Meta-analysis of temporal intestinal gene expression data to generate reference profiles

Ina Hulsegge, Henri Woelders, Annemarie Rebel, Mari Smits, Dirkjan Schokker

WAGENINGEN
UNIVERSITY & RESEARCH

Meta-analysis of temporal intestinal gene expression data to generate reference profiles

VDI-10

Ina Hulsegge[1], Henri Woelders[1], Annemarie Rebel[1], Mari Smits[1,2], Dirkjan Schokker[1]

[1] Wageningen Livestock Research, Dept. Genomics, Wageningen, The Netherlands

[2] Wageningen Bioveterinary Research, Dept. of Infection Biology, Lelystad, The Netherlands

This research was conducted by Wageningen UR Livestock Research, commissioned and funded by the Ministry of Economic Affairs, within the framework of Policy Support Research theme 'Feed4Foodure: Voeding, darmgezondheid, en Immuniteit' (BO-22.04-002-001) TKI-AF12039B.

Wageningen Livestock Research
Wageningen Bioveterinary Research
Wageningen, June 2017

Report 1058

WAGENINGEN
UNIVERSITY & RESEARCH

Ina Hulsegge, Henri Woelders, Annemarie Rebel, Mari Smits, Dirkjan Schokker, 2017. *Meta-analysis of temporal intestinal gene expression data to generate reference profiles.* Wageningen UR (University & Research centre) Livestock Research, Livestock Research Report 1058.

This report can be downloaded for free at https://doi.org/10.18174/426338 or at www.wur.nl/livestock-research (under Wageningen Livestock Research publications).

P.O. Box 338, 6700 AH Wageningen, The Netherlands, T +31 (0)317 48 39 53, E info.livestockresearch@wur.nl, www.wur.nl/livestock-research. Wageningen Livestock Research is part of Wageningen University & Research.

Wageningen Livestock Research Report 1058

Table of contents

Foreword

Feed4Foodure is a public-private partnership between the Dutch Ministry of Economic Affairs, a consortium of various organizations within the animal production chain and Wageningen UR Livestock Research. Feed4Foodure aims to contribute to sustainable and healthy livestock farming in the Netherlands, simultaneously strengthening our competitive position on the global market. The Feed4Foodure program line “Nutrition, Intestinal Health, and Immunity”, aims to contribute to a reduction in the use of antibiotics in livestock farming by increasing general health and disease resistance. The main goals are to develop innovative measuring techniques and to evaluate new feeding concepts, feed ingredients and additives to improve gut health and immunity.

Until now, not much is known with regard to the window of opportunity to modulate immune competence. Nevertheless, it is known that early life conditions are crucial for gut microbiota colonization and immune system development. The current report describes a meta-analysis of intestinal gene expression data in order to identify temporal reference profiles. Such profiles could help to understand the timing of certain biological processes in more detail. With such information it would be possible to modulate early life conditions concerning immune competence more precisely in future experiments. The data for this meta-analysis were acquired from online databases or experiments conducted within the frame work of the Feed4Foodure program line “Nutrition, Intestinal Health, and Immunity”.

For the current study, scientist of Wageningen UR Livestock Research and Central Veterinary Institute worked together with representatives from the various private partners, including Agrifirm, ForFarmers BV, Nutreco, De Heus, Denkavit, van Drie, and Darling Ingredients International. The authors thank the industry partners of the project team for their worthwhile input.

Prof. Dr. Mari Smits, leader Feed4Foodure program line “Nutrition, Intestinal Health, and Immunity”.

Dr. Dirkjan Schokker

Summary

Gene expression is measured because it reflects the status of a certain tissue in the host. Here, we targeted the jejunum, the second part of the small intestine, because of the importance of the small intestine for nutrient uptake and immune surveillance. Immediately after birth, the intestine rapidly develops morphologically, functionally, and immunologically. We acquired different transcriptomic datasets from literature measuring gene expression at different time-points (age) to perform a meta-analysis. In this meta-analysis, genes were grouped into 16 clusters according to their gene expression profiles over time, using data from the control animals of all studies. Within these reference profiles (clusters), we investigated to what extent genes contribute to similar biological processes. From this functional analysis we observed that the different temporal profiles (clusters) had different dominant processes, such as immune related processes, or barrier function. Subsequently, we superimposed transcriptomics data of piglets of the 'dietary intervention groups' of four of the used studies, in which the piglets themselves, or the sow that farrowed them, had been administered either zinc-oxide, amoxicillin, or medium chain fatty acids). In this way we could investigate which temporal profiles (and which biological processes) were modulated by the interventions. Interestingly, not all 16 temporal profiles were modulated. In conclusion, we showed that it is possible to re-use (publicly available) transcriptomics data and produce temporal reference gene expression profiles with overexpression of genes representing specific biological processes. Subsequently, by superimposing gene expression data from (dietary) intervention studies we observed deviations from some of these reference profile(s). The latter suggests that it may be possible to modulate intestinal processes in a beneficial way.

1 Background

1.1 Overall aim of VDI-10 gene expression meta-analysis

A gene-expression time-profile reflects the way in which genes are expressed in a specific place in the body, at a certain time and under the influence of certain (external) factors. This expression-profile determines the functioning of cells / tissue at that site in the body, at that time, and under the influence of given external factors. Based on the gene expression patterns and by using bioinformatics techniques we are able to (better) predict inter- and intra-cellular functions. Changes in gene-expression profiles of intestinal tissue reflect the manner in which intestinal tissue responds to changes in the intestinal lumen (e.g. diet, microbiota), including the manner in which the tissue signals to immune cells in the gut mucosa. Therefore such profiles can help identify important components for a "yardstick" for immune competence.

The primary objective is to acquire a large reference pool of small intestinal gene expression data and subsequently generate reference time-profiles and link these to biological processes. The secondary objective is to investigate to what extent (feed) interventions perturb these reference profiles.

The most suitable combination of gene expression datasets were selected taking into account age, gut location, presumed positive / negative interventions, and nature of the effects (for example, cell proliferation vs. immune regulation). First, a "reference gene expression profile" was determined on the basis of the data of "control" animals. Secondly, this reference profile was compared with the gene expression profiles of the animals with an assumed positive and negative intervention in relation to immune competence.

1.2 Introduction

After birth, a rapid development in the intestine occurs in which the intestine undergoes important morphological, functional, and immunological changes. Morphological changes include larger villi and crypts, this change is mainly due to the availability of feed for the neonate. The entire gut mucosal surface will increase and therefore the absorptive capacity will increase. The latter is necessary to take up nutrients for maintenance and growth of tissues and organs. Functional changes also occur after feed uptake, including pH change in the different intestinal segments, as well as mucin production. Immunological development occurs after birth concomitantly with the microbiota colonization. The interaction between host cells and microbes is necessary for a proper immune/intestinal development. It has already been shown that gnotobionts (germ-free reared animals) have a deprived immune system, and lack a proper response against pathogens. Taken together, after birth the gastro-intestinal tract undergoes many changes, and this period is of importance for immune competence in later life. The window-of-opportunity regarding the development of immune competence is 1) around birth and subsequent first weeks of life and 2) around the process of weaning. These two life events are important aspects of shaping the gut microbiota and the host's immune system.

Within the Feed4Foodure-program many pig studies were carried in which gene expression was measured in the gut mucosal tissue at various time points in control as well as in treated piglets, without an immune challenge. Note, these studies were carried out at different locations, with animal containing a different genetic background, and different feed, all contributing to a large reference pool. Although all these experiments had different objectives, it is still possible and worthwhile to combine these gene expression data. To further improve the power of such a meta-analysis, gene expression studies from literature were incorporated as well. By combining and analysing multiple studies simultaneously, it was possible to generate a 'reference' gene expression profile in time. To further elucidate whether such temporal gene expression profiles (clusters) also reflect biological changes in the gut system, we performed a pathway analysis on the genes underlying each temporal profile. Lastly, we superimposed gene expression data from different intervention studies, which will give more insight in which profiles are possible to modulate and to what extent, without too big adverse effects like death.

2 Material and Methods

2.1 Datasets

Two major public microarray repositories: Gene Expression Omnibus (GEO, National Center for Biotechnology Information; [1, 2]) and *ArrayExpress* (AE, European Bioinformatics Institute; [3, 4]) were searched for experiments with species *Sus scrofa*, and keyword search terms 'jejunum'. The R package *GEOmetadb* (Version: 1.28.0) [5] was used for searching GEO and the package *ArrayExpress* (Version: 1.28.1) [6] for searching *ArrayExpress*. The raw data of the control samples (samples of pigs that have not undergone any treatment) of experiments found by the search query were retrieved from the repositories (see Table 1). Furthermore datasets of 4 experiments from the Wageningen U&R VDI research programme (results not published yet) were used (see Table 1). Besides control datasets, studies 6-9 also provided data from intervention group piglets.(Table2).

Table 1 *List of studies used in this meta-analysis providing data of control piglets*

Study	Accession number	Days	Tissue	platform	Year	Publication
1	GSE13456	0 (piglet was just born), 3, 8, 14, and 21 days postnatal (dpn)	Mid-section of jejunum	Affymetrix Porcine Genome Array	2008	[7]
2	GSE13457	24, 28, 35	Jejunum	Affymetrix Porcine Genome Array	2008	-
3	GSE22596	~ 4 weeks old	Mucosa of the jejunum	Affymetrix Porcine Genome Array	2010	[8]
4	E-MEXP-2198	± 56	Jejunum	Affymetrix Porcine Genome Array	2009	[9]
5	GSE48050	25*	Jejunum	Agilent-026440 Sus scrofa (Pig) Oligo Microarray v2 (Probe Name version)	2013	[10]
6	VDI2	1, 7, 26, 30, 54	Jejunum	Agilent-035953 Sus scrofa Array	2013	
7	VDI5.1	42,51,63	Jejunum	Agilent-035953 Sus scrofa Array	2014	
8	VDI5.2	42,51	Jejunum	Agilent-035953 Sus scrofa Array	2015	
9	VDI12	1,31	Jejunum	Agilent-035953 Sus scrofa Array	2016	
10	VDI 3	2	Jejunum	Agilent-035953 Sus scrofa Array	2014	

* 8 piglets on day 25, 4 of which had been weaned on day 21 and 4 had not been weaned before being sampled.

Table 2. *List of included microarray datasets containing intervention samples*

Study	Accession number	Days	Tissue	platform	Year	Publication
6	VDI2	1,7,26,30,54	Jejunum	Agilent-035953 Sus scrofa Array	2013	
7	VDI5.1	51,63	Jejunum	Agilent-035953 Sus scrofa Array	2014	
8	VDI5.2	42,51	Jejunum	Agilent-035953 Sus scrofa Array	2015	
9	VDI12	1,31	Jejunum	Agilent-035953 Sus scrofa Array	2016	

2.2 Array Quality Check

Quality assessment for the raw data was carried out by using the R package *arrayQualityMetrics* (version 3.18.0)" [11]. Arrays with more than two potential problems (three stars or more in the summary table from the *arrayQualityMetrics* report) were excluded from further analyses.

2.2.1 Affymetrix

For the Affymetrix Porcine Genome Array arrays the '.CEL' file were imported with the usual *cdf* package based on Affymetrix mappings in the R package *affy* (version 1.46.1). Each Affymetrix Porcine Genome Array dataset was background adjusted, normalized and log2 probe-set intensities calculated using the Robust Multichip Averaging (RMA) algorithm in *affy* package [12] [13].

2.2.2 Agilent

The Agilent-026440 *Sus scrofa* (Pig) Oligo Microarray v2 (Probe Name version) and Agilent-035953 *Sus scrofa* Array arrays were background corrected (method= "normexp" and offset = 1) using functions from the R package *Limma* (Version: 3.18.13) [14, 15]. Quantile normalisation of the data was done between arrays. Duplicate probes were averaged using the function *avereps*.

2.3 Annotation

The Affymetrix Porcine Genome Array annotation file: Affymetrix Porcine Annotation, Revision 6 downloaded from http://www4.ncsu.edu/~stsai2/annotation/ (downloaded on 8-2-2016) was used [16]. For Agilent array data, an in-house manually curated revised version of Agilent-035953 *Sus scrofa* Array (GPL18045; http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL18045) was used (curated in-house May 2014). The Porcine gene symbols were converted to human gene symbols using BioMart Gene ID Converter (http://www.biomart.org/) [17], because the annotation of the human genome is more advanced than the pig genome. The human gene symbols were check on http://www.gene-names.org/cgi-bin/symbol_checker (checked on 8-3-2016).

2.4 Data integration

The first step of data set integration was to extract a set of gene names/symbols, based upon the probe sequences, common across the three different platforms (Affymetrix, Agilent-026440, and Agilent-035953). In total, 7,189 genes were common across all platforms and used for further analyses, while those absent in one of the platforms were excluded from further analysis.
By combining different experiments with different platforms, an amount of gene expression information will be lost, i.e. the non-common genes. In order to identify possible loss of functional annotations due to loss of genes for the analysis, we performed a large-scale gene function analysis (Panther (Version 10.0; release date April 25, 2015) [18]). Therefore we compared representation of the extracted genes and all the human genes, respectively, in Gene Ontology (GO) clusters within category 'Biological Process' (Fig. 1). Figure 1 shows that the 7,189 extracted genes had a proper representation of all biological functions.

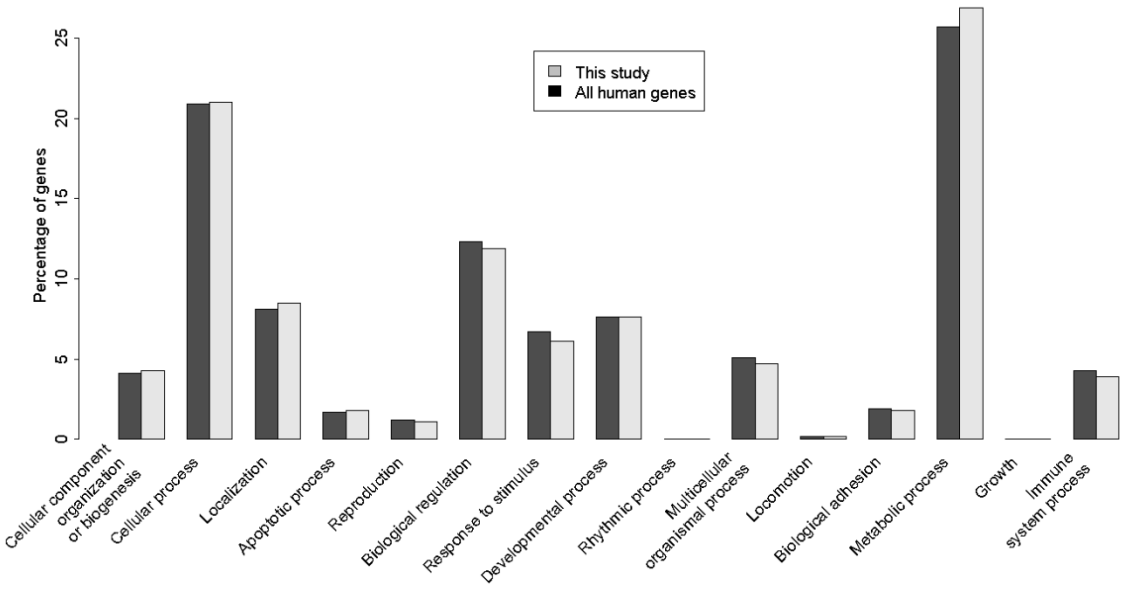

Figure 1. *Functional annotation of the genes in this study and of all the human genes.*

Secondly, for the 'common' genes, the expression values were transferred to a unified scale with the use the functions *normalizeBetweenArrays* and *removeBatchEffect* from the R package *Limma* [15].

2.5 Clustering

Expression values for the time series were soft clustered using a fuzzy c-means algorithm via the R package *Mfuzz* (Version: 2.30.0), which is suggested for microarray time-course data [19]. Since the clustering is performed in Euclidian space (encompasses the *x*, *y*, and *z* planes), the expression values of genes were standardized using the standardise function of R, so that mean expression for each gene is zero with a standard deviation of one. This ensures that vectors of genes with similar changes in expression are close in Euclidean space. Optimal cluster number (c) of 16 was determined by looking for a plateau in the minimum centroid distance using the *Dmin* function (Supplementary Fig. S1). The optimal fuzzifier (m) of 1.130857 was calculated using the *mestimate* function. A membership value between 0 and 1 for each gene gives an indication as to how closely that gene matches the cluster core.

Statistical analysis and visualization of functional profiles for the different gene clusters were carried out with the R package *clusterProfiler* (Version 1.9) [20]. First the human genes symbols in the clusters were converted to Entrez gene identifiers using the function bitr of the package clusterProfiler using Bioconductor annotation packages *org.Hs.eg.db* (Version: 3.2.3). Thereafter enriched functional categories of each gene clusters were calculated using the function *compareCluster* of the package *clusterProfiler*. The function *enrichPathway* from the package *ReactomePA* (Version: 1.14.4) [21] was used in the function *compareCluster* for comparing biological themes in Reactome pathway [22] perspective.

2.6 Time courses

The day of weaning in the various studies varied from 21-28. In order to be able to compare the post-weaning expression data from the various studies, the time axis after weaning was standardized by using day post weaning (denoted by using 'w' before the day number).

3 Results

3.1 Data

Jejunum mucosal gene expression data from ten studies were used, with a total of 103 piglets (is number of arrays), measuring gene expression at 18 different time point (Fig. 2). The number of piglets per time point from 'control' piglets ranged from 1 to 21 for Jejunum, many time-points have at least 5 arrays. Gene expression data from 'intervention group' piglets were available from four of these ten studies, for piglets at days 1, 7, 26, w4, w5,w14, w23, w28 and w35 (Fig 3.). For the intervention piglets, each time point had at least 4 piglets

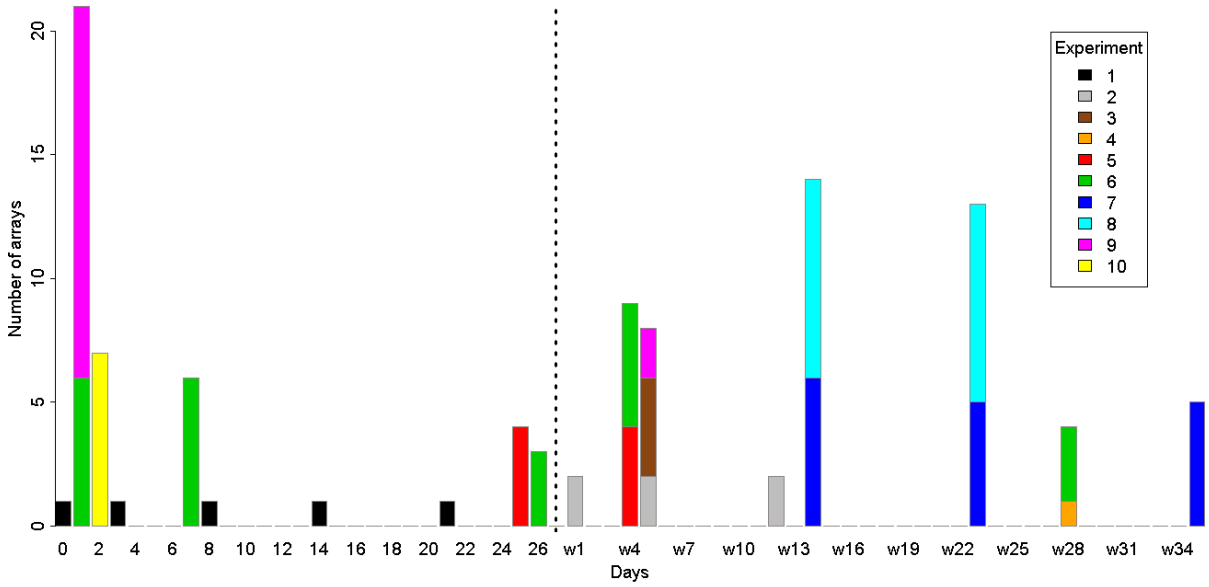

***Figure 2**. Number of piglets per time point for the control samples; colours are the different experiments*

(Black dotted line separates pre-weaning piglets ('true' day numbers 1-26, left) from post weaning piglets (days post weaning, w1- w35, right).

Figure 2 shows an overview of the time points available for the 'control' piglets and the number of piglets per time point. The figure shows an acceptable distribution, gaps between sample days is a maximum of only 9 days, of expression data over time ranging from day zero (day of birth) to day w35 (day 35 after weaning). However, most experiments cover only a few time points and most time-points were from only one experiment; 6 of the 18 time-points had data from two or three experiments.

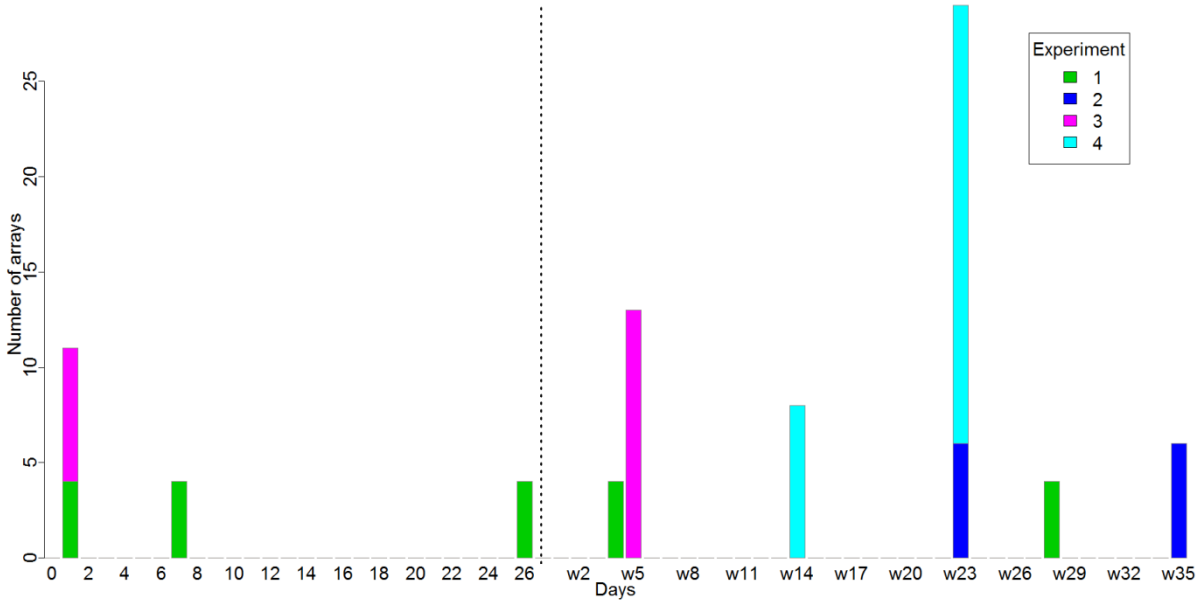

***Figure 3**. Number of piglets per time point for the 'intervention' groups.*

3.2 Clustering

The 7,189 'common' genes (see chapter 2.4) were grouped into 16 clusters, bringing together genes with similar expression time patterns during the whole time series (Fig. 2). The cluster represented by the highest number of genes was cluster 5, which contained 561 genes. The cluster with the lowest number of genes was cluster 15, represented by 337 genes (Table 3). Figure 4 shows the gene expression time patterns of the 16 clusters for individual genes as well as for the average per cluster (black line). Figure 5 emphasizes the average expression pattern for each cluster combined with the information of how many arrays were performed on each time-point.

***Table 3.** Number of genes in each cluster.*

Cluster	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
# genes	450	439	403	560	561	468	539	375	392	531	389	424	432	337	537	352

The result of comparing the gene clusters by their enriched Reactome Pathway are shown in Figure 6. The clusters 1, 3, 4, 6, 7, 8, 9, 13, 14 and 15 did not show enriched Reactome Pathways, with the strict cutoff of p-values < 0.05.

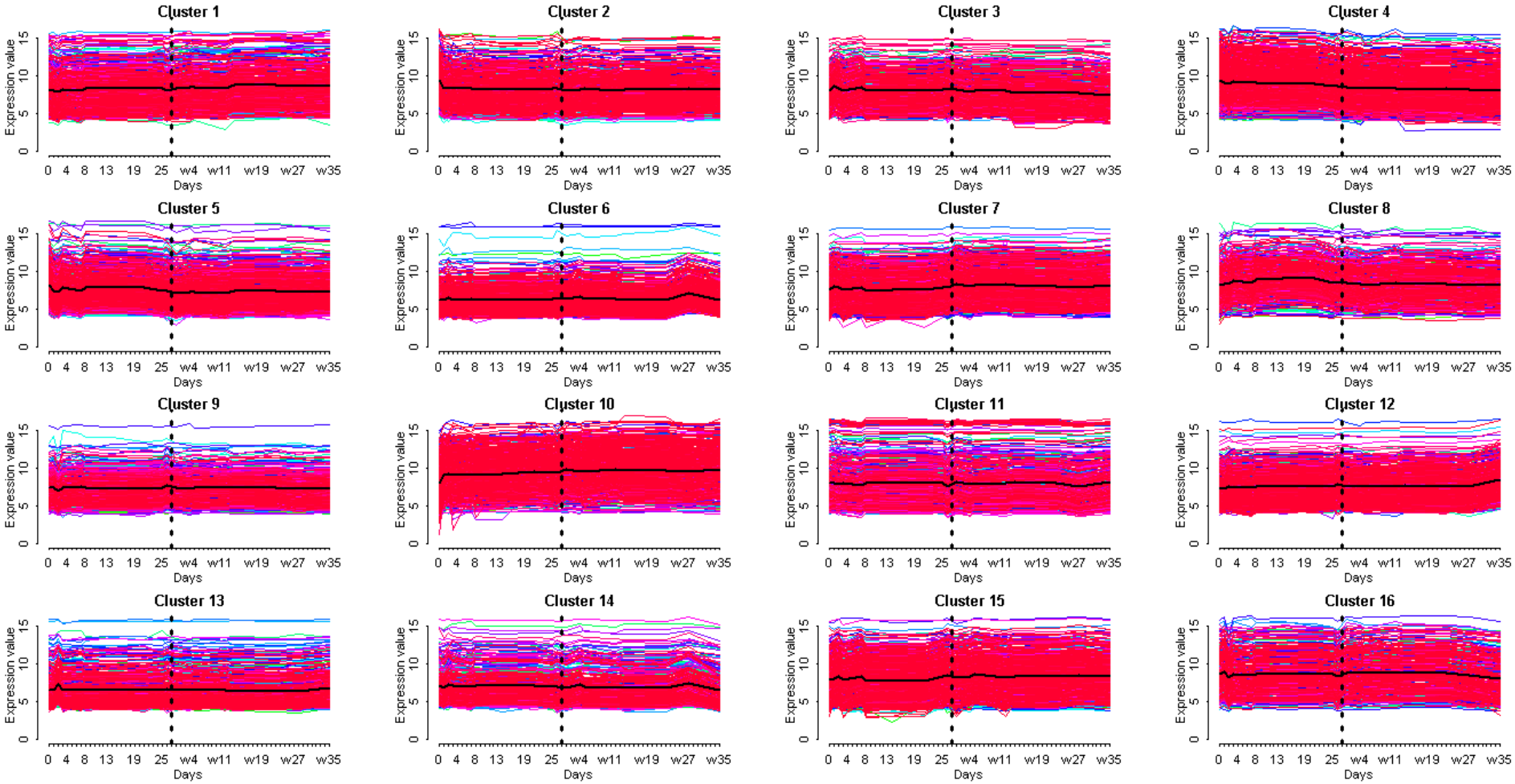

Figure 4. *Clusters of gene expression data of pigs that have not undergone any treatment (control)*
The black solid line is the mean expression pattern of a cluster. Black dot line is time of weaning.
Red and purple lines correspond to genes with a high membership value, whereas yellow or green lines correspond to genes with a low membership value.

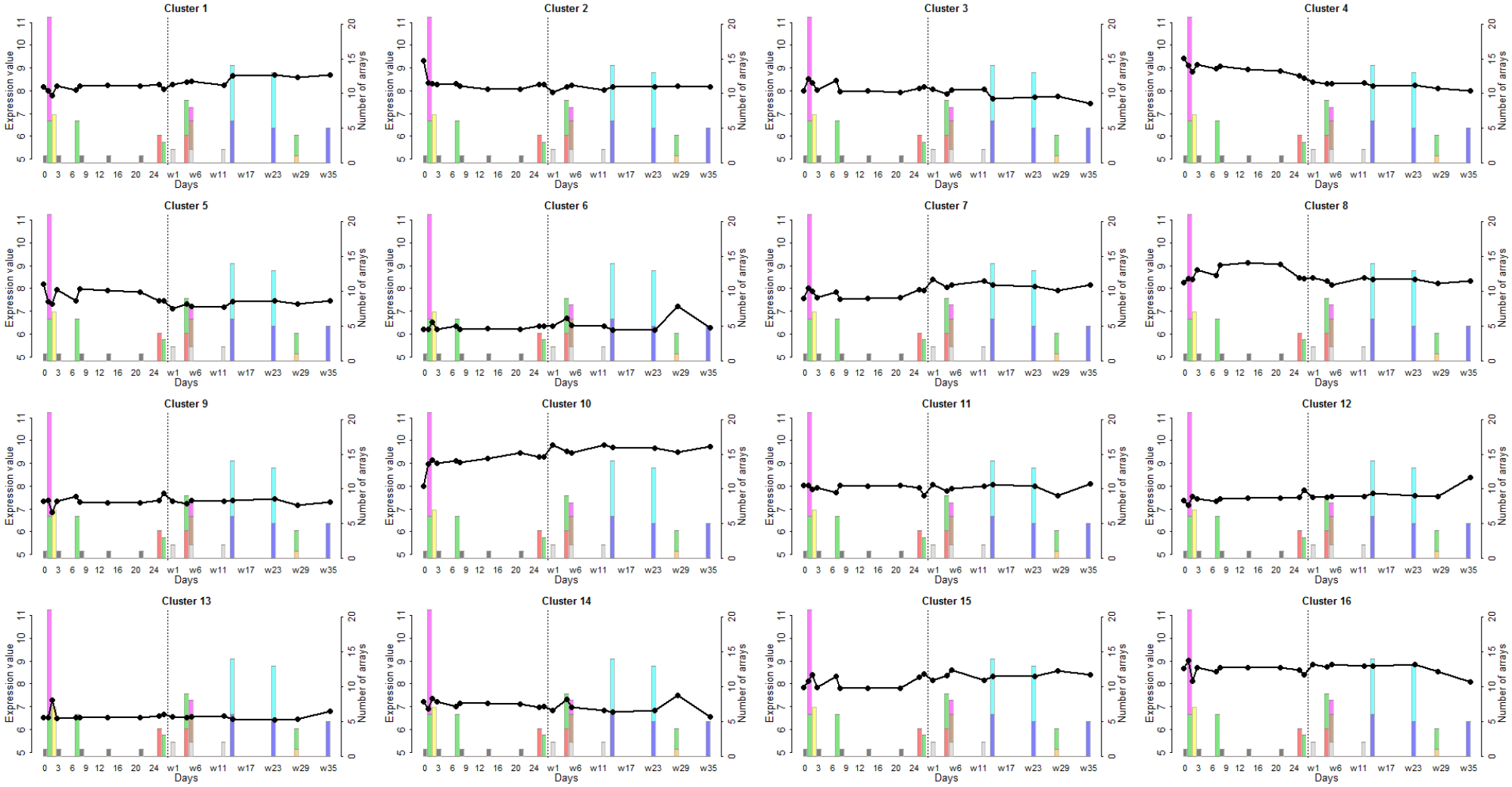

Figure 5. *Sixteen different time-profiles (clusters) based on the average gene expression*
In each graph the x-axis depicts time in days, where 'w' stand for weaning. The y-axis denotes the (normalized) expression value. On the background of each cluster (i.e. time-profile) the number of arrays (piglets) per time point is shown.

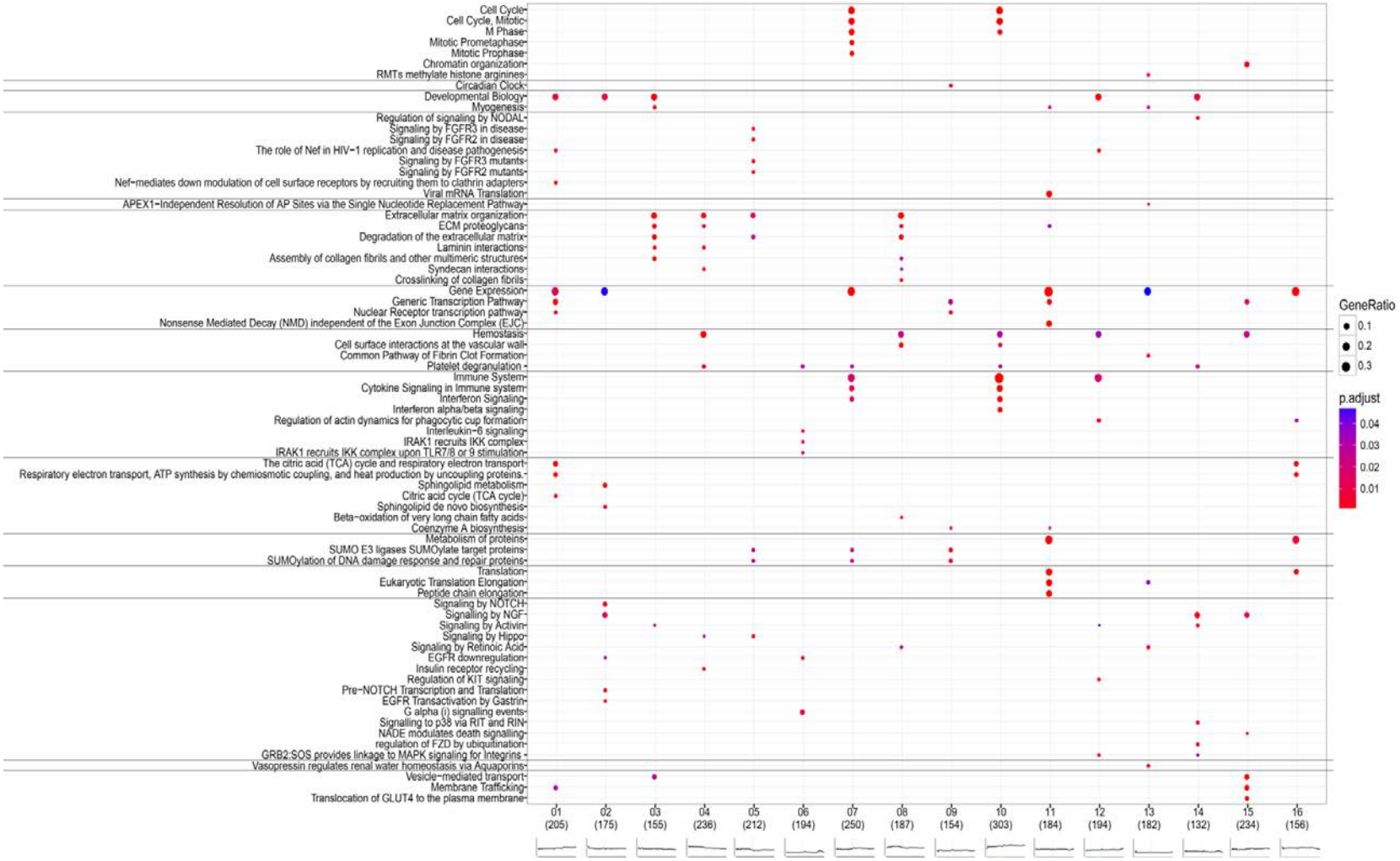

Figure 6. *Pathway enrichment (Reactome) analysis of each gene cluster*
The dot size is proportional to the percentage of genes in the cluster belonging to that particular pathway. Coloured dots correspond to the p-value (< 0.1), where blue corresponds to 0.05 and red to below 0.01.

3.3 Superimposing gene expression date from intervention studies

Within the VDI-programme we conducted four studies in which (feed) interventions were administered to the pigs, i.e. studies 6-9. Study 6 investigated Effect of maternal antibiotic intervention in sows on gut development and microbiota in offspring, study 7 investigated the effects of a high level of dietary zinc on intestinal microbiota and mucosal gene expression in piglets, whereas study 8 investigated the latter over different post weaning periods, and study 9 investigated the effect of neonatal and maternal dietary interventions, here medium-chain fatty acids (MCFAs), on gut health of piglets. The gene expression of these four studies were superimposed onto the reference profiles (Figure 7). Only small deviations are observed from the reference profile. Often these deviations are of similar nature for the different studies, i.e. up- or down-regulation. For example compared to the reference profile all interventions show higher gene expression in cluster 3 and lower expression in clusters 2, 5, 7, 8, 10, 11, and 16. Some clusters show no deviation of gene expression due to the different interventions, like cluster 4, 9, and 13.

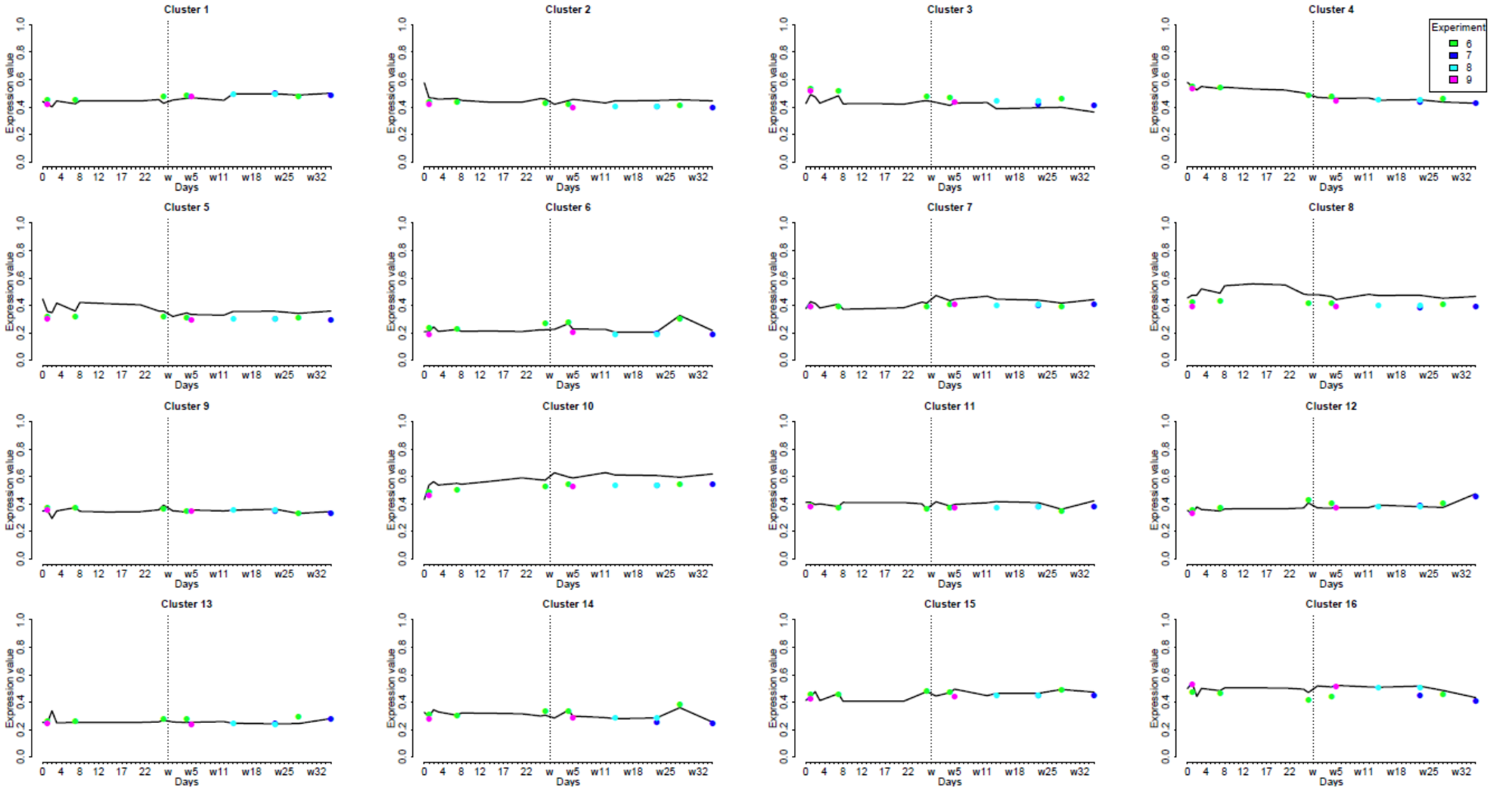

Figure 7. *Intervention gene expression data of four experiments superimposed on references time-profiles*

All 16 clusters showing the reference profiles are depicted. In each graph the x-axis depicts time in days, where 'w' stand for weaning. The y-axis denotes the (normalized and scaled) expression value. The four experiments were 6 (green), 7 (blue), 8 (cyan), and 9 (magenta).

4 Discussion

In this study we have combined 15 datasets with gene expression data of intestinal mucosal tissue (jejunum). Subsequently, we have performed a meta-analysis on all these datasets and identified clusters of genes that have similar expression patterns in time. Such clusters, harbouring many genes were subjected to functional analysis in order to identify which biological processes, represented as pathways, were enriched. This resulted in a few clusters with clear associated function(s), including cell cycle, immunology, or extracellular matrix.

4.1 Current limitations of meta-analysis

Meta-analyses of gene expression studies often focus on static data, comparing two states, e.g. 'control' versus diseased animals, on a single time point. Focussing available resources on a single time point increases the power and allows one to draw (more) unbiased conclusions about specific genes involved in a particular disease [23], but does not capture developments or reactions over time. To our knowledge, the present study is the first meta-analysis aimed at providing temporal expression profiles of intestinal mucosa in pigs. We excluded datasets that provided data on time points beyond 100 days, because we considered the available data too scant to provide enough time points in that upper range of pig ages. In the considered time period of days 0-63 post-natal in the current study, the coverage of time-points was not overwhelming, despite the fact that we collected 10 gene expression datasets for this meta-analysis. Therefore, gene expression levels of specific time points may have been skewed by a single experiment. This also occurred in our results, therefore to show the link between average gene expression time patterns per cluster with the underlying data, we plotted both the average gene expression and the contributing datasets per time-point.

For this meta-analysis we generated a new-analysis-pipeline, by combining different existing tools. The first part of the pipeline mainly focuses on the technicalities with respect to the transcriptomics data, including normalization, batch effects, and gene annotation. These methods and the combination of methods are not new and straightforward in transcriptomic analyses, nevertheless they are necessary for the comparison between and across studies.

The second part of the pipeline consists of different tools to analyse the temporal aspect of the data, including clustering and functional annotation (pathways and Gene Ontology). The temporal aspect is important in relation to the (intestinal) development. The currently used analysis grouped genes expressing similar temporal profiles and subsequently defined the dominant functional process of these groups. By identifying these processes for similar expression profiles more insight was gained towards the timing during (intestinal) development.

4.2 Time-profiles per cluster

Here, we only focus on the clusters that show the most prominent time-dependent profiles in relation to immune competence.

In cluster 10, immune related processes were dominant and the temporal profile showed an initial sharp increase directly after birth, followed by a steady rise in expression level over time. Furthermore, this cluster had an overall 'high' expression level (intensity signal of 9) before and after weaning. This is in agreement with other studies in literature. For instance, in chickens it has been shown that genes involved in intestinal immune development are increasing immediately post-hatch [24]. In addition, this 'high' base level of immune related processes is in line with the fact that one of the gut's most important task is to constantly monitor the environment [25]. Cluster 7 also follows this increase in time to some extent, from birth to weaning a small increase is observed, and after weaning a small sharp increase and subsequent plateauing is observed. This cluster is also involved in immune system processes, as

well as cell cycle related processes. Taken together, these clusters show an increase in gene expression as function of time, reflecting the immune system programming. Because after birth, piglets encounter feed (milk) for the first time and simultaneously microbial colonization occurs. The (sharp) increase around weaning could be due to the fact that solid feed is introduced and that they are separated from their mothers, i.e. less maternal antibodies, thus the piglets' immune system needs to more active, monitoring 'new' antigens/microbes and react appropriately.

Clusters 2, 3, 4, and 8, showed a basal high expression level before weaning and low basal levels of expression after weaning, genes in these clusters were dominantly associated to developmental and morphological processes. This is in line with the fact that immediately after birth, the gut will have its first contact with feed (colostrum/milk) will enter the gut, as well as simultaneously colonization with bacteria, consequently initiating morphological development of the intestine. This includes rapid proliferation of the intestinal tissue, meaning that the cellular structure (i.e. extracellular matrix) will be modified to support this morphological change. It was expected that the gene expression of these processes would decrease in time, because of this maturation process. In conclusion, these 5 clusters are most important to investigate further to possibly develop a 'yardstick' for immune competence. Especially, if it would prove to be possible to modulate these temporal expression patterns with nutritional interventions.

4.3 Superimposing intervention data

In order to investigate to what extent the reference expression profiles could be modulated. We superimposed several datasets from the VDI-programme in which nutritional interventions were administrated to the animals. The following (dietary) interventions were superimposed, Zinc oxide (post-weaning; studies 7 and 8), amoxicillin (maternal-lactation period; study 6), and medium-chain fatty acids (MCFAs; maternal – lactation period; study 9). All of these interventions can be categorized as anti-bacterial and therefor it could explain that the deviation from the reference profile were similar. An observation is that in cluster 10, associated to immune processes, all intervention data-points are below the reference line. The latter could be due to this anti-bacterial effect, fewer bacteria in the gut could result in less inflammation and/or less activity in positive feedback loops for more immune surveillance. Another striking observation was that only higher gene expression values were observed for cluster 3, implicating that all these anti-bacterial interventions dampen the other reference gene expression profiles in the small intestine or the reference profiles can not be modulated.
This meta-analysis has shown that it is possible to re-use (publicly available) transcriptomic data and investigate temporal profiles in light of important biological processes in the gut. Furthermore, by superimposing transcriptomic data from intervention experiments we have shown that certain processes could be modulated, whereas others could not. The next step in future research will be to further interpret these functional changes, mainly with respect to gut health and the implications it could bring with.

Gene expression of clusters 4, 9, and 13, were not modulated due to interventions. Cluster 4 genes are mainly involved in extracellular matrix and cluster 13 to generic gene expression processes. Whereas, genes from cluster 9 are involved in Small Ubiquitin-like Modifier (SUMO) processes. SUMOylation, i.e. modifications of proteins by a SUMO, this modification is important in functions like protein stability, transport from the nucleus to the cytosol, and transcriptional regulation. That these processes were not modulated by the (feed) interventions may suggest that these processes are important for survival of the cell/tissue. The latter may also be reflected in the time-profiles which are relatively stable throughout the time-series.

4.4 Conclusion(s)

Meta-analysis of intestinal gene expression data was technically successful. Subsequent functional analysis of clusters, that harbour genes displaying similar temporal profiles, showed functional processes coupled to the specific temporal profiles. For example, cluster 10, showing a sharp increase in the first days after birth and an overall high basal expression level was coupled to immune related processes.
Another observation was that independent of the (dietary) intervention the accompanying gene expression profiles were unanimously lower, similar, or higher compared to our reference gene expression profile. This similar effect on the reference profiles could be due to the fact that all these interventions exert some form of antimicrobial properties.

References

1. Barrett, T., et al., *NCBI GEO: Mining tens of millions of expression profiles - Database and tools update.* Nucleic Acids Research, 2007. **35**(SUPPL. 1): p. D760-D765.
2. Edgar, R., M. Domrachev, and A.E. Lash, *Gene Expression Omnibus: NCBI gene expression and hybridization array data repository.* Nucleic Acids Research, 2002. **30**(1): p. 207-210.
3. Parkinson, H., et al., *ArrayExpress - A public database of microarray experiments and gene expression profiles.* Nucleic Acids Research, 2007. **35**(SUPPL. 1): p. D747-D750.
4. Parkinson, H., et al., *ArrayExpress - A public repository for microarray gene expression data at the EBI.* Nucleic Acids Research, 2005. **33**(DATABASE ISS.): p. D553-D555.
5. Zhu, Y., et al., *GEOmetadb: Powerful alternative search engine for the Gene Expression Omnibus.* Bioinformatics, 2008. **24**(23): p. 2798-2800.
6. Kauffmann, A., et al., *Importing ArrayExpress datasets into R/Bioconductor.* Bioinformatics, 2009. **25**(16): p. 2092-2094.
7. Sun, Y., et al., *Paradigm of time-sequence development of the intestine of suckling piglets with microarray.* Asian-Australasian Journal of Animal Sciences, 2012. **25**(10): p. 1481-1492.
8. Uribe, J.H., et al., *Transcriptional analysis of porcine intestinal mucosa infected with Salmonella Typhimurium revealed a massive inflammatory response and disruption of bile acid absorption in ileum.* Veterinary Research, 2016. **47**(1).
9. Hulst, M., et al., *Oral administration of Lactobacillus plantarum 299v modulates gene expression in the ileum of pigs: prediction of crosstalk between intestinal immune cells and sub-mucosal adipocytes.* Genes and Nutrition, 2015. **10**(3): p. 1-13.
10. Zhu, L.H., et al., *Gene expression profiling analysis reveals weaning-induced cell cycle arrest and apoptosis in the small intestine of pigs.* Journal of Animal Science, 2014. **92**(3): p. 996-1006.
11. Kauffmann, A., R. Gentleman, and W. Huber, *arrayQualityMetrics--a bioconductor package for quality assessment of microarray data.* Bioinformatics, 2009. **25**(3): p. 415-6.
12. Irizarry, R.A., et al., *Exploration, normalization, and summaries of high density oligonucleotide array probe level data.* Biostatistics (Oxford, England), 2003. **4**(2): p. 249-264.
13. Gautier, L., et al., *Affy - Analysis of Affymetrix GeneChip data at the probe level.* Bioinformatics, 2004. **20**(3): p. 307-315.
14. Smyth, G.K., *Linear models and empirical bayes methods for assessing differential expression in microarray experiments.* Statistical Applications in Genetics and Molecular Biology, 2004. **3**(1).
15. Ritchie, M.E., et al., *Limma powers differential expression analyses for RNA-sequencing and microarray studies.* Nucleic Acids Research, 2015. **43**(7): p. e47.
16. Tsai, S., et al., *Annotation of the Affymetrix1 porcine genome microarray.* Animal Genetics, 2006. **37**(4): p. 423-424.
17. Guberman, J.M., et al., *BioMart Central Portal: An open database network for the biological community.* Database, 2011. **2011**.
18. Mi, H., et al., *PANTHER version 10: expanded protein families and functions, and analysis tools.* Nucleic Acids Research, 2016. **44**(Database issue): p. D336-D342.
19. Kumar, L. and M. E. Futschik, *Mfuzz: A software package for soft clustering of microarray data.* Bioinformation, 2007. **2**(1): p. 5-7.
20. Yu, G., et al., *clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters.* OMICS : a Journal of Integrative Biology, 2012. **16**(5): p. 284-287.
21. Yu, G. and Q.-Y. He, *ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization.* Molecular BioSystems, 2016. **12**(2): p. 477-479.
22. Joshi-Tope, G., et al., *Reactome: a knowledgebase of biological pathways.* Nucleic Acids Res, 2005. **33**(Database issue): p. D428-32.
23. Sorlie, T., et al., *Repeated observation of breast tumor subtypes in independent gene expression data sets.* Proc Natl Acad Sci U S A, 2003. **100**(14): p. 8418-23.
24. Schokker, D., et al., *Gene expression patterns associated with chicken jejunal development.* Dev Comp Immunol, 2009. **33**(11): p. 1156-64.
25. MacDonald, T.T., et al., *Regulation of homeostasis and inflammation in the intestine.* Gastroenterology, 2011. **140**(6): p. 1768-75.

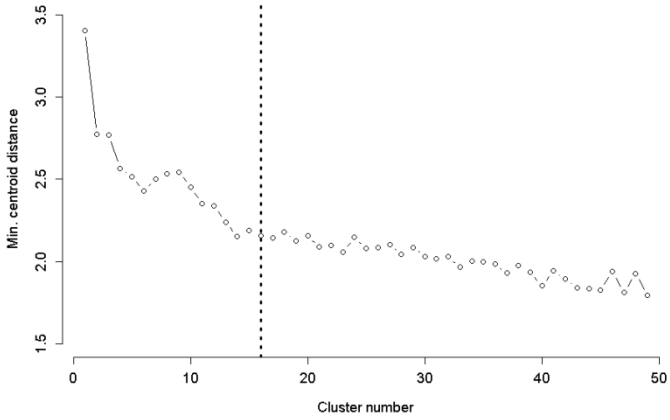

Figure S1. Determining an appropriate cluster number using minimum centroid distance.

Based on this figures the number of clusters was set to 16 (dotted line).

To explore the potential of nature to improve the quality of life

Wageningen Livestock Research
P.O. Box 338
6700 AH Wageningen
The Netherlands
T +31 (0)317 48 39 53
E info.livestockresearch@wur.nl
www.wur.nl/livestock-research

Wageningen Livestock Research creates science based solutions for a sustainable and profitable livestock sector. Together with our clients, we integrate scientific knowledge and practical experience to develop livestock concepts for future generations.

Wageningen Livestock Research is part of Wageningen University & Research. Together we work on the mission: 'To explore the potential of nature to improve the quality of life'. A staff of 6,500 and 10,000 students from over 100 countries are working worldwide in the domain of healthy food and living environment for governments and the business community-at-large. The strength of Wageningen University & Research lies in its ability to join the forces of specialised research institutes and the university. It also lies in the combined efforts of the various fields of natural and social sciences. This union of expertise leads to scientific breakthroughs that can quickly be put into practice and be incorporated into education. This is the Wageningen Approach.