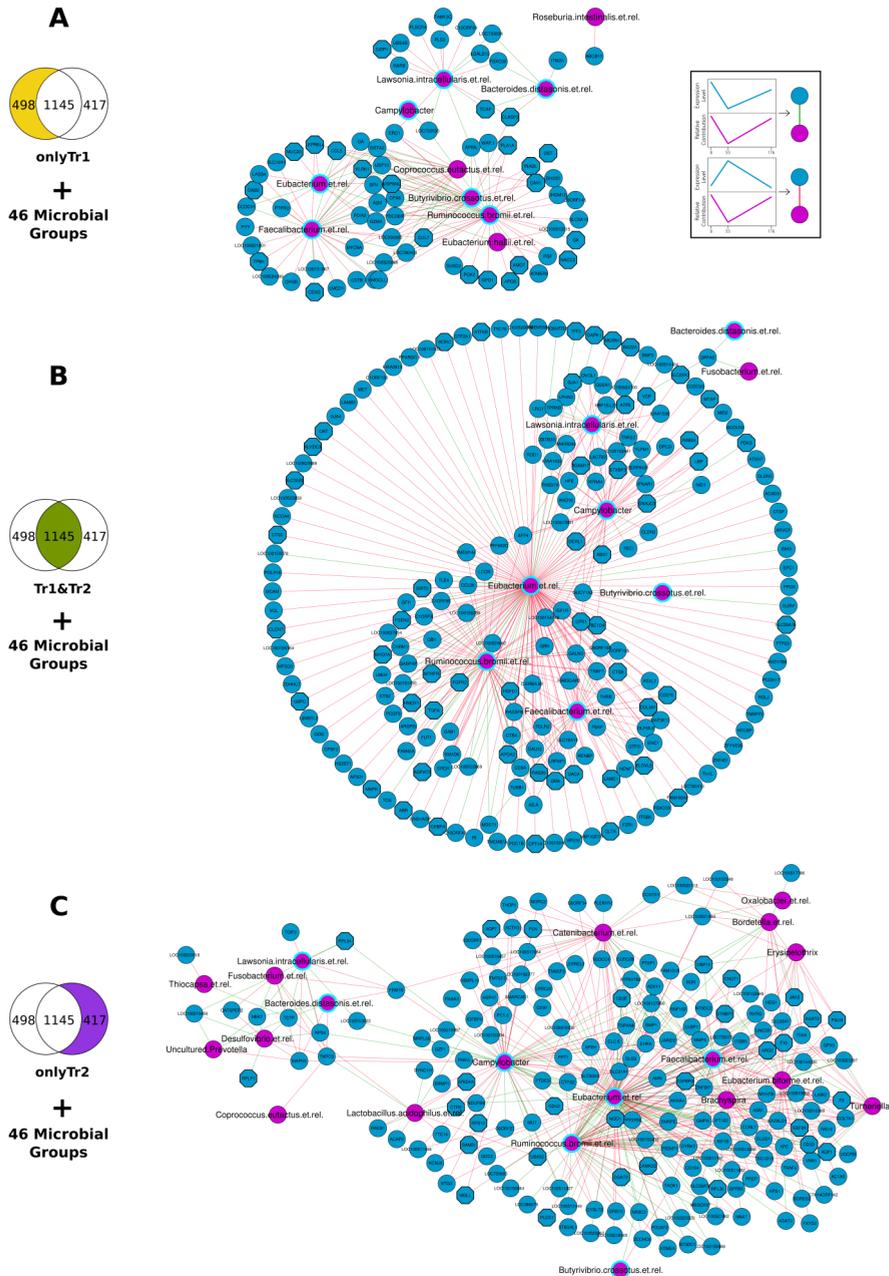
2). The genes in the ubiquitin cluster are from the **Tr1&Tr2** and **OnlyTr2** networks; they code for ubiquitin itself (UBA52) and also for proteins that perform the conjugation and ligation of ubiquitin to other proteins. These hubs in the **Tr1&Tr2** network are in the module for protein ubiquitination; this module has the highest number of hubs. Other hubs in this network are from four modules as shown in Figure 2. In the **OnlyTr2** network, there are two hubs, UBA52 and STAT1.

### 3.2. Microbiota temporal changes

The relative contribution of the microbiota was filtered as described in Material & Methods and was left with 125 microbial groups which represent 99% of the population. The relative contribution values of these 125 groups were used in the regression analysis to identify groups that can be further related to gene expression levels. After a two-way ANOVA analysis in maSigPro, there were 46 microbial groups that showed significant differences over time or treatment compared to the control. Some bacterial groups like *Brachyspira* show different contributions in different conditions but some groups like *Bacillus* et rel show major differences only at one time point and one treatment. Supplementary Figure 6 shows these temporal changes for all the 46 groups.

### 3.3. Integration of microbiota and gene expression analysis

By performing statistical integration of both the microbiota and host gene expression datasets, (mixOmics R package), we tracked changes in jejunal gene expression which follow the changes in microbial populations as determined on the same location in the gut. The changes will reflect the similarity or dissimilarity of a pair of data points across time. The resulting similarity matrix was used to connect the microbial groups and the expression of genes into a network. The first network, **OnlyTr1**, was built with 498 genes in 6 conditions (Ctrl, Tr1 with 3 time-points each) and 46 level2 microbial groups in the same conditions. The second network (**Tr1&Tr2**) was built with 1,145 genes and the same 46 microbial groups, and the conditions were the control and both Tr1 and Tr2 at the 3 time-points, this gave rise to 9 conditions. The third one had again 6 conditions (Ctrl and Tr2 in 3 time-points) with 417 genes and the same 46 bacterial groups. The networks represent the correlation between the microbiota and the gene expression data and are shown in Figure 4.



**Figure 4: Correlation networks of changes in gene expression patterns and microbiota composition:** Blue nodes represent genes and the pink ones represent bacterial groups; pink nodes with a cyan boundary are nodes common in the three networks. The edges represent positive (green) and

negative (red) correlation between a gene and a bacterial group. Networks A, B and C were built by correlating the gene lists **OnlyTr1**, **Tr1&Tr2** and **OnlyTr2** respectively with the 46 microbial groups resulting from the regression analysis. All the nodes (bacterial groups and genes) have a significantly different expression profile in time or treatment compared to the control.

There are a total of 22 bacterial groups involved in the three networks, among these, 7 are found in all three networks (Table 2). Most of these bacterial groups share the same temporal pattern of relative contribution (shown in Supplementary Figure 6) and are characterized by an increase in abundance over time. Of the 22 bacterial groups in the three correlation networks, only six are known to be possible pathobionts and their abundance, though increasing with time, remains low at all time-points and treatments. The other 16 bacterial groups are known for being beneficial to the intestine by producing short chain fatty acids or reducing toxic substances (Table 2). Four of the bacterial groups show a decrease in abundance, and are found in the **OnlyTr2** correlation network. Three bacterial groups had consistently large number of gene neighbours: *Eubacterium* et rel, *Faecalibacterium* et rel and *Ruminococcus bromii*. These three groups also share several gene neighbours in all the three networks and are quite central in the network as indicated by network parameters (Supplementary File 3).

**Table 2: Summary of the three correlation networks with information on bacterial groups.**

The bacterial groups that are part of the networks are listed along with the number of genes with which they are highly correlated. Several genes are shared between bacterial groups. Both the genes and the bacterial groups are significantly different either in time or treatment between the control. The first seven groups are common between the three networks. Some general information on these bacterial groups is also given in the last column.

Bacterial Groups	Number of correlated genes in			Information
	Only	Tr1&	Only	
	Tr1	Tr2	Tr2	
<i>Eubacterium</i> et rel	14	165	128	Anaerobic, Gram positive, broad range of species
<i>Ruminococcus bromii</i> et rel	39	65	63	Anaerobic, Gram positive, keystone species in gut, digestion of resistant starch
<i>Faecalibacterium</i> et rel	35	40	43	Obligate anaerobe, Gram negative, produces butyrate, has anti-inflammatory properties

<i>Campylobacter</i>	1	35	62	Microaerophilic, Gram negative, known pathogen in humans
<i>Lawsonia intracellularis</i> et rel	15	22	9	Gram positive, Porcine intracellular enteropathogen
<i>Butyrivibrio crossotus</i> et rel	22	1	2	Anaerobic, Gram positive, butyrate producer
<i>Bacteroides distasonis</i> et rel	5	1	5	Obligate anaerobe, Gram negative, reclassified to <i>Parabacteroides distasonis</i> , produces acetate and succinate, possible pathogen
<i>Fusobacterium</i> et rel		1	9	Anaerobic, Gram negative, generally butyrate producers, possible pathogen, potent lipopolysaccharide
<i>Eubacterium hallii</i> et rel	3			Anaerobic, uses lactate and produces butyrate
<i>Roseburia intestinalis</i> et rel	1			Anaerobic, Gram positive, butyrate producer
<i>Coprococcus eutactus</i> et rel	16		1	Anaerobic, Gram positive, mostly acetate producer
<i>Brachyspira</i>			43	Anaerobic, Gram negative, some known pathogens
<i>Eubacterium bifforme</i> et rel			36	Anaerobic, involved in lipid metabolism
<i>Catenibacterium</i> et rel			32	Obligate anaerobes, Gram positive, digest simple sugars
<i>Turneriella</i>			18	Obligate anaerobes, Gram negative
<i>Bordetella</i> et rel			15	Obligate anaerobes, Gram negative, known to cause respiratory diseases
<i>Erysipelothrix</i>			11	Facultative anaerobes, porcine pathogens for skin diseases
<i>Lactobacillus acidophilus</i> et rel			10	Microaerobic, Gram positive, ferments sugars into lactic acid, probiotic
<i>Desulfovibrio</i> et rel			8	Aerotolerant, Gram negative, reduce toxic substances like sulphate
Uncultured <i>Prevotella</i>			4	Gram negative, pathogens in respiratory tract, breaks down proteins and carbohydrates
<i>Oxalobacter</i> et rel			2	Anaerobic, oxalate reducing
<i>Thiocapsa</i> et rel			2	Mostly anaerobic, oxidise sulphur, photosynthetic

Network topological features also revealed that in all three networks, 40% to 55% of the host genes were neighbours of a single bacterial group. The rest of the genes highly correlate with more than one bacterial group, and a maximum of 8 bacterial groups. About 60% of the genes are shared between at least two bacterial groups. The maximum radiality (indicative of how connected the nodes are to all the other nodes, with values ranging from 0 to 1) was 0.8 to 0.9. This is apparent from the visualization of the networks in Figure 4 which shows only a few nodes in the centre connected to most of the other nodes. In the three correlation networks, 60% – 80% of the edges have negative correlation values

The GO terms that appeared enriched in the gene list of each of the correlation networks are mostly related to four broad functions, metabolic processes, transport of substances, translation, and some immune processes (shown in Supplementary File 3). Most of these genes are negatively correlated with the bacterial groups; this indicates general reduction of the expression levels of these genes across time. But these genes exhibit significant treatment effects, especially long term effects which are not reflected in the temporal patterns of the bacterial groups with which they are highly correlated.

In each correlation network, 20% of the genes are found in the FI networks from gene expression data alone. A separate topGO analysis shows that the genes in each of the correlation networks are enriched for GO terms that are related to four broad categories of biological processes. These categories are, Metabolic Process, Transport, Translation and Immune Processes. In the **OnlyTr1** and **Tr1&Tr2** correlation networks, the genes associated with the most significant GO terms are mostly (70% and 80%) negatively correlated with the microbial groups. But in the **OnlyTr2** correlation network, these genes are mostly (60%) positively correlated with the microbial groups. More than 50% of the genes in the three networks have significant long term differences from the control.

#### 4. Discussion

In this paper we describe a workflow and a set of methods capable of analyzing and integrating different types of data in two dimensions, treatment and time. These methods were then used for the identification of gut system components that contribute to the induction and maintenance of long-lasting effects in the GI tract as induced by perturbations at a young age. To this end we combined multiple, freely available tools and advanced statisti-

cal analytical tools. We used temporal gene expression patterns, obtained using microarray technology, to identify genes and biological processes that are affected over time by the early life perturbation with antibiotic and/or stress treatments. In addition, we used community-scale analysis of gut microbiota from the same location of the gut to identify changes in microbiota profiles over time due to the perturbations.

2

With this approach we have taken the first steps in unravelling the genomic and microbial networks that contribute to long-lasting responses of early life perturbations in the gut. The results reveal that there are significant long-lasting differences in the system of the GI tract between the different perturbations groups, mainly at the gene expression level. The data presented are consistent with the hypothesis that observed long lasting effects on gene expression are most probably due to differences in the programming of the gut immune system as induced by the temporary early life changes in the composition and/or diversity of microbiota in the gut. Furthermore, we were able to identify potential key regulator genes (hubs) for the long-lasting effects and we have identified microbial groups that are potentially associated with the observed changes in intestinal gene expression.

In the following sections of this Discussion we explain the rationale behind choosing these methods for our particular biological question. In the sections “Biological relevance of the hubs”, “Crosstalk between host and microbe components” and “Understanding the long lasting changes in the host” we examine the biological relevance of the results. The other sections mainly deal with methodology aspects.

### **Experimental approach**

Intestinal gene expression and microbiota profiles were generated on three different time-points; day 8 (neonatal), 55 (young adult), and 176 (adult). Day 8 was chosen because it is known that immediately after perturbations, as used here, changes occur in the pattern of microbial colonizing of the gut as well as in the GI tract mucosal gene expression (Willing, Russell, and Finlay 2011; Niewold 2007; Collins and Bercik 2009). The second measurement at day 55 was chosen because the microbiota would have stabilised by then after weaning (around day 28). Weaning is an extremely turbulent process (Pluske, Le Dividich, and Verstegen 2003; Konstantinov et al. 2004; Lallès et al. 2007 that brings

about several temporary changes which are not the focus of our study. The last measurement was on day 176 after birth which coincides with the time of slaughter of these pigs.

Intestinal bacterial composition and gene expression patterns change over the lifetime of an animal due to changes in nutritional, environmental and physiological factors (Elahi et al. 2013; Belkaid and Hand 2014; Konstantinov et al. 2004). Obviously, these normal developmental changes could also be detected in this study; however we were particularly interested in deviations from the normal developmental patterns due to the early life perturbations. Previous work on subsets of these data (Schokker et al. 2014; Schokker et al. 2015) showed that there are significant differences between the control and treatment groups at each time point. In the analysis described here, we incorporated the information across time and between treatments by the use of a quadratic regression analysis (maSigPro R package). In addition, we studied the behaviour of the microbiota over time and looked for correlations with gene expression patterns using the mixOmics R package, because it is known that the microbiota and gut mucosal tissue respond to each other in various ways (L. V. Hooper, Littman, and Macpherson 2012; Nicholson et al. 2012).

### **Analysis of treatment effect on temporal gene expression**

For analysis of data derived from three time-points and two treatments, a regression analysis is better suited than a pairwise comparison with linear models, PCOA or hierarchical clustering which was done in earlier work by Schokker *et al* (Schokker et al. 2014; Schokker et al. 2015). The resulting gene lists from Tr1vsCtrl and Tr2vsCtrl are separated into three lists to be able to look at the biological processes common to both the gene lists which is expected to be due to the effect of the antibiotic. With the resulting three gene lists, we performed GO enrichment analysis with two different methods: topGO and Reactome. An important difference between the two methods is that Reactome is manually curated with experimentally verified information; whereas topGO only uses gene annotation information that is, in many cases, automatically generated without further manual verification. As expected, the Reactome analysis resulted in richer information and encompassed most of the results from the topGO analysis. The network format of Reactome also allows for more informative visualization and further analysis using the topology. For these reasons we decided to concentrate our gene expression analysis on the enrichment results from Reactome where ever possible.

---

## Network analysis

In the FI networks **OnlyTr1** and **OnlyTr2**, more than half of the nodes have a significant long-lasting expression pattern difference in the treatment versus control. This is significant for most of the nodes in the **Tr1&Tr2** FI network which is representative of the action of antibiotic. This reveals that most of the long lasting effects are due to the antibiotic effect in the treatment. These genes with long lasting differences are spread over all the modules which indicates that all the biological processes in the network (Figure 2) are different over time. Several of these processes are essential cellular processes and some are immune related. This could lead to situations that the treated animals respond differently to external stimuli compared to control animals. In order to look further into these long lasting differences we analyse the hubs of the network.

### Analysis of hubs

Investigating the hubs of networks gives insight into the entire network function (Barabasi and Oltvai 2004 and the nodes in our networks are also found to be important for several biological functions (Albert, Jeong, and Barabasi 2000; Jeong et al. 2001; Barabasi and Oltvai 2004 as mentioned in the Results. In most network analysis the degree itself (Ekman et al. 2006 or the degree distribution is used to choose the most connected hubs (I. W. Taylor et al. 2009. We chose the top 40% of the degree distribution as hubs, which is a more lenient threshold than usual because of the large differences in degree distribution within the three networks. Nevertheless, we still were able to identify multiple hubs per network and extract biological functions for these hubs (Table 1). Of the three categories that the functions of the hubs can be classified into, the cell cycle/proliferation category is more abundant than the other two.

It is intriguing to note that most of the hubs of the three networks are known for being high level regulators for several biological processes, which supports the relevance of this network analysis. The hubs from the overlapping network, **Tr1&Tr2** are involved in diverse biological processes and signalling/cascading events. The expression of all the hubs in the **OnlyTr1** and the **Tr1&Tr2** networks is significantly different over time in comparison with their expression pattern in the control group. We speculate that the treatments directly or indirectly affect these hubs which then regulate genes to bring about the long

term effects of the treatments. The hubs in the **OnlyTr2** network conspicuously do not have significant expression level differences over the time span of the experiment when compared to the control group. This suggests that the stress component of the treatment does not affect the animal in the long term as much as the antibiotic alone does.

### **Biological relevance of the hubs**

Among the hubs, there are two STAT genes, STAT3 and STAT1, in the **OnlyTr1** and **OnlyTr2** networks respectively. The STAT family of genes are well studied and have a role in signalling events but also act as transcription factors (Levy and Darnell 2002). The STAT1, STAT2, STAT3, STAT4 genes, found to have significantly different temporal expression patterns in the treatments versus the control, are especially related to immune reactions (Levy and Darnell 2002; Levy and Lee 2002; Poe et al. 2013). This is in agreement with the results of our functional analysis of the networks. STAT3 and STAT1 are known to be up-regulated in cases of bacterial infection (Levy and Darnell 2002; Lad et al. 2005; Kernbauer et al. 2012; Avalle et al. 2012). In addition, expression of STAT1 has been reported to be reduced in the presence of antibiotics (Lad et al. 2005; Levy and Lee 2002; Fryknäs et al. 2007; Avalle et al. 2012). The latter is reflected in our data by the temporal expression patterns shown in Figure 3. Though the overall expression patterns of STAT1 among the three groups are similar, antibiotic treatment and the concomitant change in microbial communities, has a clear effect on STAT1 expression, especially at the first two time-points. Expression of STAT1 on day 8 in Tr2 is, however, quite high compared to the Ctrl, this could indicate that the stress applied at day 4 counteracts the effect of the antibiotic. STAT3 is known to be important for survival of embryos (Ihle 2001; Levy and Darnell 2002; Levy and Lee 2002), it is important during the early stages of development but is not found to be essential in adult tissue (Levy and Lee 2002). This is reflected in the observed decreasing expression levels of STAT3 over the three time-points (Figure 3). The observed contrasting expression pattern of the STAT1 and STAT 3 hubs, may be related to the previous observations that STAT1 and STAT3 counteract each other (Qing and Stark 2004; Avalle et al. 2012).

---

## Data integration

With regard to the microbiota composition and diversity, this study revealed a more prominent effect of time than that of treatment. This is in accordance with previous observations that indicate that the diversity and composition of intestinal microbiota change considerably over time (Konstantinov et al. 2006; Inoue et al. 2005). As expected, the treatment effects are strongest at the first time point, taking into account the efficacy of the used antibiotic (Arrieta et al. 2014; “EMA: Tulathromycin” 2004). As soon as the antibiotic wears off, the microbiota returns to the “normal” state. This is consistent with known literature about the resilience of microbial communities (Jernberg et al. 2010; Antonopoulos et al. 2009).

Since the gene expression data as well as the microbiota data change with treatment and time, this presents an opportunity to correlate these different types of information and look into probable relationships and interactions in the jejunum between microbiota and gene expression of the cells in the mucosal layer. We assumed that the **OnlyTr1** and **Tr1&Tr2** networks represent the effect of the antibiotic on the cross-talk between microbiota and mucosal cells. The **OnlyTr2** correlation network could be interpreted as a representation of the cross-talk between the host and microbiota in response to stress, although we did not look at the effect of physical stress alone. The bacterial groups with the most gene connections are common to the three correlation networks. This indicates that the two treatments do not have drastic effects on the possible host microbiota interactions. The **OnlyTr2** correlation network has more bacterial groups that correlate with genes which suggests that the physical stress factor alters the cross-talk more than the antibiotic alone does. The number of known pathobionts increases in the latter network which could be an indication of a tendency towards pathology in stressful conditions.

## Crosstalk between host and microbe components

There are three bacterial groups that correlate with a high number of genes in all the three networks: *Eubacterium*; *Ruminococcus bromii*; and *Faecalibacterium*. *Eubacterium* is a genus with very diverse species. *Ruminococcus bromii* is a species that has been extensively studied for its ability to breakdown resistant starch (Ze et al. 2012) to produce acetate on which other bacterial groups can survive. Another beneficial bacterial group is

*Faecalibacterium* which is a relatively new genus with only one species documented so far, *F. parusnitzii* and it has been reported to be extremely beneficial to the host. *Ruminococcus bromii*; and *Faecalibacterium* are reported to be most abundant in the anaerobic environment of the colon (Abell et al. 2008; Zheng et al. 2009). Nevertheless, there is evidence that the small intestine is populated with quite a few of such fermenters (Zoetendal et al. 2012). In our analysis the abundance of these bacterial groups in the jejunum correlates to the expression level of a considerable number of genes. This suggests that they are major players in the cross-talk in the jejunum. *Lawsonia intracellularis* and species of *Campylobacter* are found to be correlated with several genes. These groups are known mammalian pathogens of the small intestine. It could be speculated that, in these healthy animals, they are involved in balancing between immune tolerance and immune responses.

### **Understanding the long lasting changes in the host**

The results of the regression analysis on gene expression alone show that there are differences between the three treatment groups for a long time after the perturbation. These differences are reflected in immune functions and cell proliferation related functions. It has been suggested that stress and antibiotics compromise the immune system, but in different ways (Söderholm and Perdue 2001; Konturek, Brzozowski, and Konturek 2011; Willing, Russell, and Finlay 2011). The action of antibiotics could change the delicately balanced signalling between microbiota and the intestinal cells. This change during early stages of development is expected to influence the development and programming of the immune system with consequences for the later life functioning of the immune system. It has been proposed that macrolides (the type of antibiotic given in this experiment) work by recruiting immune cells to carry the antibiotic to the afflicted tissue (X. Wang et al. 2012; Benchaoui et al. 2004). This mode of action will bring about temporary changes in the immune system but cannot explain the observation as described here. Stress is known to cause structural changes in the intestinal tissue (Söderholm and Perdue 2001; Bhatia and Tandon 2005; Taché et al. 2001) and may affect the microbial populations (Collins and Bercik 2009).

The above described changes are not expected to last over a long period of time; intestinal cells have a large turnover (Creamer, Shorter, and Bamforth 1961; van der Flier and Clevers 2009 almost every 4 to 5 days. Yet we find changes that persist for 51 and even 172 days after the perturbation of the gut system. This suggests that the programming and development of the immune system occurred differently between the different treatment groups. Since the microbial composition at day 8 significantly differed between the treatments groups we believe that the difference in immune programming and development is due to differences in the early life crosstalk in the gut between microbes and host. This line of thinking is agreement with the conclusions in several other recent studies in this area (Hoskin-Parr et al. 2013; Cox et al. 2015. It is also in line with our own observation that the microbiota composition returns to the normal state briefly after the perturbation. Memory cells of the mucosal immune system could play an important role in the programming of the mucosal immune system (Ahmed and Gray 1996; Schokker et al. 2015; Belkaid and Hand 2014. Furthermore even though stress was not expected to play a very prominent role in changing the system in the long term, we see that in some instances, it counteracts the effects of the antibiotic. The mechanism behind this is yet to be fully understood.

Pigs are regarded as a useful model for research into modulation of the human gut, especially with regard to microbiota-host immune interactions (Heinritz, Mosenthin, and Weiss 2013. This experiment was designed to disturb the pattern of early colonization of the gut by microbiota with known effects later in life (Arrieta et al. 2014. The factors used for perturbation were, treatment with an antibiotic or an antibiotic with stress at day 4 after birth. Both human neonates as well as piglets in intensive husbandry systems may be exposed to such factors. Therefore our results may be relevant for both humans and piglets.

## **Conclusions**

We used an early life intestinal perturbation in piglets and demonstrate long lasting effects, as measured on the intestinal gene expression level. We provide substantial evidence that several biological processes of the gut mucosal tissue are in a different state between the experimental groups over a long period of time. However, the regression

analysis did not identify significant differences among the temporal patterns of the bacterial species, and the treatments do not seem to affect long term changes in the microbiota. We conclude that the difference in immune programming and development is due to differences in the early life crosstalk in the gut between microbes and host, resulting from the perturbations. Since we identified potential high-level regulators of long term changes in the gut, we are one step closer to identifying the underlying mechanisms. Although the treatments did not lead to phenotypical manifestations, like weight differences between the animals, the exposure of pigs to stressors like pathogen challenges could bring out immune variations between the groups. Studying these subtle differences may help to develop strategies to modulate the process of immune development and programming. The results of this paper are supportive for the recent notion that antibiotics should be used more carefully in neonatal humans and animals.

### **Availability of data**

The data used in this research article is available on Gene Expression Omnibus with the accession number GSE53170, through this link <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE5317>.

### **Ethics statement**

The experiment was approved by the institutional animal experiment committee “Dier Experimenten Commissie (DEC) Lelystad” (2011077.b), in accordance with the Dutch regulations on animal experiments.

### **Acknowledgements**

The authors would like to acknowledge the fruitful discussions with dr. ir. Henk Parmentier in interpreting the results of the analysis. This work has been financially supported by the Systems Biology Investment Programme of Wageningen University, KB-17-003.02-022.

---

## Supplementary Figures and Files

All the Supplementary files and figures can be found at:

<http://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-015-1733-8#Declaration>

**Supplementary Figure 1: Experimental Design.** The figure shows the timeline of the experiment, and each line shows a different group of animals. The red line represents the control group, the blue line the Antibiotics alone (Tr1) and the green line the group given Antibiotics and Stress (Tr2). The pink block represents the intervention of the groups and the light blue boxes represent times of sampling. The boxes at the bottom represent the distribution of the groups among the sows, the pigs were housed in the same manner after weaning.

**Supplementary Figure 2: Microarray Analysis.** Workflow of the microarray analysis with the specific details.

**Supplementary Figure 3: Types of Differences in Time Profiles.** Each graph depicts the temporal expression pattern of a single gene. These temporal changes are shown under three different conditions: Ctrl (red line), Tr1 (green line), and Tr2 (blue line). The x-axis indicates the time in days, the y-axis has expression values scaled such that the average expression of each gene is 0 and the standard deviation is 1.

**Supplementary Figure 4: Expression of FI network hubs.** Expression profiles of 17 hubs of the three FI networks. Each graph depicts the temporal expression pattern of a single gene. These temporal changes are shown under three different conditions: Ctrl (red line), Tr1 (green line), and Tr2 (blue line). The x-axis indicates the time in days, the y-axis has expression values scaled such that the average expression of each gene is 0 and the standard deviation is 1.

**Supplementary Figure 5: Original Tr1&Tr2 FI network.** Network formed with genes that have significantly different time profiles between both treatments vs the control group.

**Supplementary Figure 6: Temporal changes in composition of bacterial groups used in the correlation networks.** Each graph represents the compositional changes of selected bacterial groups over time. The bacterial groups were chosen based on an ANOVA analysis. The time in days is on the x-axis, the log values of the contribution of the bacterial groups is on the y-axis. Each of the 46 graphs have a different scale on the y-axis due to the extreme differences between contributions of each of the bacterial groups.

**Supplementary Figure 7: OnlyTr1 Reactome FI network**

**Supplementary Figure 8: Tr1&Tr2 Reactome FI network**

**Supplementary Figure 9: OnlyTr2 Reactome FI network**

**Supplementary Figure 10: OnlyTr1 correlation network**

**Supplementary Figure 11: Tr1&Tr2 correlation network**

**Supplementary Figure 12: OnlyTr2 correlation network**

**Supplementary File 1: List of significant genes in different conditions.**

**Supplementary File 2: Information on the Reactome FI networks.**

**Supplementary File 3: Information on the correlation networks.**



# Chapter 3

## Multi-level integration of environmentally perturbed internal phenotypes reveals key points of connectivity between them

**Nirupama Benis\***; Host Microbe Interactomics, Wageningen University & Research.

**Soumya Kanti Kar\***; Host Microbe Interactomics, Wageningen University & Research.

**Vitor AP Martins dos Santos**; Systems and Synthetic biology, Wageningen University & Research; Lifeglimmer GmbH, Berlin, Germany.

**Mari A Smits**; Host Microbe Interactomics; Wageningen Livestock Research; Wageningen Bioveterinary Research, Wageningen University & Research.

**Dirkjan Schokker~**; Wageningen Livestock Research, Wageningen University & Research.

**Maria Suarez-Diez~**; Systems and Synthetic biology, Wageningen University & Research.

\*, ~ - These authors contributed equally to the work

**Published in**

Frontiers in Physiology 2017 8:388



**Abstract**

The genotype and external phenotype of organisms are linked by so-called internal phenotypes which are influenced by environmental conditions. In this study, we used five existing –omics datasets representing five different layers of internal phenotypes, which were simultaneously measured in dietarily perturbed mice. We performed ten pair-wise correlation analyses verified with a null model built from randomized data. Subsequently, the inferred networks were merged and literature mined for co-occurrences of identified linked nodes.

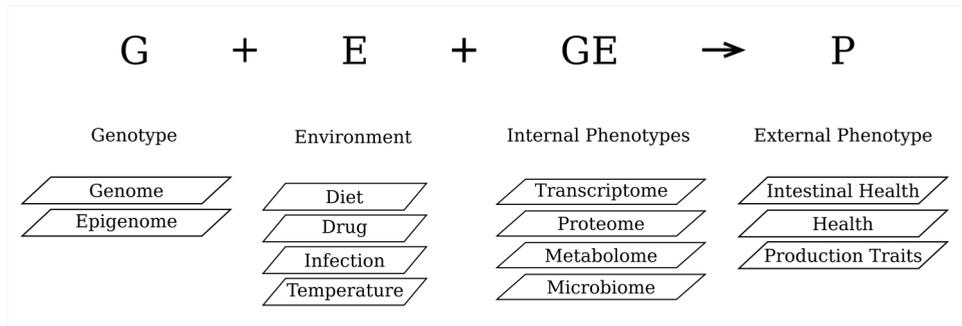
Densely connected internal phenotypes emerged. 45 nodes have links with all other data-types and we denote them ‘connectivity hubs’. In literature, we found proof of 6% of the 577 connections, suggesting a biological meaning for the observed correlations. The observed connectivity’s between metabolite and cytokines hubs showed higher numbers of literature hits as compared to the number of literature hits on the connectivity’s between the microbiota and gene expression internal phenotypes. We conclude that multi-level integrated networks may help to generate hypotheses and to design experiments aiming to further close the gap between genotype and phenotype. We describe and/or hypothesize on the biological relevance of four identified multi-level connectivity hubs.

**Keywords:** Data integration, Internal phenotype, Transcriptomics, Proteomics, Metabolomics, Microbiota, Gastrointestinal tract, Systems Biology.

---

## Introduction

The information encoded in the genome (genotype) and the external quantitative traits or characteristics (phenotype) of an organism are linked to each other by several layers of so-called, intermediate (Fontanesi 2016; Leuchter et al. 2014) or internal (Houle, Govindaraju, and Omholt 2010) phenotypes. Several of these internal phenotypic layers are shown in Figure 1 that visualizes the conceptual relationship between the external phenotype (P), the genotype (G), the environment (E), and the G&E interactions. The epigenome is tightly associated with the genome and represents the programming of gene expression which is not dependent on the DNA code itself. The transcriptome layer represents direct effects of the environment on the gene expression of the (epi-)genome. Translation of the transcriptome into proteins represents the next internal phenotype. The subsequent layer is represented by complex metabolite profiles. The organism-associated microbiota, especially those in the gut, can be regarded as a separate internal phenotypic layer, because it is not only dependent on the host genome but also heavily influenced by its environment, particularly by nutrition (Montiel-Castro et al. 2013; Schwartz et al. 2012). Although for several traits the quantitative effects of the environment on the external phenotypes are known (Cani et al. 2008; de Wit et al. 2011; Gentry et al. 2004), the specific effects of the environment on the internal phenotypes are largely unknown. Furthermore, it is obvious to assume that the various layers of internal phenotypes are connected to each other and that their joint profiles ultimately determine the external phenotype (Fontanesi 2016; Leuchter et al. 2014). Unfortunately, most of these assumptions are not based on solid evidence and at best represent oversimplifications of the dynamic nature of processes involved in determining external phenotypes. It, furthermore, partly explains the knowledge gap that exists between the genotype and the external phenotype.



**Figure 1:** Relationship between the external phenotype (P), the genotype (G), the environment (E), and the G&E interactions. The internal phenotypic layers and the environmental factor with a darker outline are included in the present study.

Therefore, the objective of this study was to develop methodologies to identify components in the internal phenotypic layers that are connected to components in other internal phenotypic layers. To this end, we integrated multi-scale quantitative (-omics) data using a regression approach. The used data sets were derived from a single experiment with inbred mice which were exposed to five different dietary interventions as a means to perturb the different internal phenotypes. With a data-driven approach we were able to identify a large number of potential connections between the various intermediate phenotypes and for several we found proof of causal relationships in literature. The results of this study provide a basis to understand how various internal phenotypic layers are connected to each other. The identified connections may be crucial for the identification of causal relationships (Civelek and Lusis 2014 between various biological scales and to uncover mechanisms involved in determining external phenotypes.

## Materials and Methods

### Origin of data

We used data from an experiment with six-week old inbred mice that were fed for four weeks with six different semi-synthetic diets (Kar et al. 2016. In brief: thirty-six 21-day-old C57BL/6J mice (Harlan Laboratories, Horst, the Netherlands) were divided into 6 groups and housed in pairs with ad libitum access to diet and water. After adaptation for one week to a standard diet, the mice were fed semi-synthetic diets containing 300 g/kg

(as fed basis) of one of the alternative protein sources for 28 days: soybean meal; casein; partially delactosed whey powder; spray dried plasma protein; wheat gluten meal and yellow meal worm. At the end of the experiment, mice were sacrificed to collect ileal tissue to acquire gene expression data, ileal digesta to study changes in microbiota, blood serum to profile cytokines and chemokines and blood and urine to profile amine metabolites. All procedures were approved by the animal experimentation board at Wageningen University & Research Center (accession number 2012062.c) and carried out according to the guidelines of the European Council Directive 86/609/EEC dated November, 1986. Multi-omics data were obtained with regards to: whole genome gene expression profiles of ileal tissue as measured with Affymetrix GeneChip mouse gene 1.1 ST microarrays (Affymetrix, Santa Clara, CA, USA); community scale microbiota composition of ileal digesta by targeted-amplicon DNA sequencing of the bacterial 16S rDNA V3 region on an Illumina Mi-Seq sequencer; 23 serum cytokine and chemokine concentrations (pg/ml) using a Bio-Rad Mouse 23-plex kit (Bio-Rad, Hercules, CA, USA); and amine metabolic profiles of serum and urine using an ACQUITY UPLC system coupled online with a Xevo Tandem quadrupole mass spectrometer (Waters) operated using QuanLynx data acquisition software (version 4.1; Waters) (Kar et al, in preparation). The data from the ileum reflects the local effects of the dietary interventions, the other three data assess the systemic effects.

### **Pre-processing and selection of data**

An overview of the five types of data and their specifics are given in Table 1. Each dataset was pre-processed in a similar way using the R package limma (Smyth 2005) to find the differentially significant data-points. The data is first log transformed and then this data is fitted to a linear model using the function lmFit (Phipson et al. 2016) which will give back information on the differences between the genes in different arrays and subsequently different comparisons of control vs treatment. Then we used the function eBayes (Phipson et al. 2016) which applies an empirical Bayes method to compute p-values for a t-statistic under the assumption that only 1% of the genes are differentially regulated among all the genes in the arrays. This p-value is then subjected to a Benjamini-Hochberg (Benjamini and Hochberg 1995) multiple testing, also known as a False Discovery Rate (FDR).

**Table 1: Pre-processing and specificities of each data-type.** Details of the site of sampling and data dimensions before and after pre-processing are indicated. The first number indicates the number of variables in the data and the second number denotes the number of samples.

Properties	Transcriptomics	Microbiota	Cytokine	Metabolomics Serum	Metabolomics Urine
Sampling	Ileum	Ileum	Serum	Serum	Urine
Before pre-processing	16,410 * 33	148 * 33	23 * 36	41 * 36	16 * 28
After pre-processing	52 * 33	22 * 33	13 * 36	26 * 36	16 * 28

This analysis was done by comparing the data of each dietary group against the data of the dietary group that received soy bean meal as protein source, which is the most common source of protein in animal diets. The FDR value of the data, is used to gauge significance and data-points that were significant in at least one of the five comparisons of the diets were included in the integration analysis. Except for the Cytokine and Metabolomics Serum (using the amine measurement), all the data-types had some samples thrown out due to quality control. Two types of metabolomics measurements were done on the sampled urine; Amine and Acyl-carnitine, the amine dataset did not have sufficient statistically significant data-points so was discarded. We only work with the Acyl-carnitine measurement in urine.

### **Data integration, network generation and network assessment.**

All significantly different data-points were used in the integration which was initially performed with two datasets at a time, so that from the 5 datasets 10 integrated networks were generated. The integration was performed using the function sPLS (sparse Partial Least Squares) in regression mode with  $ncomp = 5$ , from the R package mixOmics (Dejean et al. 2011; González et al. 2012; Lê Cao, González, and Déjean 2009). The regression mode is used to model causal relationship between variables in both datasets by identifying combinations of variables between both datasets. Weight vectors used in the regression modelling are termed loading vectors. sPLS is used to perform simultaneous

variable selection in the two datasets to be integrated and employs LASSO (Least Absolute Shrinkage and Selection Operator) penalization (Tibshirani 2011) on the loading vectors. This approach requires one data set,  $\mathbf{X}$  with  $n_x$  elements, to be designated the predictor and the other,  $\mathbf{Y}$  with  $n_y$  elements, the response. As an output, the approach produces a matrix  $Ma(\mathbf{X}, \mathbf{Y})$  of size  $n_x \times n_y$  representing the relevant correlations between both datasets, so that:

$$ma_{ij} = \begin{cases} 0, & \text{if } Y_j \text{ independent of } X_i \\ \text{cor}(X_i, Y_j), & \text{if } Y_j \text{ dependent on } X_i \end{cases}, \text{ with } i \in \{1, \dots, n_x\} \text{ and } j \in \{1, \dots, n_y\}$$

Where  $\text{cor}(X_i, Y_j)$  is Pearson's correlation between elements  $i$  and  $j$  from datasets  $\mathbf{X}$  and  $\mathbf{Y}$  respectively. The correlation is computed across all available samples (here corresponding to dietary exposures).

Since it is not trivial to determine the predictor and response with biological data, we swapped the two types of data to compute  $Mb(\mathbf{Y}, \mathbf{X})$ , a matrix of size  $n_y \times n_x$  where the roles of  $\mathbf{X}$  and  $\mathbf{Y}$  have been interchanged. Both matrices,  $Ma$  and  $Mb$  were combined into a final matrix  $M(\mathbf{X}, \mathbf{Y})$  size  $n_x \times n_y$ , using

$$M(\mathbf{X}, \mathbf{Y}) = Ma(\mathbf{X}, \mathbf{Y}) + t(Mb(\mathbf{Y}, \mathbf{X}))$$

where  $t$  represents matrix transposition. Thus, non-null elements of the matrix  $M(\mathbf{X}, \mathbf{Y})$  represent correlations between data types that have been deemed associated. This matrix can be seen as a weighted adjacency matrix representing a network where two nodes  $X_i$  and  $Y_j$  are connected via an edge if a non-null weight can be assigned to the edge. This weight is represented by the matrix value  $m_{ij}$ .

To further prune the network of (possibly) spurious interaction two additional thresholds ( $th_l < 0$ ; and  $th_h > 0$ ) were imposed to obtain an unweighted adjacency matrix  $A(\mathbf{X}, \mathbf{Y})$  of size  $n_x \times n_y$ ,

$$A_{ij} = \begin{cases} 1 & \text{if } m_{ij} \geq th_h \text{ or } m_{ij} \leq th_l \\ 0 & \text{if } |m_{ij}| < |th_l| \text{ and } m_{ij} < th_h \end{cases}$$

where  $|x|$  represents the absolute value.  $th_l$  and  $th_h$  were selected for each network so that only top 5% of the highest (positive) and lowest (negative) weights were kept for building the networks.

Networks represented by these adjacency matrix were transformed into the edge-list format, a two column table of the connected nodes in a network where each row represents an edge and visualized in Cytoscape (Ono et al. 2015; Shannon et al. 2003).

For each pair of integrated datasets, a null model of the association networks was constructed using a strategy based on random permutations of measured values (Saccenti et al. 2014). Measured data-points were randomly permuted over samples before data integration to obtain randomized datasets that still retained the same value distribution for each variable. The randomized datasets were then used for data integration following the aforementioned approach thereby generating randomized associations networks. The process was iterated  $N_{it} = 1,000$  times for each pair of datasets; for each iteration,  $k$ , the values of the dynamic cut-offs ( $th_{lk}$  and  $th_{hk}$ ) (5% of the highest and lowest correlation) were recorded. For the ten pairwise combinations of datasets, the values obtained for  $th_l$  and  $th_h$  obtained using the unpermuted dataset, were compared with the distribution of values of  $th_{lk}$  and  $th_{hk}$  with  $k = \{1, \dots, N_{it}\}$  to get networks from the random data to compare to the networks from the biological data.

### **Network merging and topological analysis**

The ten networks arising from pair-wise data integration of the 5 data sets were merged in a combined network including all the nodes and edges of the 10 networks. This network is then restricted by only including nodes present in at least two of the separate networks. We used the *igraph* R package (Csardi and Nepusz 2006) to further analyze the network, which was treated as non-directed, since no particular directionality was assigned to the edges. We obtained values for the following topological properties of the merged network (Barabasi and Oltvai 2004; Csardi and Nepusz 2006): Degree: number of neighbors of a given node, that is the number of nodes connected to it. Clustering coefficient of a node is the ratio of the number of connections between the neighbors of a node and the total number of possible connections between said neighbors. Characteristic path length: median of the average distance between a node and all the rest. Network density: ratio between the

total number of existing edges and the total number of possible edges (given the number of nodes in the network). Connected components maximal subgraphs in a network such that each node is connected to all the rest by means of network paths. For node level metrics, such as degree or clustering coefficient average values were computed over all nodes. Cytoscape was used for network visualization.

### **Literature mining**

To investigate the co-occurrence of the names of the connected nodes in the correlation network, we used the R package *rentrez* (Winter 2016). This package searches for selected keywords in PubMed abstracts while making use of the MeSH (Medical Subject Headings) thesaurus to maximize results via the API from NCBI. The search was not restricted to a specific tissue type or organism. These results were examined, although not exhaustively, to find literature evidence of established relationships between nodes connected through identified edges; these were then considered as true positive search results.

The script used to generate all these results will be made available on request. All the above mentioned operations were performed using existing functions from R packages.

## **Results**

### **Analysis of the individual datasets**

A dietary intervention was performed on mice where the protein content was changed and multi-omics data were obtained with regard to: whole genome gene expression profiles of ileal tissue (Transcriptomics), community scale microbiota composition of ileal digesta (Microbiota), 24 different cytokine levels in blood serum (Cytokine), and protein-associated metabolic profiles of serum (Metabolomics Serum) and urine (Metabolomics Urine). These data were pre-processed and analyzed separately by fitting a linear model on the data-points and looking for differentially expressed readouts in each treatment versus the control. Each dataset had its own p-value (corrected for multiple testing with the Benjamini-Hochberg method) threshold, ranging from 0.001 to 0.1 for difference between the tested and reference diets. The highest number of statistically significant entities was

found in Transcriptomics. Furthermore, all the measured variables in Metabolomics Urine were found to be significantly different in at least one comparison.

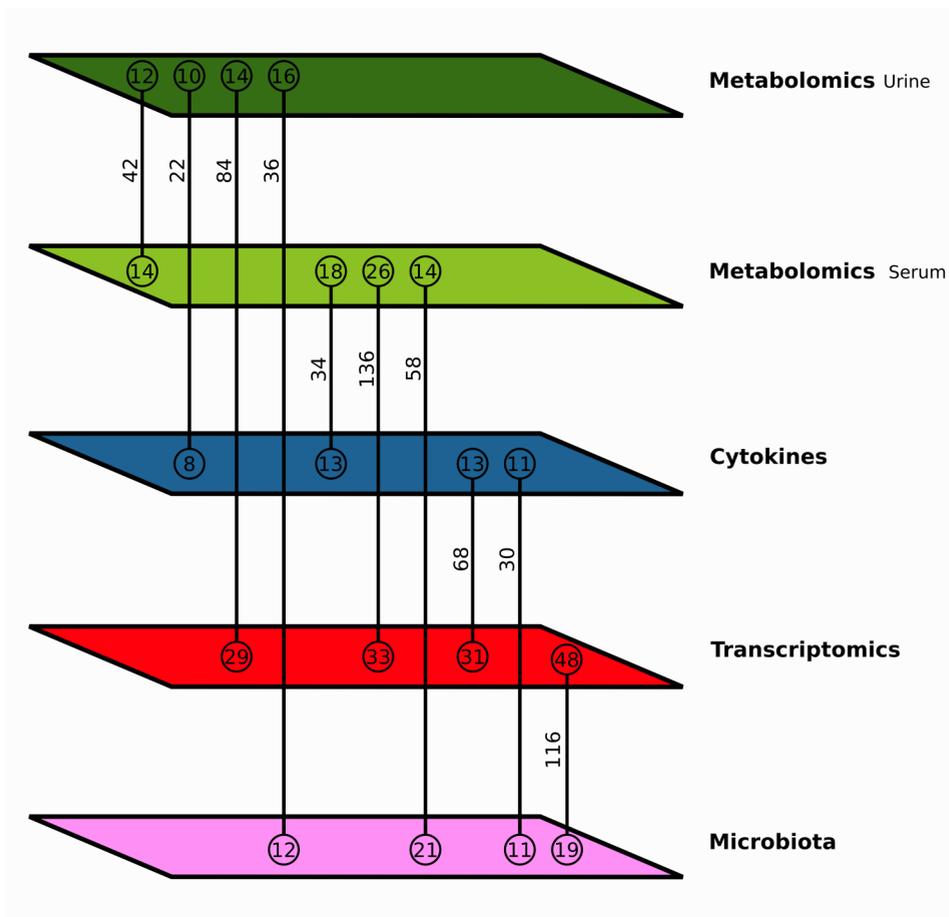
### **Pairwise data association and network generation**

We performed the integration by linking two data-types at a time and in such a way that after the pairwise analysis all the observed correlation data could be combined to build a multi-level interaction network. Therefore, each data-type was integrated with the other four types of data, resulting in ten correlation networks. The topological characteristics of all these ten networks are given in Table 2 and Figure 2, and the network graphs are available in Supplementary Figure S1 as an image. Supplementary File 1 has the networks in a format that can be uploaded into Cytoscape in order to further explore the connectivity's of these networks by simply clicking on these nodes. Table 2 shows the positive and negative thresholds that were used separately for the correlation network. Connections between pairs of data points with correlation values between the threshold values, i.e. Low Threshold (negative threshold) and High Threshold (positive threshold) as indicated in Table 2, were discarded and the corresponding edges removed from the final network. There were two disconnected sub-graphs in five of the networks while the other five have only a single, fully connected graph.

The largest network, in terms of nodes, is the Microbiota & Transcriptomics network. This seems logical as it represents the most comprehensive datasets and spacial interactions between the two data-types are known to occur. Overall, networks involving Transcriptomics data had higher number of nodes than other networks. The smallest network with 18 nodes and 22 edges was the Metabolomics Urine & Cytokine network.

**Table 2: The 10 individual correlation networks.** Each row represents one of the 10 correlation networks. Low Threshold and High Threshold represent the thresholds used for the correlation values. The 3<sup>rd</sup> and 4<sup>th</sup> columns have the number of nodes in the network that belong to the first and second data, respectively. The last column displays the number of connected graphs in the network.

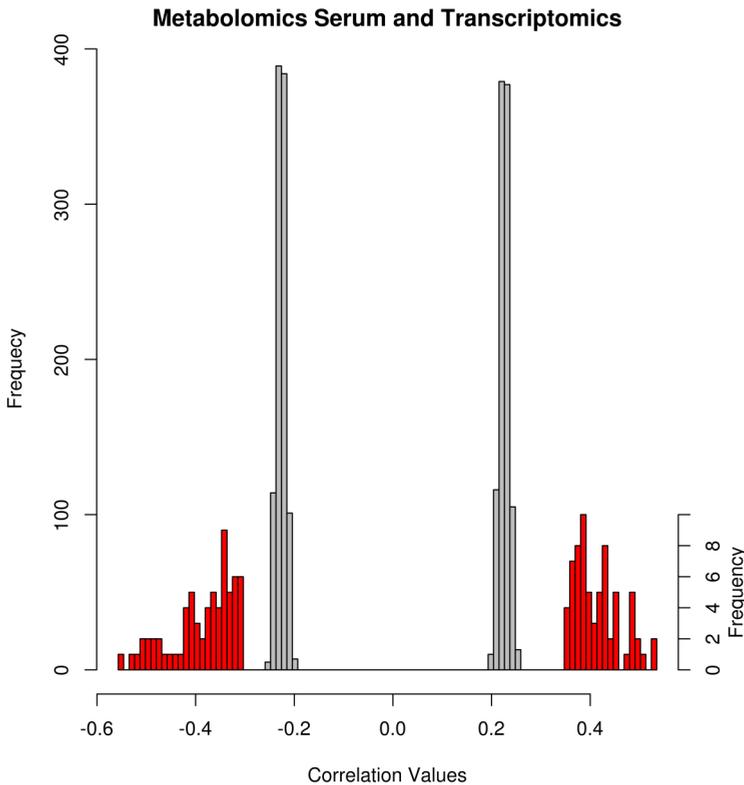
<b>Network Names (data A &amp; data B)</b>	<b>Low Threshold</b>	<b>High Threshold</b>	<b># of nodes (A)</b>	<b># of nodes (B)</b>	<b>Connected graphs</b>
Metabolomics Serum & Metabolomics Urine	-0.51	0.6	14	12	2
Metabolomics Serum & Microbiota	-0.38	0.3	21	14	1
Metabolomics Serum & Transcriptomics	-0.31	0.35	26	33	2
Metabolomics Serum & Cytokine	-0.33	0.5	18	13	2
Metabolomics Urine & Microbiota	-0.28	0.42	16	12	1
Metabolomics Urine & Transcriptomics	-0.55	0.54	14	29	1
Metabolomics Urine & Cytokine	-0.32	0.55	10	8	2
Microbiota & Transcriptomics	-0.28	0.27	19	48	1
Microbiota & Cytokine	-0.38	0.35	11	11	1
Transcriptomics & Cytokine	-0.27	0.34	31	13	2



**Figure 2: Multilevel integration.** This schematic image shows the number of connections between each internal phenotypic level with the other levels in a merged network. The colours of the parallelograms denote the internal phenotypic level to which the data-types belong. Green is Metabolomics (light green – Metabolite from Serum and dark green – Metabolite from Urine), blue is Cytokines, red is Transcriptomics and pink is Microbiota. Each line connects two levels and the vertical number above the line indicates the number of edges in the correlation network between those two phenotypic levels. The number of connected nodes in each level is given in circles above and below the connecting lines.

### Technical validation of pairwise integration networks by random permutation

We performed the same method of integration on the five different data-types after randomly permuting the measured data, this process was iterated a 1000 times. In this way, the networks obtained from random permutations are considered a null model with no biological information, and used to assess the significance of the results obtained with the non-permuted data. Figure 3 shows the spread of correlation values for the integration of Metabolomics Serum and Transcriptomics. The thresholds for network reconstruction were selected so that only the 5% highest and lowest correlations were kept. The separation between the values obtained for the integrated data and the randomly permuted datasets indicates the high significance of the edges in the integration networks. In this way, selection of the 5% highest and lowest correlations and significant limits the number of spurious correlations that could be due to chance alone while retaining maximum information in the networks.



**Figure 3: Distribution of network correlations and random network cut-offs of the Metabolomics Urine and Transcriptomics networks.** The x-axis depicts the range of correlation values and the y-axis shows its frequency. The grey bars denote the distribution of the thresholds of the 1,000 random correlation networks with frequency on the left y-axis. The red bars are distributions of the correlation values of the inferred network with frequency on the right y-axis.

Similar results were obtained for most of the integration networks (Supplementary Figure S2). In three of the networks, there is an overlap between the correlation values from the inferred network and the values arising from the randomly generated networks. The overlaps are in the networks Metabolomics Urine & Microbiota, Metabolomics Urine & Cytokine and Transcriptomics & Cytokine network. The highest overlap appears in the first two and mostly affects edges with negative correlations.

### **Merged Network**

All the 10 integration networks (Supplementary File 1) were merged and only nodes linked with nodes of at least two other data-types were kept (see Table 3). The gene expression data has the highest number of nodes in the merged network. However, nodes with the highest degree (number of connecting edges) arise from the microbiota data, with S24-7 having 57 neighbors and Bifidobacterium having 47 neighbors. The merged network encompasses 45 nodes that are connected to all the other types of data. For that reason we denote them ‘Connectivity hubs’ and they are included in Table 3 and Supplementary Table 1.

**Table 3. Characteristics of the merged network.** Characteristics of the merged correlation network. The number of nodes from each data-type are given in rows three to seven. Between brackets the number of connectivity hubs is indicated.

<b>Network Statistics</b>	
Total number of nodes	112 (45)
Total number of edges	577
Number of Metabolomics Urine nodes	15 (8)
Number of Metabolomics Serum nodes	24 (11)
Number of Cytokine nodes	13 (7)
Number of Transcriptomics nodes	43 (12)
Number of Microbiota nodes	17 (7)
Degree range	2 to 57
Average number of neighbours	10.35
Clustering coefficient	0.20
Network Diameter	4
Network radius	3
Characteristic path length	2.31
Network density	0.09
Connected components	1

### Functional validation of merged network by text mining

A PubMed literature search for co-occurrence of linked nodes gave results for 6% of the links corresponding to 37 edges. We further investigated reported causality effects between the nodes in question. Most of the retrieved results are related to metabolites and

cytokines measurements whereas a few results confirming causal relationships were found involving gene nodes. We were able to find literature confirmation pertaining to associations for six out of the ten pair-wise connections between phenotypes, as summarized in Table 4 and Supplementary Table 2. Among the nodes with literature results, four are from Microbiota, two from Transcriptomics, 15 from Metabolomics Serum, three from Metabolomics Urine and six from Cytokines. The node with the highest number of hits in literature is Tnfa which co-occurs 8,563 times with nine metabolites from the Metabolomics Serum data and one bacterial group (Bifidobacterium).

**Table 4: Overview of text mining results.** The first column shows the types of data that are connected by the edges that were found in the PubMed literature search.

Data Connections	PubMed Ids	Distinct edges
Cytokines & Metabolomics Serum	9,554	16
Metabolomics Serum & Metabolomics Urine	906	6
Microbiota & Metabolomics Serum	254	7
Microbiota & Cytokines	250	5
Transcriptomics & Microbiota	83	3
Metabolomics Serum & Transcriptomics	59	2

Of the 30 data-points from all the types of data that have literature results, 15 are connectivity hubs. One such connectivity hub is Glutathione (GSH) which has 21 direct neighbors from four data-types as shown in Figure 4. This hub is especially interesting because six of the connected nodes (Carnitine, Tnfa, Il-1b, Il17c, Bifidobacterium and Dapk2) have textual co-occurrences found by the text mining algorithm. The terms GSH and Tnfa were found 2,231 times in the abstracts of Pubmed indexed articles. Full text inspection shows that some of the connections are causal relationships as one of the connected nodes activates or inhibits the other.

**Figure 4: Glutathione sub-network.** This figure shows the 21 connections of the node Glutathione in the merged network. The different colours of nodes indicate the data-type of internal phenotypic level of that node, pink is Microbiota, red is Transcriptomics, blue is Cytokines, and green is Metabolites (light green – Metabolites from Serum and dark green – Metabolites from Urine). Oval nodes are connectivity hubs. Dotted lines show un-validated edges and continuous, thicker edges show connections also present in the results retrieved from scientific literature. Edge colour, yellow and purple, indicates positive and negative correlations, respectively.

### Discussion

In this study we developed and used a set of computational methods to identify components in internal phenotypic layers that are connected to components in other internal phenotypic layers of an organism. We successfully integrated multi-scale quantitative (-omics) data, derived from a single experiment with inbred mice and which were exposed to five different diets. Since the data originated from an animal experiment that was not designed for the detection of genetically and/or dietarily induced differences in external phenotypes, we only focused on the connectivity between 5 intermediate phenotypic levels. Some studies have reported pairwise data integration of two (Benis et al. 2015; Lu et al. 2014; Rajasundaram et al. 2014 or three data sets (Adourian et al. 2008. But this is, to the best of our knowledge, the first time that an integration of such heterogeneous data-types from different tissues, arising from a single experiment, has been reported. The ap-

proach as described here could, in principle, be applied on any number and type of datasets, as long as they are from the same experiment, from samples at the same time-point and have comparable dimensions of differentially regulated data.

### **Internal phenotypic data and pairwise data integration**

Each used data-type represents a different internal phenotype and a different layer of the system that (co-) drives the manifestation of external phenotypes. We subjected each data-type to a separate analysis in order to correlate only those changes induced by the dietary intervention. Nodes with significantly different values could easily be identified in each of the sampled tissues and fluids (ileum, blood and urine) thereby representing the local and systemic effects of the interventions and the need of a multi-scale approach.

In order to investigate connections between the five data-types we used sPLS, an integration method that can be applied to several types of data, two at a time. This method can also handle the dimensionality problem of biological datasets where the number of variables is usually higher than the number of samples. sPLS has been previously used for integration of microbiota with gene expression data (Benis et al. 2015; Steegenga et al. 2016, and measurements on cell wall polysaccharides of fibers with phenotypic characterizations of fibers in cotton balls (Rajasundaram et al. 2014).

We performed pairwise integration of the datasets, resulting in ten networks with varying spreads of correlation values. Deciding on a threshold to distinguish genuine from spurious correlations is a major bottleneck for the definition of correlation networks. While a 0.8 threshold (absolute value) has been suggested for gene expression data (Schäfer and Strimmer 2005, other authors suggested smaller values (0.6) in metabolomics data sets (Camacho, De La Fuente, and Mendes 2005. The correlation values greatly depend on the biological dataset under study and its dimensionality. There are several methods to choose a threshold based on the data: use assigned p-values as threshold; use network characteristics of the correlations; or use a percentage of the correlation distribution. When evaluated by Borate et al (Borate et al. 2009 they concluded that threshold selection methods based on network properties such as the clustering coefficient are best for gene co-expression networks. This would not work here because the generated networks always induce con-

nections between data points of different type and as a result they have a zero clustering coefficient for every node. While integrating two types of metabolomics datasets with gene expression of the tissues in which they were measured Adourian et al (Adourian et al. 2008 assigned p-values to the correlation values and then set a threshold. In this study, we used the top 5% of the correlation values because this dynamic threshold (separate for the positive and negative values) eliminates bias towards the size of the datasets.

We further validated the observed correlations by comparing them with a null model obtained by randomly permuting the data along the samples (Eguiluz et al. 2005; Saccenti et al. 2014. In two of the networks, Metabolomics Urine & Microbiota and Metabolomics Urine & Cytokine (the smallest network), the significance of the negative correlation values could not be established as we observed a considerable overlap between the negative correlation values of this network and the negative thresholds of the random networks. This calls for caution when biologically interpreting these networks. For five of the networks we observed a very clear separation of the random thresholds and the start of the correlation values in the network (Supplementary Figure S2). The other networks showed slight overlaps between the random threshold distribution and the network correlation distribution. This extra validation step reassured us that the observed correlations are rooted in biological phenomena. To our knowledge this technical validation step is not common in current studies of this type.

The edges of the inferred networks, indicate significant computationally-determined correlations between values of connected nodes. Our approach does not require a mechanistic model on how the associations are established and in each network these associations may be caused through entirely different mechanisms. In some cases the associations would be due to causal relationships between the connected nodes, such as increased expression levels of a cytokine gene linked to increased cytokine levels. However, in many cases, the associations could be indirect, mediated by elements that have not been measured in the experimental set up. In a formal mathematical model, they are considered hidden variables. Such would be the case of, for example, the changes in the metabolite levels of urine. These changes might have been caused by the colonic microbiota, in turn affected by the ileal microbiota. Since we only used the ileal microbiota data, we observe correlations

between the ileal microbial populations and the urine metabolite levels which could be in reality, indirect relationships mediated by the colonic microbiota.

### **Network of connected internal phenotypes**

The pair-wise integration method allowed us to merge the ten individual networks into a single network. Correlations within a dataset were deliberately excluded from this study because we only wanted to focus on connections between different internal phenotypes, where little work has been done. Thus in the ten networks, all detected connections are between two different data types and every node has a zero clustering coefficient. However, in the merged network, a non-zero clustering coefficient emerges as a result of nodes connecting to multiple data types (Table 2). This emphasizes the biological relevance of this method because the ten networks were built without any information on cross-linking. Thus, we identified individual nodes that directly or indirectly participate in processes of the other four individual networks. Because they seem to connect different internal phenotypes, we denoted them ‘Connectivity Hubs’. Starting the procedure as developed and applied here with networks with non-zero clustering coefficients (correlating within a dataset) would, however, not alter the connections between internal phenotypes.

### **Functional validations of phenotype connections**

Results of the text-mining were used to validate some of the identified links. This revealed insights into the mechanistic relationships between the variables predicted to be linked to each other. 37 of the 577 (6%) computational inferred links have already been described in literature as detected by our text-mining approach, which was not exhaustive because it focused only on text in journal abstracts. This indicates that our method identifies currently known biological interactions. The rest of the predicted links have not been discovered and investigated yet, have not been mentioned in abstracts, or do not exist in the biological system. Furthermore, by inspecting some of the retrieved abstracts and corresponding articles, we were even able to find causal relationships between some of the computational identified nodes where one of the nodes was used as an experimental perturbation and the other node was measured as a response parameter. Several indirect associations were also validated through reports on experiments where nodes, found to be connected in this study, were measured in response to another perturbation. During text-mining, in order to

retrieve as many results as possible, search terms were matched against the MeSH thesaurus, irrespective of the organism, and all the synonyms were included in the search. The downside to this approach is the inclusion of several false textual associations. The most striking case is that of the identified association between Glutathione and I117c. In the literature results, the reported association is between Glutathione and I117a and not I117c. Through the thesaurus, I117c was mapped to I117 and subsequently to I117a thereby giving rise to that falsely identified association in literature.

In order to increase the precision and recall of text mining searches, and overcome problems associated to the use of a thesaurus, one needs to move from mining text, to mining the knowledge embedded in the text and the use of data hidden in public databases. Such an approach requires the use of knowledge management tools and representations that can be automatically accessed (Antezana, Kuiper, and Mironov 2009). Semantic web technologies represent a new class of tools that include natural language processing, ontologies, machine learning algorithms and much more to facilitate integration knowledge from heterogeneous sources. The expansion of the use of semantic technologies in the life sciences domain will allow associating concepts such that inferences on causality, regulation, organism or tissue can be made using high-throughput methods and automated reasoning.

Among the interactions retrieved from the automated literature search, a high prevalence of associations involving cytokines and/or metabolites was observed. In fact, such type of interactions represent 97% of the retrieved results. This probably highlights the extraordinary amount of work that has been done in these types of data in the past. On the opposite extreme, only 8% of the retrieved interactions involved associations between the expression of genes, reflecting the fact that most of the available gene expression data originates from genome-wide techniques. In such type of experiments, papers, especially abstracts, usually report on systems behaviors and pathways and less frequently on the individual behavior or role of individual genes and connected response nodes.

**Validated connectivity hubs**

Even though we only performed integrations of two datasets at a time, we find data-points (metabolites, cytokines, genes or microbial groups) that correlate with different types of data. We identified 45 connectivity hubs in the merged network that seem to have associations with all four types of data. More than 30% of them are involved in links that were retrieved in literature. To further support the biological relevance of identified multi-level connectivity's we discuss the implications of two of the 15 biologically validated connectivity hubs as examples. The two connectivity hubs were chosen because of the large amount of literature results for these hubs. The first hub, *Tnfa* has the highest number of literature results among all the nodes in the network and the other hub, Glutathione, has literature validations to the most number of data-types.

*Tnfa* is a connectivity hub in the merged network, with links to several neighbors belonging to the four other types of data. The position of this cytokine in our merged network shows that it plays a role in processes of the other internal phenotypes. The literature validated links are between *Tnfa* and two other types of data (Metabolomics Serum, Microbiota). Many of the validated links represent causal relationships. With regards to immune responses and as a drug target, *Tnfa* has been studied in great detail (Cicha and Urschel 2015). The un-validated edges show that *Tnfa* could be a regulator of other internal phenotypes as well, than currently known.

The metabolite Glutathione (GSH) was measured in the serum and in the merged network is a connectivity hub proving that it is vital part of the system that connects several internal phenotypes. Among the 15 connectivity hubs with functionally validated links, GSH is the only one that has validated links to all other data-types based on our literature mining. These results support our claim of GSH being a connectivity hub, a biological component influencing several internal phenotypes. Several PubMed results for GSH are from in-vivo studies where GSH was administered to alleviate symptoms of a disease. Our literature results show that GSH has been studied in relation to all different types of data. Of the six validated links in our merged network, five represent proven causal relationships (see Figure 4 and discussion of the functional validation). These neighboring nodes in the merged network are mostly related to immune and homeostatic mechanisms. GSH is a tripeptide,

ubiquitously distributed in living cells and plays an important role in the intracellular defense mechanism against oxidative stress (Couto, Wood, and Barber 2016; Diaz-Vivancos et al. 2015). It is known that GSH metabolism is very important for the antioxidant and detoxifying action of the intestine. It is also essential for the maintenance of the luminal thiol-disulfide ratio involved in regulation mechanisms of the protein activity of epithelial cells (Iantomasi et al. 1997 which could be important since the intervention is changes in protein. Our results also demonstrate the manifold and central role of GSH when it comes to proteins, peptides and amino acids in nutrition. These observations indicate that the presented merged network represents, at least in part, associations of biological phenomena.

### **Potential relevance of selected connectivity hubs**

There are 30 connectivity hubs in the merged network that do not co-occur with their connected nodes in our literature search. However, the prominence of these nodes in our merged network indicates that they could represent potential relevant interactions with components of the other internal phenotypes. In order to demonstrate how the results of this study may be used to hypothesize on functional relationships between different molecular components, we here describe the potential biological relevance of two highly linked connectivity hubs, Tmem72 and S24-7. Both hubs are not yet described in literature abstracts in conjunction with other data-types.

The high number of connectivity hubs in the Transcriptomics layer suggest that the expression of several intestinal genes is involved in many more interactions than currently known. None of the observed Transcriptomics connectivity hubs popped-up in our literature mining results. The most highly connected Transcriptomics node, Tmem72 (Transmembrane Protein 72), has only been studied in the kidney so far (Habuka et al. 2014 and not much information is available on it. But in the merged network this node has 27 links to other data-types (can be visualized in Supplementary File 1), mostly to metabolites from both the metabolomics datasets. Based on this, we hypothesize that Tmem72 is not specific to the kidney and that it has some sort of communication function in intestinal mucosa as well. The fact that Tmem72 is a transmembrane protein is supportive for this. Given its observed links with different microbiota, metabolites and cytokines, it might be

involved in diverse interactions with other internal phenotypes. Based on such a hypothesis, targeted experimental designs may be developed in order to investigate the hypothesized “communication” function of Tmem72 in intestinal mucosal tissue.

The most highly linked node of the merged network is the bacterial family classification, S24-7, suggesting an important role for this species in gut functionality. In some of the inferred individual correlation networks we already found it to be linked to a high number of nodes. Unfortunately, this node is not represented in literature abstracts together with the here observed neighbors. However, there is compelling literature that shows this microbial classification to be a significant part of the gut microbial community structure (Harris et al. 2014; Jakobsson et al. 2015). This family classification does not have a good functional definition, yet several studies show that it could be an important player in the functionality of the gut (Evans et al. 2014; Harris et al. 2014; Rooks et al. 2014). The latter claims are in line with the high number of neighbors that S24-7 has in our merged network. The current technical inability to cultivate S24-7 is most certainly due to the absence of knowledge on S24-7 interactions. However, a recent in-silico study (Ormerod et al. 2016) shows that S24-7 species have the ability to survive on different types of carbohydrate sources, similar to the genus Bifidobacteria. In the merged network, the connectivity hubs S24-7 and Bifidobacteria, share the highest number of neighbors (directly linked nodes). Among them are 16 genes, and neither S24-7 nor Bifidobacteria have literature results with any of these genes. An enrichment analysis on these shared network gene neighbors shows that they are involved in functions related to linoleic and linolenic acid metabolism (data not shown). It is known that these fatty acids are produced by Bifidobacteria (Teran et al. 2015) and that they are involved in the maintenance of the epidermal barrier function (Muñoz-García et al. 2014). The observation that in our network these genes are shared between S24-7 and Bifidobacteria underscores the here hypothesized importance of S24-7 and indicates that these two bacterial groups are indeed closely related in function as hypothesized before (Ormerod et al. 2016).

From the results described in this paper, we conclude that we successfully developed methodologies to identify components in internal phenotypic layers that are connected to components in other internal phenotypic layers. By integrating multi-scale quantitative (-omics) data using a regression approach, we were able to provide provisional insight into

potential ways internal phenotypic layers are connected to each other, including those between local and systemic layers. By a technical and functional validations, we underscored the relevance of our findings. Based on data generated by this type of integrated approaches, hypothesis driven and targeted research may be developed to identify causal relationships between various biological scales in order to diminish our knowledge gap between genotype and external phenotype. In addition, by expanding comparable approaches by incorporating data on genetic diversity and/or variation in external phenotypes, this knowledge gap may be even further closed down. The analysis pipeline that we developed is very general and can easily be applied to any other type or number of data sets.

### **Availability of data and materials**

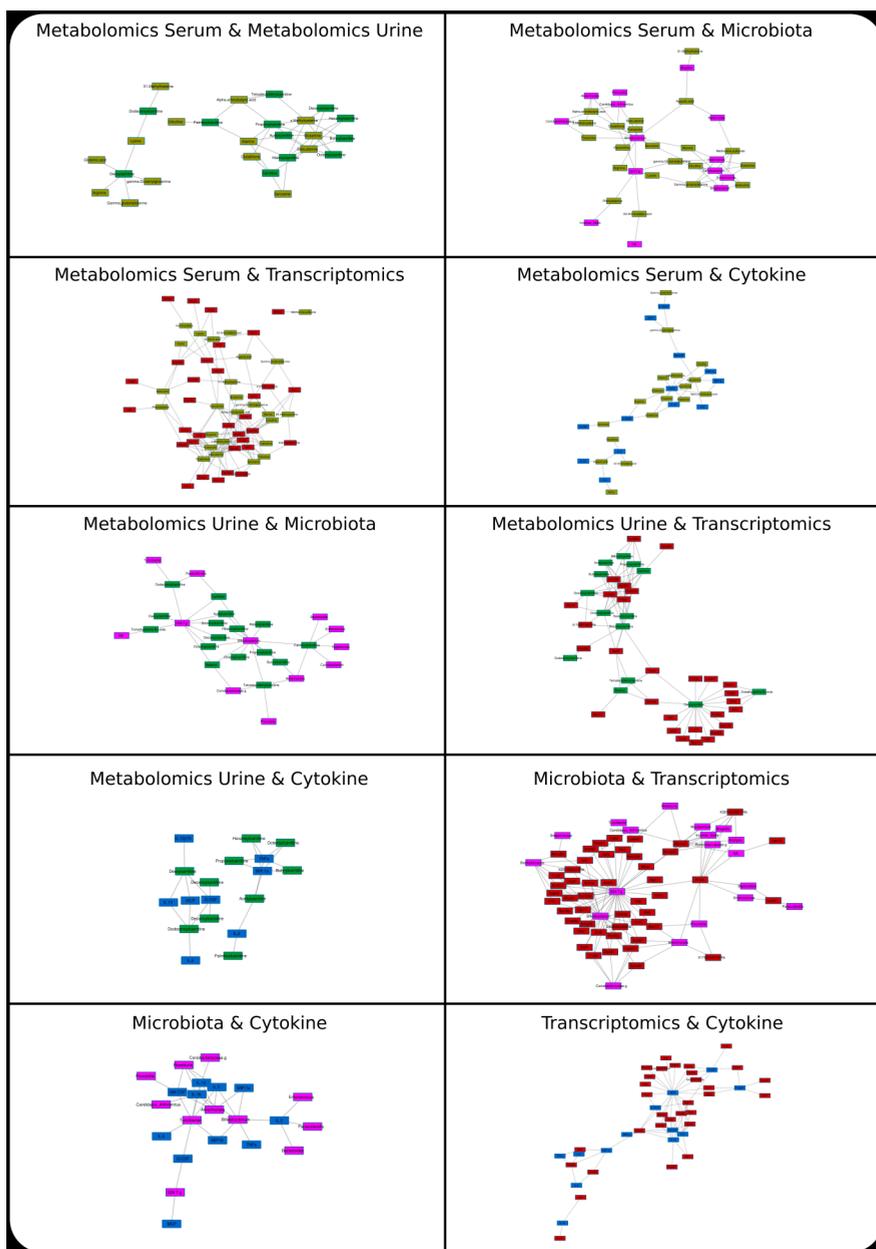
Transcriptomics data has been uploaded into GEO with the accession number GSE84442.

The microbiota data, the two metabolomics datasets and the cytokine data are available on request.

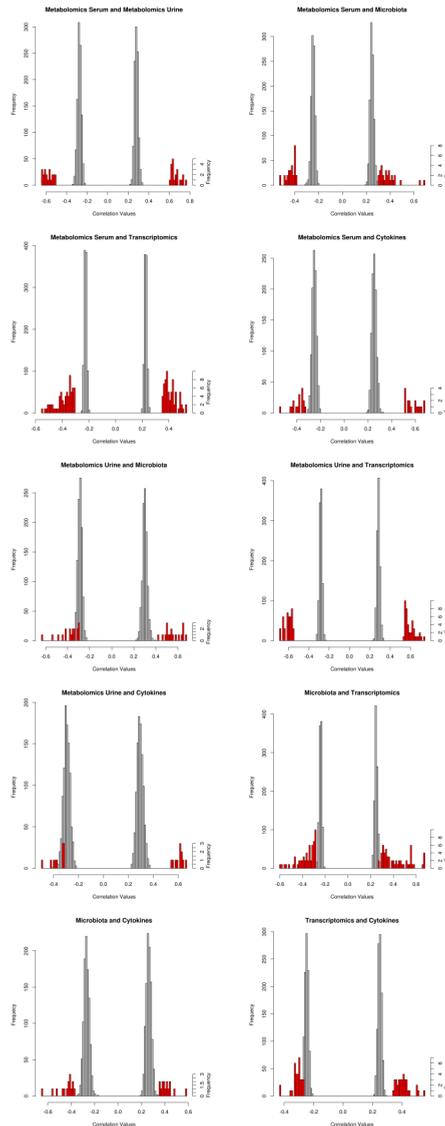
The R scripts using functions from existing R packages are also available on request.

### **Supplementary files:**

All supplementary files are also available at <http://semantics.systemsbiology.nl/index.php/data-sets-used-in-publications/>



**Supplementary Figure S1. 10 correlation networks.** The images show the 10 networks from pairwise integration of 5 datasets. The colours indicate different types of data, pink is Microbiota, red is Transcriptomics, blue is Cytokines, and green is Metabolites (light green – Metabolite Serum and dark green – Metabolite Urine).



**Supplementary Figure S2: Distribution of network correlations and random network cut-offs of all 10 networks.** The x-axis of all these graphs has the range of correlation values, the y-axis the frequency of the correlation values. The grey bars denote the distribution of the cut-offs of the thousand random correlation networks. The red bars are distributions of the correlation values of the real networks. The y-axis on the left has the frequency of the random network cut-offs and the y-axis on the right has the frequency of the real network correlations.

### **Supplementary File 1**

This file is a zipped file that contains the 10 individual networks and the merged network that can be loaded into Cytoscape.

**Supplementary Table 1: Connectivity hubs.** This table has the names and details of 45 nodes from the merged network that have connections to the other four types of data. Their degree in the network is also shown in the third column.

**Supplementary Table 2: Literature connections.** This table shows the 37 literature connections that are found in the co-occurrence search of the merged network edges.

All the supplementary data is available at <http://semantics.systemsbiology.nl/index.php/data-sets-used-in-publications/>



# Chapter 4

## Building pathway graphs from BioPAX data in R

**Nirupama Benis**; Host Microbe Interactomics, Wageningen University & Research.

**Dirkjan Schokker**; Wageningen Livestock Research, Wageningen University & Research.

**Mari A Smits**; Host Microbe Interactomics; Wageningen Livestock Research; Wageningen Bioveterinary Research, Wageningen University & Research.

**Maria Suarez-Diez**; Systems and Synthetic biology, Wageningen University & Research.

**Published in**

F1000Research 2016, 5:2414



### **Abstract**

Biological pathways are increasingly available in the BioPAX format which uses an RDF model for data storage. One can retrieve the information in this data model in the scripting language R using the package `rBiopaxParser`, which converts the BioPAX format to one readable in R. It also has a function to build a regulatory network from the pathway information. Here we describe an extension of this function. The new function allows to build graphs of entire pathways, including regulated as well as non-regulated elements, and therefore provides a maximum of information. This function is available as part of the `rBiopaxParser` distribution from Bioconductor.

### **Keywords:**

BioPAX, RDF, R, `rBiopaxParser`, Reactome

---

## Introduction

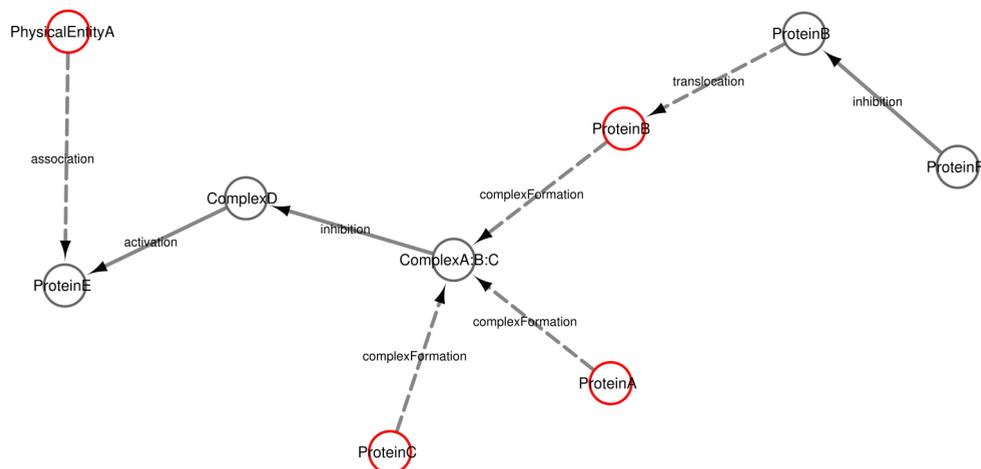
Biological pathways represent signalling and/or metabolic events involving protein and non-protein molecules. They are increasingly used in gene and protein expression studies to provide an aggregate score for gene sets encoding for defined biological events (Mitrete et al. 2013). Several pathway databases, either curated or not, have adopted the BioPAX [RRID:SCR\_009881] (Biological Pathway Exchange) language as a standard for pathway representation using the RDF (Resource Description Framework) data model (Demir et al. 2010).

The structure of BioPAX is founded upon groupings, called classes, for physical entities and interactions with hierarchical networks of their sub-classes. Interactions between physical entities are represented such that conjoint interactions may form a specific pathway with defined, but different types of interactions between the involved physical entities. The BioPAX format is being actively developed, with BioPAX level 2 format focusing on metabolic pathways and BioPAX level 3 introducing full support for signaling pathways.

SPARQL (Simple Protocol And RDF Query Language) is a query language able to retrieve and manipulate data stored in RDF. Pathway information is often combined with statistical data analysis using tools such as R (Kramer, Bayerlová, and Beißbarth 2014). The `rBiopaxParser` [RRID:SCR\_002744] 4 is an R package to retrieve data stored in a BioPAX RDF format. It comes with several options that are useful to probe the data and extract specific information from it, for example participants of a pathway, stoichiometric conditions to be fulfilled for an interaction, etc.

One such option is the `pathway2RegulatoryGraph` (P2RG) function that converts a pathway into a graphical structure. This is extremely useful for visual representation and subsequent graph-based network analysis. The P2RG function returns the parts of a pathway that are regulated (activated or inhibited) by proteins or protein complexes which is important to understand the role of regulated proteins in pathways. Here we present an adaptation of P2RG, denoted `pathway2Graph` (P2G) which can be used to build a graph of the entire pathway, including the regulated as well as the non-regulated elements. This new function expands P2RG and can be used to investigate all different types of processes and

connections of pathways instead of only studying the regulated elements of pathways. P2RG retrieves regulatory interactions, such as inhibitions and activations, the new P2G additionally retrieves protein modifications, such as translocations or complex formation, like the ones shown in Figure 1.



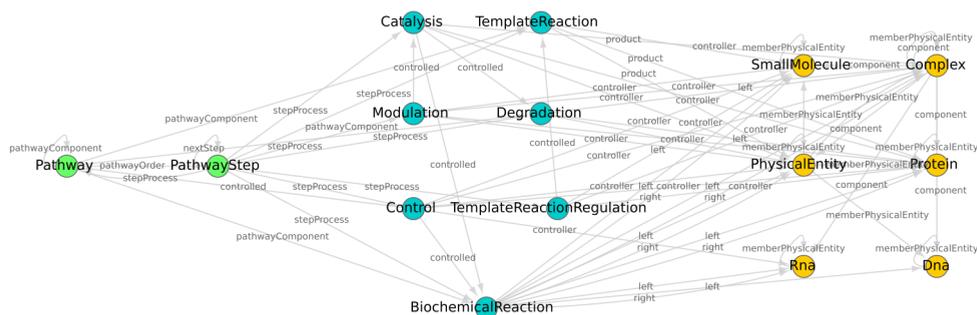
**Figure 1. Hypothetical pathway.** This cartoon of a pathway shows examples of nodes and edges that could be encountered in a BioPAX database. The nodes are proteins, complexes or other physical entities and the edges are interactions between the nodes, that represent either interactions among proteins or protein modifications. The solid edges are those detected by the P2RG function and the solid and dashed edges are detected by the P2G function.

P2G is specifically aimed at retrieving results from Reactome BioPAX level 3. In this paper we describe detailed information on this function which, we believe, will help rBiopaxParser users to better understand the graphs generated from pathway information. We have verified P2G results by directly querying the original BioPAX data using SPARQL.

## Methods and Results

The classes of PhysicalEntity and Interaction that are used in Reactome v51 to represent information on pathways are shown in Figure 2. This graph was generated using the tool RDF2Graph 5 on the Reactome Level 3 RDF file. The nodes in Figure 2 represent classes and the edges show the possible relationships, called predicates, these classes could have in the database. As depicted in Figure 2 the node Pathway could have one or more Path-

waySteps that consist of different types of Interaction sub-classes. All the Interaction nodes shown in Figure 2 describe interactions between PhysicalEntities, hence are connected to them by particular types of predicates as indicated in the edge labels. The Interaction classes are interconnected because they can be dependent on each other. The Control interaction and its sub-classes (Catalysis and Modulation) represent signaling events. They regulate BiochemicalReaction and Degradation interactions which mostly represent metabolic reactions.



**Figure 2. Interplay of classes in Reactome BioPAX.** This figure shows a network of the Interaction and PhysicalEntity classes that are a part of a pathway in Reactome v51 BioPAX level 3. Nodes are classes and the directed edges are links between them in the database. The green nodes are the Pathway and PathwayStep classes, the blue nodes are Interaction classes and orange nodes are PhysicalEntity classes.

To create a regulatory graph, the P2RG function starts with the Control, Catalysis and Modulation interactions that are either activating or inhibiting other interactions. This method provides a graph with plenty of information on the regulatory components of the pathway. The nodes of this graph are physical entities like Proteins or SmallMolecules and the directed edges are either activation or inhibition events. An example of such a reconstruction is shown in Figure 1, where P2RG is able to retrieve the black nodes and the continuous edges. However, interactions can be missed if they are not regulated by the Control interactions and could result in the loss of valuable information in the graphical representation of the pathway.

The new function P2G can start with any type of interaction in order to obtain a graph with all possible physical entities involved in the pathway. Similar to the result of the

P2RG function, the P2G function gives a graph with nodes that are physical entities, but the edges are not strictly activation or inhibition events. The directed edges could represent several types of events like translocation of a protein or cleavage of DNA, these are shown as dis-continuous edges in the cartoon in Figure 1. The P2G function recognizes the continuous and the dis-continuous edges and thus retrieves the black as well as the red nodes shown in Figure 1. In some cases there is more than one documented connection between the same physical entities. In this case only the first connection is used as an edge in the final pathway graph.

### **Comparison of two methods: P2G vs P2RG**

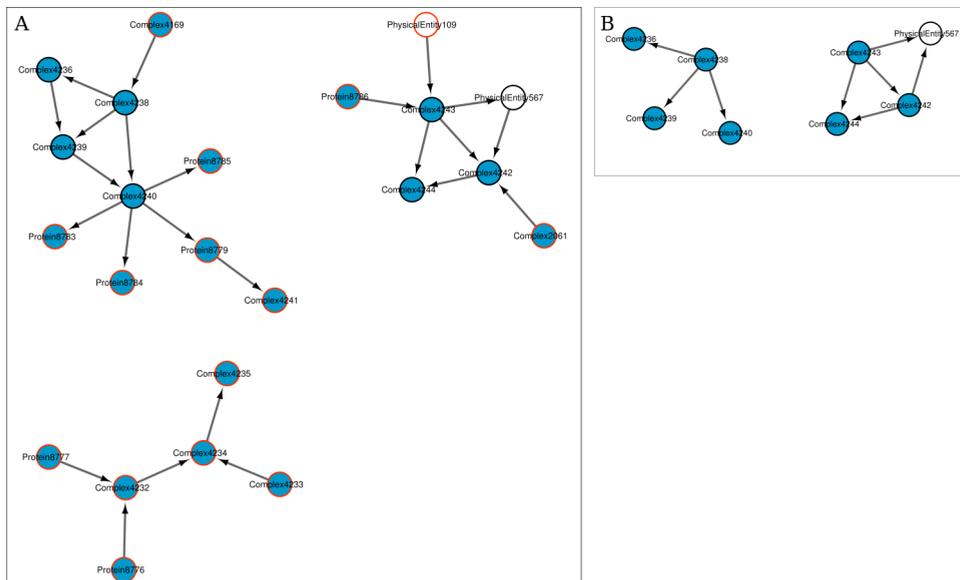
The Reactome database (v51) categorizes pathways into 27 branches. Here we worked with pathways that have more than one interaction, which resulted in 1,666 pathways. Using P2RG, graphs for 1,548 pathways were retrieved. By using the new P2G function, we were able to retrieve information on all 1,666 pathways. The highest number of pathways were obtained, using either method, in the “Disease” category (P2RG: 3,396 pathways, P2G: 4,888 pathways). In 85% of the cases, pathways retrieved using P2G consisted of more physical entities (nodes) than those retrieved using P2RG. 19% of the P2G retrieved pathways have at least twice the number of nodes, and 60% have at least twice the number of interactions between nodes (edges) as compared to the P2RG version, Figure 3 is an example of this difference. Total numbers of nodes and edges in major Reactome categories are given in Table 1. Missing information causes the appearance of disconnected graphs when reconstructing pathways. By using the new P2G function, the percentage of disconnected pathways is reduced by 9%. Additionally, P2G also has the option of only retrieving the largest connected component, for example with this option, in Figure 3.A only the top left part of the graph will be retrieved and the disconnected graphs discarded. The pathways have directed edges because most of the interactions have direction. Edges without a direction are represented as bidirectional edges in the output of P2G.

**Table 1. Numbers of nodes and edges.** The number of nodes and edges of ten different pathways (Reactome Categories) are indicated as obtained after application of P2RG and P2G on the same set of BioPAX RDF information.

Reactome Categories	P2RG Nodes	P2RG Edges	P2G Nodes	P2G Edges
Binding and Uptake of Ligands by Scavenger Receptors	0	0	68	56
Cell-Cell communication	13	14	142	142
Disease	3,396	5,878	4,888	12,159
Gene Expression	652	900	1,110	2,450
Immune System	1,431	2,233	2,419	5,045
Membrane Trafficking	86	121	181	382
Metabolism	3,082	5,922	3,479	11,289
Signaling Pathways	2,069	3,274	3,430	7,131
Steroid hormones	72	147	81	333
Transcription	281	420	623	1,324

As an example, we discuss here the ‘Apoptosis induced DNA fragmentation’ pathway, which is a pathway in the ‘‘Programmed Cell Death’’ category (Figure 3). When the information in the BioPAX file is reconstructed with the P2RG function, the pathway has seven nodes (Figure 3.B); with the P2G function the same pathway has 16 more nodes (Figure 3.A). Detailed information on these nodes, as retrieved with P2G and P2RG, is given in Table 2 and demonstrates the additional information retrieved by P2G. The node ‘Complex4169’, which is found in the cytosol, translocates to the nucleus where it is called ‘Complex4238’. However, this information is only available from the P2G function because the node ‘Complex4169’ does not regulate any other interaction or node. The presence of extra nodes in the P2G retrieved graph (Figure 3) also visualize that ‘Complex4240’ breaks up into its’ individual components after being cleaved by Caspase-3 (‘Complex4238’). This extra information is very useful for researchers analyzing the

phenomen represented by the pathway. In case P2G retrieved pathway graphs are used for analysis (e.g, differential gene expression analysis) the presence of these extra nodes may improve biological interpretation of experimental data.



**Figure 3. Graphs of the pathway ‘Apoptosis induced DNA fragmentation’.** Both graphs were extracted from the same BioPAX file. A) Graph recovered using the new P2G function; B) Graph recovered using P2RG function. In both panels blue nodes are proteins or protein complexes, white nodes are non-protein entities. Black encircled nodes are found in both graphs and red encircled nodes are only detected with the new P2G function. Names of the nodes are in Table 2.

**Table 2: Node names and locations of the “Apoptosis induced DNA fragmentation” pathway.** The first column has the names of the nodes in the pathways depicted in Figure 3. The second column has the actual name of the node and the third column the cellular location of the node. All this information is represented as given in Reactome version 51. The nodes shown with a black outline in Figure 3 are shown here in bold font.

<b>Node</b>	<b>Name</b>	<b>Location</b>
Protein8776	DFFB	Cytosol
Protein8777	DFFA	Cytosol

Complex4232	DFFA : DFFB	Cytosol
Complex4233	Importin alpha : Importin beta	Cytosol
Complex4234	DFF : associated with Importin alpha : Importin beta	Cytosol
Complex4235	DFF : associated with Importin alpha : Importin beta	Nucleoplasm
Complex4169	Active CASP3	Cytosol
Complex4238	Active CASP3	Nucleoplasm
Complex4236	DFFA : DFFB	Nucleoplasm
Complex4239	Caspase cleaved DFFA	Nucleoplasm
Complex4240	Caspase cleaved DFFA : DFFB	Nucleoplasm
Protein8779	DFFB	Nucleoplasm
Protein8784	DFFA fragment	Nucleoplasm
Protein8785	DFFA fragment	Nucleoplasm
Protein8783	DFFA fragment	Nucleoplasm
Complex4241	DFFB homodimer	Nucleoplasm
PhysicalEntity567	DFFB homodimer/homooligomer	Nucleoplasm
Complex2061	Histone H1 bound chromatin DNA	Nucleoplasm
Complex4242	DFFB associated with chromatin	Nucleoplasm
Protein8786	HMGB1/HMGB2	Nucleoplasm
PhysicalEntity109	DNA	Nucleoplasm
Complex4243	HMGB1/HMGB2 – bound chromatin	Nucleoplasm
Complex4244	DFF cleaved DNA	Nucleoplasm

## **Conclusion**

P2G is a useful addition to the rBiopaxParser package because it retrieves all the components of a pathway from the database and provides complete graphical information for both signaling as well as metabolic pathways. The P2G function (pathway2Graph) is currently available in the rBiopaxParser package in the Bioconductor 3.4 release.

## **Data availability**

The input data for this package is the BioPAX format of any pathway database. We used the Reactome database which is freely available for download in different formats from the website [www.reactome.org](http://www.reactome.org). A subset of this database is given as Supplementary file 1.

## **Software availability**

The function pathway2Graph is available in the latest version of the R package rBiopaxParser and can be installed from Bioconductor.

Archived source code as at the time of publication: <http://dx.doi.org/10.5281/zenodo.616186>

Software license: GPL-2

## **Grant information**

This work has been financially supported by the Systems Biology Investment Programme of Wageningen University, KB-17-003.02-022. Frank Kramer's work is funded by the German Ministry of Education and Research (BMBF) grants FKZ01ZX1508 and FKZ031L0024A.

**Supplementary material****Subset of Reactome database.**

This .owl file contains information on four pathways from the Reactome v51 BioPAX level 3 database. This format can be loaded into the R environment using the rBiopaxParser package and used to test the P2G function and obtain graphs which were used as the basis for Figure 2. More information on loading and processing this file format can be found in the package documentation.

Supplementary material can be accessed at <https://f1000research.com/articles/5-2414/v2>

# Chapter 5

## Functional plasticity of murine intestinal mucosal tissue.

**Nirupama Benis**; Host Microbe Interactomics, Wageningen University & Research.

**Mari A Smits**; Host Microbe Interactomics; Wageningen Livestock Research; Wageningen Bioveterinary Research, Wageningen University & Research.

**Jerry M Wells**; Host Microbe Interactomics, Wageningen University & Research.

**Soumya Kanti Kar**; Host Microbe Interactomics, Wageningen University & Research.

**Bart van der Hee**; Host Microbe Interactomics, Wageningen University & Research.

**Maria Suarez-Diez\***; Systems and Synthetic biology, Wageningen University & Research.

**Dirkjan Schokker\***; Wageningen Livestock Research, Wageningen University & Research.

\* - These authors contributed equally

**Submitted for publication**



**Abstract:**

The mammalian intestine is a complex biological system that plays an important role in the processing of a high diversity of nutritional and environmental stimuli. These stimuli may alter the functioning of the intestine and elicit different responses from the intestinal mucosal tissue to maintain gut homeostasis. This variety of responses requires a high functional plasticity for the intestinal mucosal tissue. In order to improve our understanding of the underlying mechanisms of this plasticity, we performed a high-level data integration of 14 whole-genome transcriptomics datasets from samples of intestinal mouse mucosae. The data integration was not performed at the level of individual gene expression but on pathway expression. The gene expression datasets were obtained from open access databases and derived from 14 independent experiments that satisfied all our selection criteria. We used a customized version of the pathway analysis tool CePa (Centrality based Pathway Analysis), along with pathway information extracted from the Reactome database. The results of pairwise comparisons were combined to obtain an integrated picture of gut mucosal responses in mice, orally challenged with biological, chemical or dietary agents which are expected to perturb intestinal homeostasis.

The results of this high-level integrated pathway analysis show that the mucosal tissue uses a common set of pathways to respond to the three classes of stimuli. Based on these results we designed and executed an *in vitro* experiment on intestinal organoids to verify that the pathway modulation is the result of local changes. We conclude that the functional plasticity of intestinal mucosal tissue can be explained, at least partially, by variation in the pathway sets that are regulated by different challenges and by the induction of intra-pathway specific gene expression patterns.

**Keywords:**

Gastrointestinal tract, gene expression, functional plasticity, pathway analysis

---

**Introduction**

The mammalian gastro-intestinal tract is crucial for the digestion and absorption of nutrients, while maintaining a peaceful co-existence with the microbial symbionts and avoiding infection from enteric pathogens. The intestinal epithelium plays an important role in orchestrating innate defences (Wells et al. 2017; Peterson and Artis 2014) and signalling to the numerous cells of the immune system located underneath the epithelial layer (Wells, Loonen, and Karczewski 2010; Akira and Takeda 2004; Sharma, Young, and Neu 2010). The chemical and physical defences of the intestinal mucosa, such as antimicrobial peptides, REG3 proteins, mucins and secreted mucus, often referred to as the “gut barrier,” provide the first line of defence against colonisation and invasion by intestinal bacteria (Lodish et al. 2000; Henderson et al. 2011; Peterson and Artis 2014; Wells et al. 2017; Rossi, van Baarlen, and Wells 2011). Besides these innate defence mechanisms, a number of tolerance mechanisms exist to maintain homeostasis and prevent chronic inflammatory responses to antigens from food/feed or commensal microbes (O’Hara and Shanahan 2006). Chronic perturbations in homeostasis can lead to hyperpermeability of the gut barrier, contributing to the pathogenesis of several gastrointestinal disorders, including IBD, celiac disease, IBS, and food allergy (Bischoff et al. 2014).

The intestinal mucosa is exposed to a variety of different components: diverse types of nutrients (including proteins, fats, carbohydrates, minerals, vitamins); commensal microbiota (bacteria, viruses); (opportunistic) pathogens and orally taken drugs such as antibiotics. After passage through the stomach, food is digested by pancreatic and small intestinal enzymes. Microbiota in the small intestine also metabolise nutrients generated by the digestive process and produce certain vitamins absorbed by the host. Changes in the diet composition, e.g. high fat diets, can drastically change the community structure of the intestinal microbiota (Serino et al. 2012), but also the mucosal gene expression of the host (de Wit et al. 2011). Dietary fibres which are not digested in the small intestine are fermented by a consortia of microbiota in the colon, resulting in the production of short-chain fatty acids and other metabolites which impact the intestinal physiology and immunity (Van den Abbeele et al. 2011; C. J. Kelly et al. 2015; Burger-van Paassen et al. 2009; P. M. Smith et al. 2013; Segain et al. 2000). Gastrointestinal physiology is also known to be sensitive to the administration of certain drugs (Willing, Russell, and Finlay 2011;

Niewold 2007; Schokker et al. 2014; Schokker et al. 2015, emotional stress (Emeran A Mayer, Savidge, and Shulman 2014; Collins and Bercik 2009, and many other environmental factors.

The gut functionalities as described above, cannot be explained by the activity of single genes alone, they are often attributed to groups of genes (Mootha et al. 2003 based on their annotation. Such gene-sets (Ashburner et al. 2000 or pathways (Kanehisa and Goto 2000; Croft et al. 2014; Fabregat et al. 2016 can be regarded as representations of functional building blocks that, in different combinations, may be used to respond to challenges such as infections or toxins, and restore homeostasis once the threat has subsided. We hypothesized that each type of challenge requires a different combination of pathways, each at a specific expression level, to restore homeostasis.

An adequate response of the intestinal mucosa to the exposure to diverse components, requires a high-level of functional plasticity that is expected to be much higher than the liver and kidney which are specialised for a particular function. This theory is supported by data present in the human protein atlas (Uhlén et al. 2015; Fagerberg et al. 2014, which provides an overview of gene expression profiles in different tissues. Genes mainly expressed in specialized tissues like the heart, liver and kidney are related to a narrow range of functions. The heart tissue specifically expresses genes related to contraction and regeneration functions (Lindskog et al. 2015. Genes with higher expression in the kidney than other measured organs are involved in transport and absorption functions (Habuka et al. 2014 and genes with higher expression in the liver compared to other organs are involved in immune and metabolic processes and growth (Kampf et al. 2014. The number of genes, specifically expressed in gut mucosal tissues is higher than that in heart, liver, kidney and they are involved in a wider variety of functions, nutrient breakdown, transport and metabolism, enteroendocrine system, host protection and maintenance of tissue morphology (Gremel et al. 2015.

Responses of intestinal mucosal tissues to individual stimuli or perturbations have been documented widely in literature in terms of gene expression data (Kar et al. 2016; Reikvam et al. 2011; Schokker et al. 2014. However, little is known yet with regard to the differences and commonalities in the functional genomic information used by the mucosal

tissue to display its extreme functional plasticity. A major driver in the understanding of this plasticity is the current trend to develop (dietary) interventions to stimulate intestinal health. Therefore, it might be useful to investigate the functional plasticity of mucosal tissues at the genomic level in terms of pathways. Such an approach may help to identify key sets of biosynthetic and signalling pathways involved in the various mucosal responses and to identify the commonalities and differences in the use of pathways between the various responses.

The objective of this study was to identify pathways expressed in intestinal mucosal tissue that are regulated by different types of challenges (e.g. dietary, infectious/immune, and drugs) in order to identify commonalities and differences in the use of pathways to display the challenge-specific tissue responses. The ultimate aim is to understand the functional plasticity of the mucosal tissue in more detail. To perform this study, we used publicly available gene expression datasets generated from mouse intestinal tissues exposed to orally administered challenges.

## Materials and Methods

### Datasets

With the R tool *GEOMETadb* (Zhu et al. 2008) we searched Gene Expression Omnibus (GEO) (Edgar, Domrachev, and Lash 2002; Barrett et al. 2013) for publicly available datasets generated from intestinal samples of mice. Among the 450 experiments available (as of 07-07-2015) we only selected 14 experiments (17 GEO datasets) on wild-type mice where the tissue of any section of the intestine was sampled after weaning. These experiments all used single channel microarrays with at least 3 biological replicates and were published on GEO between 2006 and 2014. Except for 3 of the datasets that were measured using Illumina microarrays, all the others were performed on different versions of Affymetrix. Most of the experimental data were obtained from inbred C57BL6J mice, but there were two experiments that used BALB/c mice, one that used the 129S1/svlnj mouse strain and another that used IQI mice.

There were 2 experiments where the fat content of the diet was increased by reducing the carbohydrate portion of the diet (de Wit et al. 2011; Desmarchelier et al. 2012). The former

experiment gave 3 different levels of fat (20%, 30% and 45% of total energy in the diet) to the mice and measured the response in 3 sections of the small intestine. The latter used fat for 60% of the energy from the diet and measured 2 sections of the small intestine. One experiment tested 5 different fibres which were substituted for part of the corn starch in the diet and the transcriptomics data was generated from the colon (Lange et al. 2015). The rest of the dietary interventions were additives or supplements to the mouse diet. In one experiment, dietary heme was added to a high fat diet and the response was measured in the colon (Ijssennagger et al. 2012). In another experiment quercetin was added to the standard diet and the response was measured in 2 parts of the small intestine (jejunum and ileum), as well as in the colon (Mutch et al. 2006). Another experiment added a synthetic PPAR $\alpha$  agonist to the diet and a microarray analysis was performed on the small intestine (Bünger et al. 2007). One experiment tested a probiotic on 2 strains of mice and 2 sections of the intestine (small and large intestine) (Mariman et al. 2015). There are 3 experiments where a drug was administered to the animals. In one of these experiments a herbal drug was tested on wild-type and specific pathogen free mice, and the response measured in the small intestine and colon (Munakata et al. 2012). The other 2 drug experiments involved administration of antibiotics. One study utilised a mix of several antibiotics to strongly deplete the abundance of gut bacteria, and investigated gene expression in the colon (Reikvam et al. 2012). In the other experiment with antibiotics a low dose of penicillin was administered daily to the animals from an early age and also changed the diet to measure the response in the small intestine (Cox et al. 2015). The Immune challenge experiments involved infectious challenge with *Salmonella typhimurium* (Liu et al. 2010) or the parasite *Giardia Duodenalis* (Tako et al. 2013). Additionally, the effect of flagellin from *Salmonella enterica* serovar Typhi (Carvalho et al. 2012) on severity of DSS-induced colitis was investigated (Ahmad et al. 2014). In the experiment involving *G. Duodenalis* the small intestine was sampled, whereas the colon was sampled for the other immune challenges. More details on the experimental conditions and the control conditions can also be found in Supplementary Table 1.

### **Data pre-processing**

Using *GEOQuery* (Davis and Meltzer 2007), we downloaded normalised datasets of pre-selected experiments from GEO. In each experiment the normalization was performed with

one of the following methods, GCRMA, RMA, MAS5 or quantile normalization. Our high-level integration approach does not require uniform normalization, so in each case we preferred the method chosen by the authors of the original study. The probes were mapped to mouse Entrez identifiers using the annotation files from the platform that was used for microarray analysis. After a quality check using Principle Component Analysis plots, these mice gene identifiers were then mapped to their human homologs using the NCBI HomoloGene database [RRID:SCR\_002924].

### Pathway Database

The analysis of all the datasets was done using pathways from the Reactome database [RRID:SCR\_003485], a freely accessible and a manually curated database available in different formats. Pathways from Reactome were downloaded in the BioPAX [RRID:SCR\_009881] (Biological Pathway Exchange) (Demir et al. 2010 format (version 51) from the official website. These pathways were then converted to a pathway catalogue object in R that can be used by the pathway analysis algorithm. This was accomplished by using the `pathway2Graph` function from the R package *rBiopaxParser* [RRID:SCR\_002744] (Benis et al. 2016; Kramer et al. 2013).

### Pathway analysis

We used a modified version of the algorithm CePa (Centrality based Pathway analysis) which uses pathways as networks where the nodes in a pathway could be small molecules (compounds), macro-molecules (proteins or RNA) or complexes (more than one protein). The topological information of the pathway is used to assign weights to each node using centralities. The user can choose between one of 4 centrality measures, in-degree (number of edges that are directed towards the node), out-degree (number of edges that are directed outwards from the node), in-reach (longest path that brings information to the node) and out-reach (longest path that directs out of the node), with another option of giving equal weight to all nodes. This centrality information is used along with the expression data to give a list of significantly enriched pathways for given conditions.

The gene expression data is mapped to the nodes of the pathways, when the node is a protein the expression value of the corresponding gene is used as such. When the node is a

complex, the largest component from a Principle Component Analysis of the expression values of all the corresponding proteins is assigned as the node expression value. Subsequently, these expression values are inputted in a t-statistic to obtain a differential expression value for each node, which can be positive or negative based on the up- or down-regulation of that protein. This differential node value is multiplied with the centrality based weight of the node to obtain a final node value. This calculation is performed for each of the nodes in the pathway and all these values are averaged to obtain a pathway level score. The pathway level score is then tested for significance by substituting random expression values in the same pathway calculations. This p-value calculation was modified from the original CePa function which randomizes the samples of the tested conditions. As we work with a minimum of 3 samples per condition, we modified this calculation to be able to handle smaller sample sizes. Because the hierarchical nature of the pathway database implies dependence between the pathways, we decided not to perform a multiple testing correction.

We weighted the nodes with the in-reach and out-reach centrality calculations, because they assign higher weights to the nodes down-stream and up-stream of the pathway respectively. In this way, we can capture signalling pathways, where the effectors are more likely to be down-stream of the pathway. However, we did not rule out the metabolic pathways where the enzymes are generally up-stream in a pathway. The threshold of the p-values was set at 0.01 to compensate for the lack of multiple testing correction.

### **Intestinal organoid cultures**

3-D organoids grown from murine small intestine as described before in (T. Sato et al. 2009; T. Sato et al. 2011; Dekkers et al. 2013). Briefly, a 2cm duodenal section was opened longitudinally and washed in ice-cold phosphate-buffered saline solution (PBS). After scraping excess villi, the tissue was transferred to PBS containing 2.5 mM EDTA and incubated for 30 minutes. Following incubation, the sections were washed with PBS and remaining residue was passed on a 70  $\mu$ m cell strainer, pelleted at 300 x g for 5 minutes, and suspended in matrigel basement membrane (Growth factor reduced, Corning) at a density of 50-100 crypts per 50  $\mu$ l. After inverted polymerization at 37  $^{\circ}$ C for > 10 minutes, 600  $\mu$ l basal culture medium (DMEM/F12) was added, enriched with mouse

EGF, Hepes 1M (Invitrogen), N-acetylcysteine (Sigma), B-27 (Thermo-Fisher), Noggin, and R-spondin. The culture was passaged 1:4 every 7 days by mechanical disruption, and re-suspended in fresh matrigel. All experiments were performed after 2 passages of the organoid cultures.

### Stimulation of organoids and reverse transcriptase-quantitative PCR.

3-D organoids were stimulated with TNF $\alpha$  (10 ng/ml), a PPAR $\alpha$  agonist (WY14643 0.1% v/v), and flagellin (200 ng/ml) for 6 hours before total RNA was extracted with the Qiagen mini-kit according to manufacturer's instructions along with a 15 minutes DNase step. Purity and integrity measurements were performed on a DS-11 spectrophotometer (DeNovix) and 1  $\mu$ g total RNA was reverse transcribed into cDNA using a QScript kit (Quantabio). Quantification of gene expression (RT-qPCR) was performed using a Rotor-gene Q2 plex RT-cycler (Qiagen) on primers specified in Table 1 with the rotor-gene SYBR green PCR kit. These genes were selected based on their representative contribution to the pathway 'Regulation of Complement Cascade'. Relative expression levels were calculated following methods described in (Schmittgen and Livak 2008 using individual amplification values, with 18S and  $\beta$ -Actin as endogenous control genes for normalization.

**Table 1. Primers used for RT-qPCR.** The information on the primers used to quantify 10 genes is given in each row along with the publication from which this sequence was obtained.

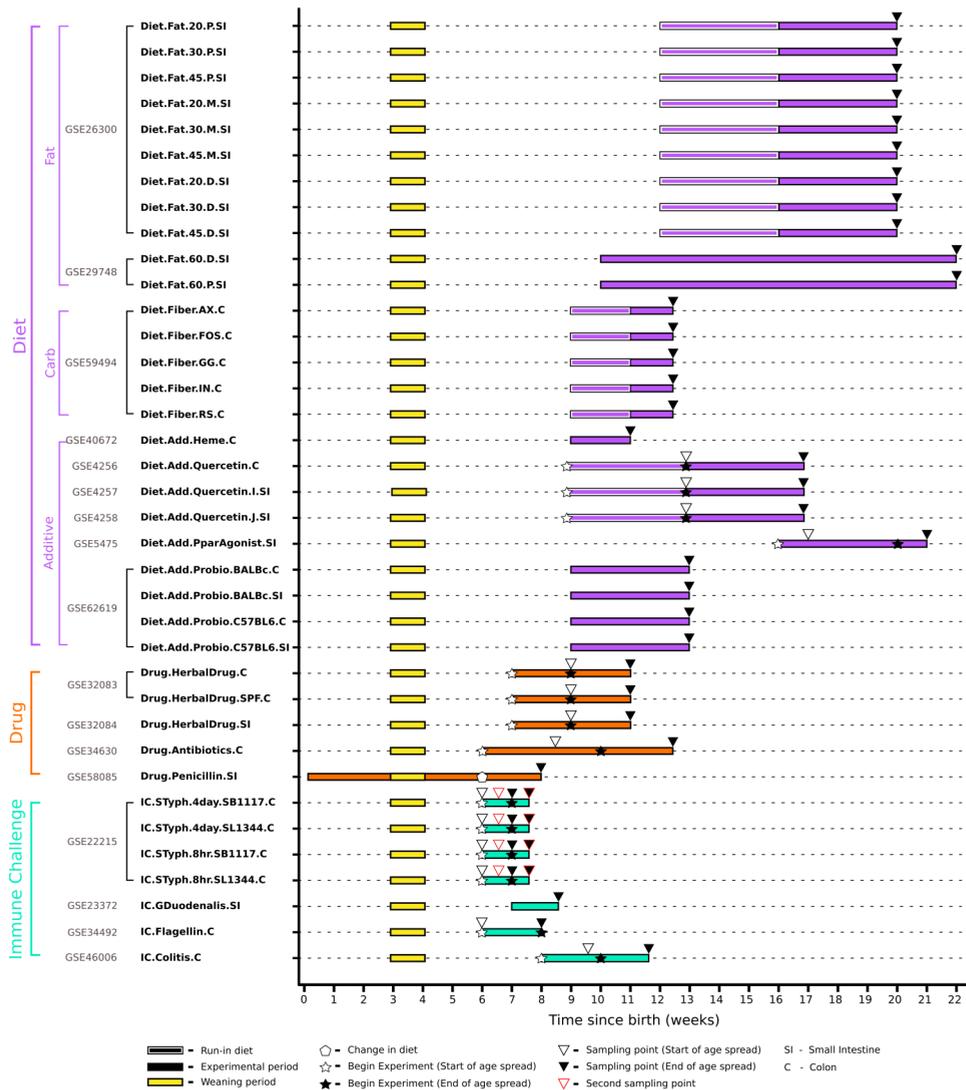
<i>Gene</i>	<i>Forward</i>	<i>Reverse</i>	<i>AT (<math>^{\circ}</math>C)</i>	<i>Amplicon (bp)</i>	<i>Ref</i>
C2	CTCATCCGCGTTTA CTCCAT	TGTTCTGTTCGATG CTCAGG	60	178	(Li et al. 2016)
C3	AGAGGTCATCAAGT CAGGC	GATGTAGCTGGTGT TGGGCT	60	167	(Li et al. 2016)
C5	AGGGTACTTT- GCCTGCTGAA	TGTGAAGGTGCTCT TGGATG	60	173	(Li et al. 2016)

<i>Cfh</i>	CGTGAATGTGGTG- CAGATGGG	AGAATTTCCACACAT CGTGGCT	60	248	(Li et al. 2016)
<i>Cfi</i>	TTCCAC- TGGGTGTTCGTGAC	TAAAGGCACAC- TCCGCCAAA	60	126	(Li et al. 2016)
<i>Cd46</i>	CCAGGGCCAGA- TAAGTTTTTC	TATTTCCGCCAGCTCC TGATA	60	153	(Li et al. 2016)
<i>Cd55</i>	CTCTGTT- GCTGCTGTCCC	CGAATAA- TATGCCGGTTG	60	477	(Nie et al. 2015)
<i>Cd59</i>	TAAGTGAG- TTCTTGCAACC	AGGGCCTGTGAA- GATTATGA	60	152	(Li et al. 2016)
<i>Cr2</i>	CCTGCTCCTCTCTG- TAAACT	GATCTGACTGCTTCC ACTCA	60	162	(Li et al. 2016)
<i>C8g</i>	CTGGTCCTT- GTGGCTGT	CGAAACTCTGGTAG- TCGGTCTC	60	257	Author

## Results

### Intestinal gene expression datasets

Based on the stimuli in the 14 experiments, we classified the interventions into three broad stimulation categories: Diet; Drug; and Immune Challenge. Seven of the experiments had dietary interventions with changes either in the composition of the diet or with additives (non-immune and non-drug) in the diet and were placed in the Diet category. In the ‘Drug’ category there were 3 experiments, one with a Japanese herbal drug and two with antibiotics. The Immune Challenge category consisted of 4 experiments performed with infectious or inflammatory stimuli. Information on the age of the animals during the intervention/challenge period and sampling time is indicated in Figure 1. These 14 experiments comprised of 37 experimental conditions, where a condition is defined as a unique combination of an inbred mouse strain, a stimulus and intestinal tissue sectioned at a specific time point, as depicted in Figure 1.

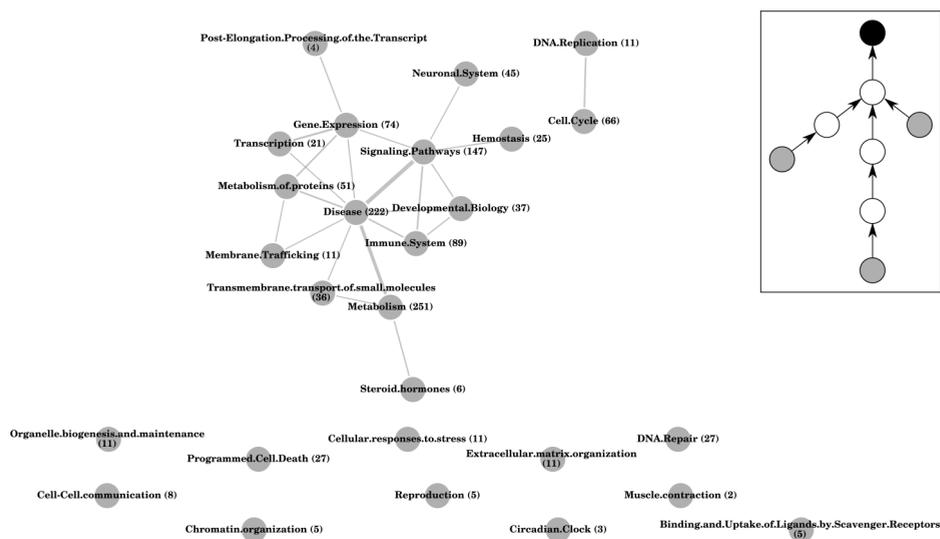


**Figure 1. Experimental datasets:** The 37 conditions from 14 experiments (with 17 GEO accession numbers) used in this study are detailed in a timeline of the age of the mice in the experiment. Mice of varying ages can be selected for the same experimental group if they weigh the same, this variation in ages is shown with empty and filled stars and triangles. The stars and triangles denote the start and end of the interventions respectively. Challenges have been divided in three categories (colour coded): Diet, Drug and Immune challenge. The names given for each dataset are abbreviated to show the challenge category in the first part of the name, the tissue sampled at the end (SI: small intestine; C: colon) and the text in the middle indicates the nature of challenge. Additional detailed

explanations for the abbreviated condition names and the control conditions are given in Supplementary Table 1.

## Pathway database

We employed the freely available and manually curated pathway database Reactome version 51. All pathways are made up of several interactions between different entities including proteins, complexes, and DNA. In the Reactome database, pathways are arranged in a hierarchy, larger ‘root’ pathways consist of more and more specific pathways, ending in several ‘leaf’ pathways. This hierarchy is depicted in a simplified cartoon in the inset of Figure 2. The main image in Figure 2 shows a network of all the root pathways in Reactome version 51. These 27 root pathways contain 1,639 pathways within their hierarchy, of which 950 are leaf pathways. All individual pathways can belong to more than 1 root pathway. These shared pathways are shown as edges in Figure 2. Disease and Metabolism are root pathways that have the largest number of leaf pathways. There are several other roots with smaller numbers of member pathways as shown in Figure 2.



**Figure 2. Network of Reactome root pathway.** The nodes in this network represent the 27 root pathways as present in Reactome v51 and the edges indicate the ‘leaf’ pathways shared by connected root pathways. The thickness of the edges indicates the number of leaf pathways shared by the nodes. The nodes are labelled with the names of the root pathways and the number of enclosed leaf

pathways is given between brackets. The inset shows a simplified example of root and leaf pathways, where the cartoon has one root pathway with 3 leaf pathways.

### Significant pathway results for all the datasets

Pathway analyses were performed on the gathered datasets using the modified version of the CePa algorithm in order to identify pathways that are specifically regulated by the challenges. Each analysis was performed on single comparisons, where a comparison is made between a stimulated condition versus the control in that experiment. In the collection of datasets used in this study, we compared 37 stimulated conditions against their respective controls. Most of the conditions (25) belong to the Diet category, 5 are from the Drug category, and the remaining 7 from the Immune Challenge category (Figure 1). Leaf pathways enriched in differentially expressed were considered significantly regulated and were compared between challenge categories. Over all 37 conditions, 1,351 pathways were significantly enriched and 710 are leaf pathways belonging to 21 root pathways. The majority of these leaf pathways (84%) were significant in both centrality measures whereas about 11% were significant only in the ‘in reach’ centrality and 5% significant in only the ‘out-reach’ centrality. An overview of these significant leaf pathways, irrespective of the centralities, is shown in the heatmap in Figure 3. The number of significantly regulated pathways for each condition is highly variable, with the maximum being 377 from the experiment with Heme (Diet.Add.Heme.C) and the minimum being 37 in the experiment where 20% of the energy in the diet came from fat (Diet.Fat.20.P.SI). The average number of significantly regulated pathways is highest in the conditions belonging to the Diet category with 132 pathways, second highest is Immune Challenge with 101 pathways followed by Drug with 87 pathways on average. Many of these results were concentrated in 3 root pathways, ‘Disease’, ‘Metabolism’ and ‘Signalling Pathways’ (see Figure 3) which have the highest number of leaf pathways in the database. The root ‘Steroid Hormones’ is significantly regulated in only 7 of the 37 conditions.

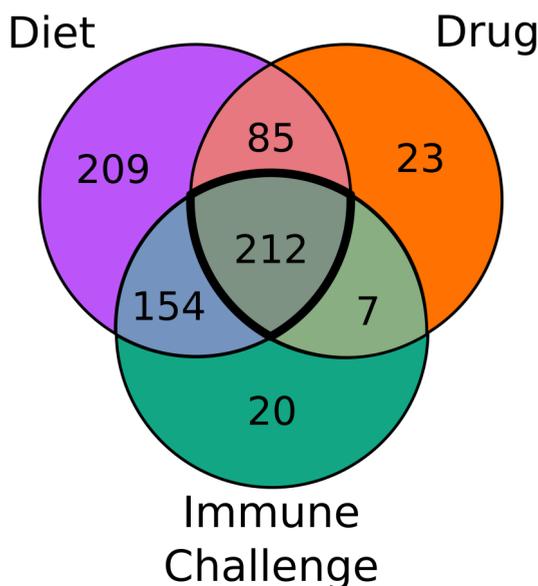


---

### Comparison of the three categories

The results of all the experimental conditions within a category were compared to the results of the other challenge categories. There was a large overlap in the results between the three classes (Figure 4). In addition, there were several pathways unique to each challenge category, the largest number of pathways were identified in the Diet category, which had the largest number of experimental conditions. The pathways ‘PKA activation’ and ‘DARPP-32 events’ are two examples of pathways which were only significant in experiments in the Diet category. These pathways are from the roots Signalling Pathways, with PKA activation also belonging to the roots Disease and Immune System. Another example of a pathway unique to Diet is ‘Glutathione synthesis and recycling’ which is significant in 7 of the 25 dietary conditions.

The number of pathways unique to Drug and Immune Challenge was less than in the Diet category. The root pathway Neuronal System has only some significant pathways that are unique to the Drug category. Two recurring pathways in experiments within the Drug category were ‘Glutamate neurotransmitter release cycle’ and ‘Acetylcholine neurotransmitter cycle’, which share several interactions. The Immune System pathway ‘IRF3-mediated induction of type1 IFN’ and ‘Synthesis of dolichyl-phosphate mannose’ from the Disease and Metabolism of Proteins pathway were unique to the Immune Challenge category.



**Figure 4: Number of differentially expressed leaf pathways in the three challenge categories.**

The three circles are indicative of the significant leaf pathways in the datasets belonging to one of the three challenge categories Diet, Drug, or Immune Challenge. The 212 common pathways of all the three classes are indicated in the centre.

There is a large number (212) of leaf pathways that are shared among all challenge categories (Figure 4). These 212 leaf pathways belong to 24 of the 27 root pathways. These 24 shared or common root pathways are statistically significant in at least 3 of the 37 conditions, but the number of experimental conditions in which the pathway is significant varies. The distribution of the 212 leaf pathways among the 24 root pathways is shown in Table 2. The roots with the largest number of results are Metabolism, Disease, and Signaling pathways, but the proportion of common leaf pathways is similar to the proportion of all the leaf pathways in the roots with which we started.

**Table 2. The 24 root pathways common to the three challenge categories.** The names of the root pathways are given in the first column. The second column shows the number of leaf pathways in a particular root and this number is shown as a percentage of all the leaf pathways in the root pathway in the database.

<b>Root pathways</b>	<b>Common Leaf pathways</b>	<b>Ratio of common leaf pathways in the root pathway</b>
Metabolism	54	22%
Disease	53	24%
Signalling Pathways	47	32%
Immune System	29	33%
Cell Cycle	20	30%
Gene Expression	13	18%
Hemostasis	11	44%
Programmed Cell Death	7	26%
Transcription	6	29%
Metabolism of proteins	4	8%
Chromatin organization	3	60%
Circadian Clock	3	100%
Developmental Biology	2	5%
DNA Replication	2	18%
Extracellular matrix organization	2	18%
Membrane Trafficking	2	18%
Transmembrane transport of small molecules	2	6%
Binding and Uptake of Ligands by Scavenger Receptors	1	20%
Cell-Cell communication	1	13%
Cellular responses to stress	1	9%

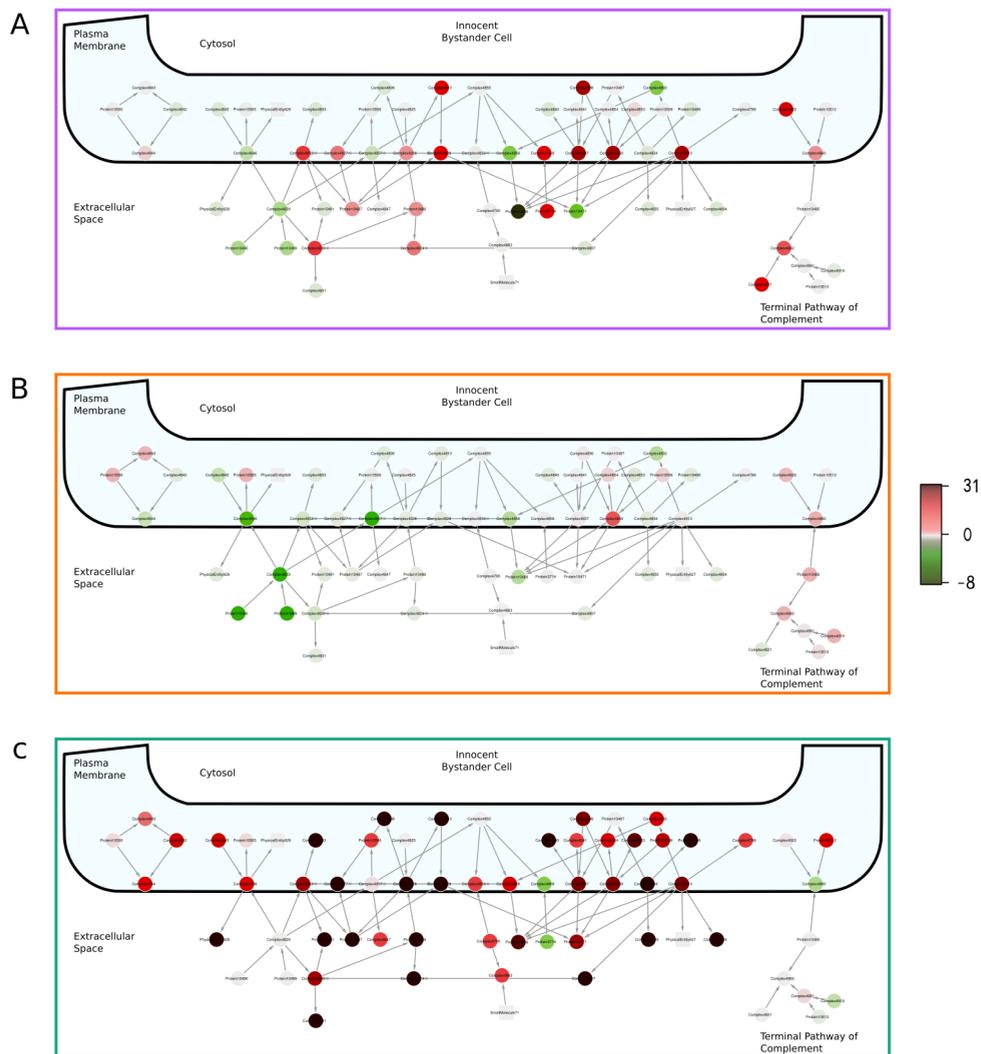
DNA Repair	1	4%
Muscle contraction	1	50%
Organelle biogenesis and maintenance	1	9%
Post-Elongation Processing of the Transcript	1	25%

### Use of common pathways under specific conditions

When the differential expression values (or node values) of individual experimental conditions in these pathways are superimposed on the nodes, we observed that under different challenges the same pathway had different expression patterns (number of genes expressed and levels of gene expression). In Figures 5 to 7 we show graphical representations of three pathways, ‘Regulation of Complement Cascade’ (Figure 5): the pathway with highest differences between gene expression of nodes in the three challenge categories, ‘Digestion of Dietary Lipid’ (Figure 6): a common pathway that has similar node expression profiles over all the experimental conditions where it is significant, and ‘Laminin Interactions’ (Figure 7): a common pathway representing a mechanism that is important for proliferation, migration and differentiation of the intestinal epithelium (Khalfaoui et al. 2013). The 3 pathways are arranged based on the cellular location information provided in the Reactome database. Each figure panel shows the same pathway; however, it is superimposed with the pathway gene expression measurements for 3 different conditions, one from each challenge category. Here we consider a pathway to be up-regulated when most of its nodes have a positive differential value, and down-regulated when most of the nodes have negative differential value. By simplifying the response to up- or down-regulation we observed that 95% of the pathways were up-regulated compared to the controls in all the experiments. When we investigated the responses of the pathways across all experimental conditions within each challenge category, we observed that 85% (181 pathways) of the common pathways were up-regulated in the same direction in the three categories. About 15% of the pathways common to the three challenge categories were regulated in different directions, either up- or down-regulated depending on the challenge category.

‘Regulation of Complement Cascade’ is an Innate Immune pathway for expression of complement factors and proteins that protect host cell membranes from microbial activation of the complement cascade. The ‘Regulation of Complement Cascade’ pathway is further linked to the pathways ‘Terminal Pathway of Complement’ and ‘Alternative Complement Activation’ in the Reactome database. This pathway is significantly regulated in 17 conditions (11 Diet, 3 Drug and 3 ImmuneChallenge conditions) and has the most divergent differential expression values between the 3 categories. The responses of ‘Regulation of Complement Cascade’ pathway for all 17 significant experimental conditions are shown in Supplementary Figure 1.

In most of the Diet conditions (see Supplementary Figure 1) the pathway is up-regulated except for the Diet.Add.Ppar.Agonist.SI and Diet.Add.Probio.BALBc.C conditions, the graphical pathway representation is dominated by down-regulated (green nodes in Figure 5 and Supplementary Figure 1) genes. When Penicillin is administered (Drug.Penicillin.SI), the genes in the pathway are down-regulated, whereas in the other Drug conditions involving the herbal drug (Drug.HerbalDrug.SI, Drug.HerbalDrug.SPF.C), the pathway genes are up-regulated. All the significant ImmuneChallenge conditions have highly up-regulated the pathway, with large differential values for the pathway genes (red nodes in Figure 5 and Supplementary Figure 1).



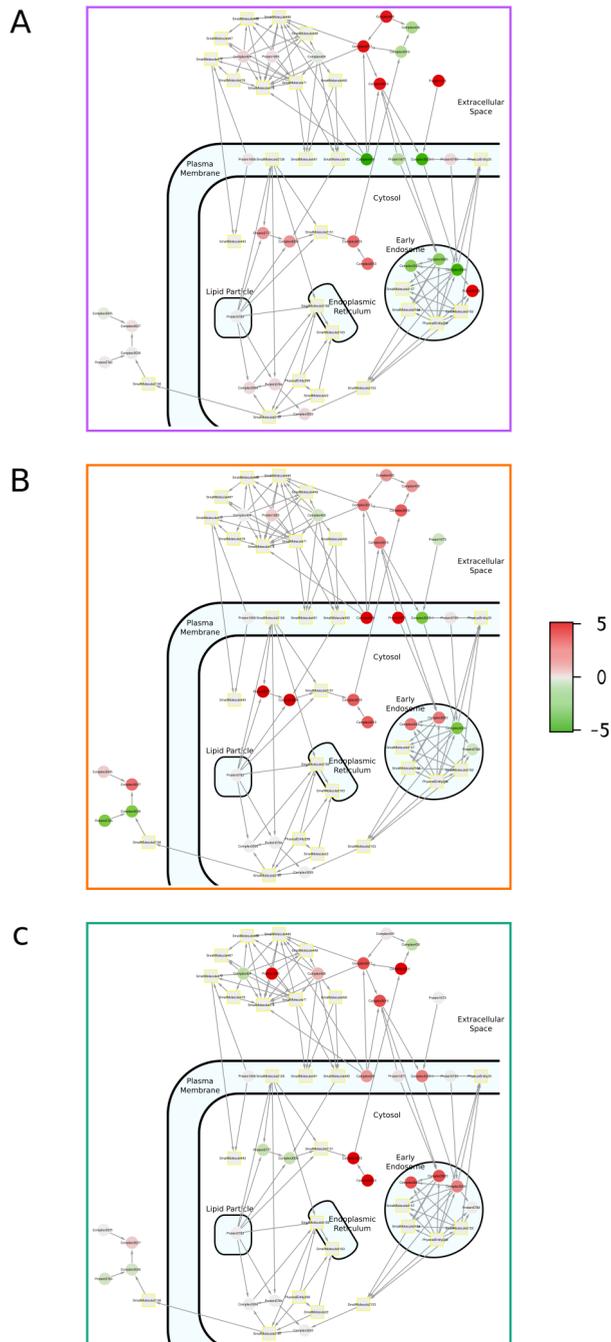
**Figure 5. Expression profile of nodes within the complement cascade pathway.** Expression values in each condition were superimposed on the graphical representation of the pathway, where panel A represents the condition Diet.Fat.45.P.SI, panel B Drug.Penicillin.SI, and panel C IC.STyph.4day.SL1344.C. The nodes represent expressed gene products or small molecules in different states that are involved in the pathway. The round nodes are proteins or complexes which are coloured according to their differential values and the square nodes are non-protein nodes like metabolites or ATP. Darker reds indicate a more positive differential gene expression and darker greens indicate a more negative differential gene expression. The edges represent interactions be-

---

tween the nodes as given in the Reactome database. The legend indicates the range of colours allocated to nodes based on their differential values from t-statistic calculations.

‘Digestion of Dietary Lipid’ is a Metabolic pathway that is significant in 15 conditions (13 – Diet, 1 – Drug and 1 – ImmuneChallenge). This pathway is known to be active in the stomach and in the small intestine that uses lipases to digest dietary lipids for absorption by enterocytes. Figure 6 shows that the genes of this pathway are expressed in a similar way all over the 3 challenge categories (details of all significant experimental conditions are shown in Supplementary Figure 2).

There are 11 Diet.Fat conditions where ‘Digestion of Dietary Lipid’ pathway is significant, these conditions modulate the gene expression profile of the pathway genes in a very similar way. Only a few genes of the pathway are regulated differently in the Diet.Fat conditions, but overall the pathway is up-regulated in these conditions. However, in the Diet.Fiber conditions (Diet.Fiber.AX.C (Figure 6) and Diet.Fiber.RS.C), the pathway genes are down-regulated. When an antibiotic was administered to deplete intestinal bacteria (Drug.Antibiotics.C), the pathway behaves quite similarly to the Diet.Fat conditions. The significant ImmuneChallenge condition, IC.GDuodenalis.SI shows a tendency towards up-regulation of the pathway genes.

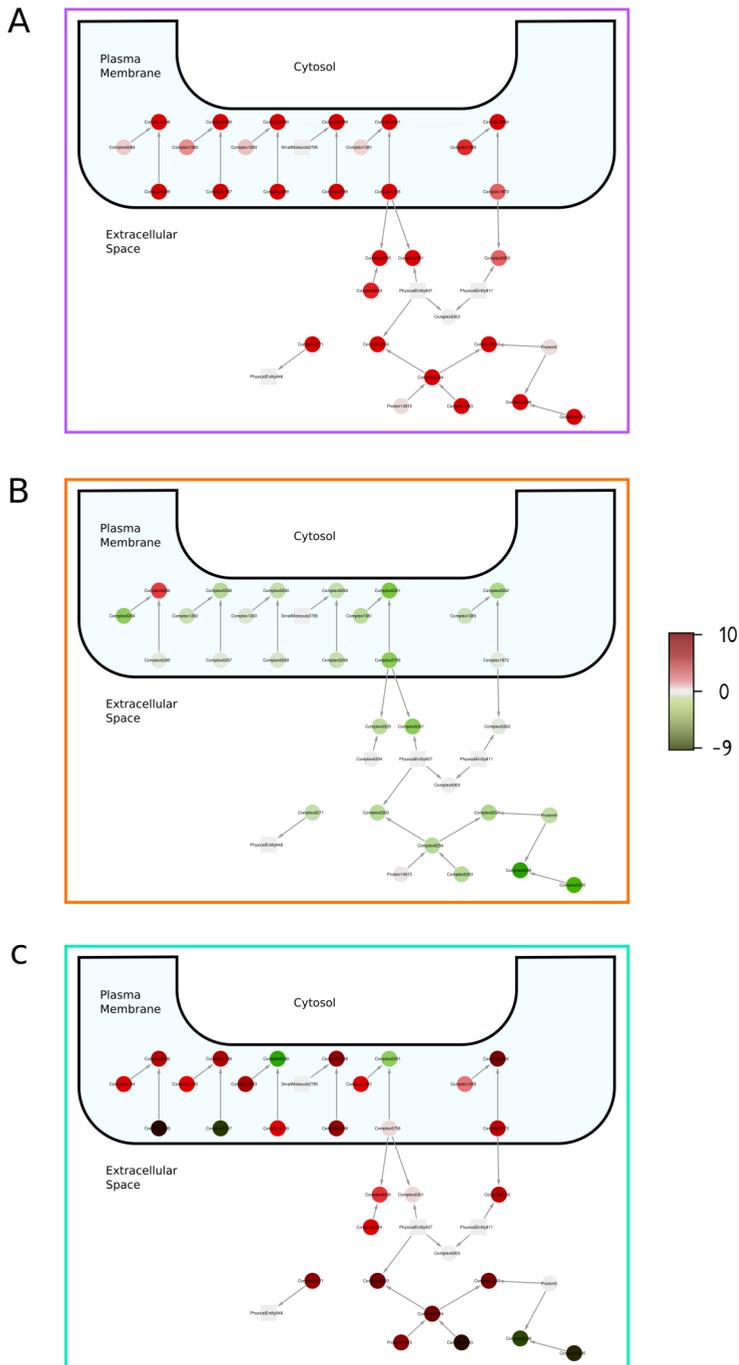


**Figure 6. Digestion of dietary lipids.** Expression values in each condition were superimposed on the graphical representation of the pathway, where panel A represents the condition Di-

et.Fiber.AX.C, panel B represents Drug.Antibiotics.C, and panel C represents IC.GDuodenalis.SI. The nodes represent expressed gene products or small molecules in different states that are involved in the pathway. The round nodes are proteins or complexes which are coloured according to their differential values and the square nodes are non-protein nodes like metabolites or ATP. Darker reds indicate a more positive differential gene expression and darker greens indicate a more negative differential gene expression. The edges represent interactions between the nodes as given in the Reactome database. The legend indicates the range of colours allocated to nodes based on their differential values from t-statistic calculations.

‘Laminin Interactions’ is a biologically relevant pathway representing interactions of a family of laminin binding proteins, (including integrins and non-integrins) on intestinal epithelial cells with laminin, a major intestinal basement membrane glycoprotein. These interactions regulate cell proliferation and adhesion, two crucial functions for intestinal epithelial homeostasis. ‘Laminin Interactions’ is a leaf pathway from the root ‘Extracellular Matrix Organization’ and is also a common pathway in all the three challenge categories. The genes in the pathway react very differently to the different types of exposures (Figure 7). It is significantly regulated in 18 experimental conditions, 16 of which are from the category Diet and 1 experimental condition from both the Drug and ImmuneChallenge categories (shown in Supplementary Figure 3).

Most of the Diet conditions up-regulate the ‘Laminin Interactions’ pathway genes. The exceptions being the Diet.Fat.60.D.SI and Diet.Add.Heme.C which down-regulate the pathway genes (see Supplementary Figure 3), the latter has a stronger suppressing effect on pathway expression than the former condition. The single Drug condition in which this pathway is significantly regulated is Drug.Antibiotics.C where the pathway genes are overall down-regulated (see Figure 7). The ImmuneChallenge condition IC.STyph.4day.SB1117.C, strongly up-regulated some genes in the ‘Laminin Interactions’ pathway but also strongly down-regulated some others (see Figure 7).



**Figure 7. Laminin Interactions.** Expression values in each condition were superimposed on the graphical representation of the pathway, where panel A represents the condition Diet.Fat.30.M.SI,

panel B represents Drug.Antibiotics.C, and panel C represents IC.STyph.4day.SB1117.C. The nodes represent expressed gene products or small molecules in different states that are involved in the pathway. The round nodes are proteins or complexes which are coloured according to their differential values and the square nodes are non-protein nodes like metabolites or ATP. Darker reds indicate a more positive differential gene expression and darker greens indicate a more negative differential gene expression. The edges represent interactions between the nodes as given in the Reactome database. The legend indicates the range of colours allocated to nodes based on their differential values from t-statistic calculations.

### **Stimulation of organoids and reverse transcriptase-quantitative PCR of gene expression in the ‘Regulation of Complement Cascade’ pathway.**

To investigate which signalling pathways might regulate genes of the Regulation of Complement Cascade’ pathway and verify the results obtained with some of the conditions *in vivo* we performed experiments on intestinal organoids derived from mice. Organoids generated from dissociated intestinal crypt cells contain all the differentiated types of epithelial cells found in the intestinal location from which they were isolated (Fatehullah, Tan, and Barker 2016). These untransformed cells have a global transcriptome profile that more closely resembles that of the tissue of origin and overcome some of the limitations associated with aneuploidy cancer cell lines.

As stimuli, we used TNF $\alpha$ , an inflammatory cytokine induced by infection or activation of inflammatory pathways, flagellin an agonist of TLR5 used as one of the challenge conditions in the Immune category and a pharmacological agonist of PPAR $\alpha$  which was used as one of the challenge conditions in the Diet category. Untreated 3-D organoids were used as control to obtain relative gene expression values of the 10 chosen genes when stimulated by the three treatments. All data were analysed using Prism statistical software (v5.0, Graphpad, San Diego, US), measured for normality using the Kolmogorov-Smirnov test, and shown in Box and Whisker plots in Supplementary Figure 4. A t-statistic test was performed on the RT-qPCR results of the 10 genes using the same methods used on the nodes in the gene set pathway analysis of CePa, see Table 3. The results of the validation experiment are shown in Supplementary Figure 4 as box and whisker plots.

These genes were chosen to represent the activity of the pathway ‘Regulation of Complement Cascade’, five of them (C3, C5, CD46, CD55, and CD59) were chosen to observe the activity of the complement system and the host defence against the complements. The other five (C8GH, CFI, CFH, CR1, and C2) were chosen because they are part of nodes that change the most in the 17 experimental conditions where the ‘Regulation of Complement Cascade’ pathway is significant. The expression of the 10 homologous mice genes is not changed much when stimulated by the virulence factor flagellin but respond significantly to the inflammatory cytokine TNF $\alpha$  and moderately to the PPAR $\alpha$  agonist. Overall the genes that change most in the pathway analysis show more differential changes compared to the ones chosen from a biological perspective.

**Table 3. Differential expression values from t-statistic calculations of the 10 genes in stimulated organoids and in the 17 experimental conditions where the 'Regulation of Complement Cascade' pathway is significant.** Columns 2 to 11 contain differential values of the genes in the conditions given in the rows. The experimental conditions of the Diet category are coloured in purple, the Drug category conditions are coloured in orange and the ImmuneChallenge category conditions are coloured green. The results of the experimental validation on organoids is in black. The first five genes were chosen based on biological information and the last five were based on the data analysis

Experimental Condition	C3	C5	CD46	CD55	CD59	C2	C8G	CFI	CFH	CR1
Flagellin	-0.58	-1.17	0.53	-0.27	0.66	0.8	-0.57	-0.75	-0.59	-1.73
TNF $\alpha$	17.38	-2.96	1.03	6.99	0.4	1.73	0.37	5.02	6.4	-5.05
PPAR $\alpha$	1.81	-1.08	0.92	2.98	0.74	1.96	0.37	7.42	4.92	-3.03
Diet.Fat.20.Distal.SI	1.39	0.74	-0.56	1.53	-0.42	0.43	0.67	0.6	1.22	-0.56
Diet.Fat.45.Middle.SI	1.54	1.08	-0.26	0.97	-0.33	2.18	0.2	1.03	2.61	-0.26
Diet.Fat.20.Proximal.SI	16.27	1.18	-0.68	13.25	1.18	20.35	0.67	2.29	22.2	-0.68
Diet.Fat.30.Proximal.SI	11.43	-2.05	0.9	4.59	-2.52	9.73	0.15	0.78	8.33	0.9
Diet.Fat.45.Proximal.SI	5.44	-0.83	0.73	3.61	-0.97	3.97	1.81	1.63	2.52	0.73
Diet.Fibre.Arabinosylan.C	1.61	0.54	-0.6	2.06	2.23	2.31	0.92	2.94	3.81	-0.6

Diet.Fiber.Inulin.C	-1.34	1.61	1.56	1.05	1.64	2.55	-1.24	-4.32	-1.6	1.56
Diet.Additive.Quercetin.Ileum.SI	1.01	-0.31	-0.51	1.99	3.54	7.51	4.57	0.39	-1.91	-0.51
Diet.Additive.PparAgonist.SI	1.73	2.9	-0.52	1.84	1.68	1.88	0.9	1.68	-0.21	-0.52
Diet.Additive.Probiotic.BALBc.C	2.04	2.48	-0.67	-0.82	5.35	3.7	2.93	0.51	2.22	-0.67
Diet.Additive.Probiotic.BALBc.SI	0.97	-0.99	-0.47	-0.34	1.99	8.1	3.53	-1.85	2.8	-0.47
Drug.HerbalDrug.SPFree.C	6.97	2.22	-1.6	2.32	2.28	8.74	-0.73	0.59	2.25	-1.6
Drug.HerbalDrug.SI	-0.09	-1.3	-2.36	1.12	-0.49	-0.15	-0.1	0.15	1.94	-2.36
Drug.Penicillin.SI	3.72	1.92	-3.12	-2.08	-1.34	3.84	-0.59	-1.8	3.71	-3.12
ImmuneChallenge.S.Typhimurium. 4day.SB1117.C	5.61	1.23	2.05	-3.55	1	8.41	-0.07	2.85	2.38	2.05
ImmuneChallenge.S.Typhimurium. 4day.SL1344.C	0	-1.21	0	-3.84	-0.57	2.26	0.43	-3.84	0	0
ImmuneChallenge.Colitis.C	0	1.66	0	1.64	0.83	3.21	-0.09	1.64	0	0

---

## Discussion

The results of this study showcase examples of the functional plasticity of the gut mucosal tissue at the level of pathway biosynthesis and signalling. By integrating the results of experiments in which intestinal homeostasis was perturbed by completely different challenges, we were able to investigate the plasticity of the gastrointestinal tissue in terms of pathway use. To the best of our knowledge, this type of study, focusing on one tissue and different types of challenges, has not yet been performed on this scale. By grouping the different challenge conditions in categories according to whether they were drugs, dietary ingredients or inflammatory agents, we grouped the responses of the mucosal tissue to facilitate broader comparisons. We observed that some pathways are regulated by experimental conditions in all categories. In addition, we observed large differences in the expression profiles of pathway genes between different exposure conditions, in some cases, irrespective of the challenge category. By focussing on the commonly regulated pathways, we show that the gut mucosa employs similar pathway systems, however, in different combinations and with different intra-pathway gene expression profiles, to respond to different exposures. Experimental validation of one of these pathways shows that these responses are due to local responses of the tissue to the administered challenge.

### High-level data integration

5 The most important criteria for dataset selection was the age of the mice when sampled, because the mucosal immune system and intestinal microbiota of mammals is known to change dramatically around weaning (Cera et al. 1988. This holds true for mice (X. Gu, Li, and She 2002; D. Kelly, Smyth, and McCracken 1991, therefore we only included datasets where the mice were sampled 2 or more weeks after weaning. The inclusion criteria provided datasets that are comparable, but they still differed in many aspects such as the use of microarray platforms, in sampling of the tissue and in the control conditions. Therefore, we used a high-level data integration method that started with the identification of differently expressed pathways as detected within an individual experiment and/or experimental condition.

### **Pathway level analysis**

We reasoned that genome-wide gene expression data is more easily interpreted, in terms of biological processes, when the analysis is performed on a functional level with pathways or gene-sets as functional units or building blocks of biological processes, rather than on individual genes (Emmert-Streib and Glazko 2011). We used the pathway database Reactome because it is manually curated and regularly updated. Both signalling and metabolic pathways are represented in the same format with separate nodes for proteins (we use gene expression values here) and metabolites. The nodes are also annotated with details of the cellular location.

It has been theorized that using the topology of the pathways will enhance the results of the enrichment analysis. Therefore we chose the CePa algorithm that considers pathway topology although we modified the algorithm to accommodate smaller sample sizes. Based on the biological information behind the pathways, we decided to use two centrality calculations, 'In-reach' and 'Out-reach' to capture regulation of pathways down-stream (important for signalling pathways) and up-stream (for metabolic pathways) respectively. Most of the results were significant in both the centrality calculations irrespective of the type of pathway. This apparent indifference in topology is also observed by Bayerlova *et al.* (Bayerlová et al. 2015 in a different pathway database using a variety of algorithms).

### **Several pathways are regulated by all three challenge categories**

The results of the integrated pathway analysis show that there is a notable overlap in the pathway response between the 3 challenge categories. The Diet category has the highest number of leaf pathway results, and also the highest number of experimental conditions. Surprisingly, most of the results from Drug and Immune Challenges were also shared by the Diet class. This is not expected since the drug and immune challenges are regarded as more severe perturbations of homeostasis than a change in the diet. However, the results demonstrate that there is a group of common pathways (building blocks) that are always regulated when homeostasis is perturbed. These common pathways contribute to a major extent to the capability of the intestinal mucosa to display a high-level of functional plasticity. That is the reason we focussed our subsequent research on these common pathways.

Most of the transcriptomics data used in this study came from intestinal scrapings which greatly enriches for different types of epithelial cells involved in innate immunity and cross talk with the immune cells in the lamina propria (Wells et al. 2011). The different functions of these epithelial cells contribute to the functional plasticity of the epithelium. The results of this study revealed another layer of plasticity which is based on the specific use of a common set of pathways. These common pathways are significantly differentially expressed compared to controls in at least one experimental condition in each category. Simplifying the regulatory output of a whole pathway is difficult due to issues like up-regulation of the expression of stimulatory or inhibitory molecules. For the sake of simplicity, in this paper we assume that if the majority of the nodes are up-regulated, then the pathway as a whole is up-regulated and vice versa.

### **Examples of pathways that respond differently to the classes of exposure**

Three of the 212 common pathways which respond differently to the challenge categories are discussed further below. The first pathway, ‘Regulation of Complement Cascade’, was chosen because it showed the largest difference in node expression profile between the three challenge categories (based on t-values of the nodes) and because the bulk of soluble complement proteins are produced by the liver with some being produced by only a few cell types, suggesting they may have specific functions. Hence, we also investigated whether stimuli for specific signalling pathways that might be associated with the challenges induced complement pathway genes in intestinal organoid models. The second pathway, ‘Digestion of Dietary Lipids’, was chosen because it is part of the majority of pathways that display an almost similar intra-pathway gene expression profile between the three challenge categories. The third pathway, ‘Laminin Interactions’, was chosen because it is part of the ‘Extra Cellular Matrix Organization’ root pathway, which regulates diverse functions, including proliferation, migration and differentiation and also morphogenesis of the intestine (Khalfaoui et al. 2013).

#### *Regulation of Complement Cascade*

The complement system consists of several inactive pre-proteins produced in the liver that circulate in the blood. It is part of the innate immune system and activation of the complement cascade plays a key role in the opsonisation of microorganisms to increase phagocy-

tosis by macrophages and neutrophils at the sites of infection (Ricklin et al. 2010). Activation of the complement cascade also promotes the development of the pore forming membrane attack complex on microorganisms. The triggers for activation of this cascade are the formation of the C1 complex pathogen-binding antibodies (classical pathway), spontaneous hydrolysis of C3 and deposition of C3b on the surface of microorganisms or binding of mannose-binding lectin and activation of serine proteases acting on complement factors (Merle et al. 2015). A family of regulatory proteins act as regulators of complement activation to protect host cells from homologous complement attack. This prevents the membrane attack complex formed during complement activation, from lysing nearby host cells.

It is expected that the regulation of complement cascade pathway is modulated in the Drug category by administration of antibiotics and ImmuneChallenge conditions. The fact that the ‘Regulation of Complement Cascade’ pathway was modulated by several conditions (five Diet.Fat, two Diet.Fibre and five Diet.Additive conditions) in the Diet category that are not considered to be extreme perturbations, was surprising and interesting. However, there has been research linking elements of the complement system to lipid metabolism by acylation-stimulating protein/complement C3adesArg (ASP/C3adesArg) which has been shown to stimulate triglyceride (TG) synthesis in both mature adipocytes and pre-adipocytes (Munkonda et al. 2012; Mamane et al. 2009; Paglialunga et al. 2007).

The Diet and ImmuneChallenge conditions up-regulate the ‘Regulation of Complement Cascade’ pathway and the experimental conditions in the Drug category mostly down-regulated the pathway. The regulation of the pathway is the strongest in the ImmuneChallenge conditions, however our results indicate that several Diet conditions induce similar regulation of the pathway, although expression is lower than with two conditions (IC.STyph.4day.SL1344.C, IC.STyph.4day.SB1117.C) in the Immune Challenge category. The expression of the regulators of complement cascade that protect host cells from potentially damaging effects of complement in response to infectious agents seems logical. Mucosal damage due to infection might lead to vascular damage and leakage of plasma into the lamina propria. Here activation of the complement cascade would necessitate up-regulation of the host regulators of complement cascade (CD46, CD55) to protect its own cell membranes. Additionally, the production of blood complement components by

epithelial cells would contribute to the local pool. This could be an important mechanism in the periphery of the intestine which might have poor access to circulating complement components from the blood. The Diet category of challenges which strongly upregulated this pathway were from high fat diets (Figure 5), which are known to cause low-grade inflammation and a leaky gut (Bischoff et al. 2014; Chassaing and Gewirtz 2014). Interestingly depletion of the microbiota with penicillin, reduced pathway expression suggesting that the microbiota contribute to “tonic” stimulation of the complement related pathways via stimulation of innate immunity.

To investigate this hypothesis and identify the signalling pathways associated with altered expression of this pathway we stimulated small intestinal mouse organoids with TNF $\alpha$ , flagellin, an agonist of PPAR $\alpha$ , or culture medium as a control and compared the relative transcript abundance of several complement pathway genes by reverse transcription polymerase chain reaction (RT-PCR). The RT-PCR measurements were made on mice homologs of the human genes in the pathway, and these mice genes are denoted in italics with only the first letter of the gene symbol capitalized. Stimulation with TNF $\alpha$  a cytokine secreted by activation of NF- $\kappa$ B in tissue and immune cells simulates inflammation of the intestine. The PPAR $\alpha$  agonist simulates high-fat dietary conditions and the flagellin an agonist of TLR5, was used to investigate the role of TLR signalling. Transcription of complement factors *C2*, *C3* and regulatory proteins *Cfi*, *Cd46* and *Cd55* were significantly up-regulated by TNF $\alpha$  whereas expression of *C5*, *C8* and *Cfh* was not significantly changed and *Cd59* was significantly down-regulated. The agonist of PPAR $\alpha$  significantly up-regulated transcription of regulatory proteins *Cd55*, *Cr2*, *Cd46*, and *Cfi* and significantly down-regulated expression of *Cd59*. Surprisingly flagellin (which was shown to activate TLR5 in reporter cells) did not significantly alter expression of any of these genes. These results suggest that activation of inflammatory pathways in intestinal organoids via TNF $\alpha$ -receptor mediated pathway upregulates expression of host membrane proteins (CD46 and CD55) involved in protection of membranes against complement activation as well as factor I (CFI) which regulates complement activation by cleaving cell-bound or fluid phase C3b. Upregulation of C2 required for generation of the C3 convertase upon complement activation and C3 which is crucial component for activation and the first stages of complement cascade were also upregulated by TNF $\alpha$ .

This finding suggests that activators of NF- $\kappa$ B pathway (e.g. IL-1 $\beta$ , and agonists of NLRs and TLRs) would also induce expression of the complement related genes. However, addition of flagellin to organoids did not induce significant changes in expression of these genes. The reasons for this are unclear but may be due to lack of expression of membrane TLRs in organoid cells or strong negative regulation of TLR signalling pathways. It has been proposed that TLR signalling is tightly controlled in epithelial cells to avoid chronic inflammatory responses to bacterial MAMPs from the intestinal lumen (Sham et al. 2013; Biswas et al. 2011). Interestingly, activation of the PPAR $\alpha$  pathway also increased expression of protective factors *Cd55*, *Cd46*, *Cfi* and *Cr2* which allows the complement system to play a role in B cell activation and maturation. This links PPAR $\alpha$  to regulation of complement cascade in the gut and the effects of high fat diets on this pathway *in vivo*.

The variations in the expression of the genes measured in the validation experiment reflect the plasticity in the response of the ‘Regulation of Complement Cascade’ pathway that is observed in the high-level data integration results.

### *Digestion of Dietary Lipids*

The ‘Digestion of Dietary Lipids’ pathway describes the digestion of dietary lipids by pancreatic and gastric lipases which are known to increase with increase in fat consumption (Deschodt-Lanckman et al. 1971; Corring 1980). This pathway is regulated in the upwards direction under all the experimental conditions where it is significant (featured in Supplementary Figure 2). All the Diet conditions (11 experimental conditions) with an increase in fat content show up-regulation of this pathway which is consistent with existing literature. However, up-regulation is also seen in four other experimental conditions, two of which are substitutions of fibres in the diet. Dietary fibres have been documented to change lipid levels (Ulmius, Johansson, and Önning 2009; Brown et al. 1999) depending on the type of fibre, but the fibres used in the original study (Arabinosyln from wheat and Resistant Starch from tapioca) have no documented effects on lipid levels. Antibiotics that severely affect the intestinal microbiota have been observed to reduce lipid absorption in mammals (H. Sato et al. 2016; Rotenberg and Andersen 1982). Bacterial pathogens of the genus *Giardia* have an interesting relationship with dietary lipids which can kill them in *in vitro* conditions (Rayan, Stenzel, and McDonnell 2005; Das et al. 1988; Yichoy et al.

2011 but only reduces the severity of the infection in *in vivo* conditions (Suh, Belosevic, and Clandinin 2004; Yichoy et al. 2011). In our analysis, we see that even though these 15 experimental conditions have very different effects on lipid levels in the organism, they modulate the ‘Digestion of Dietary Lipids’ pathway in the same direction.

### **The “Laminin Interactions’ pathway as indicator for gut health**

Our analysis enables unbiased analysis of the response of pathways to experimental conditions. Therefore, besides analysing significant pathways it is possible to investigate minutely at the behaviour of specific pathways to understand their regulation by any of the 37 different experimental conditions. One biologically relevant pathway for the tissue is the ‘Laminin Interactions’ pathway. It is a leaf pathway that is part of the root pathway Extracellular Matrix Organization which is important for intestinal health (Frantz, Stewart, and Weaver 2010). This class of pathways encompass biological processes related to the extracellular matrix, which is made up of several types of fibrous proteins that are linked to provide structural support to the cells. The biological process of maintaining the extracellular matrix is crucial to the barrier function of the gut and intestinal regeneration (Blikslager et al. 2007; Göke, Zuk, and Podolsky 1996). There are 11 leaf pathways in this root, and two of them were part of the common significantly regulated pathways: the ‘Integrin Cell Surface Interactions’ pathway and the ‘Laminin Interactions’ pathway. The latter is a pathway that describes the interactions of the laminin family of proteins with each other and other fibrous proteins. Laminins have been extensively studied (Domogatskaya, Rodin, and Tryggvason 2012) and have been implicated in, among other things, wound repair (Iorio, Troughton, and Hamill 2015) and cell adhesion. These are functions that are of eminent importance to the gut in keeping harmful substances from breaching the epithelium.

Our analysis shows that this pathway is up-regulated in most of the 18 conditions in which it is significant. The Diet conditions modulate the pathway in mostly the same way except for the experiment when 60% fat is added to the diet. In the latter experiment two parts of the small intestine were sampled, the proximal and distal halves. The ‘Laminin Interactions’ pathway is modulated in opposite directions in each of these sections, and this dif-

ference between the two tissue sections has also been described by the researchers who conducted this experiment (Desmarchelier et al. 2012).

In the Drug condition where several antibiotics were administered, almost all the nodes of the ‘Laminin Interactions’ pathway are down-regulated. This treatment depletes the intestinal microbe population the presence of which stimulates epithelial proliferation and the involvement of the Laminin pathways in epithelial remodelling (Hörmann et al. 2014). This down-regulation of the ‘Laminin Interactions’ pathway in the Drug.Antibiotics.C condition is consistent with the reports from an experiment with similar conditions (Reikvam et al. 2011 where there was reduced epithelial turnover. ‘Laminin Interactions’ shows strong signals in the ImmuneChallenge condition IC.STyph.4day.SB1117.C, some nodes are highly up-regulated and some highly down-regulated. The paper on this immune challenge (Liu et al. 2010 shows that there is an up-regulation of genes involved in ‘cell growth and proliferation’. This modulation of the ‘Laminin Interactions’ pathway shows that laminins are reactive to different types of conditions and play an important role in the functional plasticity of the gut.

## Conclusions

In this study we performed a high-level data integration of a number of datasets representing whole genome gene expression patterns of intestinal mucosa’s that have been exposed to different types of stimulations. The results highlight the functional plasticity of the gut is high and point towards different layers of plasticity. Firstly, the same pathways can be used in different combinations to generate extremely different physiological responses. Secondly, genes within pathways are not uniformly regulated which contributes further to the plasticity. The large number of the so-called common pathways that are regulated by the different classes of stimulations are proof of plasticity on the functional level. Furthermore, these pathways also show intra-pathway variations in response to the different stimulations which reveals another layer of plasticity in the pathways themselves. The latter calls for caution when analysing expression data as pathway enrichment cannot be equated to a uniform functional response. We conclude that the functional plasticity of intestinal mucosal tissue can be explained, at least partially, by variation in the pathway sets that

are regulated by different stimuli and by the induction of intra-pathway specific gene expression patterns.

### Acknowledgements

The authors would like to acknowledge the fruitful discussions with Prof. Vitor Martins dos Santos in guiding the direction of the analysis. This work has been financially supported by the Systems Biology Investment Programme of Wageningen University, KB-17-003.02-022.

### Supplementary files

**Supplementary Table 1. Description of the 37 conditions.** This table describes the abbreviated names of experimental conditions as used in the paper and the control conditions in each of the experiments.

Names	Description of Experimental conditions	Description of the control condition
Diet.Fat.20.P.SI	20% of the energy in the diet comes from fat, proximal section of the small intestine	10% of the energy in the diet comes from fat, proximal section of the small intestine
Diet.Fat.30.P.SI	30% of the energy in the diet comes from fat, proximal section of the small intestine	
Diet.Fat.45.P.SI	45% of the energy in the diet comes from fat, proximal section of the small intestine	
Diet.Fat.20.M.SI	20% of the energy in the diet comes from fat, middle section of the small intestine	10% of the energy in the diet comes from fat, middle section of the small intestine

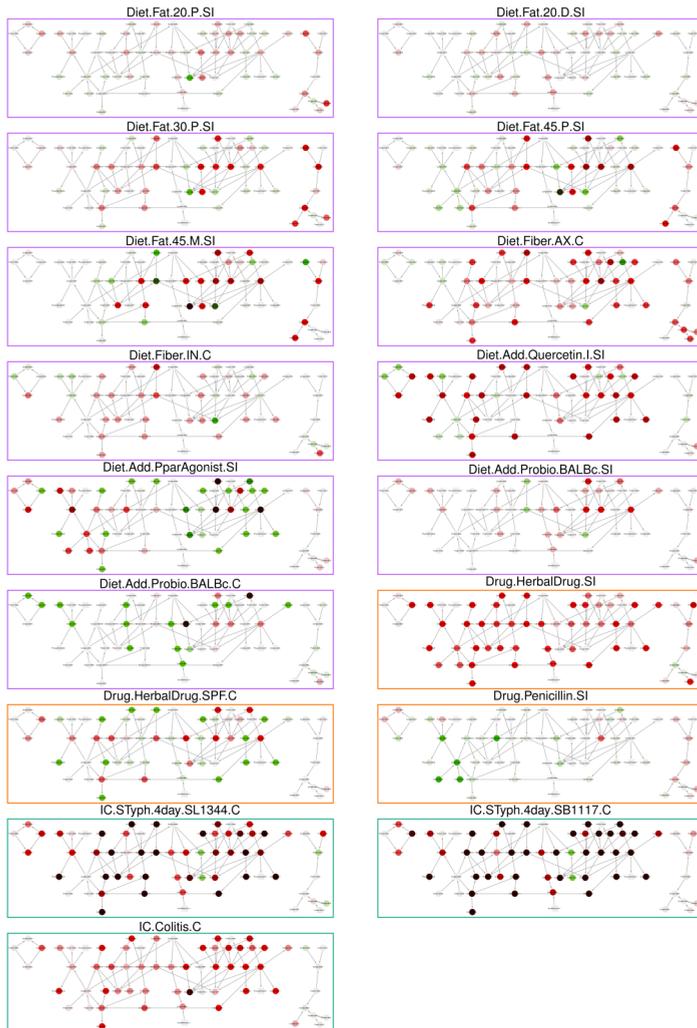
Diet.Fat.30.M.SI	30% of the energy in the diet comes from fat, middle section of the small intestine	
Diet.Fat.45.M.SI	45% of the energy in the diet comes from fat, middle section of the small intestine	
Diet.Fat.20.D.SI	20% of the energy in the diet comes from fat, distal section of the small intestine	10% of the energy in the diet comes from fat, distal section of the small intestine
Diet.Fat.30.D.SI	30% of the energy in the diet comes from fat, distal section of the small intestine	
Diet.Fat.45.D.SI	45% of the energy in the diet comes from fat, distal section of the small intestine	
Diet.Fat.60.D.SI	60% of the energy in the diet comes from fat, distal section of the small intestine	11% of the energy in the diet comes from fat, distal section of the small intestine
Diet.Fat.60.P.SI	60% of the energy in the diet comes from fat, proximal section of the small intestine	
Diet.Fiber.AX.C	Control diet with 10% of corn starch substituted with Arabinoxylan, sampled at the colon	Standard semi-synthetic low fat diet with corn starch as carbohydrate source, sampled at the colon
Diet.Fiber.FOS.C	Control diet with 10% of corn starch substituted with Oligofructose, sampled at the colon	
Diet.Fiber.GG.C	Control diet with 10% of corn starch substituted with Guar gum, sampled at the colon	

Diet.Fiber.IN.C	Control diet with 10% of corn starch substituted with Inulin, sampled at the colon	
Diet.Fiber.RS.C	Control diet with 20% of corn starch substituted with Resistant starch sampled at the colon	
Diet.Add.Heme.C	Control diet supplemented with 0.5 $\mu\text{mol/g}$ heme, sampled at the colon	Westernized control diet (40% fat (mainly palm oil) low calcium (30 $\mu\text{mol/g}$ )), sampled at the colon
Diet.Add.Quercetin.C	RM3 (E) 801710 Soya-free powdered diet + 7mg quercetin, sampled in the colon	RM3 (E) 801710 Soya-free powdered diet, sampled in the colon
Diet.Add.Quercetin.I.SI	RM3 (E) 801710 Soya-free powdered diet + 7mg quercetin, sampled in the ileum	RM3 (E) 801710 Soya-free powdered diet, sampled in the ileum
Diet.Add.Quercetin.J.SI	RM3 (E) 801710 Soya-free powdered diet + 7mg quercetin, sampled in the jejunum	RM3 (E) 801710 Soya-free powdered diet, sampled in the jejunum
Diet.Add.PparAgonist.SI	Standard laboratory chow (RMH-B), sampled at the small intestine	Standard laboratory chow (RMH-B), sampled at the small intestine
Diet.Add.Probio.BALBc.C	SSNIFF R/M-H feed pellet diet or $3 \times 10^8$ CFU VSL#3 suspended in 200 mL PBS, strain BALBc mice, sampled at the colon	SSNIFF R/M-H feed pellet diet given oral gavage with PBS, strain BALBc mice, sampled at the colon

Diet.Add.Probio.BALBc.SI	SSNIFF R/M-H feed pellet diet with $3 * 10^8$ CFU VSL#3 suspended in 200 mL PBS, strain BALBc mice, sampled at the small intestine	SSNIFF R/M-H feed pellet diet, oral gavage with PBS, strain BALBc mice, sampled at the small intestine
Diet.Add.Probio.C57BL6.C	SSNIFF R/M-H feed pellet diet with $3 * 10^8$ CFU VSL#3 suspended in 200 mL PBS, strain C57BL6 mice, sampled at the colon	SSNIFF R/M-H feed pellet diet, oral gavage with PBS, strain C57BL6 mice, sampled at the colon
Diet.Add.Probio.C57BL6.SI	SSNIFF R/M-H feed pellet diet with $3 * 10^8$ CFU VSL#3 suspended in 200 mL PBS, strain C57BL6 mice, sampled at the small intestine	SSNIFF R/M-H feed pellet diet, oral gavage with PBS, strain C57BL6 mice, sampled at the small intestine
Drug.HerbalDrug.C	Orally treated with JTX solution (1.0g/kg body weight), strain BALBc, sampled at the colon	Treated with water (1.0g/kg body weight), strain BALBc, sampled at the colon
Drug.HerbalDrug.SPF.C	Orally treated with JTX solution (1.0g/kg body weight), strain IQI, sampled at the colon	Treated with water (1.0g/kg body weight), strain BALBc, sampled at the colon
Drug.HerbalDrug.SI	Orally treated with JTX solution (1.0g/kg body weight), strain BALBc, sampled at the small intestine	Orally treated with JTX solution (1.0g/kg body weight), strain BALBc, sampled at the colon

Drug.Antibiotics.C	Oral gavage of mixture of antibiotics and in drinking water, sampled in the colon	Drinking water, sampled in the colon
Drug.Penicillin.SI	Low dose penicillin was added to the drinking water, sampled at the small intestine	Normal chow and drinking water, sampled at the small intestine
IC.STyph.4day.SB1117.C	Streptomycin treatment followed by oral gavage of $1 \times 10^7$ CFU of <i>S. Typhimurium</i> (100 $\mu$ l suspension in HBSS), <i>S.Typhimurium</i> strain SB1117, after 4 days sampled in the colon	Oral gavage of HBSS, sampled in the colon at day 4
IC.STyph.4day.SL1344.C	Streptomycin treatment followed by oral gavage of $1 \times 10^7$ CFU of <i>S. Typhimurium</i> (100 $\mu$ l suspension in HBSS), <i>S.Typhimurium</i> strain SL1344, after 4 days sampled in the colon	
IC.STyph.8hr.SB1117.C	Streptomycin treatment followed by oral gavage of $1 \times 10^7$ CFU of <i>S. Typhimurium</i> (100 $\mu$ l suspension in HBSS), <i>S.Typhimurium</i> strain SB1117, after 8 hours sampled in the colon	Oral gavage of HBSS, sampled in the colon at 8 hours
IC.STyph.8hr.SL1344.C	Streptomycin treatment followed by oral gavage of $1 \times 10^7$ CFU of <i>S. Typhimurium</i> (100 $\mu$ l suspension in HBSS), <i>S.Typhimurium</i> strain SB1117, after 8 hours sampled in the colon	

IC.GDuodenalis.SI	Oral gavage of $1 \times 10^6$ parasites in 0.1 ml phosphate-buffered saline (PBS), sampled in the small intestine	Sampled in the small intestine
IC.Flagellin.C	Flagellin (FliC) from WT <i>Salmonella enterica</i> serovar typhimurium (SL3201, fljB <sup>-</sup> ) was given 10 $\mu$ g per mouse in 0.2 ml PBS) intraperitoneally, sampled in the colon	0.2 ml PBS was given intraperitoneally, sampled at the colon
IC.Colitis.C	Dextran Sodium Sulfate (DSS: 4% w/v) in drinking water, sampled at the colon	Normal drinking water, sampled at the colon

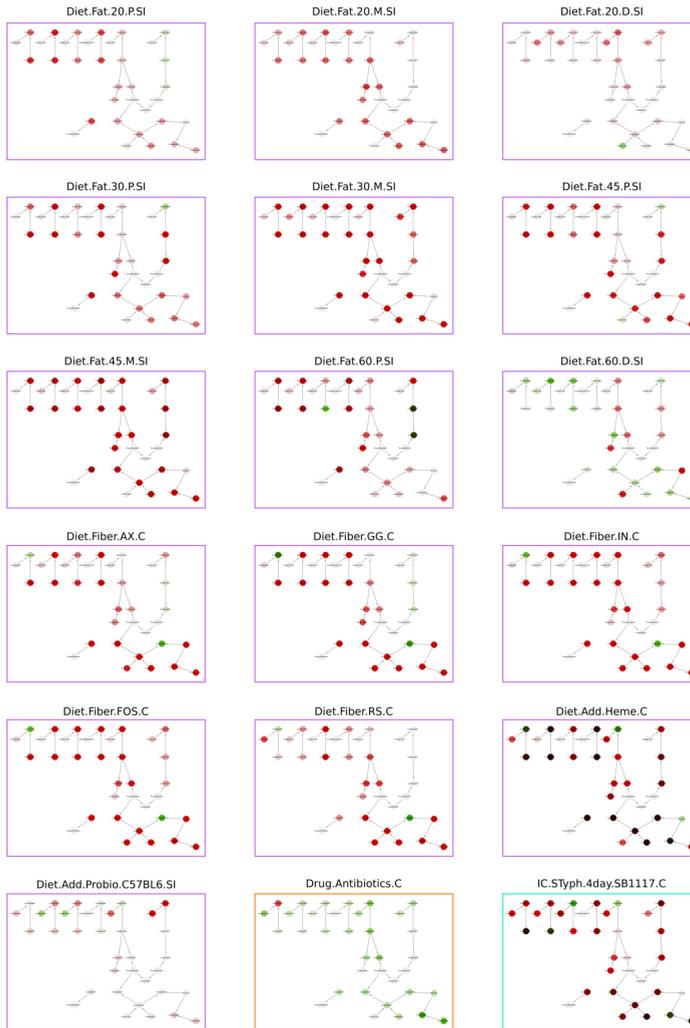


**Supplementary Figure 1: Pathway ‘Regulation of Complement Cascade’ in 17 conditions.** This figure contains the pathway as a graph, the nodes are proteins, complexes or other physical entities and the edges are interactions between them. Circular nodes represent proteins or protein complexes, square nodes represent small molecules like metabolites or ATP. The nodes are coloured based on the differential expression values calculated by the algorithm CePa, darker reds indicate a higher up-regulation of the node and darker greens indicate a higher down-regulation. Each section of the figure has the same pathway overlaid with expression data from the experimental condition with which it is labelled, a subset of this figure is in Figure 5.

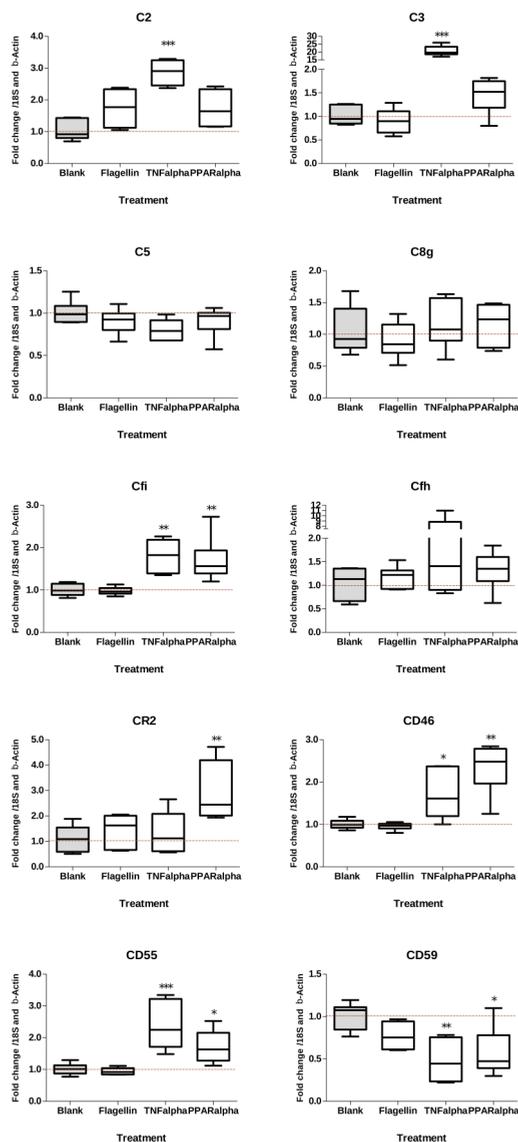


**Supplementary Figure 2: Pathway ‘Digestion of Dietary Lipids’ in 15 conditions.**

This figure contains the pathway as a graph, the nodes are proteins, complexes or other physical entities and the edges are interactions between them. Circular nodes represent proteins or protein complexes, square nodes represent small molecules like metabolites or ATP. The nodes are coloured based on the differential expression values calculated by the algorithm CePa, darker reds indicate a higher up-regulation of the node and darker greens indicate a higher down-regulation. Each section of the figure has the same pathway overlaid with expression data from the experimental condition with which it is labelled, a subset of this figure is in Figure 6.



**Supplementary Figure 3: Pathway ‘Laminin Interactions’ in 18 conditions.** This figure contains the pathway as a graph, the nodes are proteins, complexes or other physical entities and the edges are interactions between them. Circular nodes represent proteins or protein complexes, square nodes represent small molecules like metabolites or ATP. The nodes are coloured based on the differential expression values calculated by the algorithm CePa, darker reds indicate a higher up-regulation of the node and darker greens indicate a higher down-regulation. Each section of the figure has the same pathway overlaid with expression data from the experimental condition with which it is labelled, a subset of this figure is in Figure 7.



**Supplementary Figure 4. Expression of 10 chosen genes from the ‘Regulation of Complement Cascade’ pathways with significance calculated with ANOVA.** Each graph contains information on different genes, the x-axis contains information on the treatment of the organoids and the y-axis has the fold change of the control genes. All data were considered significantly different from the blank when  $P < \alpha$  (0.05), and indicated with \* ( $P < 0.05 = *$ ,  $P < 0.01 = **$ ,  $P < 0.001 = ***$ ).



# Chapter 6

## Modelling the impact of diet on immunity parameters using high-throughput data

**Nirupama Benis**; Host Microbe Interactomics, Wageningen University & Research.

**Maria Suarez-Diez**; Systems and Synthetic biology, Wageningen University & Research.

**Dirkjan Schokker**; Wageningen Livestock Research, Wageningen University & Research.

**Mari A Smits**; Host Microbe Interactomics; Wageningen Livestock Research; Wageningen Bioveterinary Research, Wageningen University & Research.

**Manuscript in preparation**



**Abstract**

The mucosal immune system of the intestine balances its tolerance towards feed/food derived antigens and commensal microbes and its response to suppress the activity of pathogenic microbes. Changes in the diet are known to affect immune parameters like Immunoglobulin A and cytokine levels, both at the intestine and at the systemic level. These immune parameters help shape the external immune phenotype of the organism. In order to investigate the effects of dietary changes on functional immunity parameters, we analysed and compared the gene expression profiles of intestinal mucosal tissues of mice fed with various different diets. Such a molecular profile may be regarded as an internal phenotype. In addition, we obtained information on 36 functional immunity parameters that are linked to the external immune phenotype based on Genome Wide Association Studies (GWAS) as deposited in the NCBI dbGaP database. Not less than 1354 genes associated with at least one of the 36 selected functional immunity parameters. These genes are indicated here as Functional Immunity Parameters Genes (FIPGs).

Data from 9 different dietary intervention studies were analysed both at the gene and the pathway level to compare the intestinal mucosal transcriptomes of the experimental animals versus the control animals for each study. These results of these analyses were then linked to the selected functional immunity parameters via the FIPGs associated with them. Out of the 36 functional immunity parameters, 31 were found to be linked to changes in diet by either differentially expressed genes or differentially regulated pathways. With this, we discovered potential links between dietary interventions and functional immunity parameters.

**Keywords:** Dietary intervention, Gene expression, Functional immunity parameter, Gastrointestinal tract, Immune system

## Introduction

The intestinal mucosal immune system maintains a delicate balance between immunogenicity against invading pathogens and tolerance against feed/food antigens and the commensal microbiota. Changes in diets have been shown to affect mucosal as well as systemic parameters of immunity, such as levels of Immunoglobulin A antibodies and cytokines like Interleukin 10 (Y.-S. Wu et al. 2016; Gomaa and El-Aziz 2017). Variations in this type of immunity parameters correlate with external immune-related phenotypes such as the inflammatory status of an organism (Ding et al. 2010; Munkonda et al. 2012) or its susceptibility to an infectious disease (Alex et al. 2009). Still, limited knowledge is available on the relationships between dietary components, immunity parameters and external immune phenotypes.

An improved understanding of these relationships may be obtained by studying the effect of diets or dietary components on so-called internal or intermediate phenotypes (Houle, Govindaraju, and Omholt 2010; Leuchter et al. 2014). Such internal phenotypes can be profiled in specific tissues or body fluids on different biological levels by the use of transcriptomic, proteomic, metabolomic and/or metagenomic measurements. Measurements with these techniques quantify all the components of a particular type (mRNA, protein, metabolites, microbiota) in the tested sample. The activity and the interactions of several of these components, both intra- component-type as well as inter- component-type, may contribute to the establishment of external phenotypes which is based on the genotype interactions with the environment.

Since complex external phenotypes, such as immunological resistance, are not always easy to quantify, it is common to identify quantifiable functional-related parameters that are known to be indicators of a particular external phenotype. Several of these indicators have been identified for external phenotypes, especially those associated with a diseased state, for example the insulin serum concentration as an indicator for diabetes.

In a previous study, we have identified several immune related genes and pathways that are affected in intestinal mucosal tissues by dietary interventions in mice (Benis *et al.*, in

preparation). It is not known yet to what extent these changes contribute to the expression of immunity parameters and to the establishment of immune related external phenotypes. Therefore, the aim of this study is to develop and build a concept of a quantitative model to predict the effects of dietary changes on the expression of a selected set of immunity parameters that are functional related to the immune phenotype (immune competence) of an organism.

As a first step, we developed and built the concept of the model using microarray derived transcriptome data because this data is comparable between various mice experiments and because it covers the expression of whole genomes. Moreover, several dedicated databases exist that store such gene expression data (Gene Expression Omnibus (Barrett et al. 2013, Array Express (Kolesnikov et al. 2015) thereby enabling the integration of results arising from multiple independent experiments. To link gene expression data with functional immunity parameters and immune related external phenotypes, we use databases with information on the linkage between genes and such immunity parameters. The database of Genotypes and Phenotypes (dbGaP) (Tryka et al. 2014 and the National Human Genome Research Institute Genome Wide Association Studies Catalog (NHGRI GWAS Catalog) (Welter et al. 2014 contain results from studies on interactions between genotypes and phenotypes, most often arising from Genome Wide Association Studies (GWAS) experiments. dbGaP and NHGRI GWAS Catalog contain information on genes associated to the functional immunity parameters.

## **Datasets and Methods**

### **Microarray datasets**

We acquired microarray data from 9 mice studies where the diet of the animals was altered and a section of the intestine was sampled for gene expression data with microarray technology. In three of these studies relative levels of dietary components were changed: three different levels of fat were tested with gene expression measured in three sections of the small intestine (de Wit et al. 2011; a high fat diet transcriptomics measured in two sections of the small intestine (Desmarchelier et al. 2012; and five different types of fibres with responses measured in the colon (Lange et al. 2015). The six other studies added supplements / additives in the diet: dietary heme was added and the colon sampled for gene

expression (Ijssennagger et al. 2012; quercetin was added and two sections of the small intestine and the colon were sampled (Mutch et al. 2006; a Ppar alpha agonist was given to the mice and the small intestine was sampled (Bünger et al. 2007; and a probiotic namely, VSL#3, was administered to the mice and the small intestine and colon of two strains of mice were sampled for gene expression (Mariman et al. 2015. These nine studies comprise a total of 25 experimental conditions. Here an experimental condition is defined as a dietary change measured in a section of tissue in a strain of animal. Each study also contained data corresponding to control conditions in the same tissue and strain of animal. More details on these 25 conditions are given in Supplementary Table 1 which also contains the abbreviations used in this paper to denote these conditions.

### **Differential gene expression analysis**

Gene expression was analysed by comparing each of the 25 experimental conditions to their respective controls using linear modelling for microarrays (LIMMA) (Smyth 2005. Significance of gene differential expression was evaluated with p-values after correcting for multiple testing with the Benjamini & Hochberg adjustment. Genes were considered differentially expressed when False Discovery Rate (FDR) less than 0.05 and log<sub>2</sub> fold change more than 1.3 or less than -1.3 to only capture highly variant genes.

### **Differential pathway analysis**

Using the topological pathway enrichment tool CePa (Z. Gu and Wang 2013 and pathways from Reactome version 51 (Croft et al. 2014; Fabregat et al. 2016 with a modified version of the GSA (Gene Set Analysis) function we analysed the data for each experimental condition versus the control in each experiment as described in Benis *et al* (manuscript in preparation). Pathways with p-values less than 0.01 were considered significant for that comparison, p-values were not adjusted for multiple comparisons because the pathways are not independent of each other.

### **Functional parameter to gene association**

We extracted genotype to phenotype associations from the interface Phenotype Genotype Integrator (PheGenI) (Ramos et al. 2014 which accesses data from dbGaP and NHGRI

GWAS Catalog. The information from dbGaP is based on over 800 GWAS studies submitted to NCBI and the NHGRI GWAS Catalog has information from over 2,000 GWAS publications. GWAS studies primarily investigate variations in genes that could be the cause of health or disease traits. These variations are generally studied in the form of Single Nucleotide Polymorphisms (SNPs) that differ between experimental and control cohorts. GWAS studies aim to identify a relation between the presence of a SNP and the occurrence or risk of a disease or another external phenotype, such as feed efficiency in livestock sciences. Generally, if the SNP is found on a gene or in an intergenic region, and this association passes a significance threshold, the disease or phenotype is attributed to these genes. Data can be retrieved from dbGaP with the browser PheGenI where it is possible to search for data pertaining a selected phenotype or trait. The terminology describing these traits is taken from the thesaurus Medical Subject Headings which has precise definitions for the terms and thus reduces ambiguity. In addition to the association between the trait and the gene where the SNP is located, PheGenI also returns provenance information related to study that identified the association.

We hand-picked 36 immunity parameters such that the parameters are easily measured and are functionally related to immune relevant external phenotypes. These functional immunity parameters were queried in PheGenI and downloaded all the genes associated with these parameters. For simplicity, we refer to these genes as Functional Immunity Parameters Genes (FIPG). dbGaP was accessed with PheGenI in 4 different batches on 09-Jan-2017.

### **Pathways associated with functional parameters**

The functional parameters were associated to different pathways by using the Over Representation Analysis (ORA) method in the pathway analysis tool CePa. All the genes associated with a functional immunity parameter (FIPGs) were used as a gene list and all the genes in the pathways were used as a background gene list. The ORA analysis was run five times with 5,000 random iterations each and the pathways with p-values less than 0.01 in all the runs were considered to be significant. These significant pathways were assigned to each functional parameter.

## Associations of diets and functional parameters

The dietary conditions were linked to functional parameters using two different methods. In the first, the *gene method*, we use the differentially expressed genes from each dietary condition. A dietary condition and a functional immunity parameter were linked if the FIPGs associated to the parameter are seen to be differentially expressed in the diet exposure.

In the second approach, the *pathway method*, we first associated pathways to diets whenever the pathways were significantly enriched (p-value < 0.01) in the experimental condition versus the control. Pathways were then associated to functional immunity parameters if the pathway genes showed enrichment in each set of FIPGs. Finally, dietary conditions and functional immunity parameters were linked whenever they were associated to common pathways.

## Literature validation of Diet-Parameter associations

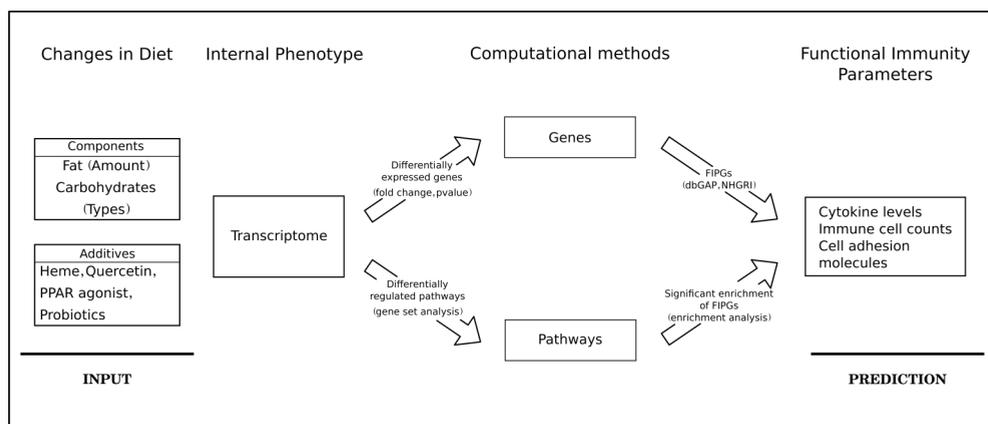
In order to verify whether the observed associations between diets and functional immunity parameter are already described in literature, we performed a literature search for co-occurrence of dietary changes and functional parameters that are associated in either methods. We used the R package *rentrez* (Winter 2016) to perform the search on PubMed abstracts. We reduced the names of the dietary changes to more general search terms; all the ‘Fat’ conditions were converted to “High Fat”, all the ‘Fibre’ conditions were converted to names of the fibres used, “Arabinoxylan”, “Oligofructose”, “Guar gum”, “Inulin” and “Resistant Starch”. The ‘Additive’ experimental conditions were searched as “Dietary heme”, “Quercetin”, “PPAR agonist” and “Probiotic” respectively.

## Results

### Concept of the model

The developed model describes associations between dietary interventions and functional immunity parameters pertinent to external immune related phenotypes by the use of die-

tary-induced genome-wide transcription data. We used two types of differential analysis on the gene expression data, one is based on the differential expression of individual genes and the other on the differential regulation of biological pathways. The differentially expressed genes and pathways were subsequently linked to the functional immunity parameters (Figure 1). From the interface PheGenI, we obtained Genes associated with Functional Immunity Parameters, termed FIPGs for simplicity. The hypotheses behind these methods is that the dietary-induced differential expression of FIPGs, or the differential expression of gene sets that are functionally linked to FIPGs, affect the same functional immunity parameters as the genetic variation of the FIPG.



**Figure 1. Linking dietary changes to functional immunity parameters related to the external immune phenotype via transcriptome data.** Scheme representing the model concept to link dietary conditions to immune functional parameters using transcriptome data from gut mucosal tissue. Model inputs and predictions are shown on the left and right, respectively. The middle part describes the two methods here presented, the *gene method*, based on differential expression of individual genes, and the *pathway method* that uses significantly regulated pathways. The arrows contain information on the analysis methods. FIPG denotes the genes associated with the functional immunity parameters as obtained from PheGenI.

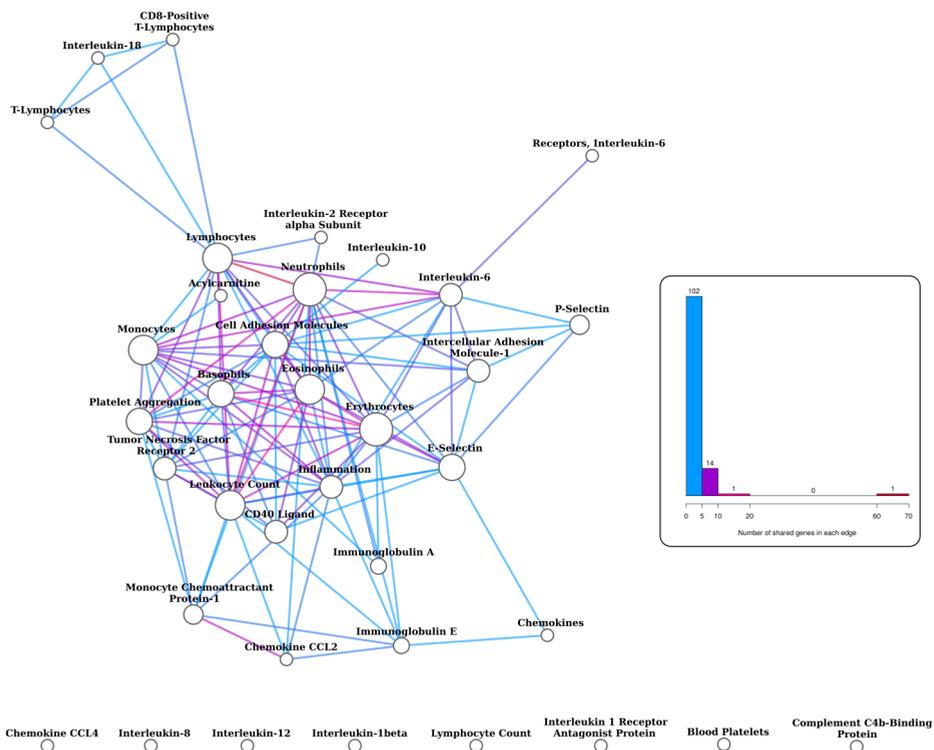
### Gene to Functional Immunity Parameter associations from GWAS studies

From the traits linked to genes in GWAS studies we chose 36 traits corresponding to easily measured parameters linked to immunity. Twelve of them are related to immune cell counts and behaviour, another twelve are cytokines or chemokines or their receptors, two

are immunoglobins, four are adhesion related proteins, three are proteins expressed on immune cells, one is a metabolite (Acylcarnitine) and one other is a physiological state (Inflammation). More details on these functional immunity parameters are given in Supplementary Table 2.

The chosen 36 functional immunity parameters were associated with 1354 Functional Immunity Parameter Genes (FIPGs) in the results from PheGenI. These trait-to-gene associations arise from SNPs being assigned to a particular gene based on the location of the SNPs in relation to the chromosomal boundaries of that gene. The gene assignments are performed in the individual GWAS the results of which are submitted to dbGaP or extracted from literature in NHGRI GWAS Catalog. The basis for gene assignment is given as a functional code in the results of PheGenI. Among the 4001 results downloaded from dbGaP, half of the associations have the functional code 'Intergenic' which indicates that the SNP associated with the trait was found more than 2 kilo base-pairs up-stream and more than 500 base-pairs down-stream of the genes flanking this intergenic region. Among all the FIPGs, 86% have been associated with traits with SNPs that are found in intergenic regions. There are eight other functional codes among the results from dbGaP: Intron, Missense, NearGene-3, NearGene-5, UTR (untranslated region) -3, UTR-5, Cds (coding DNA sequence) synon and STOP-GAIN which are explained in Supplementary Table 3.

There are four functional immunity parameters (Interleukin-10, Interleukin-1beta, Interleukin-12, Complement C4 binding protein) with only one FIPG, whereas the parameter Erythrocytes has the largest number of FIPGs with 184. From the 36 functional immunity parameters selected, 28 were found to share FIPGs, as can be seen in Figure 2. The parameter Neutrophils shares FIPGs with 17 other parameters, with a maximum of 70 FIPGs with the parameter Lymphocytes. The number of genes shared among the functional immunity parameters is limited as shown in the insert of Figure 2; 102 of the 117 connections have a maximum of five shared FIPGs. The gene DARC is the most shared FIPG and is associated with 6 functional immunity parameters (Neutrophils, Monocyte Attractant Protein -1, Leukocyte count, Chemokine CCL2, Immunoglobulin E and Inflammation).



**Figure 2. Connections between functional immunity parameters.** The nodes in the network represent functional immunity parameters with the node size corresponding to the number of FIPGs associated with them in dbGaP. The edges represent shared FIPGs between the connected functional immunity parameters and are coloured according to the number of shared FIPGs. Edges that represent less than 5 FIPGs are coloured blue, shared FIPGs more than 5 and less than 10 are coloured purple, FIPGs with more than 10 genes are coloured pink. There is one edge coloured red which represents 70 shared FIPGs. The same colours are used in the inset which shows the distribution of shared FIPGs in the network.

### Differential analysis of dietary conditions on gene and pathway levels

Each dietary condition in the nine studies was compared to the control in the corresponding study (details given in Supplementary Table 1) and differential gene expression was computed. Out of the 25 dietary conditions there were 18 conditions that had significant differentially expressed genes, as is shown in Table 1. A total of 602 differentially ex-

pressed genes were identified. The condition with the highest number of differentially expressed genes is Add.PparAgonist.SI with 199 significant genes.

The differential gene expression analysis was also performed on a pathway level where the expression of the pathway nodes in each of the 25 dietary conditions was compared to the pathway expression in the control conditions. All the 25 conditions had significantly regulated pathways, Add.Heme.C has the highest number of significant regulated pathways (882) as is shown in Table 1. A total of 1289 Reactome pathways were found significant in at least one of the dietary conditions.

**Table 1: Differential expressed gene and pathway linked to functional immunity parameters.**

The first column has the names of the dietary conditions, the second has number of differentially expressed genes and the fourth column has the number of differentially regulated pathways. The third and fifth columns show the number of associations between the dietary condition and the functional immunity parameter followed by the number of those associations validated in a literature search. Occasions where there were no literature results for uncovered diet-parameter associations are marked in red.

Dietary conditions	Number of Differential Expressed Genes	Number of associated functional immunity parameters in model / number found in literature ( <i>Gene Method</i> )	Number of Regulated Pathways	Number of associated functional immunity parameters in model / number found in literature ( <i>Pathway Method</i> )
Fat.20.D.SI	0	0	128	11 / 10
Fat.30.D.SI	6	0	238	17 / 15
Fat.45.D.SI	30	3 / 3	260	14 / 13
Fat.20.M.SI	0	0	292	16 / 13
Fat.30.M.SI	3	0	369	21 / 19
Fat.45.M.SI	37	3 / 3	349	17 / 15
Fat.20.P.SI	0	0	120	17 / 15
Fat.30.P.SI	1	0	207	15 / 13
Fat.45.P.SI	4	0	243	14 / 12
Fat.60.D.SI	62	3 / 3	230	23 / 20

Fat.60.P.SI	33	1 / 1	438	21 / 19
Fibre.AX.C	3	0	532	23 / 0
Fibre.FOS.C	0	0	669	25 / 1
Fibre.GG.C	2	0	669	26 / 0
Fibre.IN.C	9	1 / 0	356	25 / 7
Fibre.RS.C	4	0	528	23 / 2
Add.Heme.C	118	7 / 3	882	26 / 3
Add.Quercetin.C	0	0	199	15 / 7
Add.Quercetin.I.SI	0	0	126	11 / 8
Add.Quercetin.J.SI	0	0	131	14 / 8
Add.PparAgonist.SI	199	10 / 4	344	20 / 6
Add.Probio.BALBc.C	2	0	98	9 / 3
Add.Probio.BALBc.SI	143	5 / 4	166	21 / 14
Add.Probio.C57BL6.C	3	1 / 0	166	17 / 9
Add.Probio.C57BL6.SI	170	9 / 8	68	10 / 8

### Diet-Gene-Parameter associations

Out of the 18 dietary conditions that have significantly differentially expressed genes, ten of them are related to at least one functional parameter (Table 1). These associations are represented in Figure 3.A which is also used as a ranking method (from the top-right). The top three Diet – Parameter associations (ranked from the top right) from the *gene method* are, Add.Probio.C57BL6.SI, Add.PparAgonist.SI and Add.Heme.C.



### **Diet-Pathway-Parameter linkage**

We linked all the Reactome pathways to the functional immunity parameters with the gene associations obtained from the dbGaP results. Each set of FIPGs is used for pathway enrichment analysis. For a p-value threshold of 0.01 in 5 iterations, there were a total of 283 pathways linked to 29 functional immunity parameters. Each functional immunity parameter is significantly associated with an average of 13.7 pathways while the 283 pathways are linked to 1.4 parameters on average. A majority (71%) of the significant pathways are associated with only one functional immunity parameter.

Subsequently, with these Pathway-Parameter associations, we link the dietary conditions to functional immunity parameters by seeking out the differentially regulated pathways in which the FIPGs are over represented. We discovered that 217 pathways were significant in both the differential analysis and in the over-representation analysis. These 217 pathways were linked to 25 dietary conditions and 29 functional immunity parameters, as is shown in Table 1. Based on the number of significant pathways and number of functional immunity parameters linked to those pathways (see Figure 3.B), we ranked the dietary conditions. The top three dietary conditions in this ranking are Add.Heme.C, Fibre.GG.C and Fibre.FOS.C, see Figure 3.B.

### **Functional immunity parameters associated with dietary conditions**

Overall, combining the results of both methods, we observed associations between 25 diets and 31 functional parameters. No links were observed for T-lymphocytes, CD8 positive T-lymphocytes, Interleukin-8 and Interleukin-12. Among the functional immunity parameters that are associated with dietary changes, there are three with only one FIPG (Interleukin-1, Interleukin-10 and Complement C4b-binding protein). The parameter Interleukin-1 is associated with 3 dietary conditions in the *gene method* alone. The parameters Interleukin-10 and Complement C4b-binding protein are linked to 18 (through 1 significantly over represented pathway) and 14 (through 11 significantly over represented pathways) dietary conditions, respectively. The functional immunity parameter that is linked to the most dietary conditions is Immunoglobulin E which has 17 FIPGs and is associated with 23 dietary conditions. The functional immunity parameter Erythrocytes which

has the largest number of FIPGs is linked to 22 dietary conditions through the *pathway method* and 5 dietary conditions by the *gene method*.

### Screening literature for identified Diet-Parameter associations

We performed an automated search for co-occurrence of associated diets and functional immunity parameters in PubMed abstracts. The dietary changes were converted into more general search terms, such as “High fat”, “Dietary Heme”, and the results are shown in Table 1. The results of the literature search (PubMed identifiers) are given in Supplementary File 1. All the ‘Fat’ to functional parameter associations found in the *gene method* are represented in literature. The one ‘Fibre’ association from the *gene method* does not have any results in literature, neither do two ‘Fibre’ associations from the *pathway method* even though there are several associations uncovered by the *pathway method*. The functional immunity parameters associated with the ‘Fat’ dietary conditions have the most literature results.

### Highest ranked dietary intervention

In both the *gene* and the *pathway method*, the dietary condition Add.Heme.C is among the top three conditions. Table 2 shows all the connections that this dietary condition has with the functional parameters as found in this analysis. In the *gene method* Add.Heme.C is associated with six functional parameters and is ranked third, in the *pathway method* it is ranked highest with the highest number of pathway associations (882) and parameter associations (26).

**Table 2. Functional parameters associated with the Add.Heme.C dietary condition.** Each row contains information on the functional immunity parameters that are associated with the ‘Additive’ dietary condition. Columns three and four have the number of genes and pathways that link the functional immunity parameter to the dietary condition via the *gene method* and the *pathway methods* respectively.

Broad functionality	Functional immunity parameter	Number of linking genes in the <i>Gene method</i>	Number of linking pathways in the <i>Pathway method</i>
Immune Cell	Basophils	0	3

Immune Cell	Eosinophils	1	2
Immune Cell	Erythrocytes	2	9
Immune Cell	Leukocyte Count	1	7
Immune Cell	Lymphocyte Count	0	12
Immune Cell	Lymphocytes	0	2
Immune Cell	Monocytes	0	10
Immune Cell	Neutrophils	0	3
Immune Cell	Platelet Aggregation	0	6
Cytokine	CCL2	0	2
Cytokine	Interleukin 1 Receptor Antagonist Protein	0	14
Cytokine	Interleukin-6	0	5
Cytokine	Interleukin-6 Receptors	0	11
Cytokine	Interleukin-10	0	8
Cytokine	Interleukin-18	0	11
Cytokine	Monocyte Chemoattractant Protein-1	0	4
Antibody	Immunoglobulin A	0	10
Antibody	Immunoglobulin E	0	9
Cell adhesion molecule	Cell Adhesion Molecules	3	3
Cell adhesion molecule	E-Selectin	0	5
Cell adhesion molecule	P-Selectin	0	5
Cell adhesion molecule	Intercellular Adhesion Molecule-1	1	2
Others	CD40 Ligand	1	29
Others	Complement C4b-Binding Protein	0	1
Others	Inflammation	2	11
Others	Tumour Necrosis Factor Type II Receptors	0	6

---

## Discussion

In this study, we developed and built a concept of a model that links dietary changes to immunity parameters that are functionally linked to immune related external phenotypes. This is a first-generation model and it is based only on transcriptome data because it is readily available for reuse from online repositories like Gene Expression Omnibus. We used two methods, both of which rely on gene expression data obtained from gut mucosal tissues. Extension of the model by inclusion of other types of data is possible, provided that suitable new links to functional immunity parameters can be identified for the new data types. For example, it may be possible to directly link proteome data to the functional parameters using a similar method.

### Functional parameters from dbGaP

We obtained genes associated with selected functional immunity parameters (FIPGs) from the interface PheGenI (Ramos et al. 2014). Genes were associated with these selected functional immunity parameters with varying levels of significance gauged by a p-value. Since the data were derived from a variety of studies, we decided to use all the gene associations regardless of the p-value and the location of the variation (SNP) on the gene. In the GWAS results obtained from dbGaP there are several ways in which the genes are associated with a SNP and thus a trait. Most of the association were found in intergenic regions of the genome. There have been concerns about assigning traits to genes that are close to the SNPs when it is possible that the intergenic region regulates genes much further away. But we decided to work with the gene assignments as deposited in dbGaP without excluding any gene to trait associations based on SNP location. There have been several studies that integrate gene expression data with GWAS data using expression Quantitative Trait Loci's. It has been described for several traits that SNP in DNA regions that regulate the expression of specific genes, significantly contribute to that trait (Nicolae et al. 2010). Hence in this study, we used transcriptomic data to associate experimental conditions to traits under the assumption that differential expression of a gene will impact the same functions that are attributed to variations in that gene.

---

**Diet-Parameter associations and literature validations**

We obtained fewer diet to functional immunity parameter associations through the differentially expressed genes than associations through significantly different expressed pathways. This is in line with our expectation since the number of differentially expressed genes is lower than the number of significant pathways in each dietary condition as shown in Table 1. Gene set enrichment analysis is often recommended for gene expression analysis with the aim to identify differences in biological pathways, as groups of genes are more likely to be responsible for a functionality than a single gene (Emmert-Streib and Glazko 2011). This is reflected not only in the number of connections between dietary conditions and functional parameters, but also in the literature validation results as there are more literature results for the Diet-Parameter associations obtained via the *pathway method*. Literature results are highly dependent on the topics that are currently studied. Among the ‘Fat’ and ‘Fibre’ dietary conditions there are many more literature results for the Fat-Parameter connections, indicating dedicated experiments to measure the effect of changes in the fat content on aspects of the immune system. For example, Hogan et al (Hogan et al. 2017) discovered changes in the levels of E-Selectin and Interleukin 6 in high fat fed mice. Several of these experiments were performed in order to understand the development of obesity and the immune implications of the disease. Other dietary components, carbohydrates and proteins, have not been implicated in diseases, hence changes in the immune system are not the focus of these dietary intervention studies (Lange et al. 2015, which results on a lower recovery rate for the identified Fibre-Parameter connections. Among the ‘Additive’ dietary conditions, the probiotic conditions have the most literature validations as they have been extensively studied in the context of gut health benefits. Thus, literature validation of the observed associations resulting from our model, points to several gaps in knowledge in the effects of diets on the immune system. Several connections that we identified between dietary interventions and functional immunity parameters appear not to be investigated yet. As shown in this study dietary changes do bring about several changes in the immune system hence should be investigated in dietary intervention studies. Besides recording the changes induced by the dietary interventions in internal phenotypes, the changes induced in immune related parameters and external phenotypes must be recorded more often.

---

### Increased Heme in the diet

We specifically concentrated on the dietary intervention Add.Heme.C since this Diet-Parameter connection ranked among the top three in both the gene and the *pathway methods*. In this intervention, dietary heme was added to the diet in order to simulate a human diet rich in red meat. We have identified several inflammation related parameters linked to this dietary perturbation: inflammatory cytokines Interleukin-6, Interleukin-18, the CD40 ligand, and the Inflammation parameter itself (see Table 2). Also, several immune cell types appear associated with the dietary condition: Neutrophils, Basophils, Eosinophils and Monocytes. Among the associations between “heme addition to the diet” and functional immune parameters, three correspond to connections that are already described in literature, the link to Erythrocytes, Inflammation and Interleukin-6. These literature results further validate the identified links between dietary heme and inflammation (Le Leu et al. 2013). Heme addition to the diet has also been shown to damage the colonic mucosal layer and increased proliferation of epithelial cells (IJssennagger et al. 2012). However, IJssennagger *et al* report that they observed no changes in certain inflammatory marker genes, which are different from the FIPGs we found to be related to inflammation.

### Uses of a Diet-Parameter model

With the associations between dietary components and immunity parameters functionally involved in immune-related phenotypes that we identified in this study, it is possible to generate new hypothesis on how changes in diets may affect the immune system and its ability to defend the host against challenges. This model integrates both experimental data and already existing information in databases and can be expanded to build a more comprehensive model. In order to achieve a higher level of prediction, there needs to be an iterative cycle of model driven experimentation, with *in vivo* experiments specifically designed to measure associations between functional immunity parameters and dietary interventions.

Availability of an extensive enough data compendium will enable a statistical model which would treat the functionalities of the organism or gut as an unknown variable or a black box and only predict the changes in functional parameters given changes in diet. Although this model would serve the purpose of predicting immune functional changes, it

would not provide insights on the mechanisms in the gut that leads to changes. Our modelling approach explicitly includes measurements of internal phenotypes (i.e. transcriptomics or proteomics) to uncover associations between the environment (here the diet) and the external immune phenotype, therefore it also has the potential to unravel the mechanisms behind the functional changes in the immune system induced by dietary changes.

### Supplementary Files

**Supplementary Table 1. Description of the 25 conditions.** This table describes the abbreviated names of experimental conditions as used in the paper and the control conditions in each of the experiments.

	Names	Description of Experimental conditions	Description of the control condition
<b>F a t</b>	Fat.20.P.SI	20% of the energy in the diet comes from fat, proximal section of the small intestine is sampled	10% of the energy in the diet comes from fat, proximal section of the small intestine is sampled
	Fat.30.P.SI	30% of the energy in the diet comes from fat, proximal section of the small intestine is sampled	
	Fat.45.P.SI	45% of the energy in the diet comes from fat, proximal section of the small intestine is sampled	
	Fat.20.M.SI	20% of the energy in the diet comes from fat, middle section of the small intestine is sampled	10% of the energy in the diet comes from fat, middle section of the small intestine is sampled
	Fat.30.M.SI	30% of the energy in the diet comes from fat, middle section of the small intestine is sampled	
	Fat.45.M.SI	45% of the energy in the diet comes from fat, middle section of the small intestine is sampled	

<b>F a t</b>	Fat.20.D.SI	20% of the energy in the diet comes from fat, distal section of the small intestine is sampled	
	Fat.30.D.SI	30% of the energy in the diet comes from fat, distal section of the small intestine is sampled	
	Fat.45.D.SI	45% of the energy in the diet comes from fat, distal section of the small intestine is sampled	
	Fat.60.D.SI	60% of the energy in the diet comes from fat, distal section of the small intestine is sampled	11% of the energy in the diet comes from fat, distal section of the small intestine is sampled
	Fat.60.P.SI	60% of the energy in the diet comes from fat, proximal section of the small intestine is sampled	11% of the energy in the diet comes from fat, proximal section of the small intestine is sampled
<b>F i b r e</b>	Fibre.AX.C	Control diet with 10% of corn starch substituted with Arabinoxylan, sampled at the colon	
	Fibre.FOS.C	Control diet with 10% of corn starch substituted with Oligofructose, sampled at the colon	
	Fibre.GG.C	Control diet with 10% of corn starch substituted with Guar gum, sampled at the colon	
	Fibre.IN.C	Control diet with 10% of corn starch substituted with Inulin, sampled at the colon	
	Fibre.RS.C	Control diet with 20% of corn starch substituted with Resistant starch sampled at the colon	

A d d i t i v e	Add.Heme.C	Control diet supplemented with 0.5 $\mu\text{mol/g}$ heme, sampled at the colon	Westernized control diet (40% fat (mainly palm oil) low calcium (30 $\mu\text{mol/g}$ )), sampled at the colon
	Add.Quercetin.C	RM3 (E) 801710 Soya-free powdered diet + 7mg quercetin, sampled at the colon	RM3 (E) 801710 Soya-free powdered diet, sampled at the colon
	Add.Quercetin.I.SI	RM3 (E) 801710 Soya-free powdered diet + 7mg quercetin, sampled at the ileum	RM3 (E) 801710 Soya-free powdered diet, sampled at the ileum
	Add.Quercetin.J.SI	RM3 (E) 801710 Soya-free powdered diet + 7mg quercetin, sampled at the jejunum	RM3 (E) 801710 Soya-free powdered diet, sampled at the jejunum
	Add.PparAgonist.SI	Standard laboratory chow (RMH-B), sampled at the small intestine	Standard laboratory chow (RMH-B), sampled at the small intestine
	Add.Probio.BALBc.C	SSNIFF R/M-H feed pellet diet or $3 \times 10^8$ CFU VSL#3 suspended in 200 mL PBS, strain BALBc mice, sampled at the colon	SSNIFF R/M-H feed pellet diet given oral gavage with PBS, strain BALBc mice, sampled at the colon
	Add.Probio.BALBc.SI	SSNIFF R/M-H feed pellet diet with $3 \times 10^8$ CFU VSL#3 suspended in 200 mL PBS, strain BALBc mice, sampled at the small intestine	SSNIFF R/M-H feed pellet diet, oral gavage with PBS, strain BALBc mice, sampled at the small intestine
	Add.Probio.C57BL6.C	SSNIFF R/M-H feed pellet diet with $3 \times 10^8$ CFU VSL#3 suspended in 200 mL PBS, strain C57BL6 mice, sampled at the colon	SSNIFF R/M-H feed pellet diet, oral gavage with PBS, strain C57BL6 mice, sampled at the colon
	Add.Probio.C57BL6.SI	SSNIFF R/M-H feed pellet diet with $3 \times 10^8$ CFU VSL#3 suspended in 200 mL PBS, strain C57BL6 mice, sampled at the small intestine	SSNIFF R/M-H feed pellet diet, oral gavage with PBS, strain C57BL6 mice, sampled at the small intestine

**Supplementary Table 2. Functionalities of immunity parameters.** The names of the functional immunity parameters are in column one along with the most well-known functions of those parameters in the immune system. The last column shows the number of genes associated with the functional immunity parameters in the results of GWAS studies from dbGaP and NHGRI. The functional immunity parameters in bold are not found to be associated with dietary changes in either the *gene method* or the *pathway method*.

Broad functionality	Functional immunity parameter	Functional information	Number of associated genes (FIGs)
Immune Cell	Basophils	Leukocyte, involved in phagocytosis in inflammatory response	117
Immune Cell	Blood Platelets	Anucleate cellular fragments, first responders to wound to begin clotting, also have TLR receptors	2
Immune Cell	<b>CD8-Positive T-Lymphocytes</b>	T-cell that expresses CD8 on the surface, two sub-types; cytotoxic T-cells and suppressor T-cells	2
Immune Cell	Eosinophils	Leukocyte, triggered by cytokines, immunoglobins and complements, pro-inflammatory.	124
Immune Cell	Erythrocytes	Most abundant cell type in the blood, participates in antigen-antibody interaction, coats pathogens to mark for phagocytosis	184
Immune Cell	Leukocyte Count	Number of white blood cells in venous blood, immune cells	137
Immune Cell	Lymphocyte Count	Number of lymphocytes in venous blood	2
Immune Cell	Lymphocytes	Immune cells occurring in the lymphatic system, subtypes are natural killer cells, T-cells and B-cells	124

Immune Cell	Monocytes	Leukocyte produced in the bone marrow, matures to be a macrophage that can phagocytose pathogens	121
Immune Cell	Neutrophils	Leukocyte, highly motile and is among the first immune cells at site of stimulation	182
Immune Cell	Platelet Aggregation	Clumping of platelets activated by thrombin as a first step of blood clotting	103
Immune Cell	<b>T-Lymphocytes</b>	Lymphocyte produced by the thymus gland, two types of cells; cytotoxic T-cells that kill invaders and helper T-cells that activate other immune cells	2
Cytokine	CCL2	Chemokine that attracts monocytes or macrophages from blood into tissues in response to inflammation	7
Cytokine	<b>CCL4</b>	Chemokine of family CC, produced by monocytes and macrophages	4
Cytokine	Chemokines	Class of cytokines that are able to induce chemotaxis in responsive cells, classified into CXC, CC, CX3C and XC sub-families	9
Cytokine	Interleukin-1beta	Cytokine that is produced by activated macrophages, plays a role in inflammation, cell proliferation and differentiation	1
Cytokine	Interleukin 1 Receptor Antagonist Protein	Cytokine that dampens inflammatory response	3
Cytokine	Interleukin-2 Receptor alpha Subunit	Pluripotent cytokine, control regulatory T cells	3
Cytokine	Interleukin-6	Cytokine, pro-inflammatory action, produced at sites of inflammation	41

Cytokine	Interleukin-6 Receptors	A complex essential for the activity of Interleukin-6	3
Cytokine	<b>Interleukin-8</b>	Chemokine produced by tissues and blood, targets neutrophils	2
Cytokine	Interleukin-10	Cytokine synthesis inhibitory factor, is anti-inflammatory by limiting immune response	1
Cytokine	<b>Interleukin-12</b>	Cytokine produced by phagocytic cells, pro-inflammatory action by inducing cytokine production	1
Cytokine	Interleukin-18	Cytokine, pro-inflammatory action, regulates innate and adaptive immunity	4
Cytokine	Monocyte Chemoattractant Protein-1	Chemokine that attracts monocytes or macrophages from blood into tissues in response to inflammation	22
Antibody	Immunoglobulin A	Antibody, secreted from intestinal epithelial cells to control the growth of bacteria at the luminal surface of the gut	17
Antibody	Immunoglobulin E	Antibody, important in cross talk between the innate and adaptive arms of immunity by coating parasites	17
Cell adhesion molecule	Cell Adhesion Molecules	Different molecules that help cell bind to other cells or to the extracellular matrix, also involved in wound healing	108
Cell adhesion molecule	E-Selectin	A cellular adhesion molecule, responsible for leukocytes adhering to vascular walls	81
Cell adhesion molecule	P-Selectin	Cell adhesion molecule, stored in platelets and secreted by endothelial cells, facilitates binding of platelets to leukocytes and allows movement of leukocytes on vascular surface	28

Cell adhesion molecule	Intercellular Adhesion Molecule-1	Cellular adhesion molecule that is induced by cytokines, participates in T-cell mediated inflammatory responses	43
Others	Acylcarnitine	Intermediate metabolite, diagnostic marker for fatty acid oxidation diseases	2
Others	CD40 Ligand	Belongs to the TNF receptor family, is expressed on activated T and B lymphocytes	50
Others	Complement C4b-Binding Protein	Serum protein, regulates the classical complement activation pathway	1
Others	Inflammation	Pathological process where cells or tissues are injured or destroyed, is an immune response to stimulus	73
Others	Tumour Necrosis Factor Type II Receptors	Expressed on immune cells, is involved in cell proliferation and cell death	43

**Supplementary Table 3. SNP function codes when assigned to genes.** This table shows the possible ways in which SNPs are assigned to genes as used by the NCBI database dbSNP and is only limited to the results that we work with. These codes are also referenced in the results of PheGenI when linking traits to genes.

Function Codes	Explanation
cds-synon	synonymous change.
intron	intron. E.g. rs249.
nearGene-3	within 3' 0.5kb to a gene. E.g. rs3916027 is at NT_030737.9 pos7669796, within 500 bp of UTR
nearGene-5	within 5' 2kb to a gene. E.g. rs7641128 is at NT_030737.9 pos7641128, with 2K bp of UTR starts

intergenic	variant between two genes, outside of 2Kb upstream and 500bp downstream of a gene.
STOP-GAIN	changes to STOP codon. E.g. rs328, TCA->TGA, Ser to terminator.
missense	alters codon to make an altered amino acid in protein product. E.g. rs300, ACT->GCT, Thr->Ala.
UTR-3	3 prime untranslated region. E.g. rs3289.
UTR-5	5 prime untranslated region. E.g. rs1800590.

**Supplementary File 1. Results of literature search.** Two zipped files one for the *gene method* and one for the *pathway method*, contains several text files, one for each Diet – Immunity parameter connection found in literature. The text files contain PubMed identifiers where the dietary condition and the immunity parameter co-occur in the abstract. This file can be found at <http://semantics.systemsbio.nl/index.php/data-sets-used-in-publications/>

# **Chapter 7**

## **General Discussion**

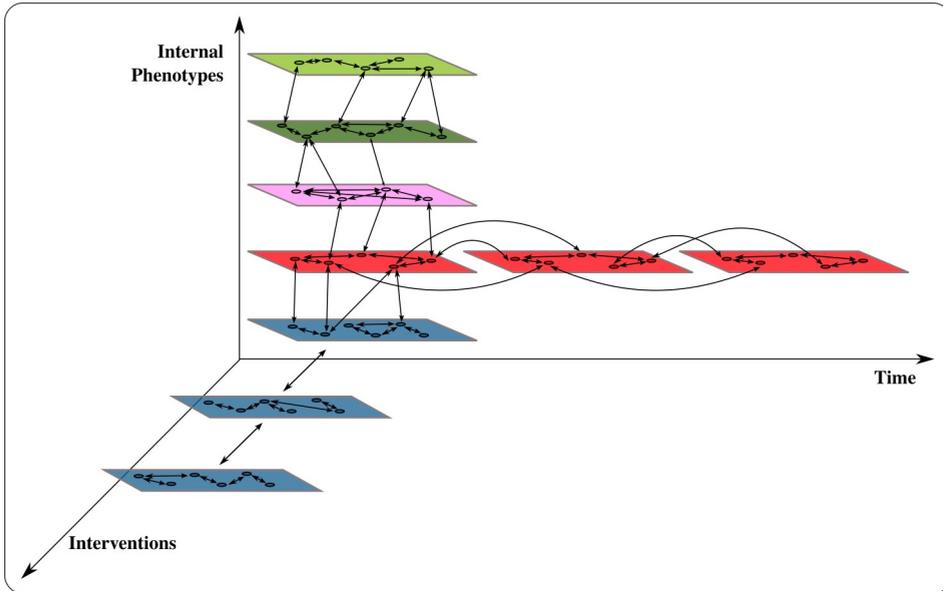


## **Background**

Health has been a major research interest for years, both related to humans and to animals of economic value, such as farm animals. Since the gastrointestinal tract (GIT) plays a major role in the health of organisms, it has been the subject of many in-depth investigations. However, even though there have been undertakings to define gut health (Kogut and Arsenault 2016; Bischoff 2011), the definition of precise parameters to quantify and define the concept of gut health has remained elusive. Knowledge is also lacking on the factors that influence health-related functionalities of the GIT and the underlying mechanisms. This is due to difficulties in studying the complex system of the GIT with precise measurements to capture time and spatial dynamics and fluctuations under different environmental conditions.

## **Objectives and research approaches**

The functioning of the GIT results from the simultaneous fulfilment of different biological processes and their mutual interactions. The various processes are heavily influenced by external factors such as dietary components and the use of (oral) drugs. In order to better understand the various processes of the gut, how they are linked together and how they are affected by different external factors, a holistic research approach is required, so that the different processes and their relationships are analysed together. Systems biology advocates for such an approach (Kitano 2002), as it aims at the holistic understanding of living systems. As a first step in obtaining a systems' view on the pleiotropic functionalities of the gut, this thesis describes the development and application of several computational methods to get an improved understanding of the gut as a system. The approach I used aggregates high throughput quantifications of intermediate biological levels (mRNA, proteins, metabolites, microbiota) connecting the genotype of the organism to the external phenotype. Throughout this thesis these intermediate levels are denoted as internal phenotypes (Houle, Govindaraju, and Omholt 2010; Leuchter et al. 2014). Their quantifications have been collected from in-house experiments, from literature and by the extraction of relevant information from accessible databases. Subsequently, computational methods and strategies have been deployed in order to integrate the multiple datasets (Figure 1).



**Figure 1. Overview of used research methodologies and data integration.** This figure is a representation of the multiple measurements of internal phenotypes and the types of data integration as described in this thesis. The colours on the parallelograms indicate different types of internal phenotypes which are represented along the vertical axis. On these parallelograms, circles represent different components of that internal phenotype e.g., transcripts, proteins, and microbes and arrows indicate interactions between them discovered from analysis of individual data sets. Arrows between layers of different colours represent interactions discovered by the use of vertical data integration methods. On the horizontal axis, internal phenotypes are indicated which are measured at different time points. Arrows between them represent interactions discovered using horizontal data integration methods. The third lateral axis represents the variation in the experimental conditions or interventions. Layers along this axis represent measurements of the same internal phenotype under different experimental conditions and the arrows between them represent the interactions discovered by the use of lateral data integration methods.

Two types of integration approaches have been used depending on whether the datasets to be integrated are related to the same internal phenotype (horizontal, also termed lateral, data integration) or they are related to different internal phenotypes (vertical data integration). These approaches are schematically indicated in Figure 1. The colours on the paral-

lelograms represent different internal phenotypes and the interactions within their components are determined by analysing individual data types. The connections between the internal phenotypes shown along the vertical axis are discovered with vertical data integration techniques used in **Chapters 2** and **3**. The parallelograms along the horizontal and lateral axes represent multiple measurements of the same internal phenotype. Along the horizontal axis are measurements across time, with the arrows indicating connections found by data integration techniques. This type of data integration across time was done in **Chapter 2** using techniques that can simultaneously handle several time points. The measurements across the Interventions axis are integrated with lateral data integration techniques, which was performed in **Chapters 5** and is the basis for the model in **Chapter 6**. This lateral data integration was performed with a method that can handle several datasets. For clarity, the spatial dimension has been left out in Figure 1. In this thesis, the spatial dimension corresponds to the different parts of the intestine, such as the small intestine and the colon, or may relate to local (GIT) or systemic internal phenotypes. Examples of such systemic internal phenotypes are metabolite and cytokine concentrations in bio fluids. The integration of local and systemic internal phenotypes represents an example of vertical data integration. It should be emphasized here that the links I have identified relate to the healthy status of the tested system. The majority of the experimental conditions or interventions used in this thesis research are dietary in nature which do not deteriorate health. It is my belief that many of these links might be affected in pathological conditions or disease states, which have not been explored here.

### **Time-dependent dynamics of processes in the gut after an early life perturbation**

There are several specifics that need to be accounted for while studying the gut as a whole. An important aspect is the time-dependent dynamics of the diverse biological processes in the GIT system since it is known that the functioning and morphology of the gut changes in time as does the composition and diversity of the luminal microbiota. In **Chapter 2**, changes in the gut of growing animals were investigated after the gut system was perturbed early in life. With a two-dimensional analysis of time and treatment together, it was found that with an interval of about 6 months, the early life perturbation still had a significant effect on the mucosal gene expression in the GIT of the animals. By using this approach, several high-level regulatory genes (linked to the functions cell cycle, cell pro-

liferation, immune processes and ubiquitination) were identified which are most likely drivers of long term changes in the gut. Such high-level regulators would not have been discovered by application of one-dimensional approaches, which is currently most common in gene expression analyses. Such high-level regulators may be important targets for the development of strategies to modulate GIT associated processes over the life of the animal.

In **Chapter 2**, not only the dynamics of the gene expression of the small intestine was studied but also, the dynamics of the microbial community structures along the horizontal axis of Figure 1. Initially, two-dimensional analysis was performed on the gene expression and the microbiota separately. Subsequently, the two types of dynamic data were integrated with each other i.e. integration of the horizontal and vertical axes of Figure 1. This time dependent integration of two different types of internal phenotypes provided information on possible microbe-host interactions that sustain the changes induced by the early life perturbation. The results showed that there were a few bacterial groups that consistently highly correlated with gene profiles that exhibited long lasting changes in response to the early life perturbation. Some of these bacterial groups (*Ruminococcus bromii*, *Faecalibacterium*) are known to have beneficial effects on the host but they have not been studied over the course of time interacting with the host. Such bacterial groups could be interesting targets for dietary prebiotics in order to modulate GIT homeostasis.

A dynamic analysis is also possible with other types of data, provided that the measurements are made at the same time points. In **Chapter 2**, only three time points with intervals of 47 and 121 days were used. However, the mechanistic information that results from this type of multi-dimensional analyses would benefit from shorter time intervals between measurements and an increased sample size. Decreasing the time intervals between sampling may provide more information on the dynamic changes that can be induced by other environment changes, for example seasonal and dietary changes or changes due to fertility cycles or movements between different compartments of husbandry systems. In addition, with large sample sizes, the statistical significance of the dynamic interactions can be gauged more accurately. However, acquiring such data from animal experiments poses several practical limitations and ethical considerations. The data used in **Chapter 2** comes from the small intestine which is almost impossible to sample without

sacrificing the animal, hence samples at different time points come from different animals undergoing the same treatment, introducing individual variation in the data. On the other hand, following the same animal over time is only possible with minimally invasive techniques, e.g., sampling blood and stool, which negatively impacts the mechanistic insights into the gut.

### **Interactions between internal phenotypes**

An important aspect that needs to be resolved for a holistic systems' understanding, is the interaction between various internal phenotypes. In **Chapter 2** a possible cross-talk between microbes and host cells was investigated through the identification of significant correlations between two internal phenotypes, the gene expression profiles and the microbiota profiles. These correlations were expected to occur, since the early life antibiotic treatment is known to affect the intestinal microbiota and the subsequent cross-talk between microbiota and host cells in the intestinal mucosa. The results of the data integration pinpointed several key microbial groups which are potentially involved in the observed long term changes in the intestinal mucosa. In **Chapter 3**, similar methods were used to investigate the interactions between five different internal phenotypes as measured at a single time point. Of these five types of data, two were based on gut associated parameters and three were based on systemic measurements. The latter were expected to be connected to biological processes in the gut system. By using a newly developed analysis pipeline, it was indeed possible to identify several biological components (genes, proteins, metabolites and microbes) that are linked to components that belong to another internal phenotypic layer. Strikingly, components were found with connections to all other internal phenotypic layers. Such components were denoted “connectivity hubs” and they can be regarded as “super system regulators”.

The presence of such regulators shows that some components play a bigger role in exchanging information between internal phenotypes. The results of **Chapter 3** are based on correlations which only points to a possible interaction between components but does not give information on causality. However, further investigation into the role of these connectivity hubs may reveal the types of interactions and the mechanisms behind their interactions. Designing new experiments to further study these interactions is not a trivial task

given the spatial differences in the measurements. In **Chapter 3** we searched existing literature for the discovered connections between the components of the internal phenotypes. Several literature results showed that the addition of one of the identified components indeed affected the connected component. Such studies give an idea of the causality between the connected components which can be investigated further. Thus, the systems approach used in **Chapter 3** successfully generated hypothesis for targeted research to identify causal relationships between the various biological scales. Expanding hypothesis driven studies by incorporating the dynamics of time dependent changes in the profiles of the various molecular and cellular components, will further reduce the knowledge gap on gut functionalities.

There are several considerations before measurements of different internal phenotypes can be integrated. It is important that all the different data being integrated originate from the same organism, from the same time point. Furthermore, it is useful to measure parameter types that are expected to influence each other. In **Chapter 2**, we integrated gene expression profiles of the jejunum and the composition of bacterial groups as measured in the same location of the small intestine. This proximity lends credibility to the possible interactions between the host and microbes that are selected up by a correlation analysis. In **Chapter 3** we integrated five data-types, two were from the same location in the small intestine and the other three were systemic measurements in serum and urine. Due to the nature of the intervention, changes in protein sources, there is reason to believe that the changes in the systemic measurements are linked to the changes in the intestine, i.e. the concentrations of amine and acyl-carnitine metabolites in the urine and blood are linked to the digestive capabilities in the small intestine. The differential cytokine concentrations in the blood are also most probably linked to the mucosal immune system of the intestine which is in direct contact with the changed diet.

The integration of five different datatypes in **Chapter 3** allowed the combined analysis of local and systemic measurements of internal phenotypes which is demands vertical integration approaches. This multi-omics integration approach is valuable and can be extended to not only explore dietary interventions, as shown in my thesis, they can also be applied to explore the effect of systemic perturbations such as stress where the majority of changes are brought about by signalling from the nervous system (Braun et al. 2013).

## Functional plasticity of the gut

In contrast to other organs like the heart and liver, the gut displays a very high level of functional plasticity. This phenomenon is reflected in the ability of the gut to adapt to the changing environment to which it is exposed. Functional plasticity of the gut can best be investigated by comparing gut responses after a variety of challenges under the same experimental conditions. Such studies with multiple challenges have not been described in literature and/or the data are not available via public databases. Therefore, in **Chapter 5** the transcriptional response data of mucosal tissues of mice, derived from 14 different individual studies, were compared. The experimental perturbations varied from mild changes in diet to an infection by pathogens and we stratified the challenges into dietary, drug based and immune challenge classes. By integrating these transcriptional responses of the gut to a variety of challenges, several biological processes (pathways) were found to be regulated irrespective of the challenge category. The consistently regulated processes fall in different classes, several metabolic processes, some immune processes and some signalling cascades, among others. The diversity of these processes indicates the number of sub-systems that are involved in restoring the gut to homeostasis after a perturbation of any sort.

Performing horizontal integration of intestinal responses to different challenges gives an overview of the processes that are common in the mucosal response towards different types of challenges and the processes that are uniquely modulated by a single type of challenge. With these results, it is possible to answer questions like “which dietary challenges regulate the same processes as the processes regulated by an infection of the intestine with a pathogenic organism?”. Identification of these common processes may form the knowledge base for the development of nutritional strategies that antagonize intestinal infections. In **Chapter 5**, datasets derived from drug-based and immune challenge studies were underrepresented as compared to the dietary challenges. Availability of larger numbers of datasets derived from drug-based and immune challenge studies may improve this knowledge-base.

The horizontal high-level data integration in **Chapter 5** was performed at a pathway level. Pathways are gene-sets that also have information on the interactions between the member

genes of a pathway. A pathway represents a biological process and is decided on by domain experts. The pathway information that was used in **Chapter 5** is from Reactome (Fabregat et al. 2016), a regularly updated and manually curated database. This pathway information was retrieved using the tool developed in **Chapter 4**. It was decided to use pathways as a functional unit because pathway gene-sets are more likely to regulate biological processes than single genes. However, in **Chapter 5** it was found that pathways are not uniformly regulated by experimental conditions that belong to different challenge categories. A majority of the pathways were up-regulated by the experimental conditions irrespective of the challenge categories. However, the expression patterns of the gene-nodes in the pathways were found to be different between experimental conditions. This plasticity in the intra-pathway response to challenges calls for caution when interpreting the pathway results.

Another way of analysing the results of the horizontal integration of data across perturbations is to critically analyse, from a biologic perspective, the regulation of a particular processes under different conditions. This can lead to the generation of hypotheses that can be tested in subsequent wet lab experiments on animals or *in vitro* as was done in **Chapter 5**. On the basis of the computational results, experts in the field of mucosal immunity hypothesised that the reason behind the regulation of several common processes in the gut under such a variety of conditions was the induction of a low-grade inflammatory condition. The pathway ‘Regulation of Complement Cascade’ was observed to be part of the processes implicated in functional plasticity of the gut. This was surprising because the Complement Cascade process is not expected to be active in gut tissue. The hypothesis of the low-grade inflammation was tested on the pathway ‘Regulation of Complement Cascade’ by designing an *in vitro* experiment that only exposed small intestinal cells to inflammatory conditions. Mice small intestinal organoids were stimulated with an inflammatory cytokine (TNF alpha), a PPARalpha agonist and flagellin to study the effects of these challenges on the local system. Among the three stimulations, the cytokine elicited the maximum response from genes in the ‘Regulation of Complement Cascade’ pathway. Besides proving that this pathway is indeed expressed in the intestinal tissue, this validation experiment showed that this ‘Regulation of Complement Cascade’ pathway is indeed regulated under low inflammatory conditions. Although experimental validation was not done, another pathway regulated by all the three categories of challenges, ‘Laminin Inter-

actions' is also known to be stimulated by inflammatory cytokines. Thus, by using the results of integrative computational approaches as a basis for the generation of hypotheses on biological processes, improved insights into the mechanisms underlying these processes in the gut may be obtained. This process is sometimes called a 'wet-lab – dry-lab' cycle in systems biology as described by Kitano (Kitano 2002) which is a data-driven design of an experiment.

### **Immune modulation by dietary changes**

The results of the lateral data integration in **Chapter 5** showed that dietary interventions have an impact on the activity of components that belong to the immune system. Although changes in the immune system have been observed for certain dietary changes like and increase in fat intake, most dietary intervention studies do not analyse the effects of the dietary change on the immune system. In **Chapter 6** we aimed to find links between dietary changes and specific immune parameters functionally related to the immune system using experimental data from **Chapter 5** and information that is present in publicly available databases. With these combined resources, we developed a first-generation model representing potential links between dietary changes and the expression of relevant immunity parameters. This model predicts the functional parameters that could be influenced by certain changes in the diet while it also provides insights into mechanisms causing these changes. With literature validations, it was proved that some of the predicted connections between dietary changes and the functional immunity parameters already have been described in literature. This underscores the biological relevance of the computational output of the model. The model predicts that several changes in diets induce an unexpectedly high level of changes in the immune system.

To improve the predictability of the model and to allow the use of more detailed information on the dietary input side, this immunity parameter model needs to be validated with more data. Predictive and simulation models always need to go through a step of validation using data (Kohavi 1995; Sargent 2013). But the work in **Chapter 6** did not have this validation step due to a lack of proper data. This validation is usually done with data from similar conditions as the data used to build the model. For example, the connections between the dietary condition where the fibre Arabinoxylan is added to the diet and

---

23 functional immunity parameters can be validated by using another dataset with a similar dietary change. The new dataset is subject to the same analysis to discover connections to the functional immunity parameters, and the overlapping connections will have higher confidence as predictions. But this type of validation could not be performed since there were a limited number of datasets that fit the selection criteria as mentioned in **Chapter 6**. Sometimes a few samples of the same dataset are left out of building the model in order to be used for validation. This was also not possible with the data that we had because of the small sample sizes in the experiments. Besides data validation, the model can also be validated with specifically designed experiments based on the ‘wet-lab – dry-lab’ cycle. New experiments on animals could be designed to test the connections found in the Diet-Immunity model by introducing changes in the diet and measuring the internal phenotypes and the functional immunity parameters. The results of such experiments could prove or disprove the predicted connections between dietary changes and parameters of the immune system.

While building this model, we used the transcriptome internal phenotype to link it to functional immunity parameters. This was done by using data that result from Genome Wide Association Studies (GWAS) that use Single Nucleotide Polymorphisms (SNPs) to link certain traits to genes. By linking dietarily perturbed genes or gene-sets to these trait-associated SNPs and the location of the SNPs near genes, it was possible to link the diets to chosen traits. Due to the expected increase in knowledge on SNP variation and their relationships with trait characteristics it will be possible to further refine the model. Introducing a stricter threshold for SNP to gene associations from GWAS databases, could also increase the confidence of the connections between dietary changes and the functional immunity parameters.

The Diet – Immunity parameter model was built with both gene and pathway information from transcriptomic data. In the horizontal integration as described in **Chapter 5** we observed a plasticity in the response of the pathway to stimulations. This plasticity can be used to further refine the model by using the pathway response pattern as an indicator of the changes in the functional immunity parameter. Even though the model was limited to only account for changes in functional parameters associated to immunity, the method developed can be expanded to encompass a wider range of external phenotypes.

The model developed in **Chapter 6** was built to predict the effects of dietary changes on the immune system. An expanded and quantitatively validated version of this model may in future be used to find important modulators (genes, proteins or metabolites) of the immune system. When the diet – immunity parameter model can predict, with reasonable accuracy, the effects of dietary changes on the immune system, it would be possible to take the model a step further. With enough data on dietary compositional changes, internal phenotypes and functional immunity parameters, this model could, theoretically, be used to predict the organism's immune competence. This objective is not easy to reach since immune competence depends on a wide number of known and unknown factors, including individual genetic variation and early life programming.

### **Computational methods**

During most of the work presented in this thesis, I was able to develop tailor made data analysis pipelines through combinations of already existing and tested methods. This was enabled by the existence of strict requirements regarding software interoperability and documentation. All the pipelines were developed in R, an open-source scripting language. On one occasion, I had to modify the inner workings of an algorithm (CePa (Z. Gu and Wang 2013)) so that it would suit the data and the analysis requirements. The software tool presented in **Chapter 4** is a function that would retrieve full pathways, while ensuring inter-operability with the rest of the tools in the software package. All these pipelines were made possible because the code was accessible and it was possible to edit it to suit the specific research question at hand. This is due to open source initiatives that allow code to be publicly shared.

I also used several freely accessible databases for gathering both data and knowledge. Some of the databases I used are Reactome, Gene Expression Omnibus (Barrett et al. 2013), dbGaP (Tryka et al. 2014) and PubMed. Most of this information was retrieved automatically with packages from R that can interface directly with these databases. This access allows a user to make maximum use of all the information stored in databases of different types with a high-throughput data search. For example, in **Chapters 3** and **6** we used the R package *rentrez* (Winter 2016) that can gather information from the NCBI application programming interface and search through PubMed abstracts for co-occurrences

---

of specific search terms. With this method, it was possible to quickly ascertain the amount of literature on a subject and delve into specific publications as required, even if only the abstracts were publicly available.

An important aspect of systems biology is to understand the interactions between different parts of the system. In this thesis, this was achieved with vertical data integration. I integrated measurements of internal phenotypes to discover interactions between them. From a computational perspective, if the different data-types are being integrated, it is important that all of them undergo the same pre-processing steps. Only this will allow a proper comparison between the different measurements. The results of a differential analysis cannot be compared to that of a time-series analysis because the premise of the two methods is different. More than two types of data are not often integrated. However, with the approach described in **Chapter 3**, I showcase a flexible method that can be used on any type of data and is built upon existing methods integrating two data types at a time. Methods that can handle more than two non-specific datasets simultaneously, like MCI (multiple co-inertia analysis) (Meng et al. 2014) compare several datasets rather than discover possible interactions between them. Other methods that can indeed integrate more than 2 data types simultaneously are specific to certain types of data and are sometimes cumulative, i.e. two datasets are integrated and a third dataset is integrated to the results of the first integration.

The presented approaches for data integration most often rely on correlations. Correlation analysis is a technique often used in individual data analysis and in data integration and has often been used in this thesis for data integration. The results of correlation studies must be interpreted with caution as correlation and causation relationships are frequently confused. A major opposition to correlation analysis is that it might often lead to the identification of spurious associations due to indirect associations. The fact that one variable (e.g. gene) changes along with another variable (e.g. metabolite) does not necessarily mean that one controls the changes in the other. Sometimes there is an independent agent causing these changes in both the variables but is not captured due to the inability to measure it or due to time delays in the measurements.

The methods used in **Chapters 2** and **5** to find significant correlations between the analysed datasets were specifically devised to diminish the possible impact of these spurious associations. In both chapters, sparse Partial Least Squares (sPLS) was chosen to minimize the possible number of false positives in the data integration result. This method penalizes the inclusion of interactions, thereby effectively minimizing them. Disentangling causative effects following the approaches presented here could be done by using so-called time delayed correlations, in which the correlations between both data types is done by introducing a shift to account for the different time points at which the measurement was taken. However, these methods can only be applied when time series data with sufficient resolution is available, for example measurements made within hours of each other. Here the only chapter in which time data series data was used was **Chapter 2** but these time points are too far apart, with intervals of several weeks, to allow for this type of analysis.

In **Chapter 5** the validity of the sPLS correlation approach on multi-layered integration was tested using a permutation test. A permutation test enables precise measurements of the power of an algorithm to detect a signal. The test compares the data from the biological system (measurements of internal phenotypes) to permuted data that contains no information. This is a powerful approach that also lies at the core of the selected method for pathway enrichment analysis when gauging the significance of the pathway score. There I permuted the gene expression values in the pathway nodes in order to remove biological information from the pathway before computing the pathway score. This randomised score is computed several times and the likelihood of the occurrence of the original score is calculated. This value is used to gauge significance of the pathway score.

Data integration over different perturbations provided information on the functional plasticity of the gut. With the horizontal integration performed in **Chapter 5**, changes in gene expression of the intestinal epithelial cells were detected that contribute to the functional plasticity of the gut. However, the transcriptional profile is only one aspect of the functioning of the gut. The integration of other types of data, like microbial or protein composition, may contribute to a further understanding of the functional plasticity of this complex system.

---

## Requirements to improve results of integrative research approaches

During the computational research as described in this thesis, I have gathered knowledge on the current limitations in the performance of integrative research approaches. These limitations often boil down to lack of sufficient detail on the descriptions of experiments and most often they could be overcome by an improved annotation of information in databases.

### *1. Availability of experimental data*

Often animal experiments are not comparable because they have different measurements of biological parameters based on the research questions in that experiment. There are a few measurements that are almost always performed like weighing the animal and estimating food and water consumption. The data most often used in this thesis are derived from genome-wide transcription measurements using microarrays. The decrease in prices of the use of sequencing technologies and the ever-growing speed in sequence acquisition has provoked a sharp rise in datasets related to complete genome structures (Stephens et al. 2015), gene expression patterns and information on the composition, diversity genetic potential and activity of microbiomes. Mass spectrometry based techniques for proteomics and metabolomics are increasingly being used but the costs incurred by these measurements limit the studies that perform them. The decrease in prices of sequencing techniques and the increased interest in the role of the microbiota have increased availability of microbiota measurements.

Several repositories exist where datasets generated from experiments can be deposited such that they are freely available to the public. It is important that these repositories are stable to ensure that the data is available over a long period of time. In the case of sensitive data (e.g., patient information) there are several cyber security measures to only make parts of the data public. Some databases are specifically for one type of data, like Gene Expression Omnibus (GEO) and Array Express (Kolesnikov et al. 2015), which are for gene expression from either microarrays or RNA sequencing techniques. NCBI and EBI have several databases for different types of data, like protein sequences, but there are also repositories like Zenodo that can handle any type of data. These general purpose databases

are often used for metabolomics data since there are very few dedicated databases to store metabolomic data from experiments, e.g. Metabolights (Haug et al. 2013).

## 2. *Data description*

Repeatability of an analysis and reusability of shared data by an independent party requires good data description. There are well-known standards that need to be followed while reporting an experiment or depositing the data from an experiment in a database. Reports on animal experiments in publications are required to follow the Animal Researching: Reporting on In Vivo Experiments (ARRIVE) guidelines set by National Center for the Replacement Refinement and Reduction of Animals in Research (Kilkenny et al. 2010). The meta-data in databases is required to adhere to a general set of guidelines for minimal information that should be provided for each experiment the – Minimum Information for Biological and Biomedical Investigations (MIBBI) (C. F. Taylor et al. 2008). These guidelines detail the information to be included such as information on ethical permissions, size of experimental and control groups, precise description of experimental and surgical procedures and sample description, among others.

Ambiguity of the experimental descriptions can be reduced by using the meta-data description terms from an ontology (agreed upon terms used in a particular field). Using ontologies also facilitates using meta-data terms as search terms to increase sensitivity and specificity when searching for data from specific experimental conditions. This issue with meta-data was especially apparent in the work done in **Chapter 5** where several publicly available datasets were re-used for analysis from a different perspective. A few datasets were left out of the analysis due either to ambiguity or lack of detail in the metadata. The databases that store the experimental data should expand the minimal meta-data that is required before a data submission is accepted. Another practise that could be adopted by the databases is to use ontologies or thesauri like Medical SubHeadings (MeSH) for the keywords used in the description.

Both issues enumerated above related to experimental data availability and data description, greatly limit re-usability of existing data. There are two major reasons for re-using data, to verify the results of the original study or to use the same data for a different research question. Being able to verify results of a study is the idea behind publishing and

---

the cornerstone of research itself. Another frequent reason for re-using data is to perform an analysis similar to the integration performed in **Chapter 5** of this thesis.

Issues with data availability and meta-data can be addressed by making data Findable Accessible Interoperable and Reusable (FAIR) as detailed by Wilkinson et al (Wilkinson et al. 2016). Fulfilling the first two parts of FAIR, Findable and Accessible, requires that the data be publicly available so that a human user or machine can find it and use it with appropriate permissions. Carrying out an integrative or meta-analysis is often restricted to specific experimental conditions which can only be identified with distinctive search terms and extensive meta-data. Once the data has been obtained, most often, it needs to be pre-processed. For example, there could be an identifier (id) mapping step where the ids used in each experiment are mapped to each other. The Interoperable part of the FAIR principle pushes data harmonisation steps from the end user to the data generator. Besides this id mapping requirement, interoperable data must be ready not only for human understanding but also for machine access. Satisfying the first three conditions (F, A, I) helps with the last principle of making data Re-usable but also requires very detailed experimental information.

The responsibility for making data FAIR falls partly on researchers who must take the initiative to properly document their data. But the power to enforce the recommendations (FAIR, ARRIVE, MIBBI) for data descriptions resides partly with the journals that publish the results of the experiments and databases that host the data from the experiments. This drive for proper data stewardship is currently gaining momentum among large funding agencies which have the most power to enforce these guidelines.

### 3. *Annotation of data*

Another important issue when dealing with biological data is the annotation of biological entities with functions that they are observed to perform. In the transcriptomics data that were used in this thesis, there are several gene sequences on the microarray that lack functional annotation. Such un-annotated entities cannot be used for further analysis, for example, when converting the animal gene ids (mice or pigs) to human ids. This id conversion step to human ids is essential since there are several databases (or knowledgebases) that have most information on human data. The functional annotation of microbial rRNA se-

quences is limited by the lack of information on the roles of microbes in a habitat, e.g. the gut.

The research described in **Chapters 5** and **6** is based heavily on pathway information which was obtained from the Reactome database. Interactions between nodes in a pathway are assigned based on information in existing literature. This means that there could be several annotations for an interaction between nodes since these interactions are sometimes studied under various conditions. There is also the issue that there are several definitions of pathways depending on the database and there is no consensus on what constitutes a pathway. It would be easier for end users if there were mappings between the different pathway databases to connect pathways with similar information.

The datasets used in this thesis represent so-called internal phenotypes. For the model in **Chapter 6**, the transcriptional internal phenotype was used. There is, however, an ambiguity surrounding the terms external phenotype and trait, which are often used in the same context, but some researchers define a phenotype as a measurable trait (Beck et al. 2012). In several cases the traits or phenotypes in databases are linked to pathological states. There are databases that link genes to traits or external phenotypes which are specific to domains or organisms. The database used in **Chapter 6** is dbGaP which is specific for human data and links genes to traits (from MeSH terms) based on results of GWAS studies. The database Mouse Genome Informatics has plenty of information on linking genes to traits (from the Mammalian Phenotype Ontology (C. L. Smith and Eppig 2009)) based on mice studies. Traits of production animals are documented in the Animal Trait Ontology for Livestock (Golik et al. 2012) and has information on the species where the trait is present or absent. It is possible to map these ontologies to obtain maximum information from them. Such attempts can be found in the ontology repository BioPortal using the automatic matching tool LOOM. These automatic mappings definitely need manual curation especially since there is ambiguity around the words trait and phenotypes. In this thesis, the information in dbGaP was used since it contained traits that can be measured to gauge the functionality of the immune system. Also, the results of GWAS studies have been linked to the expression of the trait-associated genes, which lends credence to linking gene expression changes to GWAS results.

**Final remarks**

Most of the research as described in this thesis was performed on already existing data with available methods and extensions of existing methods. This research demonstrated that the re-use of existing data and their analysis from a systems wide perspective, provides unique insights into the functioning of the system. It indicates that hidden information is present in currently available databases and datasets, that can be extracted with methodologies that arise from the objective to understand the functioning of systems as a whole. Although we are far from a model that represents a “virtual gut” the research in this thesis demonstrates computational methods can be developed and applied to get improved insight into time-dependent dynamics of a gut systems, to obtain information on key links between internal phenotypic of different biological levels, to understand the functional plasticity of the gut mucosal tissue, and to develop a concept of a model that allows to make links between internal phenotypes and to immunity parameters functionally related to external immune-based phenotypes.

# Summary



The gastrointestinal tract is subject of much research for its role in an organism's health owing to its role as gatekeeper. A major function of the gut is to act as a barrier to challenges from the external environment. The tissue must keep out harmful substances like pathogens and toxins while absorbing nutrients that arise from the digestion of dietary components in the lumen. There is a large population of microbiota that plays an important role in the functioning of the gut, digestion and homeostasis of the immune system. All these sub-systems of the gastrointestinal tract contribute to the normal functioning of the gut. Due to its various functionalities, the gut is able to respond to different types of stimuli and bring the system back to homeostasis after perturbations. This ability is called the functional plasticity of the gut. **Chapter 1** provides more information on the morphology and functionality of the gut. The techniques for the measurements of various internal phenotypes (measurements of transcripts, proteins and metabolites, among others) are described. It describes the various methods traditionally used to study the gut and the novel methods used in this thesis to study the gut from a more holistic perspective. It also introduces the two types of data integration methods used in the thesis, vertical and horizontal integration. Vertical integration finds connections between different types of data and horizontal data integration combines several datasets of the same type of data. These data integration methods were used in the following five chapters to gain insights into the dynamics of a perturbed gut system over several time points, to find connections between different types of biological measurements (gene expressions, proteins, metabolites and microbes), to improve our understanding of the functional plasticity of the gut under various types of challenges and to link changes in the diet to changes in the local and systemic immune system.

The work in **Chapter 2** aimed to improve our understanding of the dynamics of the gut functionality that contributes to long lasting changes in the intestinal mucosal tissue. The data were obtained from a pig experiment where an antibiotic was administered at a very early age, along with stress that would be common in a farm environment. The small intestines of these animals were then sampled at three time points over the next six months to obtain the measurements of two internal phenotypes; gene expression of the intestinal mucosal tissue and microbial community data from the luminal content at the same location in the gut. The data analysis in this chapter involves both horizontal data integration (across time) and vertical data integration (integration of gene expression with microbio-

ta). The integration across time was performed along two dimensions by taking into account the time dynamics and the treatment effects simultaneously. This was done using a regression-based method on both the gene expression and microbiota data. Subsequently the results of the time dynamics and treatment effects were integrated with each other. This vertical integration was performed based on correlations to discover potential cross-talk between the host and microbiota that are affected both by treatment and time. The results show that there are several genes with consistent long term changes in their gene expression. These genes could be important regulators of gut functionalities. These regulators seem to cause the persistent changes in the tissue gene expression. The vertical data integration revealed the important roles of a few bacterial groups in maintaining the long-term changes in the gene expression. These results provide insights into the dynamic interactions between the host and its microbes.

Interactions between internal phenotypes were further studied in **Chapter 3**, where five measurements of internal phenotypes were obtained from dietary interventions in mice. The mice were given diets with different protein sources and several different internal phenotype measurements were made in the intestine, serum and urine. From the same location of the small intestine, scrapings were used to obtain gene expression profiles of the mucosal tissue and the content was used to acquire data on microbial community structures. Cytokine levels in the serum were measured along with metabolomic profiles. Metabolomic data was also obtained from urine. Pairwise correlation analysis was performed ten times to integrate each dataset to the other four and these results were combined to obtain an integrated network with elements from all the five datasets. The correlations from the pairwise integration were verified using a permutation approach that checks whether the observed correlations are based on biological information in the data. The connections between internal phenotypes were validated with a literature search for co-occurrence of the connected components. The results of this literature search revealed that several connections uncovered in this Chapter have been found previously in other studies. The integrated network was found to contain several components that were linked to all other four types of data. These components were denoted “connectivity hubs”. The presence of such hubs indicates that there are some components of the internal phenotypes that are more involved in the interactions between internal phenotypes than others. The

hypotheses generated in this study can be used to design wet-lab experiments to validate uncovered connections between the internal phenotypes.

Pathways are groups of functionally related genes with information on the interactions between the genes that fulfil a biological function. When integrating several transcriptomics datasets together (horizontal integration) the analysis is more comparable between experiments and platforms when analysed with pathway expression rather than gene expression. Pathway data from several databases is available in a Resource Description Framework format called BioPAX (Biological Pathway eXchange) which is not easy to work with for transcriptomic analysis. In order to use this information in the R environment, an open source programming language and software environment for statistical computing and graphics, the package *rBiopaxParser* converts the BioPAX object into one that can be utilised in R. One important function in this package is *pathway2RegulatoryGraph* which converts the information in a pathway to a graph of the regulatory components in the pathway. **Chapter 4** is a description of a new function *pathway2Graph* that can use all information in the pathway to build a graph of the entire pathway. This function has been incorporated into the package *rBiopaxParser* distribution.

The gut is exposed to a variety of challenges and its response to them is of utmost importance since the gut is a critical barrier to the external environment. The functional plasticity displayed by the gut allows an appropriate response of the gut towards a variety of challenges, perturbations and/or stimulations. In order to study the mechanisms behind functional plasticity, in **Chapter 5** a horizontal integration of intestinal mucosal gene expression profiles was performed after exposure to three categories of challenges: dietary, drug-based or immune challenges. These transcriptomics datasets, derived from 14 independent experiments in mice, were obtained from an online database in order to study the gut under different challenging conditions. The gene expression analysis was performed on a pathway level and combined with information present in the Reactome database obtained with the *pathway2Graph* function developed in **Chapter 4**. A high-level data integration was performed where the datasets are analysed separately and only the results are integrated. The results of the high-level data integration demonstrated that there is a considerable overlap in the significantly regulated pathways induced by the three categories of challenges. We concluded that these common processes contribute to the functional

plasticity of the gut. One of the processes common to the three challenge categories is ‘Regulation of Complement Cascade’, which is an immune process that is regulated in several of the dietary challenge experiments. Since the regulation of this process is not expected in the gut tissue, a validation experiment was designed using mice intestinal organoids to understand the conditions that lead to the regulation of this pathway. The results suggest that an inflammatory condition could be the reason behind the regulation of this pathway. Thus, by performing targeted experiments, based on the results of computational analyses, helps to better understand the mechanisms behind the functional plasticity of the gut. The high-level data integration on different types of challenges, as described in **Chapter 5**, showed that the dietary changes induced changes in immune-related processes. Therefore, in **Chapter 6**, a model was built to find connections between dietary changes and functional immunity parameters. The data from 9 dietary challenge studies was analysed for differentially expressed genes and differentially regulated pathways. In order to obtain links between genes and functional immunity parameters, information from Genome Wide Association Studies (GWAS) were used that are deposited in the NCBI database Genotype and Phenotype. GWAS links traits to Single Nucleotide Polymorphisms (SNP) that are subsequently linked to genes based on the proximity of SNP’s to genes. We retrieved this trait to gene associations for selected functional immunity parameters that can be measured and linked to the immune status of organisms. We formed associations between dietary changes and functional immunity parameters through the genes and pathways. These connections were validated through a literature search which established that there have been studies linking some dietary changes to the changes in the functional immunity parameters predicted by the model.

**Chapter 7** is a general discussion of all the five research chapters. The ways in which the research objectives were achieved are enumerated along with details of the methods used. Work that can be performed on the current results to improve the understand of the gut is also detailed. I discuss current bottlenecks and possible ways to increase our understanding of gut function. It is concluded that the re-use of existing data and their analysis from a systems wide perspective, provides unique insights into the functioning of biological system. It is also concluded that hidden information, which is present in currently available databases and datasets, can be extracted with methodologies that arise from the objective to understand the functioning of systems as a whole. Finally, it is concluded that the re-

search in this thesis demonstrates that computational methods can be developed and applied to get improved insights into the time-dependent dynamics of a gut systems, to obtain information on key links between internal phenotypes of different biological levels, to better understand the functional plasticity of the gut mucosal tissue, and to develop models that allows to make links between internal phenotypes and to immunity parameters functionally related to external immune-based phenotypes.



## References



## References

---

- Abell, Guy C J, Caroline M Cooke, Corinna N Bennett, Michael A Conlon, and Alexandra L McOrist. 2008. "Phylotypes Related to *Ruminococcus Bromii* Are Abundant in the Large Bowel of Humans and Increase in Response to a Diet High in Resistant Starch." *FEMS Microbiology Ecology* 66 (3). The Oxford University Press: 505–15.
- Adourian, Aram, Ezra Jennings, Raji Balasubramanian, Wade M Hines, Doris Damian, Thomas N Plasterer, Clary B Clish, et al. 2008. "Correlation Network Analysis for Data Integration and Biomarker Selection." *Molecular BioSystems* 4 (3). The Royal Society of Chemistry: 249–59. doi:10.1039/B708489G.
- Ahmad, R, R Chaturvedi, D Olivares-Villagomez, T Habib, M Asim, P Shivesh, D B Polk, et al. 2014. "Targeted Colonic Claudin-2 Expression Renders Resistance to Epithelial Injury, Induces Immune Suppression, and Protects from Colitis." *Mucosal Immunol* 7 (6). Society for Mucosal Immunology: 1340–53. <http://dx.doi.org/10.1038/mi.2014.21>.
- Ahmed, Rafi, and David Gray. 1996. "Immunological Memory and Protective Immunity: Understanding Their Relation." *Science* 272 (5258). American Association for the Advancement of Science: 54–60.
- Akavia, Uri David, Oren Litvin, Jessica Kim, Felix Sanchez-Garcia, Dylan Kotliar, Helen C. Causton, Panisa Pochanard, Eyal Mozes, Levi A. Garraway, and Dana Pe&apos;Er. 2010. "An Integrated Approach to Uncover Drivers of Cancer." *Cell* 143 (6): 1005–17. doi:10.1016/j.cell.2010.11.013.
- Akira, Shizuo, and Kiyoshi Takeda. 2004. "Toll-like Receptor Signalling." *Nat Rev Immunol* 4 (7): 499–511. <http://dx.doi.org/10.1038/nri1391>.
- Albert, Reka, Hawoong Jeong, and Albert-Laszlo Barabasi. 2000. "Error and Attack Tolerance of Complex Networks." *Nature* 406 (6794): 378–82. <http://dx.doi.org/10.1038/35019019>.
- Alex, Philip, Nicholas C Zachos, Thuan Nguyen, Liberty Gonzales, Tian E Chen, Laurie S Conklin, Michael Centola, and Xuhang Li. 2009. "Distinct Cytokine Patterns Identified from Multiplex Profiles of Murine DSS and TNBS-Induced Colitis." *Inflammatory Bowel Diseases* 15 (3): 341–52. doi:10.1002/ibd.20753.
- Alexa, Adrian, and Jörg Rahnenführer. 2009. "Gene Set Enrichment Analysis with topGO." Available.
- Antezana, Erick, Martin Kuiper, and Vladimir Mironov. 2009. "Biological Knowledge Management: The Emerging Role of the Semantic Web Technologies." *Briefings in Bioinformatics* 10 (4). England: 392–407. doi:10.1093/bib/bbp024.
- Antonopoulos, Dionysios A, Susan M Huse, Hilary G Morrison, Thomas M Schmidt, Mitchell L Sogin, and Vincent B Young. 2009. "Reproducible Community Dynamics of the Gastrointestinal Microbiota Following Antibiotic Perturbation." *Infection and Immunity* 77 (6): 2367–75. doi:10.1128/IAI.01520-08.

- 
- Arrieta, Marie-Claire, and Barton Brett Finlay. 2012. “The Commensal Microbiota Drives Immune Homeostasis.” *Frontiers in Immunology* 3. Frontiers Media SA.
- Arrieta, Marie-Claire, Leah T Stiemsma, Nelly Amenyogbe, Eric M Brown, and Brett Finlay. 2014. “The Intestinal Microbiome in Early Life: Health and Disease.” *Frontiers in Immunology* 5. Frontiers Media SA.
- Artis, David. 2008. “Epithelial-Cell Recognition of Commensal Bacteria and Maintenance of Immune Homeostasis in the Gut.” *Nat Rev Immunol* 8 (6). Nature Publishing Group: 411–20. <http://dx.doi.org/10.1038/nri2316>.
- Ashburner, Michael, Catherine A Ball, Judith A Blake, David Botstein, Heather Butler, J Michael Cherry, Allan P Davis, et al. 2000. “Gene Ontology: Tool for the Unification of Biology.” *Nat Genet* 25 (1). Nature America Inc.: 25–29. <http://dx.doi.org/10.1038/75556>.
- Assenov, Yassen, Fidel Ramírez, Sven-Eric Schelhorn, Thomas Lengauer, and Mario Albrecht. 2008. “Computing Topological Parameters of Biological Networks.” *Bioinformatics* 24 (2): 282–84. <http://bioinformatics.oxfordjournals.org/content/24/2/282.abstract>.
- Avalle, Lidia, Sara Pensa, Gabriella Regis, Francesco Novelli, and Valeria Poli. 2012. “STAT1 and STAT3 in Tumorigenesis: A Matter of Balance.” *JAK-STAT*. doi:10.4161/jkst.20045.
- Barabasi, Albert-Laszlo, and Zoltan N Oltvai. 2004. “Network Biology: Understanding the Cell’s Functional Organization.” *Nat Rev Genet* 5 (2): 101–13. <http://dx.doi.org/10.1038/nrg1272>.
- Barrett, Tanya, Stephen E Wilhite, Pierre Ledoux, Carlos Evangelista, Irene F Kim, Maxim Tomashevsky, Kimberly A Marshall, et al. 2013. “NCBI GEO: Archive for Functional Genomics Data Sets—update.” *Nucleic Acids Research* 41 (D1): D991–95. doi:10.1093/nar/gks1193.
- Bayerlová, Michaela, Klaus Jung, Frank Kramer, Florian Klemm, Annalen Bleckmann, and Tim Beißbarth. 2015. “Comparative Study on Gene Set and Pathway Topology-Based Enrichment Methods.” *BMC Bioinformatics* 16 (1): 334. doi:10.1186/s12859-015-0751-5.
- Beck, Tim, Robert C Free, Gudmundur A Thorisson, and Anthony J Brookes. 2012. “Semantically Enabling a Genome-Wide Association Study Database.” *Journal of Biomedical Semantics* 3 (1). BioMed Central: 9. doi:10.1186/2041-1480-3-9.
- Belkaid, Yasmine, and Timothy W Hand. 2014. “Role of the Microbiota in Immunity and Inflammation.” *Cell* 157 (1): 121–41. doi:10.1016/j.cell.2014.03.011.
- Benchaoui, H A, M Nowakowski, J Sherington, T G Rowan, and S J Sunderland. 2004. “Pharmacokinetics and Lung Tissue Concentrations of Tulathromycin in Swine.” *Journal of Veterinary Pharmacology and Therapeutics* 27 (4). Blackwell Science Ltd: 203–10. doi:10.1111/j.1365-2885.2004.00586.x.
-

- Benis, Nirupama, Dirkjan Schokker, Frank Kramer, Mari A. Smits, Maria Suarez-Diez, Nirupama Benis, Dirkjan Schokker, Frank Kramer, Mari A. Smits, and Maria Suarez-Diez. 2016. "Building Pathway Graphs from BioPAX Data in R." *F1000Research* 5 (December): 2414. doi:10.12688/f1000research.9582.2.
- Benis, Nirupama, Dirkjan Schokker, Maria Suarez-Diez, Vitor A P Martins Dos Santos, Hauke Smidt, and Mari A Smits. 2015. "Network Analysis of Temporal Functionalities of the Gut Induced by Perturbations in New-Born Piglets." *BMC Genomics* 16 (1). BioMed Central Ltd: 556. doi:10.1186/s12864-015-1733-8.
- Benjamini, Yoav, and Yosef Hochberg. 1995. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society. Series B (Methodological)* 57 (1). [Royal Statistical Society, Wiley]: 289–300. <http://www.jstor.org/stable/2346101>.
- Berthoud, Hans-Rudolf. 2008. "The Vagus Nerve, Food Intake and Obesity." *Regulatory Peptides* 149 (1–3): 15–25. doi:10.1016/j.regpep.2007.08.024.
- Bhatia, Vikram, and Tandon. 2005. "Stress and the Gastrointestinal Tract." *Journal of Gastroenterology and Hepatology* 20 (3). Blackwell Science Pty: 332–39. doi:10.1111/j.1440-1746.2004.03508.x.
- Bischoff, Stephan C. 2011. "'Gut Health': a New Objective in Medicine?" *BMC Medicine* 9 (1): 24. doi:10.1186/1741-7015-9-24.
- Bischoff, Stephan C, Giovanni Barbara, Wim Buurman, Theo Ockhuizen, Jorg-Dieter Schulzke, Matteo Serino, Herbert Tilg, Alastair Watson, and Jerry M Wells. 2014. "Intestinal Permeability--a New Target for Disease Prevention and Therapy." *BMC Gastroenterology* 14 (November). England: 189. doi:10.1186/s12876-014-0189-7.
- Biswas, Amlan, Jeanette Wilmanski, Huamei Forsman, Tomas Hrnrcir, Liming Hao, Helena Tlaskalova-Hogenova, and Koichi S. Kobayashi. 2011. "Negative Regulation of Toll-like Receptor Signaling Plays an Essential Role in Homeostasis of the Intestine." *European Journal of Immunology* 41 (1): 182–94. doi:10.1002/eji.201040479.
- Blikslager, A. T., A. J. Moeser, J. L. Gookin, S. L. Jones, and J. Odle. 2007. "Restoration of Barrier Function in Injured Intestinal Mucosa." *Physiological Reviews* 87 (2): 545–64. doi:10.1152/physrev.00012.2006.
- Borate, Bhavesh R, Elissa J Chesler, Michael A Langston, Arnold M Saxton, and Brynn H Voy. 2009. "Comparison of Threshold Selection Methods for Microarray Gene Co-Expression Matrices." *BMC Research Notes* 2 (1). BioMed Central Ltd: 240.
- Braun, Thorsten, John R Challis, John. P Newnham, and Deborah M Sloboda. 2013. "Early-Life Glucocorticoid Exposure: The Hypothalamic-Pituitary-Adrenal Axis, Placental Function, and

- 
- Long-Term Disease Risk.” *Endocrine Reviews* 34 (6). The Endocrine Society: 885–916. doi:10.1210/er.2013-1012.
- Brazma, Alvis, Helen Parkinson, Ugis Sarkans, Mohammadreza Shojatalab, Jaak Vilo, Niran Abeygunawardena, Ele Holloway, Misha Kapushesky, Patrick Kemmeren, and Gonzalo Garcia Lara. 2003. “ArrayExpress—a Public Repository for Microarray Gene Expression Data at the EBI.” *Nucleic Acids Research* 31 (1). Oxford Univ Press: 68–71.
- Brown, Lisa, Bernard Rosner, Walter W Willett, and Frank M Sacks. 1999. “Cholesterol-Lowering Effects of Dietary Fiber: A Meta-Analysis.” *Am J Clin Nutr* 69 (1): 30–42. <http://ajcn.nutrition.org/cgi/content/long/69/1/30>.
- Bruggeman, Frank J, Jorrit J Hornberg, Fred C Boogerd, and Hans V Westerhoff. 2007. “Introduction to Systems Biology.” *EXS* 97. Switzerland: 1–19.
- Bünger, Meike, Heleen M van den Bosch, Jolanda van der Meijde, Sander Kersten, Guido J E J Hooiveld, and Michael Müller. 2007. “Genome-Wide Analysis of PPAR $\alpha$  Activation in Murine Small Intestine.” *Physiological Genomics* 30 (2): 192 LP-204. <http://physiolgenomics.physiology.org/content/30/2/192.abstract>.
- Burger-van Paassen, Nanda, Audrey Vincent, Patrycja J. Puiman, Maria van der Sluis, Janneke Bouma, Günther Boehm, Johannes B. van Goudoever, Isabelle van Seuningen, and Ingrid B. Renes. 2009. “The Regulation of Intestinal Mucin MUC2 Expression by Short-Chain Fatty Acids: Implications for Epithelial Protection.” *Biochemical Journal* 420 (2): 211–19. doi:10.1042/BJ20082222.
- Camacho, Diogo, Alberto De La Fuente, and Pedro Mendes. 2005. “The Origin of Correlations in Metabolomics Data.” *Metabolomics* 1 (1). Springer: 53–63.
- Cani, Patrice D, Rodrigo Bibiloni, Claude Knauf, Aurélie Waget, Audrey M Neyrinck, Nathalie M Delzenne, and Rémy Burcelin. 2008. “Changes in Gut Microbiota Control Metabolic Endotoxemia-Induced Inflammation in High-Fat Diet-Induced Obesity and Diabetes in Mice.” *Diabetes* 57 (6): 1470–81. <http://diabetes.diabetesjournals.org/content/57/6/1470.abstract>.
- Carvalho, F A, I Nalbantoglu, J D Aitken, R Uchiyama, Y Su, G H Doho, M Vijay-Kumar, and A T Gewirtz. 2012. “Cytosolic Flagellin Receptor NLRC4 Protects Mice against Mucosal and Systemic Challenges.” *Mucosal Immunol* 5 (3). Society for Mucosal Immunology: 288–98. <http://dx.doi.org/10.1038/mi.2012.8>.
- Castro-Sánchez, Patricia, and José M Martín-Villa. 2013. “Gut Immune System and Oral Tolerance.” *British Journal of Nutrition* 109 (S2). Cambridge, UK: Cambridge University Press: S3–11. doi:10.1017/S0007114512005223.
- Cera, K R, D C Mahan, R F Cross, G A Reinhart, and R E Whitmoyer. 1988. “Effect of Age, Weaning and Postweaning Diet on Small Intestinal Growth and Jejunal Morphology in Young
-

## References

---

- Swine.” *Journal of Animal Science* 66 (2). The American Society of Animal Science: 574–84.
- Chassaing, Benoit, and Andrew T Gewirtz. 2014. “Gut Microbiota, Low-Grade Inflammation, and Metabolic Syndrome.” *Toxicologic Pathology* 42 (1). United States: 49–53. doi:10.1177/0192623313508481.
- Cho, Clara E, and Mikael Norman. 2013. “Cesarean Section and Development of the Immune System in the Offspring.” *American Journal of Obstetrics and Gynecology* 208 (4): 249–54. doi:http://dx.doi.org/10.1016/j.ajog.2012.08.009.
- Cho, Ilseung, Shingo Yamanishi, Laura Cox, Barbara A Methe, Jiri Zavadil, Kelvin Li, Zhan Gao, et al. 2012. “Antibiotics in Early Life Alter the Murine Colonic Microbiome and Adiposity.” *Nature* 488 (7413). Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.: 621–26. http://dx.doi.org/10.1038/nature11400.
- Cicha, Iwona, and Katharina Urschel. 2015. “TNF- $\alpha$  in the Cardiovascular System: From Physiology to Therapy.” *International Journal of Interferon, Cytokine and Mediator Research* 7 (July). Dove Medical Press Ltd: 9. doi:10.2147/IJICMR.S64894.
- Civelek, Mete, and Aldons J Lusk. 2014. “Systems Genetics Approaches to Understand Complex Traits.” *Nature Reviews. Genetics* 15 (1): 34–48. doi:10.1038/nrg3575.
- Collins, Stephen M., and Premysl Bercik. 2009. “The Relationship Between Intestinal Microbiota and the Central Nervous System in Normal Gastrointestinal Function and Disease.” *Gastroenterology* 136: 2003–14. doi:10.1053/j.gastro.2009.01.075.
- Conesa, Ana, María José Nueda, Alberto Ferrer, and Manuel Talón. 2006. “maSigPro: A Method to Identify Significantly Differential Expression Profiles in Time-Course Microarray Experiments.” *Bioinformatics* 22 (9). Oxford Univ Press: 1096–1102.
- Corring, T. 1980. “The Adaptation of Digestive Enzymes to the Diet: Its Physiological Significance.” *Reprod. Nutr. Dévelop.* 20 (4B): 1217–35. http://dx.doi.org/10.1051/rnd:19800713.
- Couto, Narciso, Jennifer Wood, and Jill Barber. 2016. “The Role of Glutathione Reductase and Related Enzymes on Cellular Redox Homeostasis Network.” *Free Radical Biology and Medicine* 95 (February). Elsevier Inc.: 27–42. doi:10.1016/j.freeradbiomed.2016.02.028.
- Cox, Laura M., Shingo Yamanishi, Jiho Sohn, Alexander V. Alekseyenko, Jacqueline M. Leung, Ilseung Cho, Sunghoon G. Kim, et al. 2015. “Altering the Intestinal Microbiota during a Critical Developmental Window Has Lasting Metabolic Consequences.” *Cell* 158 (4). Elsevier: 705–21. doi:10.1016/j.cell.2014.05.052.
- Creamer, B., R. G. Shorter, and J. Bamforth. 1961. “The Turnover and Shedding of Epithelial Cells: Part I The Turnover in the Gastro-Intestinal Tract.” *Gut* 2 (2): 110–16. doi:10.1136/gut.2.2.110.

- 
- Croft, David. 2013. "Building Models Using Reactome Pathways as Templates." In *In Silico Systems Biology SE - 14*, edited by Maria Victoria Schneider, 1021:273–83. Methods in Molecular Biology. Humana Press. doi:10.1007/978-1-62703-450-0\_14.
- Croft, David, Antonio Fabregat Mundo, Robin Haw, Marija Milacic, Joel Weiser, Guanming Wu, Michael Caudy, et al. 2014. "The Reactome Pathway Knowledgebase." *Nucleic Acids Research* 42 (Database issue). England: D472-7. doi:10.1093/nar/gkt1102.
- Crosas, M. 2011. "The Dataverse Network: An Open-Source Application for Sharing, Discovering and Preserving Data." *D-Lib Magazine* 17 (1/2). doi:doi: 10.1045/january2011-crosas.
- Csardi, Gabor, and Tamas Nepusz. 2006. "The Igraph Software Package for Complex Network Research." *InterJournal, Complex Systems* 1695 (5): 1–9.
- Cummings, J H, J M Antoine, F Azpiroz, R Bourdet-Sicard, P Brandtzaeg, P C Calder, G R Gibson, et al. 2004. "PASSCLAIM: Gut Health and Immunity." *Eur J Nutr* 43.
- Cummings, John H, Jean-Michel Antoine, Fernando Azpiroz, Raphaelle Bourdet-Sicard, Per Brandtzaeg, Philip C Calder, Glenn R Gibson, Francisco Guarner, Erika Isolauri, and Daphne Pannemans. 2004. "PASSCLAIM1—Gut Health and Immunity." *European Journal of Nutrition* 43 (2). Springer: ii118-ii173.
- Das, Siddhartha, David S Reiner, John Zenian, Daniel L Hogan, Michael A Koss, Chi-Sun Wang, and Frances D Gillin. 1988. "Killing of Giardia Lamblia Trophozoites by Human Intestinal Fluid in Vitro." *Journal of Infectious Diseases* 157 (6): 1257–60. doi:10.1093/infdis/157.6.1257.
- Davis, Sean, and Paul S Meltzer. 2007. "GEOquery: A Bridge between the Gene Expression Omnibus (GEO) and BioConductor." *Bioinformatics (Oxford, England)* 23 (14). England: 1846–47. doi:10.1093/bioinformatics/btm254.
- De Bie, Tijn, Léon Charles Tranchevent, Liesbeth M M van Oeffelen, and Yves Moreau. 2007. "Kernel-Based Data Fusion for Gene Prioritization." In *Bioinformatics*. Vol. 23. doi:10.1093/bioinformatics/btm187.
- de Wit, Nicole J W, Mark V Boekschoten, Eva-Maria Bachmair, Guido J E J Hooiveld, Philip J de Groot, Isabel Rubio-Aliaga, Hannelore Daniel, and Michael Müller. 2011. "Dose-Dependent Effects of Dietary Fat on Development of Obesity in Relation to Intestinal Differential Gene Expression in C57BL/6J Mice." *PLoS ONE* 6 (4). Public Library of Science: e19145. http://dx.doi.org/10.1371/journal.pone.0019145.
- Dejean, Sebastien, Ignacio Gonzalez, Kim-Anh Lê Cao, and Pierre Monget. 2011. "mixOmics: Omics Data Integration Project." *R Package Version*, 2–9.
- Dekkers, Johanna F, Caroline L Wiegerinck, Hugo R de Jonge, Inez Bronsveld, Hettie M Janssens,
-

- Karin M de Winter-de Groot, Arianne M Brandsma, et al. 2013. "A Functional CFTR Assay Using Primary Cystic Fibrosis Intestinal Organoids." *Nat Med* 19 (7): 939–45. <http://dx.doi.org/10.1038/nm.3201>.
- Demir, Emek, Michael P Cary, Suzanne Paley, Ken Fukuda, Christian Lemer, Imre Vastrik, Guanming Wu, Peter D'Eustachio, Carl Schaefer, and Joanne Luciano. 2010. "The BioPAX Community Standard for Pathway Data Sharing." *Nature Biotechnology* 28 (9). Nature Publishing Group: 935–42.
- den Besten, Gijs, Karen van Eunen, Albert K Groen, Koen Venema, Dirk-Jan Reijngoud, and Barbara M Bakker. 2013. "The Role of Short-Chain Fatty Acids in the Interplay between Diet, Gut Microbiota, and Host Energy Metabolism." *Journal of Lipid Research* 54 (9): 2325–40. doi:10.1194/jlr.R036012.
- DeSantis, T Z, P Hugenholtz, N Larsen, M Rojas, E L Brodie, K Keller, T Huber, D Dalevi, P Hu, and G L Andersen. 2006. "Greengenes, a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible with ARB." *Applied and Environmental Microbiology* 72 (7): 5069–72. doi:10.1128/AEM.03006-05.
- Deschodt-Lanckman, M., P. Robberecht, J. Camus, and J. Christophe. 1971. "Short-Term Adaptation of Pancreatic Hydrolases to Nutritional and Physiological Stimuli in Adult Rats." *Biochimie* 53 (6). Elsevier Masson: 789–96. doi:10.1016/S0300-9084(71)80120-7.
- Desmarchelier, Charles, Christoph Dahlhoff, Sylvia Keller, Manuela Sailer, Gerhard Jahreis, and Hannelore Daniel. 2012. "C57Bl/6 N Mice on a Western Diet Display Reduced Intestinal and Hepatic Cholesterol Levels despite a Plasma Hypercholesterolemia." *BMC Genomics* 13 (March). BioMed Central: 84. doi:10.1186/1471-2164-13-84.
- Diaz-Vivancos, Pedro, Ambra de Simone, Guy Kiddle, and Christine H Foyer. 2015. "Glutathione - Linking Cell Proliferation to Oxidative Stress." *Free Radical Biology & Medicine* 89 (December). Elsevier Inc.: 1154–64. doi:10.1016/j.freeradbiomed.2015.09.023.
- Ding, Shengli, Michael M Chi, Brooks P Scull, Rachael Rigby, Nicole M J Schwerbrock, Scott Magness, Christian Jobin, and Pauline K Lund. 2010. "High-Fat Diet: Bacteria Interactions Promote Intestinal Inflammation Which Precedes and Correlates with Obesity and Insulin Resistance in Mouse." *PLOS ONE* 5 (8). Public Library of Science: e12191. <http://dx.doi.org/10.1371/journal.pone.0012191>.
- Domogatskaya, Anna, Sergey Rodin, and Karl Tryggvason. 2012. "Functional Diversity of Laminins." *Annual Review of Cell and Developmental Biology* 28 (1). Annual Reviews: 523–53. doi:10.1146/annurev-cellbio-101011-155750.
- Donaldson, Gregory P, S Melanie Lee, and Sarkis K Mazmanian. 2016. "Gut Biogeography of the Bacterial Microbiota." *Nat Rev Micro* 14 (1). Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.: 20–32. <http://dx.doi.org/10.1038/nrmicro3552>.

- Dopazo, Joaquin. 2006. "Functional Interpretation of Microarray Experiments." *Omics: A Journal of Integrative Biology* 10 (3). Mary Ann Liebert, Inc. 2 Madison Avenue Larchmont, NY 10538 USA: 398–410.
- Edgar, Ron, Michael Domrachev, and Alex E Lash. 2002. "Gene Expression Omnibus: NCBI Gene Expression and Hybridization Array Data Repository." *Nucleic Acids Research* 30 (1): 207–10. doi:10.1093/nar/30.1.207.
- Eguíluz, Victor M, Dante R Chialvo, Guillermo A Cecchi, Marwan Baliki, and A Vania Apkarian. 2005. "Scale-Free Brain Functional Networks." *Physical Review Letters* 94 (1). American Physical Society: 18102. <http://link.aps.org/doi/10.1103/PhysRevLett.94.018102>.
- Ekman, Diana, Sara Light, Åsa K Björklund, and Arne Elofsson. 2006. "What Properties Characterize the Hub Proteins of the Protein-Protein Interaction Network of *Saccharomyces Cerevisiae*?" *Genome Biology* 7 (6). BioMed Central Ltd: R45.
- Elahi, Shokrollah, James M Ertelt, Jeremy M Kinder, Tony T Jiang, Xuzhe Zhang, Lijun Xin, Vandana Chaturvedi, et al. 2013. "Immunosuppressive CD71+ Erythroid Cells Compromise Neonatal Host Defence against Infection." *Nature* 504 (7478). Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.: 158–62. <http://dx.doi.org/10.1038/nature12675>.
- "EMEA: Tulathromycin." 2004.
- Emmert-Streib, Frank, and Galina V Glazko. 2011. "Pathway Analysis of Expression Data: Deciphering Functional Building Blocks of Complex Diseases." *PLoS Comput Biol* 7 (5). Public Library of Science: e1002053. <http://dx.doi.org/10.1371/journal.pcbi.1002053>.
- Evans, Christian C, Kathy J LePard, Jeff W Kwak, Mary C Stancukas, Samantha Laskowski, Joseph Dougherty, Laura Moulton, et al. 2014. "Exercise Prevents Weight Gain and Alters the Gut Microbiota in a Mouse Model of High Fat Diet-Induced Obesity." *PLoS ONE* 9 (3). Public Library of Science: e92193. <http://dx.doi.org/10.1371/journal.pone.0092193>.
- Fabregat, Antonio, Konstantinos Sidiropoulos, Phani Garapati, Marc Gillespie, Kerstin Hausmann, Robin Haw, Bijay Jassal, et al. 2016. "The Reactome Pathway Knowledgebase." *Nucleic Acids Research* 44 (D1). England: D481-7. doi:10.1093/nar/gkv1351.
- Fagerberg, Linn, Björn M Hallström, Per Oksvold, Caroline Kampf, Dijana Djureinovic, Jacob Odeberg, Masato Habuka, et al. 2014. "Analysis of the Human Tissue-Specific Expression by Genome-Wide Integration of Transcriptomics and Antibody-Based Proteomics." *Molecular & Cellular Proteomics* 13 (2): 397–406. doi:10.1074/mcp.M113.035600.
- Fatehullah, Aliya, Si Hui Tan, and Nick Barker. 2016. "Organoids as an in Vitro Model of Human Development and Disease." *Nature Cell Biology* 18 (3). Nature Research: 246–54. doi:10.1038/ncb3312.

## References

---

- Foliaki, Sunia, Neil Pearce, Bengt Björkstén, Javier Mallol, Stephen Montefort, and Erika von Mutius. 2009. "Antibiotic Use in Infancy and Symptoms of Asthma, Rhinoconjunctivitis, and Eczema in Children 6 and 7 Years Old: International Study of Asthma and Allergies in Childhood Phase III." *Journal of Allergy and Clinical Immunology* 124 (5): 982–89. doi:<http://dx.doi.org/10.1016/j.jaci.2009.08.017>.
- Fontanesi, Luca. 2016. "Metabolomics and Livestock Genomics: Insights into a Phenotyping Frontier and Its Applications in Animal Breeding." *Animal Frontiers* 6. Madison, WI: American Society of Animal Science. doi:[10.2527/af.2016-0011](https://doi.org/10.2527/af.2016-0011).
- Frantz, Christian, Kathleen M Stewart, and Valerie M Weaver. 2010. "The Extracellular Matrix at a Glance." *Journal of Cell Science* 123 (Pt 24). Company of Biologists: 4195–4200. doi:[10.1242/jcs.023820](https://doi.org/10.1242/jcs.023820).
- Fryknäs, Mårten, Sumeer Dhar, Fredrik Öberg, Linda Rickardson, Maria Rydåker, Hanna Göransson, Mats Gustafsson, et al. 2007. "STAT1 Signaling Is Associated with Acquired Crossresistance to Doxorubicin and Radiation in Myeloma Cell Lines." *International Journal of Cancer* 120 (1). Wiley Subscription Services, Inc., A Wiley Company: 189–95. doi:[10.1002/ijc.22291](https://doi.org/10.1002/ijc.22291).
- Gentleman, Robert, Vincent J Carey, Wolfgang Huber, Rafael A Irizarry, and Sandrine Dudoit. 2005. *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*. Vol. 746718470. Springer.
- Gentry, J G, J J McGlone, M F Miller, and J R Blanton. 2004. "Environmental Effects on Pig Performance, Meat Quality, and Muscle characteristics1." *Journal of Animal Science* 82. Madison, WI: American Society of Animal Science: 209–17. doi:[2004.821209x](https://doi.org/2004.821209x).
- Göke, M, A Zuk, and D K Podolsky. 1996. "Regulation and Function of Extracellular Matrix Intestinal Epithelial Restitution in Vitro." *The American Journal of Physiology* 271 (5 Pt 1): G729–40. <http://www.ncbi.nlm.nih.gov/pubmed/8944685>.
- Golik, Wiktorija, Olivier Dameron, Jérôme Bugeon, Alice Fatet, Isabelle Hue, Catherine Hurtaud, Matthieu Reichstadt, et al. 2012. "ATOL: The Multi-Species Livestock Trait Ontology." In , 289–300. Springer, Berlin, Heidelberg. doi:[10.1007/978-3-642-35233-1\\_28](https://doi.org/10.1007/978-3-642-35233-1_28).
- Gomaa, Asmaa M S, and Ebtihal A Abd El-Aziz. 2017. "Vitamin D Reduces High-Fat Diet Induced Weight Gain and C-Reactive Protein, Increases Interleukin-10, and Reduces CD86 and Caspase-3." *Pathophysiology: The Official Journal of the International Society for Pathophysiology* 24 (1). Netherlands: 31–37. doi:[10.1016/j.pathophys.2017.01.003](https://doi.org/10.1016/j.pathophys.2017.01.003).
- González, Ignacio, Kim-Anh Lê Cao, Melissa J Davis, and Sébastien Déjean. 2012. "Visualising Associations between Paired 'Omics' Data Sets." *BioData Mining* 5 (1): 19. doi:[10.1186/1756-0381-5-19](https://doi.org/10.1186/1756-0381-5-19).

- 
- Gremel, Gabriela, Alkwin Wanders, Jonathan Cedernaes, Linn Fagerberg, Björn Hallström, Karolina Edlund, Evelina Sjöstedt, Mathias Uhlén, and Fredrik Pontén. 2015. "The Human Gastrointestinal Tract-Specific Transcriptome and Proteome as Defined by RNA Sequencing and Antibody-Based Profiling." *Journal of Gastroenterology* 50 (1): 46–57. doi:10.1007/s00535-014-0958-7.
- Gu, Xianhong, Defa Li, and Ruiping She. 2002. "Effect of Weaning on Small Intestinal Structure and Function in the Piglet." *Archives of Animal Nutrition* 56 (4). Taylor & Francis: 275–86.
- Gu, Zuguang, and Jin Wang. 2013. "CePa: An R Package for Finding Significant Pathways Weighted by Multiple Network Centralities." *Bioinformatics (Oxford, England)* 29 (5). England: 658–60. doi:10.1093/bioinformatics/btt008.
- Habuka, Masato, Linn Fagerberg, Björn M Hallström, Caroline Kampf, Karolina Edlund, Åsa Sivertsson, Tadashi Yamamoto, Fredrik Pontén, Mathias Uhlén, and Jacob Odeberg. 2014. "The Kidney Transcriptome and Proteome Defined by Transcriptomics and Antibody-Based Profiling." Edited by Harald Mischak. *PLoS ONE* 9 (12). San Francisco, USA: Public Library of Science: e116125. doi:10.1371/journal.pone.0116125.
- Haenen, Daniëlle, Carol Souza da Silva, Jing Zhang, Sietse Jan Koopmans, Guido Bosch, Jacques Vervoort, Walter J.J. Gerrits, et al. 2013. "Dietary Resistant Starch Improves Mucosal Gene Expression Profile and Luminal Microbiota Composition in Porcine Colon." Wageningen University. [https://www.researchgate.net/profile/Carol\\_Silva2/publication/258386844\\_Fermentation\\_in\\_the\\_gut\\_to\\_prolong\\_satiety\\_Exploring\\_mechanisms\\_by\\_which\\_dietary\\_fibres\\_affect\\_satiety\\_in\\_pigs/links/53df79f30cf27a7b8306792b.pdf#page=121](https://www.researchgate.net/profile/Carol_Silva2/publication/258386844_Fermentation_in_the_gut_to_prolong_satiety_Exploring_mechanisms_by_which_dietary_fibres_affect_satiety_in_pigs/links/53df79f30cf27a7b8306792b.pdf#page=121).
- Haenen, Daniëlle, Jing Zhang, Carol Souza da Silva, Guido Bosch, Ingrid M van der Meer, Jeroen van Arkel, Joost J G C van den Borne, et al. 2013. "A Diet High in Resistant Starch Modulates Microbiota Composition, SCFA Concentrations, and Gene Expression in Pig Intestine." *The Journal of Nutrition* 143: 274–83. doi:10.3945/jn.112.169672.
- Harris, J Kirk, Karim C El Kasmi, Aimee L Anderson, Michael W Devereaux, Sophie A Fillon, Charles E Robertson, Brandie D Wagner, Mark J Stevens, Norman R Pace, and Ronald J Sokol. 2014. "Specific Microbiome Changes in a Mouse Model of Parenteral Nutrition Associated Liver Injury and Intestinal Inflammation." Edited by Stefan Bereswill. *PLoS ONE* 9 (10). San Francisco, USA: Public Library of Science: e110396. doi:10.1371/journal.pone.0110396.
- Haug, K., R. M. Salek, P. Conesa, J. Hastings, P. de Matos, M. Rijnbeek, T. Mahendraker, et al. 2013. "MetaboLights--an Open-Access General-Purpose Repository for Metabolomics Studies and Associated Meta-Data." *Nucleic Acids Research* 41 (D1). Oxford University Press: D781–86. doi:10.1093/nar/gks1004.
-

## References

---

- Heinritz, Sonja N, Rainer Mosenthin, and Eva Weiss. 2013. "Use of Pigs as a Potential Model for Research into Dietary Modulation of the Human Gut Microbiota." *Nutrition Research Reviews* 26: 191–209. doi:10.1017/S0954422413000152.
- Henderson, Paul, Johan E van Limbergen, Jürgen Schwarze, and David C Wilson. 2011. "Function of the Intestinal Epithelium and Its Dysregulation in Inflammatory Bowel Disease." *Inflammatory Bowel Diseases* 17 (1): 382–95. doi:10.1002/ibd.21379.
- Hogan, Meghan F, Amy W Liu, Michael J Peters, Joshua R Willard, Zaheen Rabbani, Erik C Bartholomew, Adam Ottley, and Rebecca L Hull. 2017. "Markers of Islet Endothelial Dysfunction Occur in Male B6.BKS(D)-Leprdb/J Mice and May Contribute to Reduced Insulin Release." *Endocrinology* 158 (2): 293–303. <http://dx.doi.org/10.1210/en.2016-1393>.
- Hooper, L. V., D. R. Littman, and A. J. Macpherson. 2012. "Interactions Between the Microbiota and the Immune System." *Science*. doi:10.1126/science.1223490.
- Hooper, Lora V. 2004. "Bacterial Contributions to Mammalian Gut Development." *Trends in Microbiology*. doi:10.1016/j.tim.2004.01.001.
- Hörmann, Nives, Inês Brandão, Sven Jäckel, Nelli Ens, Maren Lillich, Ulrich Walter, and Christoph Reinhardt. 2014. "Gut Microbial Colonization Orchestrates TLR2 Expression, Signaling and Epithelial Proliferation in the Small Intestinal Mucosa." Edited by Antonio Moschetta. *PLoS ONE* 9 (11). Public Library of Science: e113080. doi:10.1371/journal.pone.0113080.
- Hoskin-Parr, Lauren, Alison Teyhan, Ariel Blocker, and A J W Henderson. 2013. "Antibiotic Exposure in the First Two Years of Life and Development of Asthma and Other Allergic Diseases by 7.5 Yr: A Dose-Dependent Relationship." *Pediatric Allergy and Immunology* 24 (8): 762–71. doi:10.1111/pai.12153.
- Houle, David, Diddahally R Govindaraju, and Stig Omholt. 2010. "Phenomics: The next Challenge." *Nat Rev Genet* 11 (12). Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.: 855–66. <http://dx.doi.org/10.1038/nrg2897>.
- Iantomasi, T, F Favilli, P Marraccini, T Magaldi, P Bruni, and M T Vincenzini. 1997. "Glutathione Transport System in Human Small Intestine Epithelial Cells." *Biochimica et Biophysica Acta* 1330 (2). NETHERLANDS: 274–83.
- Ihle, James N. 2001. "The Stat Family in Cytokine Signaling." *Current Opinion in Cell Biology* 13 (2): 211–17. doi:[http://dx.doi.org/10.1016/S0955-0674\(00\)00199-X](http://dx.doi.org/10.1016/S0955-0674(00)00199-X).
- IJssennagger, Noortje, Muriel Derrien, Gerdien M van Doorn, Anneke Rijnierse, Bartholomeus van den Bogert, Michael Müller, Jan Dekker, Michiel Kleerebezem, and Roelof van der Meer. 2012. "Dietary Heme Alters Microbiota and Mucosa of Mouse Colon without Functional Changes in Host-Microbe Cross-Talk." *PLoS ONE* 7 (12). Public Library of Science: e49868. <http://dx.doi.org/10.1371/journal.pone.0049868>.

- 
- Inman, C F, K Haverson, S R Konstantinov, P H Jones, C Harris, H Smidt, B Miller, M Bailey, and C Stokes. 2010. "Rearing Environment Affects Development of the Immune System in Neonates." *Clinical & Experimental Immunology* 160 (3). Blackwell Publishing Ltd: 431–39. doi:10.1111/j.1365-2249.2010.04090.x.
- Inoue, Ryo, Takamitsu Tsukahara, Nobuo Nakanishi, and Kazunari Ushida. 2005. "Development of the Intestinal Microbiota in the Piglet." *The Journal of General and Applied Microbiology* 51 (4). 財団法人 応用微生物学研究奨励会: 257–65.
- Iorio, Valentina, Lee D Troughton, and Kevin J Hamill. 2015. "Laminins: Roles and Utility in Wound Repair." *Advances in Wound Care* 4 (4). 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA: Mary Ann Liebert, Inc.: 250–63. doi:10.1089/wound.2014.0533.
- Jacobson, L., and R. Sapolsky. 1991. "The Role of the Hippocampus in Feedback Regulation of the Hypothalamic-Pituitary-Adrenocortical Axis." *Endocrine Reviews*. doi:10.1210/edrv-12-2-118.
- Jakobsson, Hedvig E, Ana M Rodríguez-Piñeiro, André Schütte, Anna Ermund, Preben Boysen, Mats Bemark, Felix Sommer, Fredrik Bäckhed, Gunnar C Hansson, and Malin E V Johansson. 2015. "The Composition of the Gut Microbiota Shapes the Colon Mucus Barrier." *EMBO Reports* 16 (2): 164–77. doi:10.15252/embr.201439263.
- Jarillo-Luna, Adriana, Victor Rivera-Aguilar, Humberto Reyna Garfias, Eleazar Lara-Padilla, Alexander Kormanovsky, and Rafael Campos-Rodríguez. 2007. "Effect of Repeated Restraint Stress on the Levels of Intestinal IgA in Mice." *Psychoneuroendocrinology* 32 (6): 681–92. doi:http://dx.doi.org/10.1016/j.psyneuen.2007.04.009.
- Jeong, H, S P Mason, A.-L. Barabasi, and Z N Oltvai. 2001. "Lethality and Centrality in Protein Networks." *Nature* 411 (6833): 41–42. http://dx.doi.org/10.1038/35075138.
- Jernberg, Cecilia, Sonja Löfmark, Charlotta Edlund, and Janet K Jansson. 2010. "Long-Term Impacts of Antibiotic Exposure on the Human Intestinal Microbiota." *Microbiology* 156 (11): 3216–23. doi:10.1099/mic.0.040618-0.
- Ju, Feng, and Tong Zhang. 2015. "16S rRNA Gene High-Throughput Sequencing Data Mining of Microbial Diversity and Interactions." *Applied Microbiology and Biotechnology* 99 (10). Springer Berlin Heidelberg: 4119–29. doi:10.1007/s00253-015-6536-y.
- Kabat, Agnieszka M, Naren Srinivasan, and Kevin J Maloy. 2014. "Modulation of Immune Development and Function by Intestinal Microbiota." *Trends in Immunology* 35 (11). Elsevier: 507–17.
- Kampf, Caroline, Adil Mardinoglu, Linn Fagerberg, Björn M Hallström, Karolina Edlund, Emma Lundberg, Fredrik Pontén, Jens Nielsen, and Mathias Uhlen. 2014. "The Human Liver-
-

## References

---

- Specific Proteome Defined by Transcriptomics and Antibody-Based Profiling.” *The FASEB Journal* 28 (7): 2901–14. doi:10.1096/fj.14-250555.
- Kanehisa, M, and S Goto. 2000. “KEGG: Kyoto Encyclopedia of Genes and Genomes.” *Nucleic Acids Research* 28 (1). England: 27–30.
- Kar, Soumya K, Alfons J M Jansman, Nirupama Benis, Javier Ramiro-Garcia, Dirkjan Schokker, Leo Kruijt, Anja J J Taverne-Thiele, Jerry M Wells, and Mari A Smits. 2016. “Soybean Meal in Mouse Feed Reduces Expression of mTOR Signalling Pathways and Strongly Increases Abundance of Bacteroidales Family S24-7.” *Submitted*.
- Kelly, Caleb J., Leon Zheng, Eric L. Campbell, Bejan Saeedi, Carsten C. Scholz, Amanda J. Bayless, Kelly E. Wilson, et al. 2015. “Crosstalk between Microbiota-Derived Short-Chain Fatty Acids and Intestinal Epithelial HIF Augments Tissue Barrier Function.” *Cell Host & Microbe* 17 (5). Elsevier: 662–71. doi:10.1016/j.chom.2015.03.005.
- Kelly, D, and A G Coutts. 2000. “Early Nutrition and the Development of Immune Function in the Neonate.” *Proceedings of the Nutrition Society* 59: 177–85. <http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=10946785&retmode=ref&cmd=prlinks%5Cnpapers3://publication/doi/10.1017/S0029665100000197>.
- Kelly, D, J A Smyth, and K J McCracken. 1991. “Digestive Development of the Early-Weaned Pig. 2. Effect of Level of Food Intake on Digestive Enzyme Activity during the Immediate Post-Weaning Period.” *The British Journal of Nutrition* 65 (2): 181–88.
- Kernbauer, Elisabeth, Verena Maier, Dagmar Stoiber, Birgit Strobl, Christine Schneckleithner, Veronika Sexl, Ursula Reichart, et al. 2012. “Conditional Stat1 Ablation Reveals the Importance of Interferon Signaling for Immunity to *Listeria Monocytogenes* Infection.” *PLoS Pathog* 8 (6). Public Library of Science: e1002763. <http://dx.doi.org/10.1371/journal.ppat.1002763>.
- Khalfaoui, Taoufik, Jean-François Groulx, Georges Sabra, Amel GuezGuez, Nuria Basora, Patrick Vermette, and Jean-François Beaulieu. 2013. “Laminin Receptor 37/67LR Regulates Adhesion and Proliferation of Normal Human Intestinal Epithelial Cells.” Edited by Michelina Plateroti. *PLoS ONE* 8 (8). Wiley-VCH Verlag: e74337. doi:10.1371/journal.pone.0074337.
- Kilkenny, Carol, William J. Browne, Innes C. Cuthill, Michael Emerson, and Douglas G. Altman. 2010. “Improving Bioscience Research Reporting: The ARRIVE Guidelines for Reporting Animal Research.” *PLoS Biology* 8 (6). Elsevier: e1000412. doi:10.1371/journal.pbio.1000412.
- Kitano, Hiroaki. 2002. “Systems Biology: A Brief Overview.” *Science* 295 (5560): 1662 LP-1664. <http://science.sciencemag.org/content/295/5560/1662.abstract>.
- Kogut, Michael H, and Ryan J Arsenault. 2016. “Editorial: Gut Health: The New Paradigm in Food

- 
- Animal Production.” *Frontiers in Veterinary Science* 3 (August). Frontiers Media S.A.: 71. doi:10.3389/fvets.2016.00071.
- Kohavi, Ron. 1995. “A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection.” <http://robotics.stanford.edu/~ronnyk>.
- Kolesnikov, Nikolay, Emma Hastings, Maria Keays, Olga Melnichuk, Y Amy Tang, Eleanor Williams, Mirosław Dylag, et al. 2015. “ArrayExpress Update--Simplifying Data Submissions.” *Nucleic Acids Research* 43 (Database issue). England: D1113-6. doi:10.1093/nar/gku1057.
- Konstantinov, Sergey R., Christine F. Favier, Wei Yun Zhu, Barbara A. Williams, Jeannette Kluess, Wolfgang-Bernhard Souffrant, Willem M. de Vos, Antoon D.L. Akkermans, and Hauke Smidt. 2004. “Microbial Diversity Studies of the Porcine Gastrointestinal Ecosystem during Weaning Transition.” *Animal Research* 53 (4): 317–24. doi:10.1051/animres:2004019.
- Konstantinov, Sergey R, Ajay A Awati, Barbara A Williams, Bevis G Miller, Philip Jones, Christopher R Stokes, Antoon D L Akkermans, Hauke Smidt, and Willem M de Vos. 2006. “Post-Natal Development of the Porcine Microbiota Composition and Activities.” *Environmental Microbiology* 8 (7). England: 1191–99. doi:10.1111/j.1462-2920.2006.01009.x.
- Konturek, Peter C, T Brzozowski, and S J Konturek. 2011. “Stress and the Gut: Pathophysiology, Clinical Consequences, Diagnostic Approach and Treatment Options.” *Journal of Physiology and Pharmacology: An Official Journal of the Polish Physiological Society* 62: 591–99. <http://www.ncbi.nlm.nih.gov/pubmed/22314561>.
- Kramer, Frank, Michaela Bayerlová, and Tim Beißbarth. 2014. “R-Based Software for the Integration of Pathway Data into Bioinformatic Algorithms.” *Biology* 3 (1): 85–100. doi:10.3390/biology3010085.
- Kramer, Frank, Michaela Bayerlová, Florian Klemm, Annalen Bleckmann, and Tim Beißbarth. 2013. “R-BiopaxParser-an R Package to Parse, Modify and Visualize BioPAX Data.” *Bioinformatics* 29 (4): 520–22. doi:10.1093/bioinformatics/bts710.
- Lad, Sonya P, Elaine Y Fukuda, Jiali Li, M Luis, and Erguang Li. 2005. “Up-Regulation of the JAK/STAT1 Signal Pathway during Chlamydia Trachomatis Infection.” *The Journal of Immunology* 174 (11). Am Assoc Immunol: 7186–93.
- Lallès, Jean-Paul, Paolo Bosi, Hauke Smidt, and Chris R. Stokes. 2007. “Weaning — A Challenge to Gut Physiologists.” *Livestock Science* 108 (1–3): 82–93. doi:10.1016/j.livsci.2007.01.091.
- Lange, Katja, Floor Hugenholtz, Melliana C Jonathan, Henk A Schols, Michiel Kleerebezem, Hauke Smidt, Michael Müller, and Guido J E J Hooiveld. 2015. “Comparison of the Effects of Five Dietary Fibers on Mucosal Transcriptional Profiles, and Luminal Microbiota Composition and SCFA Concentrations in Murine Colon.” *Molecular Nutrition & Food Research* 59 (8). Wiley Online Library: 1590–1602. doi:10.1002/mnfr.201400597.
-

- Langrish, Claire L, Brent S McKenzie, Nicholas J Wilson, Rene de Waal Malefyt, Robert A Kastelein, and Daniel J Cua. 2004. "IL-12 and IL-23: Master Regulators of Innate and Adaptive Immunity." *Immunological Reviews* 202 (1). Wiley Online Library: 96–105.
- Le Bon, Agnes, and David F Tough. 2002. "Links between Innate and Adaptive Immunity via Type I Interferon." *Current Opinion in Immunology* 14 (4): 432–36. doi:10.1016/S0952-7915(02)00354-0.
- Lê Cao, Kim-Anh, Ignacio González, and Sébastien Déjean. 2009. "integrOmics: An R Package to Unravel Relationships between Two Omics Datasets." *Bioinformatics* 25 (21). Oxford Univ Press: 2855–56.
- Le Leu, Richard K, Graeme P Young, Ying Hu, Jean Winter, and Michael A Conlon. 2013. "Dietary Red Meat Aggravates Dextran Sulfate Sodium-Induced Colitis in Mice Whereas Resistant Starch Attenuates Inflammation." *Digestive Diseases and Sciences* 58 (12): 3475–82. doi:10.1007/s10620-013-2844-1.
- Leuchter, Andrew F, Aimee M Hunter, David E Krantz, and Ian A Cook. 2014. "Intermediate Phenotypes and Biomarkers of Treatment Outcome in Major Depressive Disorder." *Dialogues in Clinical Neuroscience* 16 (4). France: Les Laboratoires Servier: 525–37. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4336921/>.
- Levy, David E, and J E Darnell. 2002. "STATs: Transcriptional Control and Biological Impact." *Nat Rev Mol Cell Biol* 3 (9). Nature Publishing Group: 651–62. <http://dx.doi.org/10.1038/nrm909>.
- Levy, David E, and Chien-kuo Lee. 2002. "What Does Stat3 Do?" *The Journal of Clinical Investigation* 109 (9). The American Society for Clinical Investigation: 1143–48. doi:10.1172/JCI15650.
- Li, Xuejuan, Fangrui Ding, Xiaoyan Zhang, Baihong Li, and Jie Ding. 2016. "The Expression Profile of Complement Components in Podocytes." *International Journal of Molecular Sciences* 17 (4): 471. doi:10.3390/ijms17040471.
- Lindskog, Cecilia, Jerker Linné, Linn Fagerberg, Björn M Hallström, Carl Johan Sundberg, Malene Lindholm, Mikael Huss, et al. 2015. "The Human Cardiac and Skeletal Muscle Proteomes Defined by Transcriptomics and Antibody-Based Profiling." *BMC Genomics* 16 (1): 475. doi:10.1186/s12864-015-1686-y.
- Liu, Xingyin, Rong Lu, Yinglin Xia, and Jun Sun. 2010. "Global Analysis of the Eukaryotic Pathways and Networks Regulated by Salmonella Typhimurium in Mouse Intestinal Infection in Vivo." *BMC Genomics* 11 (1): 722. doi:10.1186/1471-2164-11-722.
- Lodish, Harvey, Arnold Berk, S Lawrence Zipursky, Paul Matsudaira, David Baltimore, and James Darnell. 2000. "Transport across Epithelia." In *Molecular Cell Biology*. 4th Edition, Section

- 15.7. New York: WH Freeman. <https://www.ncbi.nlm.nih.gov/books/NBK21502/>.
- Lozupone, Catherine A, Jesse I Stombaugh, Jeffrey I Gordon, Janet K Jansson, and Rob Knight. 2012. "Diversity, Stability and Resilience of the Human Gut Microbiota." *Nature* 489 (7415). Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.: 220–30. <http://dx.doi.org/10.1038/nature11550>.
- Lu, Kun, Ryan Phillip Abo, Katherine Ann Schlieper, Michelle E. Graffam, Stuart Levine, John S. Wishnok, James A. Swenberg, Steven R. Tannenbaum, and James G. Fox. 2014. "Arsenic Exposure Perturbs the Gut Microbiome and Its Metabolic Profile in Mice: An Integrated Metagenomics and Metabolomics Analysis." *Environmental Health Perspectives* 122 (3): 284–91. doi:10.1289/ehp.1307429.
- Macfarlane, G T, and S Macfarlane. 1997. "Human Colonic Microbiota: Ecology, Physiology and Metabolic Potential of Intestinal Bacteria." *Scand J Gastroenterol Suppl* 222.
- Mamane, Yael, Chi Chung Chan, Genevieve Lavallee, Nicolas Morin, Li-Jing Xu, JingQi Huang, Robert Gordon, et al. 2009. "The C3a Anaphylatoxin Receptor Is a Key Mediator of Insulin Resistance and Functions by Modulating Adipose Tissue Macrophage Infiltration and Activation." *Diabetes* 58 (9). United States: 2006–17. doi:10.2337/db09-0323.
- Mantis, N J, N Rol, and B Corthesy. 2011. "Secretory IgA's Complex Roles in Immunity and Mucosal Homeostasis in the Gut." *Mucosal Immunol* 4 (6). Society for Mucosal Immunology: 603–11. <http://dx.doi.org/10.1038/mi.2011.41>.
- Mariman, Rob, Frans Tielen, Frits Koning, and Lex Nagelkerken. 2015. "The Probiotic Mixture VSL#3 Has Differential Effects on Intestinal Immune Parameters in Healthy Female BALB/c and C57BL/6 Mice." *The Journal of Nutrition* 145 (6): 1354–61. doi:10.3945/jn.114.199729.
- Mayer, E A. 2000. "The Neurobiology of Stress and Gastrointestinal Disease." *Gut* 47 (6): 861–69. <http://gut.bmj.com/content/47/6/861.short>.
- Mayer, Emeran A, Tor Savidge, and Robert J Shulman. 2014. "Brain-Gut Microbiome Interactions and Functional Bowel Disorders." *Gastroenterology* 146 (6): 1500–1512. doi:10.1053/j.gastro.2014.02.037.
- McCarthy, D. J., Y. Chen, and G. K. Smyth. 2012. "Differential Expression Analysis of Multifactor RNA-Seq Experiments with Respect to Biological Variation." *Nucleic Acids Research* 40 (10): 4288–97. doi:10.1093/nar/gks042.
- Mei, Yongguo, Vida Abedi, Adria Carbo, Xiaoying Zhang, Pinyi Lu, Casandra Philipson, Raquel Hontecillas, Stefan Hoops, Nathan Liles, and Josep Bassaganya-Riera. 2015. "Multiscale Modeling of Mucosal Immune Responses." *BMC Bioinformatics* 16 Suppl 1. England: S2. doi:10.1186/1471-2105-16-S12-S2.

- Meng, Chen, Bernhard Kuster, Aedin C Culhane, and Amin Moghaddas Gholami. 2014. "A Multi-variate Approach to the Integration of Multi-Omics Datasets." *BMC Bioinformatics* 15 (1): 162. doi:10.1186/1471-2105-15-162.
- Merle, Nicolas S., Sarah Elizabeth Church, Veronique Fremeaux-Bacchi, and Lubka T. Roumenina. 2015. "Complement System Part I - Molecular Mechanisms of Activation and Regulation." *Frontiers in Immunology* 6 (June): 262. doi:10.3389/fimmu.2015.00262.
- Milacic, Marija, Robin Haw, Karen Rothfels, Guanming Wu, David Croft, Henning Hermjakob, Peter D'Eustachio, and Lincoln Stein. 2012. "Annotating Cancer Variants and Anti-Cancer Therapeutics in Reactome." *Cancers* 4 (4). Multidisciplinary Digital Publishing Institute: 1180–1211.
- Mitreá, Cristina, Zeinab Taghavi, Behzad Bokanizad, Samer Hanoudi, Rebecca Tagett, Michele Donato, Călin Voichița, and Sorin Drăghici. 2013. "Methods and Approaches in the Topology-Based Analysis of Biological Pathways." *Frontiers in Physiology* 4 (October). Frontiers Media S.A.: 278. doi:10.3389/fphys.2013.00278.
- Montiel-Castro, Augusto J, Rina M González-Cervantes, Gabriela Bravo-Ruiseco, and Gustavo Pacheco-López. 2013. "The Microbiota-Gut-Brain Axis: Neurobehavioral Correlates, Health and Sociality." *Frontiers in Integrative Neuroscience*. Centro Darwin de Pensamiento Evolucionista and Philosophy Department, Social Sciences and Humanities Division, Universidad Autonoma Metropolitana Iztapalapa Mexico City, Mexico. doi:10.3389/fnint.2013.00070.
- Mootha, Vamsi K, Cecilia M Lindgren, Karl-Fredrik Eriksson, Aravind Subramanian, Smita Sihag, Joseph Lehar, Pere Puigserver, et al. 2003. "PGC-1[alpha]-Responsive Genes Involved in Oxidative Phosphorylation Are Coordinately Downregulated in Human Diabetes." *Nat Genet* 34 (3): 267–73. <http://dx.doi.org/10.1038/ng1180>.
- Moss, Paul, and Mark Drayson. 2015. "Normal Lymphocytes and Non-Neoplastic Lymphocyte Disorders." In *Postgraduate Haematology*, 278–302. John Wiley & Sons, Ltd. doi:10.1002/9781118853771.ch16.
- Mowat, Allan M, and William W Agace. 2014. "Regional Specialization within the Intestinal Immune System." *Nat Rev Immunol* 14 (10). Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.: 667–85. <http://dx.doi.org/10.1038/nri3738>.
- Munakata, Kaori, Kiyoe Takashima, Mitsue Nishiyama, Naoko Asano, Akihito Mase, Kyoji Hioki, Yasuyuki Ohnishi, Masahiro Yamamoto, and Kenji Watanabe. 2012. "Microarray Analysis on Germfree Mice Elucidates the Primary Target of a Traditional Japanese Medicine Juzentaihoto: Acceleration of IFN- $\alpha$  Response via Affecting the ISGF3-IRF7 Signaling Cascade." *BMC Genomics* 13 (1): 30. doi:10.1186/1471-2164-13-30.
- Munkonda, Mercedes Nancy, Marc Lapointe, Pierre Miegueu, Christian Roy, Danny Gauvreau, Denis Richard, and Katherine Cianflone. 2012. "Recombinant Acylation Stimulating Protein

- Administration to C3-/- Mice Increases Insulin Resistance via Adipocyte Inflammatory Mechanisms." *PLoS One* 7 (10). United States: e46883. doi:10.1371/journal.pone.0046883.
- Muñoz-García, Agustí, Christopher P. Thomas, Diane S. Keeney, Yuxiang Zheng, and Alan R. Brash. 2014. "The Importance of the Lipoxygenase-Hepoxilin Pathway in the Mammalian Epidermal Barrier." *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* 1841 (3): 401–8. doi:10.1016/j.bbalip.2013.08.020.
- Mutch, David M, Vanessa Crespy, Jennifer Clough, Colin J Henderson, Sofiane Lariani, Robert Mansourian, Julie Moulin, C Roland Wolf, and Gary Williamson. 2006. "Hepatic Cytochrome P-450 Reductase-Null Mice Show Reduced Transcriptional Response to Quercetin and Reveal Physiological Homeostasis between Jejunum and Liver." *American Journal of Physiology - Gastrointestinal and Liver Physiology* 291 (1): G63 LP-G72. <http://ajpgi.physiology.org/content/291/1/G63.abstract>.
- Newburg, David S, and W Allan Walker. 2007. "Protection of the Neonate by the Innate Immune System of Developing Gut and of Human Milk." *Pediatr Res* 61 (1). International Pediatrics Research Foundation, Inc.: 2–8. <http://dx.doi.org/10.1203/01.pdr.0000250274.68571.18>.
- Newman, Mark E J. 2006. "Modularity and Community Structure in Networks." *Proceedings of the National Academy of Sciences* 103 (23). National Acad Sciences: 8577–82.
- Nicholson, Jeremy K, Elaine Holmes, James Kinross, Remy Burcelin, Glenn Gibson, Wei Jia, and Sven Pettersson. 2012. "Host-Gut Microbiota Metabolic Interactions." *Science* 336 (6086). American Association for the Advancement of Science: 1262–67.
- Nicolae, Dan L, Eric Gamazon, Wei Zhang, Shiwei Duan, M Eileen Dolan, and Nancy J Cox. 2010. "Trait-Associated SNPs Are More Likely to Be eQTLs: Annotation to Enhance Discovery from GWAS." *PLoS Genetics* 6 (4). Public Library of Science: e1000888. <http://dx.doi.org/10.1371/journal.pgen.1000888>.
- Nie, Fachuan, Dong Su, Ying Shi, Jinmei Chen, Haihui Wang, Wanxiang Qin, Yaohua Chen, Suxia Wang, and Lei Li. 2015. "A Preliminary Study on the Role of the Complement Regulatory Protein, Cluster of Differentiation 55, in Mice with Diabetic Neuropathic Pain." *Molecular Medicine Reports* 11 (3): 2076–82. doi:10.3892/mmr.2014.2896.
- Niewold, T A. 2007. "The Nonantibiotic Anti-Inflammatory Effect of Antimicrobial Growth Promoters, the Real Mode of Action? A Hypothesis." *Poultry Science* 86 (4): 605–9. <http://ps.oxfordjournals.org/content/86/4/605.abstract>.
- Nylund, Lotta, Reetta Satokari, Seppo Salminen, and Willem M de Vos. 2014. "Intestinal Microbiota during Early Life - Impact on Health and Disease." *The Proceedings of the Nutrition Society*, 1–13. doi:10.1017/S0029665114000627.
- O'Hara, Ann M, and Fergus Shanahan. 2006. "The Gut Flora as a Forgotten Organ." *EMBO Reports*

7 (7): 655–746. doi:10.1038/sj.embor.7400731.

- O'Mahony, Siobhain M, Julian R Marchesi, Paul Scully, Caroline Codling, Anne-Marie Ceolho, Eamonn M M Quigley, John F Cryan, and Timothy G Dinan. 2009. "Early Life Stress Alters Behavior, Immunity, and Microbiota in Rats: Implications for Irritable Bowel Syndrome and Psychiatric Illnesses." *Biological Psychiatry* 65 (3): 263–67. doi:http://dx.doi.org/10.1016/j.biopsych.2008.06.026.
- Odamaki, Toshitaka, Kumiko Kato, Hirosuke Sugahara, Nanami Hashikura, Sachiko Takahashi, Jin-zhong Xiao, Fumiaki Abe, and Ro Osawa. 2016. "Age-Related Changes in Gut Microbiota Composition from Newborn to Centenarian: A Cross-Sectional Study." *BMC Microbiology* 16 (May). London: BioMed Central: 90. doi:10.1186/s12866-016-0708-5.
- Olszak, Torsten, Dingding An, Sebastian Zeissig, MP Miguel Pinilla Vera, Julia Richter, Andre Franke, Jonathan N Glickman, et al. 2012. "Microbial Exposure during Early Life Has Persistent Effects on Natural Killer T Cell Function." *Science* 336: 489–93. doi:10.1126/science.1219328.Microbial.
- Ono, K, T Muetze, G Kolishovski, P Shannon, and B Demchak. 2015. "CyREST: Turbocharging Cytoscape Access for External Tools via a RESTful API [Version 1; Referees: 2 Approved]." *F1000Research* 4 (478). doi:10.12688/f1000research.6767.1.
- Ormerod, Kate L, David L A Wood, Nancy Lachner, Shaan L Gellatly, Joshua N Daly, Jeremy D Parsons, Cristiana G O Dal'Molin, et al. 2016. "Genomic Characterization of the Uncultured Bacteroidales Family S24-7 Inhabiting the Guts of Homeothermic Animals." *Microbiome* 4 (1): 1–17. doi:10.1186/s40168-016-0181-2.
- Orth, Jeffrey D, Ines Thiele, and Bernhard Ø Palsson. 2010. "What Is Flux Balance Analysis?" *Nature Biotechnology* 28 (3). Nature Research: 245–48. doi:10.1038/nbt.1614.
- Ottman, Noora, Hauke Smidt, Willem M de Vos, and Clara Belzer. 2012. "The Function of Our Microbiota: Who Is out There and What Do They Do?" *Frontiers in Cellular and Infection Microbiology* 2 (August). Frontiers Media S.A.: 104. doi:10.3389/fcimb.2012.00104.
- Pabst, O, and A M Mowat. 2012. "Oral Tolerance to Food Protein." *Mucosal Immunol* 5 (3). Society for Mucosal Immunology: 232–39. http://dx.doi.org/10.1038/mi.2012.4.
- Paglialunga, Sabina, Patrick Schrauwen, Christian Roy, Esther Moonen-Kornips, Huiling Lu, Matthijs K C Hesselink, Yves Deshaies, Denis Richard, and Katherine Cianflone. 2007. "Reduced Adipose Tissue Triglyceride Synthesis and Increased Muscle Fatty Acid Oxidation in C5L2 Knockout Mice." *The Journal of Endocrinology* 194 (2). England: 293–304. doi:10.1677/JOE-07-0205.
- Palmer, Chana, Elisabeth M Bik, Daniel B DiGiulio, David A Relman, and Patrick O Brown. 2007. "Development of the Human Infant Intestinal Microbiota." *PLoS Biol* 5 (7). Public Library of

- Science: e177. <http://dx.doi.org/10.1371/journal.pbio.0050177>.
- Pérez Gutiérrez, O. 2010. *Unraveling Piglet Gut Microbiota Dynamics in Response to Feed Additives*. [SI: sn].
- Pérez Gutiérrez, O, B van den Bogert, M Derrien, S Koopmans, D Molenaar, W M de Vos, and H Smidt. 2010. "Design of a High-Throughput Diagnostic Microarray for the Characterization of Pig Gastrointestinal Tract Microbiota (Chapter 3)." *Unraveling Piglet Gut Microbiota Dynamics in Response to Feed Additives [Dissertation]*. Wageningen: Wageningen University, 40–67.
- Peterson, Lance W, and David Artis. 2014. "Intestinal Epithelial Cells: Regulators of Barrier Function and Immune Homeostasis." *Nat Rev Immunol* 14 (3). Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.: 141–53. <http://dx.doi.org/10.1038/nri3608>.
- Phipson, Belinda, Stanley Lee, Ian J. Majewski, Warren S. Alexander, and Gordon K. Smyth. 2016. "Robust Hyperparameter Estimation Protects against Hypervariable Genes and Improves Power to Detect Differential Expression." Applications; Genomics. *The Annals of Applied Statistics* 10 (2): 946–63. doi:10.1214/16-AOAS920.
- Pluske, John R, Jean Le Dividich, and Martin W A Verstegen. 2003. *Weaning the Pig: Concepts and Consequences*. Wageningen Academic Pub.
- Poe, S L, M Arora, T B Oriss, M Yarlagadda, K Isse, A Khare, D E Levy, et al. 2013. "STAT1-Regulated Lung MDSC-like Cells Produce IL-10 and Efferocytose Apoptotic Neutrophils with Relevance in Resolution of Bacterial Pneumonia." *Mucosal Immunol* 6 (1). Society for Mucosal Immunology: 189–99. <http://dx.doi.org/10.1038/mi.2012.62>.
- Qing, Yulan, and George R Stark. 2004. "Alternative Activation of STAT1 and STAT3 in Response to Interferon-Gamma." *The Journal of Biological Chemistry* 279 (40): 41679–85. doi:10.1074/jbc.M406413200.
- Rajasundaram, Dhivyaa, Jean-Luc Runavot, Xiaoyuan Guo, William G T Willats, Frank Meulewaeter, and Joachim Selbig. 2014. "Understanding the Relationship between Cotton Fiber Properties and Non-Cellulosic Cell Wall Polysaccharides." *PLoS ONE* 9 (11). Public Library of Science: e112168. <http://dx.doi.org/10.1371/journal.pone.0112168>.
- Rajilić-Stojanović, Mirjana, Hans G H J Heilig, Douwe Molenaar, Kajsa Kajander, Anu Surakka, Hauke Smidt, and Willem M De Vos. 2009. "Development and Application of the Human Intestinal Tract Chip, a Phylogenetic Microarray: Analysis of Universally Conserved Phylogenotypes in the Abundant Microbiota of Young and Elderly Adults." *Environmental Microbiology* 11 (7). Blackwell Publishing Ltd: 1736–51. doi:10.1111/j.1462-2920.2009.01900.x.
- Ramos, Erin M, Douglas Hoffman, Heather A Junkins, Donna Maglott, Lon Phan, Stephen T Sher-

## References

---

- ry, Mike Feolo, and Lucia A Hindorff. 2014. "Phenotype–Genotype Integrator (PheGenI): Synthesizing Genome-Wide Association Study (GWAS) Data with Existing Genomic Resources." *European Journal of Human Genetics* 22 (1). Nature Publishing Group: 144–47. doi:10.1038/ejhg.2013.96.
- Rayan, Paran, Deborah Stenzel, and Pauline Ann McDonnell. 2005. "The Effects of Saturated Fatty Acids on *Giardia Duodenalis* Trophozoites in Vitro." *Parasitology Research* 97 (3). Germany: 191–200. doi:10.1007/s00436-005-1432-5.
- Reikvam, Dag Henrik, Muriel Derrien, Rejoanoul Islam, Alexander Erofeev, Vedrana Grcic, Anders Sandvik, Peter Gaustad, et al. 2012. "Epithelial-Microbial Crosstalk in Polymeric Ig Receptor Deficient Mice." *European Journal of Immunology* 42 (11): 2959–70. doi:10.1002/eji.201242543.
- Reikvam, Dag Henrik, Alexander Erofeev, Anders Sandvik, Vedrana Grcic, Frode Lars Jahnsen, Peter Gaustad, Kathy D McCoy, Andrew J Macpherson, Leonardo A Meza-Zepeda, and Finn -Eirik Johansen. 2011. "Depletion of Murine Intestinal Microbiota: Effects on Gut Mucosa and Epithelial Gene Expression." *PLoS ONE* 6 (3). Public Library of Science: e17996. http://dx.doi.org/10.1371%252Fjournal.pone.0017996.
- Relman, David A. 2012. "The Human Microbiome: Ecosystem Resilience and Health." *Nutrition Reviews* 70. Blackwell Publishing Inc: S2–9. doi:10.1111/j.1753-4887.2012.00489.x.
- Ricklin, Daniel, George Hajishengallis, Kun Yang, and John D Lambris. 2010. "Complement: A Key System for Immune Surveillance and Homeostasis." *Nature Immunology* 11 (9): 785–97. doi:10.1038/ni.1923.
- Rooks, Michelle G, Patrick Veiga, Leslie H Wardwell-Scott, Timothy Tickle, Nicola Segata, Monia Michaud, Carey Ann Gallini, et al. 2014. "Gut Microbiome Composition and Function in Experimental Colitis during Active Disease and Treatment-Induced Remission." *ISME J* 8 (7). International Society for Microbial Ecology: 1403–17. http://dx.doi.org/10.1038/ismej.2014.3.
- Rossi, Oriana, Peter van Baarlen, and Jerry M Wells. 2011. "Host-Recognition of Pathogens and Commensals in the Mammalian Intestine." In *Between Pathogenicity and Commensalism*, 291–321. Springer.
- Rotenberg, Samuel, and Jens Oluf Andersen. 1982. "The Effect of Antibiotics on Some Lipid Metabolism Parameters in Rats Receiving Corn Starch, Potato Flour or Pectin in the Diet." *Acta Agriculturae Scandinavica* 32 (3). Taylor & Francis: 321–40.
- Saccanti, Edoardo, Maria Suarez-Diez, Claudio Luchinat, Claudio Santucci, and Leonardo Tenori. 2014. "Probabilistic Networks of Blood Metabolites in Healthy Subjects as Indicators of Latent Cardiovascular Risk." *Journal of Proteome Research* 14 (2). ACS Publications: 1101–11.

- 
- Sahoo, Swagatika, and Ines Thiele. 2013. "Predicting the Impact of Diet and Enzymopathies on Human Small Intestinal Epithelial Cells." *Human Molecular Genetics* 22 (13). England: 2705–22. doi:10.1093/hmg/ddt119.
- Sanden, Suzy Van, Dan Lin, and Tomasz Burzykowski. 2008. "Performance of Gene Selection and Classification Methods in a Microarray Setting: A Simulation Study." *Communications in Statistics - Simulation and Computation* 37 (2). Taylor & Francis Group : 409–24. doi:10.1080/03610910701792554.
- Sargent, R G. 2013. "Verification and Validation of Simulation Models." *Journal of Simulation* 7 (1). Palgrave Macmillan UK: 12–24. doi:10.1057/jos.2012.20.
- Sato, Hirokazu, Linda S. Zhang, Kristina Martinez, Eugene B. Chang, Qing Yang, Fei Wang, Philip N. Howles, Ryota Hokari, Soichiro Miura, and Patrick Tso. 2016. "Antibiotics Suppress Activation of Intestinal Mucosal Mast Cells and Reduce Dietary Lipid Absorption in Sprague-Dawley Rats." *Gastroenterology* 151 (5): 923–32. doi:10.1053/j.gastro.2016.07.009.
- Sato, Toshiro, Daniel E Stange, Marc Ferrante, Robert G J Vries, Johan H Van Es, Stieneke Van den Brink, Winan J Van Houdt, et al. 2011. "Long-Term Expansion of Epithelial Organoids from Human Colon, Adenoma, Adenocarcinoma, and Barrett's Epithelium." *Gastroenterology* 141 (5). United States: 1762–72. doi:10.1053/j.gastro.2011.07.050.
- Sato, Toshiro, Robert G Vries, Hugo J Snippert, Marc van de Wetering, Nick Barker, Daniel E Stange, Johan H van Es, et al. 2009. "Single Lgr5 Stem Cells Build Crypt–villus Structures in Vitro without a Mesenchymal Niche." *Nature* 459 (7244). Macmillan Publishers Limited. All rights reserved: 262–65. <http://dx.doi.org/10.1038/nature07935>.
- Schäfer, Juliane, and Korbinian Strimmer. 2005. "A Shrinkage Approach to Large-Scale Covariance Matrix Estimation and Implications for Functional Genomics." *Statistical Applications in Genetics and Molecular Biology* 4 (1).
- Schmittgen, Thomas D, and Kenneth J Livak. 2008. "Analyzing Real-Time PCR Data by the Comparative CT Method." *Nature Protocols* 3 (6). Nature Publishing Group: 1101–8. doi:10.1038/nprot.2008.73.
- Schokker, Dirkjan. 2014. "Pig Experiment Sterksel Early Antibiotics/stress." *Gene Expression Omnibus*. <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE53170>.
- Schokker, Dirkjan, Jing Zhang, Stéphanie A Vastenhouw, H G Heilig, Hauke Smidt, J M Rebel, and Mari A Smits. 2015. "Long-Lasting Effects of Early-Life Antibiotic Treatment and Routine Animal Handling on Gut Microbiota Composition and Immune System in Pigs." *PloS One* 10 (2). Public Library of Science.
- Schokker, Dirkjan, Jing Zhang, Ling-li Zhang, Stéphanie A Vastenhouw, Hans G H J Heilig, Hauke Smidt, Johanna M J Rebel, and Mari A Smits. 2014. "Early-Life Environmental Variation
-

- Affects Intestinal Microbiota and Immune Development in New-Born Piglets.” *PLoS ONE* 9 (6). Public Library of Science: e100040. <http://dx.doi.org/10.1371/journal.pone.0100040>.
- Scholtens, Petra A M J, Raish Oozeer, Rocio Martin, Kaouther Ben Amor, and Jan Knol. 2012. “The Early Settlers: Intestinal Microbiology in Early Life.” *Annual Review of Food Science and Technology* 3 (1). Annual Reviews: 425–47. doi:10.1146/annurev-food-022811-101120.
- Schumann, Alexandra, Sophie Nutten, Dominique Donnicola, Elena M Comelli, Robert Mansourian, Christine Cherbut, Irène Corthesy-Theulaz, and Clara Garcia-Rodenas. 2005. “Neonatal Antibiotic Treatment Alters Gastrointestinal Tract Developmental Gene Expression and Intestinal Barrier Transcriptome.” *Physiological Genomics* 23: 235–45. doi:10.1152/physiolgenomics.00057.2005.
- Schwartz, Scott, Iddo Friedberg, Ivan V Ivanov, Laurie A Davidson, Jennifer S Goldsby, David B Dahl, Damir Herman, Mei Wang, Sharon M Donovan, and Robert S Chapkin. 2012. “A Metagenomic Study of Diet-Dependent Interaction between Gut Microbiota and Host in Infants Reveals Differences in Immune Response.” *Genome Biology* 13 (4): 1–16. doi:10.1186/gb-2012-13-4-r32.
- Segain, J P, D Raingeard de la Blétière, A Bourreille, V Leray, N Gervois, C Rosales, L Ferrier, C Bonnet, H M Blottière, and J P Galmiche. 2000. “Butyrate Inhibits Inflammatory Responses through NFkappaB Inhibition: Implications for Crohn’s Disease.” *Gut* 47 (3): 397–403. <http://www.ncbi.nlm.nih.gov/pubmed/10940278>.
- Sender, Ron, Shai Fuchs, and Ron Milo. 2016. “Revised Estimates for the Number of Human and Bacteria Cells in the Body.” *PLoS Biology* 14 (8). Public Library of Science: e1002533. doi:10.1371/journal.pbio.1002533.
- Serino, Matteo, Elodie Luche, Sandra Gres, Audrey Baylac, Mathieu Bergé, Claire Cenac, Aurelie Waget, et al. 2012. “Metabolic Adaptation to a High-Fat Diet Is Associated with a Change in the Gut Microbiota.” *Gut* 61 (4): 543–53. doi:10.1136/gutjnl-2011-301012.
- Sham, Ho Pan, Emily Yi Shan Yu, Muhammet F. Gulen, Ganive Bhinder, Martin Stahl, Justin M. Chan, Lara Brewster, et al. 2013. “SIGIRR, a Negative Regulator of TLR/IL-1R Signalling Promotes Microbiota Dependent Resistance to Colonization by Enteric Bacterial Pathogens.” Edited by Dana J. Philpott. *PLoS Pathogens* 9 (8). Public Library of Science: e1003539. doi:10.1371/journal.ppat.1003539.
- Shannon, Paul, Andrew Markiel, Owen Ozier, Nitin S Baliga, Jonathan T Wang, Daniel Ramage, Nada Amin, Benno Schwikowski, and Trey Ideker. 2003. “Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks.” *Genome Research* 13 (11): 2498–2504. doi:10.1101/gr.1239303.
- Sharma, R, C Young, and J Neu. 2010. “Molecular Modulation of Intestinal Epithelial Barrier: Contribution of Microbiota.” *J Biomed Biotechnol* 2010.

- 
- Sherwood L, Gorbach. 1996. "Microbiology of the Gastrointestinal Tract." In *Medical Microbiology*, edited by Samuel Baron, 4th ed. The University of Texas Medical Branch at Galveston. <https://www.ncbi.nlm.nih.gov/books/NBK7670/>.
- Smith, Cynthia L, and Janan T Eppig. 2009. "The Mammalian Phenotype Ontology: Enabling Robust Annotation and Comparative Analysis." *Wiley Interdisciplinary Reviews. Systems Biology and Medicine* 1 (3). NIH Public Access: 390–99. doi:10.1002/wsbm.44.
- Smith, P. M., M. R. Howitt, N. Panikoy, M. Michaud, C. A. Gallini, M. Bohlooly-Y, J. N. Glickman, and W. S. Garrett. 2013. "The Microbial Metabolites, Short-Chain Fatty Acids, Regulate Colonic Treg Cell Homeostasis." *Science* 341 (6145): 569–73. doi:10.1126/science.1241165.
- Smyth, G K. 2005. "Limma: Linear Models for Microarray Data." In *Bioinformatics and Computational Biology Solutions Using R and Bioconductor SE - 23*, edited by Robert Gentleman, Vincent J. Carey, Wolfgang Huber, Rafael A. Irizarry, and Sandrine Dudoit, 397–420. Statistics for Biology and Health. Springer New York. doi:10.1007/0-387-29362-0\_23.
- Söderholm, J D, and M H Perdue. 2001. "Stress and the Gastrointestinal Tract II. Stress and Intestinal Barrier Function." *American Journal of Physiology - Gastrointestinal and Liver Physiology* 280 (1 43-1): G7–13. <http://www.scopus.com/inward/record.url?eid=2-s2.0-0342960436&partnerID=40&md5=7e28463f6b40250d05c609af96d47695>.
- Steengena, Wilma T, Mona Mischke, Carolien Lute, Mark V Boekschoten, Agnes Lendvai, Maurien G M Pruis, Henkjan J Verkade, et al. 2016. "Maternal Exposure to a Western-Style Diet Causes Differences in Intestinal Microbiota Composition and Gene Expression of Suckling Mouse Pups." *Molecular Nutrition & Food Research*. doi:10.1002/mnfr.201600141.
- Steinway, Steven N, Matthew B Biggs, Thomas P Loughran Jr, Jason A Papin, and Reka Albert. 2015. "Inference of Network Dynamics and Metabolic Interactions in the Gut Microbiome." *PLOS Computational Biology* 11 (6). Public Library of Science: e1004338. <http://dx.doi.org/10.1371/journal.pcbi.1004338>.
- Stephens, Zachary D., Skylar Y. Lee, Faraz Faghri, Roy H. Campbell, Chengxiang Zhai, Miles J. Efron, Ravishankar Iyer, Michael C. Schatz, Saurabh Sinha, and Gene E. Robinson. 2015. "Big Data: Astronomical or Genomical?" *PLOS Biology* 13 (7). Public Library of Science: e1002195. doi:10.1371/journal.pbio.1002195.
- Suh, M, M Belosevic, and M T Clandinin. 2004. "Dietary Lipids Containing Gangliosides Reduce Giardia Muris Infection in Vivo and Survival of Giardia Lamblia Trophozoites in Vitro." *Parasitology* 128 (Pt 6). England: 595–602.
- Taché, Y, V Martinez, M Million, and L Wang. 2001. "Stress and the Gastrointestinal Tract III. Stress-Related Alterations of Gut Motor Function: Role of Brain Corticotropin-Releasing Factor Receptors." *American Journal of Physiology - Gastrointestinal and Liver Physiology*
-

- 280 (2 43-2): G173–77. <http://www.scopus.com/inward/record.url?eid=2-s2.0-0035008662&partnerID=40&md5=ea12c3f385aa1e64224f90de8be4b3d7>.
- Tako, Ernest A, Maryam F Hassimi, Erqiu Li, and Steven M Singer. 2013. “Transcriptomic Analysis of the Host Response to *Giardia Duodenalis* Infection Reveals Redundant Mechanisms for Parasite Control.” *mBio* 4 (6): e00660-13. doi:10.1128/mBio.00660-13.
- Tanaka, Shigemitsu, Takako Kobayashi, Prapa Songjinda, Atsushi Tateyama, Mina Tsubouchi, Chikako Kiyohara, Taro Shirakawa, Kenji Sonomoto, and Jiro Nakayama. 2009. “Influence of Antibiotic Exposure in the Early Postnatal Period on the Development of Intestinal Microbiota.” *FEMS Immunology & Medical Microbiology* 56 (1). Blackwell Publishing Ltd: 80–87. doi:10.1111/j.1574-695X.2009.00553.x.
- Taylor, Chris F, Dawn Field, Susanna-Assunta Sansone, Jan Aerts, Rolf Apweiler, Michael Ashburner, Catherine A Ball, et al. 2008. “Promoting Coherent Minimum Reporting Guidelines for Biological and Biomedical Investigations: The MIBBI Project.” *Nat Biotech* 26 (8). Nature Publishing Group: 889–96. <http://dx.doi.org/10.1038/nbt.1411>.
- Taylor, Ian W, Rune Linding, David Warde-Farley, Yongmei Liu, Catia Pesquita, Daniel Faria, Shelley Bull, Tony Pawson, Quaid Morris, and Jeffrey L Wrana. 2009. “Dynamic Modularity in Protein Interaction Networks Predicts Breast Cancer Outcome.” *Nat Biotech* 27 (2). Nature Publishing Group: 199–204. <http://dx.doi.org/10.1038/nbt.1522>.
- Teran, V, P Luna Pizarro, M F Zacarias, G Vinderola, R Medina, and C Van Nieuwenhove. 2015. “Production of Conjugated Dienoic and Trienoic Fatty Acids by Lactic Acid Bacteria and Bifidobacteria.” *JOURNAL OF FUNCTIONAL FOODS* 19 (A): 417–25. doi:10.1016/j.jff.2015.09.046.
- Tibshirani, Robert. 2011. “Regression Shrinkage and Selection via the Lasso: A Retrospective.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 73 (3). Wiley Online Library: 273–82.
- Topping, David L, and Peter M Clifton. 2001. “Short-Chain Fatty Acids and Human Colonic Function: Roles of Resistant Starch and Nonstarch Polysaccharides.” *Physiological Reviews* 81 (3): 1031 LP-1064. <http://physrev.physiology.org/content/81/3/1031.abstract>.
- Tremaroli, Valentina, and Fredrik Backhed. 2012. “Functional Interactions between the Gut Microbiota and Host Metabolism.” *Nature* 489 (7415). Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.: 242–49. <http://dx.doi.org/10.1038/nature11552>.
- Tryka, Kimberly A, Luning Hao, Anne Sturcke, Yumi Jin, Zhen Y Wang, Lora Ziyabari, Moira Lee, et al. 2014. “NCBI’s Database of Genotypes and Phenotypes: dbGaP.” *Nucleic Acids Research* 42 (D1): D975–79. <http://dx.doi.org/10.1093/nar/gkt1211>.

- 
- Tseng, George C, Debashis Ghosh, and Eleanor Feingold. 2012. "Comprehensive Literature Review and Statistical Considerations for Microarray Meta-Analysis." *Nucleic Acids Research* 40 (9). Oxford University Press: 3785–99. doi:10.1093/nar/gkr1265.
- Tsurugizawa, Tomokazu, Akira Uematsu, Eiji Nakamura, Mai Hasumura, Mariko Hirota, Takashi Kondoh, Hisayuki Uneyama, and Kunio Torii. 2009. "Mechanisms of Neural Response to Gastrointestinal Nutritive Stimuli: The Gut-Brain Axis." *Gastroenterology* 137 (1). Elsevier: 262–73. doi:10.1053/j.gastro.2009.02.057.
- Uhlén, Mathias, Linn Fagerberg, Björn M Hallström, Cecilia Lindskog, Per Oksvold, Adil Mardinoglu, Åsa Sivertsson, et al. 2015. "Tissue-Based Map of the Human Proteome." *Science* 347 (6220). <http://science.sciencemag.org/content/347/6220/1260419.abstract>.
- Ulmius, Matilda, Anna Johansson, and Gunilla Önning. 2009. "The Influence of Dietary Fibre Source and Gender on the Postprandial Glucose and Lipid Response in Healthy Subjects." *European Journal of Nutrition* 48 (7): 395. doi:10.1007/s00394-009-0026-x.
- Van den Abbeele, Pieter, Philippe Gérard, Sylvie Rabot, Aurélie Bruneau, Sahar El Aidy, Muriel Derrien, Michiel Kleerebezem, et al. 2011. "Arabinoxylans and Inulin Differentially Modulate the Mucosal and Luminal Gut Microbiota and Mucin-Degradation in Humanized Rats." *Environmental Microbiology* 13 (10). Blackwell Publishing Ltd: 2667–80. doi:10.1111/j.1462-2920.2011.02533.x.
- van der Flier, Laurens G, and Hans Clevers. 2009. "Stem Cells, Self-Renewal, and Differentiation in the Intestinal Epithelium." *Annual Review of Physiology* 71 (1). Annual Reviews: 241–60. doi:10.1146/annurev.physiol.010908.163145.
- Wang, Bo, Aziz M Mezlini, Feyyaz Demir, Marc Fiume, Zhuowen Tu, Michael Brudno, Benjamin Haibe-Kains, and Anna Goldenberg. 2014. "Similarity Network Fusion for Aggregating Data Types on a Genomic Scale." *Nat Meth* 11 (3). Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.: 333–37. <http://dx.doi.org/10.1038/nmeth.2810>.
- Wang, X, Y F Tao, L L Huang, D M Chen, S Z Yin, A Ihsan, W Zhou, et al. 2012. "Pharmacokinetics of Tulathromycin and Its Metabolite in Swine Administered with an Intravenous Bolus Injection and a Single Gavage." *Journal of Veterinary Pharmacology and Therapeutics* 35 (3). Blackwell Publishing Ltd: 282–89. doi:10.1111/j.1365-2885.2011.01322.x.
- Wells, Jerry M, Robert J Brummer, Muriel Derrien, Thomas T MacDonald, Freddy Troost, Patrice D Cani, Vassilia Theodorou, et al. 2017. "Homeostasis of the Gut Barrier and Potential Biomarkers." *American Journal of Physiology. Gastrointestinal and Liver Physiology* 312 (3). United States: G171–93. doi:10.1152/ajpgi.00048.2015.
- Wells, Jerry M, Linda M P Loonen, and Jurgen M Karczewski. 2010. "The Role of Innate Signaling in the Homeostasis of Tolerance and Immunity in the Intestine." *International Journal of*
-

- Medical Microbiology : IJMM* 300 (1). Germany: 41–48. doi:10.1016/j.ijmm.2009.08.008.
- Wells, Jerry M, Oriana Rossi, Marjolein Meijerink, and Peter van Baarlen. 2011. “Epithelial Cross-talk at the Microbiota-Mucosal Interface.” *Proceedings of the National Academy of Sciences of the United States of America*, no. Supplement 1 (March). National Academy of Sciences: 4607–14. doi:10.1073/pnas.1000092107.
- Welter, Danielle, Jacqueline MacArthur, Joannella Morales, Tony Burdett, Peggy Hall, Heather Junkins, Alan Klemm, et al. 2014. “The NHGRI GWAS Catalog, a Curated Resource of SNP -Trait Associations.” *Nucleic Acids Research* 42 (Database issue). Oxford University Press: D1001–6. doi:10.1093/nar/gkt1229.
- Weng, M, and W A Walker. 2013. “The Role of Gut Microbiota in Programming the Immune Phenotype.” *Journal of Developmental Origins of Health and Disease* 4: 203–14. doi:10.1017/S2040174412000712.
- Wilkinson, Mark D, Michel Dumontier, I Jstrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. 2016. “The FAIR Guiding Principles for Scientific Data Management and Stewardship.” *Scientific Data* 3 (March). England: 160018. doi:10.1038/sdata.2016.18.
- Willing, Benjamin P, Navkiran Gill, and B Brett Finlay. 2010. “The Role of the Immune System in Regulating the Microbiota.” *Gutmicrobes* 1 (4). Landes Bioscience Inc.: 213–23. <http://www.landesbioscience.com/journals/gutmicrobes/article/12520/>.
- Willing, Benjamin P, Shannon L Russell, and B Brett Finlay. 2011. “Shifting the Balance: Antibiotic Effects on Host-Microbiota Mutualism.” *Nature Reviews. Microbiology* 9: 233–43. doi:10.1038/nrmicro2536.
- Winter, David. 2016. “Rentrez: Entrez in R.” <https://cran.r-project.org>. <https://cran.r-project.org/web/packages/rentrez/rentrez.pdf>.
- Wu, Guanming, Xin Feng, and Lincoln Stein. 2010. “Research a Human Functional Protein Interaction Network and Its Application to Cancer Data Analysis.” *Genome Biol* 11: R53.
- Wu, Yu-Sheng, Shu-Ying Ho, Fan-Hua Nan, and Shiu-Nan Chen. 2016. “Ganoderma Lucidum Beta 1,3/1,6 Glucan as an Immunomodulator in Inflammation Induced by a High-Cholesterol Diet.” *BMC Complementary and Alternative Medicine* 16 (1). England: 500. doi:10.1186/s12906-016-1476-3.
- Xia, Jianguo, Christopher D Fjell, Matthew L Mayer, Olga M Pena, David S Wishart, and Robert E W Hancock. 2013. “INMEX--a Web-Based Tool for Integrative Meta-Analysis of Expression Data.” *Nucleic Acids Research* 41 (Web Server issue). Oxford University Press: W63-70. doi:10.1093/nar/gkt338.

- Yichoy, M, T T Duarte, A De Chatterjee, T L Mendez, K Y Aguilera, D Roy, S Roychowdhury, S B Aley, and S Das. 2011. "Lipid Metabolism in Giardia: A Post-Genomic Perspective." *Parasitology* 138 (3): 267–78. doi:10.1017/S0031182010001277.
- Yu, Jeffrey J, and Sarah L Gaffen. 2008. "Interleukin-17: A Novel Inflammatory Cytokine That Bridges Innate and Adaptive Immunity." *Frontiers in Bioscience: A Journal and Virtual Library*. Department of Microbiology and Immunology, University at Buffalo School of Medicine and Biomedical Sciences, Buffalo, NY, USA. doi:10.2741/2667.
- Ze, Xiaolei, Sylvia H Duncan, Petra Louis, and Harry J Flint. 2012. "Ruminococcus Bromii Is a Keystone Species for the Degradation of Resistant Starch in the Human Colon." *ISME J* 6 (8). International Society for Microbial Ecology: 1535–43. <http://dx.doi.org/10.1038/ismej.2012.4>.
- Zheng, G, H Yampara Iquise, J E Jones, and C Andrew Carson. 2009. "Development of Faecalibacterium 16S rRNA Gene Marker for Identification of Human Faeces." *Journal of Applied Microbiology* 106 (2). Wiley Online Library: 634–41.
- Zhu, Yuelin, Sean Davis, Robert Stephens, Paul S Meltzer, and Yidong Chen. 2008. "GEOmetadb: Powerful Alternative Search Engine for the Gene Expression Omnibus." *Bioinformatics* 24 (23): 2798–2800. doi:10.1093/bioinformatics/btn520.
- Zoetendal, Erwin G, Jeroen Raes, Bartholomeus van den Bogert, Manimozhiyan Arumugam, Carien C G M Boonjink, Freddy J Troost, Peer Bork, Michiel Wels, Willem M de Vos, and Michiel Kleerebezem. 2012. "The Human Small Intestinal Microbiota Is Driven by Rapid Uptake and Conversion of Simple Carbohydrates." *ISME J* 6 (7). International Society for Microbial Ecology: 1415–26. <http://dx.doi.org/10.1038/ismej.2011.212>.

# **Acknowledgements**



## Acknowledgements

---

I have several people to thank for the successful completion of my PhD starting from my patient supervisory team to my loving family and wonderful friends. I thank all of them for making my four years filled with laughter and joy (I know it sounds cliché but if you were there, you know this is absolutely true).

From the very first meeting Prof. Mari Smits has been very supportive and understanding. The confidence he placed in my abilities allowed me to develop them more and I am eternally grateful for that. Dirkjan Schokker, one of my daily supervisors was always very involved in my work with hours of discussion/poking holes in my theories, that tended to stray away from scientific topics several times. As his first PhD student, I would ‘strongly recommend’ him as a supervisor to my friends. Without Maria Suarez-Diez, my dry-lab supervisor, I might have given up on my PhD in the first year. From helping me turn on a computer (in my defence, it was a mac, they hide the power button), she has helped me every step of the way with quick solutions and encouraging words. Constructive criticism (or scepticism) from Prof. Vitor Martins dos Santos has always kept me on my toes. My discussions with Prof. Jerry Wells have helped me understand the biology behind my results and achieve the wet-lab – dry-lab – wet-lab cycle.

Even starting my PhD would be inconceivable without the love and support of my husband Rajaram Kaliyaperumal. From drying my tears in the first year to cooking all the meals in the last few months of my PhD, he was truly my rock. I would need a lifetime to thank him adequately. My mother Asumtha Mary, my father Philip Benis and my sister Anupama Philip bore with me very patiently as I kept forgetting to call them. Although doing a masters is not very common where I come from, and committing to a PhD even more rare, my family supported me through it all just to see me happy. This came with several sacrifices from their part for which I will always be grateful. My mother always knew the right yogasana to do for a particular ailment and never tired of helping me keep my health intact. My father's absolute confidence in me to basically do anything and my sister's kind encouragement were absolutely essential for my mental health.

All my colleagues at both HMI (Host Microbe Interactomics) and SSB (Systems and Synthetic Biology) are the best co-workers one could ask for. I had very lively officemates in both departments. Soumya Kar aka. Sam, was the first colleague I met, incidentally in a

deep dark forest that looked like a serial killer's hideout. He helped me with my orientation, where the grocery store was, how to get to Wageningen from Ede, how to pronounce Ede and so many, many things. He also did not stop directing me until the third year of my PhD, but well, if you know Sam you know this is completely normal. All of HMI made me feel included and at home and I am privileged to call them my friends. Even though several HMIers were not used to a dry-lab person before, most of them gave me the benefit of the doubt and assumed I was in fact working and not just on Facebook all the time. Aga helped me with the outsider's view into the Netherlands while also allowing me to vent out my frustrations. I had a great time with Nuning and Berdien, I just hope I did not scare them too much on what was to come with my final year stress sessions. Loes is the heart of HMI, she is always willing to lend a ear and make you feel special. Anja and Nico are the most generous people I have ever met, they have, at some point, helped every single HMIer with a kind word, a shoulder to cry on or apple pie. Although Marcela and Edoardo took a lot of time to master my name (I only gave you 4 letters by the way), we did have a lot of non-fruitful conversations, which are the best kind anyway. Linda, Rogier, Ellen and Marjolein initiated me into the Dutch way with a lot of frankness that I was too nervous to understand in the first 6 months of my studies, but I must confess, I am far too comfortable with it now. Nadya, here is one whole sentence dedicated only to you, partly devoid of content, but seriously, I admire your wry sense of humour partly because you say things out loud as I am filtering them out of my thoughts. Raka, Jonna, Simon, Jori, Bart, Jurgen, Peter and Michiel (sorry we took your office) have all made my stay at HMI pleasant and eventful and I am very thankful for that.

As Maarten once said, SSB used to be a group full of nerdy guys, they had things called LAN parties, I was hurt at not even being invited for this party until I learned what it was. Maarten, o where do I start, he did not interrupt anyone during work hours and balanced that by not staying quiet even for a minute during the breaks. Ruben 'white-knighted' me fending off many of Maarten's barbs, but Ruben's goal might have been to contradict him rather than protect me. During the ruckus Bart was very patient or he had very good headphones, but he never refused to help any way he could, even if that meant teaching me some very basic things. Dorett, the one person I could relate to in SSB initially, is a great listener and gave me permission to not write the whole acknowledgement section on her alone, for which I am very grateful, and I am not just saying this because she has embar-

## Acknowledgements

---

raising photos of me. Her wit and coffee after lunch kept me quite entertained, yes, someone else drinking coffee benefited me, its a long story. Bastian initiated our breaks, coffee and lunch, and did not take no for an answer, sometimes resorting to physical means to take people out of the office or make them stay at the break room. Benoit saved my laptop when it was 'too hot to handle', yes overheating problems, and has since helped me with shell scripts while making sure I feel thoroughly inadequate. My new office mates Linde and Nhung regaled me with their impressions of pressure which took me back to my first year. These wonderful officemates, initiated by Ruben, also kept a countdown of the days before my submission deadline, although I did not appreciate having it right in front of my face. Kalyanasundaram aka Kal, helped me pour out my 'challenges' in Tamil so that not everyone necessarily knew what I was talking about. Coffee and lunch breaks were always 'eventful' especially if Bastian, Maarten, Benoit, Robert and Nikolaos were all involved, unless one migrated to the 'other' break group of Jasper, Jesse and Niels who were a little more PG-13. But now with more of a female presence with Emma, Erika, Linde and Nhung, we can finally talk about more than games I have never heard of or Maarten's VR system. My time in SSB was made more enjoyable by my colleagues Mark, Michel, Milad, Edoardo Saccenti, Peter Schaap, Tjerko, Sudarshan and Alex.

All my collaborators on the IPOP Systems Biology project helped me bounce ideas off them as we were locked away in the most remote place organizers could find. I am especially grateful to the IPOP Virtual Gut team, Bastian, Balaji and Neeraj and their supervisors for helping me along the way.

My PhD experience was enhanced by the wonderful people I met and the amazing people who supported me. I apologise if I have missed anyone, it was not intentional. I hope to be able to carry on these friendships as I embark on my career in the 'real world'.



# **Training and Supervision Plan**



**Conferences and Symposiums (10 credits)**

ECCB2016, European Conference on Computation Biology, 2016

(Poster presentation)

BioSB2016, Dutch Bioinformatics and Systems Biology Conference, 2016

(Poster presentation)

BioSB2015 Dutch Bioinformatics and Systems Biology Conference, 2015

(Poster presentation)

SB@NL 2014 Systems Biology Symposium, Maastricht, 2014

(Poster presentation)

Symposium Systems Biology for Food, Feed, and Health, 2013,

(Oral presentation)

SB@NL 2013 Systems Biology Symposium, 2013

(Poster presentation)

NBIC2014 Netherlands Bioinformatics Conference, 2014

(Poster presentation)

**In-depth Courses (12 credits)**

Parameter Estimation Tutorial, 2013

Elementary Calculus in Systems Biology, 2014

Algorithms for Biological Networks, 2014

The Intestinal Microbiome and Diet in Human and Animal Health, 2014

Pattern recognition, 2015

Managing and Integrating Life Science Information, 2015

**Other courses (3)**

The Essentials of Scientific Writing and Presenting, 2015

Mobilising your scientific network, 2015

Information Literacy including EndNote Introduction, 2013

**Research skills training (6 credits)**

Preparing own PhD research proposal

**Management skills training (2 credits)**

WIAS Science Day organizing committee

# **List of Publications**



Benis N, Schokker D, Suarez-Diez M, Martins dos Santos V, Smidt H, Smits M A, Network Analysis of temporal functionalities of the gut induced by perturbations in new-born piglets, BMC Genomics, <http://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-015-1733-8>.

Benis N, Schokker D, Kramer F, Smits M A, Suarez-Diez M, Building pathway graphs from BioPAX data in R, [version 2; referees: 3 approved, 1 approved with reservations]. F1000Research 2016, 5:2414 (doi: 10.12688/f1000research.9582.2).

Benis N, Kar S K., Martins dos Santos V A. P., Smits M A., Schokker D, Suarez-Diez M, Multi-Level Integration of Environmentally Perturbed Internal Phenotypes Reveals Key Points of Connectivity between Them, Frontiers in Physiology (<https://doi.org/10.3389/fphys.2017.00388>)

The research described in this thesis was financially supported by the Systems Biology Investment Programme of Wageningen University. Financial support from the Wageningen Institute of Animals Sciences for printing this thesis is gratefully acknowledged.

Cover design by Nirupama Benis

Printed by Proefschriftmaken on FSC-certified paper

