Predictive biomarkers as traits for digestive efficiency in pigs

. Aarhus University, . Wageningen University And Research, . Inra, . Topigs Norsvin

FEED-A-GENE

Adapting the feed, the animal and the feeding techniques to improve the efficiency and sustainability of monogastric livestock production systems

Deliverable D2.6

Predictive biomarkers as traits for digestive efficiency in pigs

Due date of deliverable: M48

Actual submission date: M48

Start date of the project: March 1st, 2015 **Duration:** 60 months

Organisation name of lead contractor: AU

Revision: V1

Dissemination level	
Public - PU	**x**
Confidential, only for members of the consortium (including Commission Services) - CO	
Classified, as referred to in Commission Decision 2001/844/EC - Cl	

Table of contents

1. Summary

Objectives

The aim of Feed-a-Gene is to improve and adapt the different components of monogastric (i.e., pigs, poultry, and rabbits) livestock production systems to enhance their overall efficiency and to reduce their environmental impact. One instrument to reach this overall aim is to develop novel indicators of feed efficiency.

Feed efficiency has two components, relating to the digestive efficiency on the one hand and to the post-absorptive efficiency on the other hand. Therefore, new data referring to digestive efficiency and data referring to the post-absorptive efficiency (and to the overall feed efficiency) are needed to improve feed efficiency in animals. This deliverable aims to report findings on predictive biomarkers as new traits for digestive and overall feed efficiency in pigs.

Rationale:

Large variations in growth responses and nitrogen efficiency have been observed in populations of pigs. This issue has been addressed using different approaches in the Feed-a-Gene project. Differences in piglet birth weight may cause variation in the digestibility of protein and amino acids. Faecal protein digestibility and N-retention as affected by birth weight was evaluated in an experiment performed at DLO (reported in Deliverable D2.4). However, N-balance studies are time-consuming and involve a limited number of animals. Indirect indicators such as biomarkers in blood or urine would be useful for the in vivo evaluation of differences in protein and amino acid metabolism and amino acid requirements. Another approach is to study feed efficiency by selecting for feed efficiency and study differences between the most and the least efficient animals. This is traditionally done by measuring feed intake and body weight gain in individual animals. However, knowledge of differences and changes in the molecular pathways contributing to digestive or post-absorptive nutrient efficiency through the identification of biomarkers would be of value.

Pigs divergently selected for residual feed intake (RFI), also called net efficiency, can be used as an animal model to obtain large variation in traits related to feed efficiency (e.g., conversion ratio (FCR)) and different tissues and fluids can be collected to propose biomarkers of overall feed efficiency. Because of the genetic background of the animals, biomarkers can be used as (early) predictors of this trait in the population to increase genetic progress (e.g., in connection with WP5).

The measurement of small molecules involved in or generated by metabolic processes in tissues and body fluids (e.g., blood and urine) is a feasible tool to identify possible biomarkers for specific metabolic responses and traits related to feed efficiency. Molecules include mRNA produced by gene transcription, proteins produced by mRNA translation, and metabolites. Metabolites measured with untargeted metabolomics can, because of the sensitivity of the technique, detect subtle alterations in biological pathways and provide insight in the mechanisms that underlie different physiological conditions. Liquid chromatography-mass spectrometry (LC-MS) based untargeted metabolomics analysis was applied to i) blood and urine samples from the study relating to N-utilization (i.e., birth weight x dietary protein supply

or genetic capacity to deposit protein x dietary protein supply), ii) to blood samples from a study with genetic lines divergently selected for residual feed intake (RFI) and plasma cortisol concentration, and iii) to blood samples from pigs with feed efficiency extremes.

Taking advantage of large sets of transcriptomics data (mRNA concentrations) that have been acquired in i) muscle, one of the main tissues affected by feed efficiency differences, or ii) blood (a fluid that can be collected at repeated times during the growth period, large sets of molecular predictors of feed efficiency traits, such as RFI, gain-to-feed ratio (overall feed efficiency), and gain-to-energy feed ratio (energy efficiency) were also identified using machine learning methods

The assumption that it is possible to identify molecules in tissues and fluids that are able to predict feed efficiency traits was thus validated. The variety of biological functions represented by the genes and metabolites included in the predictive models confirmed the integrative and complex nature of feed efficiency in growing pigs. The lists of biomarkers are transferred to WP5 as new possible traits to be included in next selection procedures.

Teams involved:

AU
DLO
INRA
Topigs Norsvin

Species and production systems considered:

Growing pigs

2. Introduction

Feed efficiency has traditionally been a very important component of the efficiency of pork production due to the fact that feed represents 60-70% of the total costs of pork production. Feed efficiency has several components, including external components like feed spillage and behaviour, digestive efficiency, maintenance requirements, and post-absorptive energy and nutrient utilization. It has been shown that digestive efficiency has a high potential for improvements through genetic selection (Kyriazakis, 2011, Noblet *et al.*, 2013).

In production farms, feed efficiency in growing animals is generally assessed by its inverse trait, the feed conversion ratio (FCR), calculated as daily feed intake divided by daily growth rate over a defined time period. Residual feed intake (RFI) is a refinement of this trait and captures the efficiency of feed use independent from the production needs, corresponding to the net feed efficiency. The RFI can be computed at the phenotypic or genetic levels as the difference between observed feed intake and feed intake predicted from production and maintenance needs. Divergent selection experiments for RFI have been successful to generate lines with low or high RFI and large differences in FCR. Lines of pigs divergently selected for RFI could be useful animal models to identify small molecules in biological samples, which are indicative for and potentially predictive of animal variation in feed efficiency to improve our understanding of feed efficiency and increase genetic progress.

Measurement of digestive efficiency has traditionally been assessed through nutrient balance trials, which are time-consuming and constraining and may only be performed with a small number of animals. Finding biomarkers for digestive efficiency would allow screening of animals in a herd and would allow for integration of this trait in genetic selection schemes.

3. Description of the research

3.1 Metabolomics analyses on plasma and urine samples to study nitrogen digestibility, retention and nitrogen efficiency

Plasma and urine samples from the study “Effect of birth weight of piglets on the nitrogen digestibility, retention and nitrogen efficiency later in life” were analysed using a non-targeted liquid chromatography-mass spectrometry (LC-MS) based metabolomics approach. The study has been described in detail in deliverable D2.4 of the Feed-a-Gene project and had the goal to investigate differences in nitrogen metabolism between low and high birth weight pigs later in life. In short, the 40 pigs used in this study originated from a three-way cross (i.e., synthetic boar x (Large White x Landrace)). At weaning, litters were split based between high and low birth weight piglets and housed in separate pens. In a change-over design, pigs were either fed a protein-adequate (100% of the requirement) or a protein-restricted (70% of the requirement) diet at 2.8 times maintenance (458 kJ ME/kg$^{0.75}$/d). Plasma samples were collected at beginning of the experiment (at 14 weeks of age; d0), and after the first (d17), and second (d28) experimental period. Urine samples (spot samples) were collected on d16 and d27.

Analysis of the urine samples from Experiment 1 showed that it was possible to obtain separation between animals fed diets with an adequate or restricted protein supply (Figure 1).

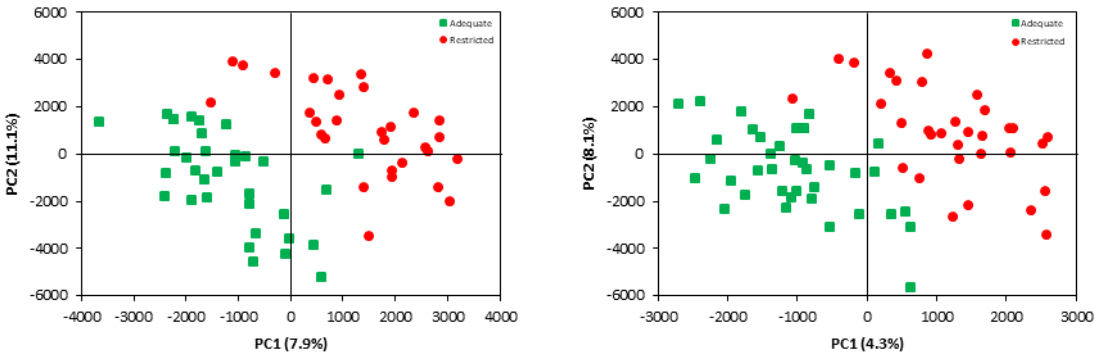

Figure 1. Partial least squares discriminant analysis (PLS-DA) scores plot of urinary metabolites in positive (left panel) and negative (right panel) mode in pigs receiving an adequate or restricted dietary protein supply.

The variance in the composition of urine samples explained by the dietary protein levels was only 7.9% and 4.3% in positive and negative mode, respectively. The very low level of explanation is an indication that factors other than the dietary protein level are important for the variability of the metabolome in urine.

Based on the corresponding loading plots, metabolites important for the discrimination between dietary protein supplies were identified (Table 1). The dietary protein level affected the level of metabolites of the tricarboxylic acid (TCA) cycle with α-ketoglutaric acid, a key molecule in the TCA cycle, playing a fundamental role in determining the overall rate of this important metabolic process, being excreted at a significantly lower level in pigs fed restricted dietary protein. Furthermore, excretion of malic acid was reduced (P=0.001) whereas the levels of citric acid and aconitic acid contributed to the separation of the dietary protein groups but the dietary protein level did not induce differences in the excretion of these metabolites. A major group of metabolites affecting the separation dietary protein levels were metabolites related to amino acid metabolism and metabolites of microbial origin. 3-hydroxy-3-methyl glutaric acid and N-acetylleucine were both excreted at significantly higher levels when feeding pigs an adequate dietary protein level, which was also the case for hydroxyphenyl-lactic acid, a tyrosine metabolite. A significant number of microbial metabolites (i.e., hippuric acid, phenylacetylglycine, and p-cresol glucuronide) contributed to the separation of the dietary protein groups as well. However, the excretion of these metabolites did not differ between dietary protein levels. Glucuronidated compounds comprised a major part of the discriminating metabolites. Glucuronidation is a mammalian chemical detoxification mechanism, that increases the solubility and thereby the excretion of metabolites. The majority of the glucuronide conjugates were identified at level 3, meaning that the compound class was established based on MS/MS spectra characteristics, but it was not possible to identify the original metabolite. It was possible to identify two compounds at level 2 (based on spectral similarities) as daidzein glucuronide and equol glucuronide. Daidzein is an isoflavone found exclusively in soybean and equol is a metabolite of daidzein. The glucuronidated compounds were excreted at higher levels when feeding adequate dietary protein.

In this experiment, it was not possible to obtain separation between urine samples based on birth weight group (results not shown).

Table 1. List of discriminating metabolites identified in urine from pigs fed adequate and restricted dietary protein levels (Exp. 1).

Metabolite	Polarity	RT	m/z	Level of identification[2]	Adequate[1]		Restricted[1]		*P*-values		
					High	Low	High	Low	Diet	BWC[3]	Diet x BWC
Creatinine	Pos	0.72	114.066	1	1.00	0.81	0.97	0.95	0.42	0.12	0.21
Malic acid	Neg	0.82	133.014	1	1.00	0.99	0.40	0.48	0.002	0.76	0.73
α-ketoglutaric acid[4]	Neg	0.85	145.014	1	1.00		0.51		0.01		
Citric acid	Neg	0.89	191.020	1	1.00	0.93	0.84	0.91	0.19	0.99	0.32
Cis-aconitic	Neg	1.12	173.009	2	1.00	0.99	0.91	1.01	0.66	0.60	0.50
3-hydroxy-3-methyl glutaric acid	Neg	1.36	161.046	1	1.00	0.93	0.65	0.66	0.001	0.74	0.64
2-Methyl-guanosine	Pos	2.01	298.115	2	1.00	0.82	0.95	0.80	0.71	0.08	0.90
Hippuric acid sulfate	Neg	2.39	258.008	2	1.00	1.40	0.82	1.21	0.16	0.004	0.95
Hydroxyhippuric acid[5]	Neg	2.64	194.046	2	1.00	0.96	0.96	0.92	0.35	0.35	0.98
Pyridyl acetylglycine	Pos	2.70	195.077	2	1.00	1.10	0.74	0.68	0.002	0.99	0.43
gamma-Glutamylleucine	Pos	2.74	261.145	2	1.00	1.17	0.83	0.84	<.0001	0.13	0.18
Hydroxyhippuric acid[5]	Neg	2.89	194.046	2	1.00	1.28	1.36	1.25	0.44	0.68	0.37
Hydroxyphenyl-lactic acid	Neg	2.93	181.051	1	1.00	0.79	0.73	0.75	0.02	0.17	0.09
2-methylbutyryl-glycine	Neg	3.04	158.082	1	1.00	1.10	0.93	1.03	0.41	0.27	1.00
Riboflavin	Pos	3.49	377.146	1	1.00	1.14	1.73	1.77	<.0001	0.56	0.74
Hippuric acid	Neg	3.55	178.051	1	1.00	0.97	1.02	0.98	0.75	0.46	0.91
Daidzein O-glucuronide[4,6]	Pos	3.82	431.098	2	1.00		0.87		0.05		
N-acetyl-leucine	Neg	3.87	172.098	1	1.00	1.02	0.74	0.82	0.01	0.53	0.67
Unknown	Neg	3.87	307.140	4	1.00	1.10	1.88	0.87	0.32	0.18	0.10
Glucuronide conjugate	Pos	3.88	461.109	3	1.00	0.95	0.81	0.74	0.0004	0.27	0.92
Phenylacetyl-glycine	Neg	3.91	192.067	1	1.00	0.96	1.04	1.01	0.14	0.27	0.78
Phenylacetylglycine	Pos	3.91	194.081	1	1.00	1.05	1.12	1.04	0.13	0.68	0.10
Sulfate conjugate	Neg	3.92	201.113	3	1.00	0.82	0.98	1.08	0.09	0.55	0.05
Unknown	Pos	4.03	251.103	4	1.00	1.13	0.67	0.67	0.004	0.63	0.63
p-cresol glucuronide	Neg	4.05	283.082	2	1.00		1.07		0.08		
Daidzein O-glucuronide[6]	Pos	4.17	431.098	2	1.00	0.98	0.88	0.81	<.0001	0.16	0.44
Daidzein O-glucuronide[6]	Neg	4.18	429.083	2	1.00	0.97	0.84	0.82	0.002	0.55	0.90
Glucuronide conjugated	Pos	4.28	461.109	3	1.00	0.89	0.81	0.71	0.0002	0.03	0.99
Hydroxy-hippuric acid[5]	Neg	4.30	194.046	2	1.00	0.91	0.97	1.04	0.42	0.91	0.18
Equol glucuronide	Neg	4.73	417.119	2	1.00	0.92	0.55	0.69	<.0001	0.70	0.12

Metabolite	Polarity	RT	m/z	Level of identification[2]	Adequate[1]		Restricted[1]		*P*-values		
					High	Low	High	Low	Diet	BWC[3]	Diet x BWC
Glucuronide conjugate	Neg	4.77	297.098	3	1.00	0.91	0.85	0.77	0.01	0.12	0.88
4-Oxododecanedioic acid	Neg	4.86	243.124	2	1.00	0.96	1.09	1.37	0.001	0.12	0.04
Azelaic acid	Neg	4.89	187.098	1	1.00	0.93	0.93	1.08	0.36	0.36	0.01
Glucuronide conjugate[4]	Neg	5.47	387.166	3	1.00		0.89		0.03		
Glucuronide conjugate	Neg	5.63	433.208	3	1.00	1.15	1.10	1.24	0.13	0.03	0.96
Glucuronide conjugate	Neg	5.92	415.197	3	1.00	1.24	0.91	1.03	0.05	0.02	0.39
Glucuronide conjugate	Neg	6.19	415.197	3	1.00	0.83	0.68	0.69	0.005	0.33	0.26

[1]Peak areas are lsmeans normalised to the area of the peak in the high birth weight group fed an adequate protein diet.
[2] Annotated features were classified on different levels of identification as suggested by Sumner *et al*. (2007).
[3]BWC: Birth weight category.
[4]The statistical model was reduced to only contain diet effect.
[5]Three isomeric forms of hydroxyhippuric acid.
[6]Two isomeric forms of daidzein glucuronide.

Plotting the plasma samples in a simple PCA scores plot showed that there was a clear difference between the runs on the LC-MS instrument. The samples were corrected for this effect by the van der Kloet method (van der Kloet *et al.*, 2009; results not shown). Further analysis of plasma samples from Experiment 1 showed that it was possible to obtain separation between animals fed diets with an adequate or restricted protein supply, but not between the birth weight classes (Figure 2).

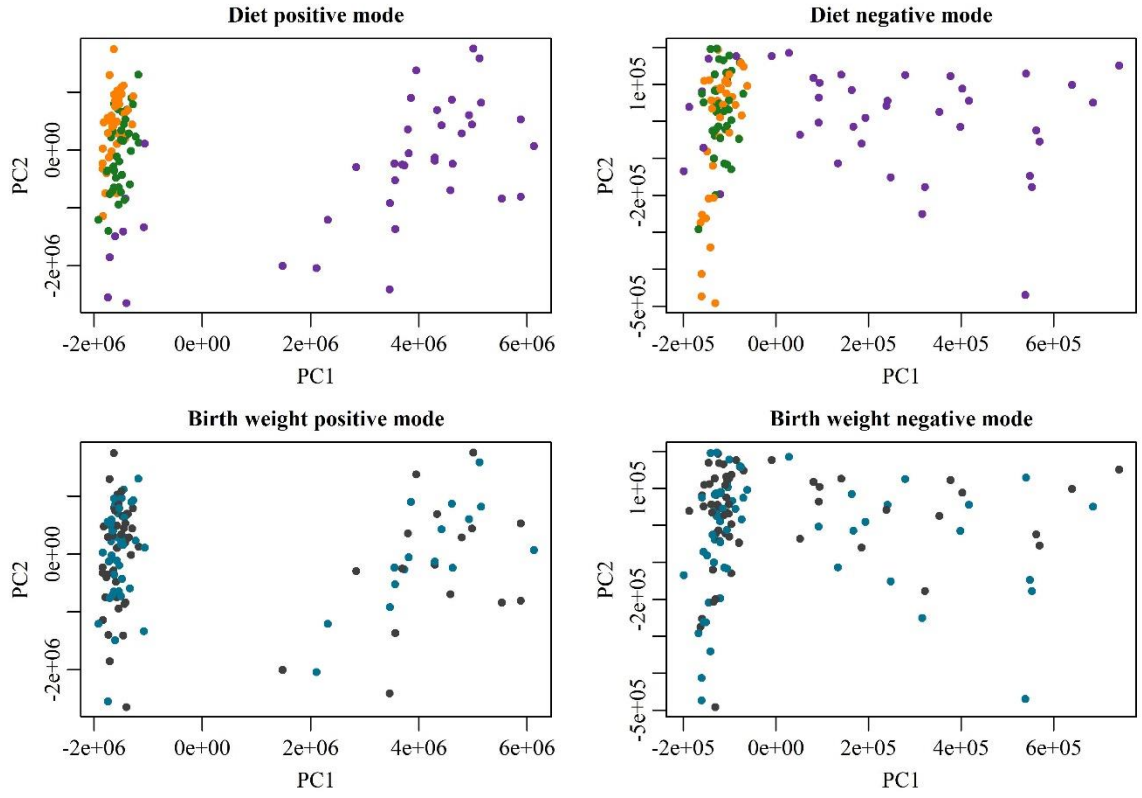

*Figure 2. Principle component analyses scores plot of plasma metabolites in positive (left) and negative (right) mode in pigs with high or low birth weight (bottom) receiving an adequate or restricted dietary protein supply (top). Diet) Green = adequate, Orange = restricted, Purple=start. Birthweight) Light blue = high birthweight, Dark blue = low birthweight. The data were normalised using the van der Kloet method (*van der Kloet *et al.*, 2009*).*

However, when using Discriminant Analysis of Principle Components (DAPC) to investigate the differences in nitrogen metabolism at the grower-finisher age (i.e., between 98-126 days of age), the difference between pigs fed different diets became even more pronounced, including a difference between the birth weight groups. Only two metabolites were sufficient to separate the pigs fed the adequate and restricted protein diet based on the DAPC in both the positive and negative mode. To separate the animals of high or low birth weight, the DAPC needed 43 metabolites in the positive mode and 28 in the negative mode. Also, within birth weight category, it was possible to separate the animals fed the different diets based on the metabolite profile. For the high birth weight pigs, two metabolites were sufficient in both the negative and positive mode. For low birth weight pigs, two metabolites were needed in the positive mode and 3 in the negative mode to separate the animals based on the diets fed. Interestingly, the most important metabolites to separate the pigs based on diet fed were different for the animals of both birth weight categories. When comparing the mass information of the metabolites to the human metabolome database, the most likely metabolic candidates

to discriminate pigs fed a protein adequate or a restricted diet were phenylalaline (both in positive and negative mode), tryptophan, and indoleacrylic acid for the high birth weight pigs. Phenylalaline and trypophan are essential amino acids, whereas indoleacrylic acid is a fragment of trypophan produced by bacterial fermentation. These three metabolites were higher in the plasma of pigs fed the restricted protein diet, indicating a possible amino acid imbalance. PC(18:2)/LysoPC(18:2), piperidine, phenylalanine, and two unknown metabolites were the most likely candidates to separate the two diets for the low birth weight pigs. Piperidine, a fermentation product of colonic protein, and phenylalanine were highest in the plasma of pigs fed the restricted protein diet, again indicating a possible amino acid imbalance. The level of PC(18:2)/LysoPC(18:2) in the plasma samples did not differ between the animals fed the two different diets even though it was the most important metabolite to discriminate the low birth weight pigs fed the two different diets.

A second experiment was performed where both birth weight and genetic capacity to deposit protein was taken into consideration when grouping the animals, otherwise the experimental design was identical to the design of the first experiment. Plasma and urine samples from this experiment were also subjected to non-targeted LC-MS based metabolomics analysis.

Analysis of the urine samples from Experiment 2 showed that it was possible to obtain separation between the dietary protein regimes (i.e., adequate and restricted dietary protein supply) and between the breeding value for protein deposition level (high vs. low) (Figure 4).

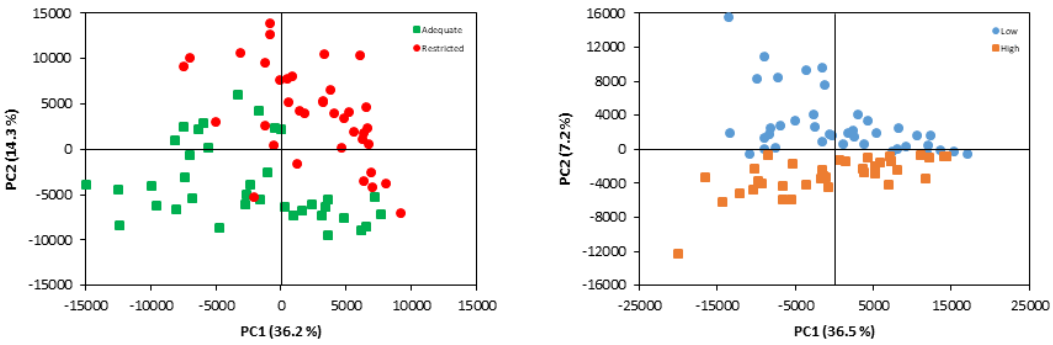

Figure 4. Partial least squares discriminant analysis (PLS-DA) scores plot of urinary metabolites in positive mode. Left panel: separation according to dietary protein level; Right panel: separation according to breeding value for protein deposition.

The separation according to dietary protein supply is both along PC1 and PC2 and the variance added up to 50.5%. In comparison, the separation according to breeding value for protein deposition was only explained by PC2, which accounted for 7.2% of the variance. The corresponding regression coefficients plots are shown in Figure 5 and 6, respectively.

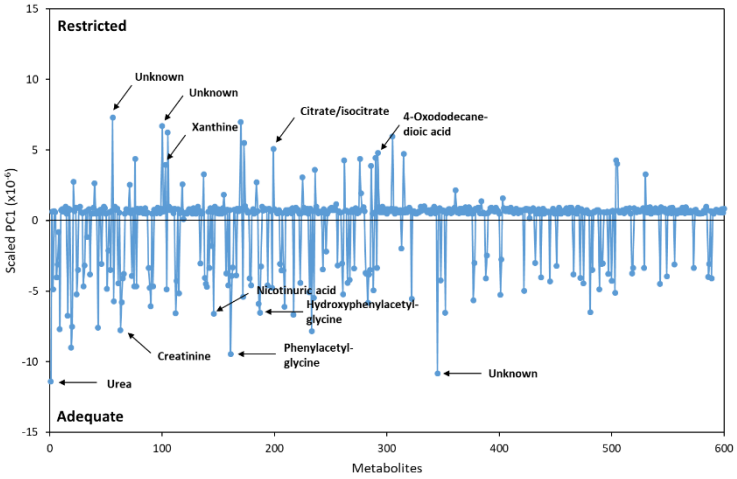

Figure 5. Regression coefficients plot for the comparison of the supply of dietary protein (adequate and restricted).

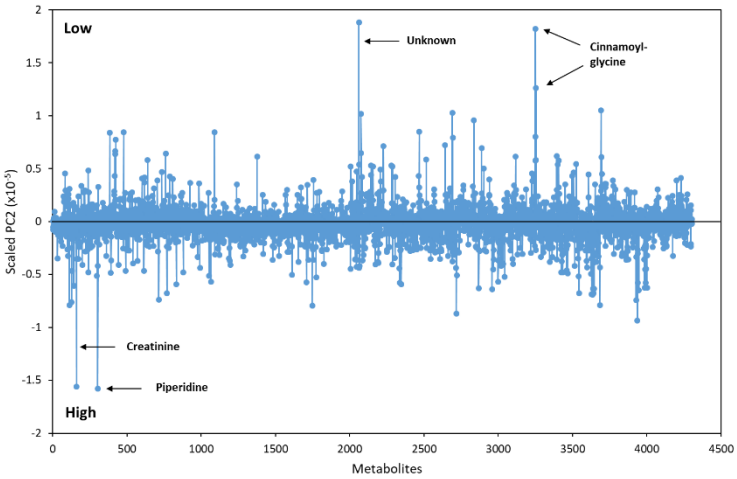

Figure 6. Regression coefficients plot for the comparison of the breeding value for protein deposition levels (high and low).

The tentative identification of metabolites responsible for the discrimination of dietary protein level in experiment 2 showed that the majority of the metabolites were similar to those found in experiment 1. In the group fed adequate dietary protein urea was the metabolite with the highest regression coefficient and metabolites of microbial origin, phenylacetylglycine and hydroxyphenylacetylglycine, had high regression coefficients as well.

The regression coefficients plot for the breeding value for protein deposition showed that two metabolites in each group were the main contributors to the separation of the groups. In the group of the low breeding value for protein deposition, an unknown metabolite (RT 3.00, m/z 160.0970) had the highest regression coefficient. Cinnamoylglycine, representing the second and third highest regression coefficients, is a product of gut bacteria metabolizing phenylalanine. Piperidine and creatinine are the metabolites mainly associated with the high breeding value for protein deposition group. Piperidine is a small molecule resulting from the fermentation by colonic bacteria of non-digestible peptides and protein. Creatinine is directly associated with energy metabolism through the creatine-phosphocreaine shuttle system and creatinine being a discriminating molecule suggests pigs in the high breeding value for protein deposition group to have a higher energy metabolism.

Analysing the untargeted LC-MS metabolomics data of the plasma samples with PLS-DA showed that it was possible to discriminate the plasma samples in negative mode both regarding to breeding value for protein deposition category and dietary protein level (Figure 7).

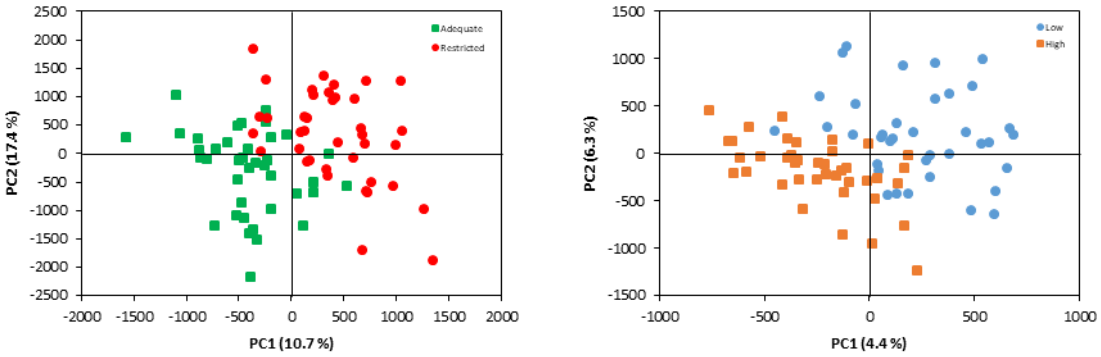

Figure 7. PLS-DA scores plots of plasma metabolites measured in negative mode. Left panel: separation according to dietary protein level; Right panel: separation according to breeding value for protein deposition.

The separation according to dietary protein level is both along PC1 and PC2 and the variance adds up to 28.1%. In comparison, the separation according to breeding value for protein deposition along PC1 and PC2 only accounts for 10.7% of the variance. The level of explained variance is low for both categories and the analysis of the data will be continued to obtain better models to explain the variance and select the metabolites important for the separation.

In positive mode, it was not possible to obtain properly validated PLS-DA scores plots and the work on this part will be continued.

3.2 Metabolomics analysis of plasma from pigs divergently selected for cortisol level or feed efficiency

Plasma samples from the study “Responses of pigs divergently selected for cortisol level or feed efficiency to a challenge diet during growth” performed at INRA in Task 5.1 (see deliverable D5.2 for details) were subjected to non-targeted LC-MS analysis. The plasma samples were collected at 15 and 23 weeks of age, and lines were selected for either divergent residual feed intake (RFI) or divergent plasma cortisol one hour after an ACTH injection.

The plasma samples (361 in total) were analysed in six batches and the first exploratory PCA plot of the samples showed a clear batch effect. This was corrected for by the van der Kloet method (results not shown). Exploratory PCA plots of the corrected dataset showed no grouping according to feed (i.e., control, cereal-based diet versus challenge, high fibre, low energy and low digestible amino acid diet), line (i.e., high RFI (G9+FG), low RFI (G9-FG), high cortisol (G2+FG), low cortisol (G2-FG)), or sampling week (15 or 23). The use of DAPC and PLS-DA showed that it was possible to obtain separation based on the genetic lines. The PLS-DA scores plot is shown in Figure 8, and the metabolites responsible for the separation in the models are shown in Table 2 and 3.

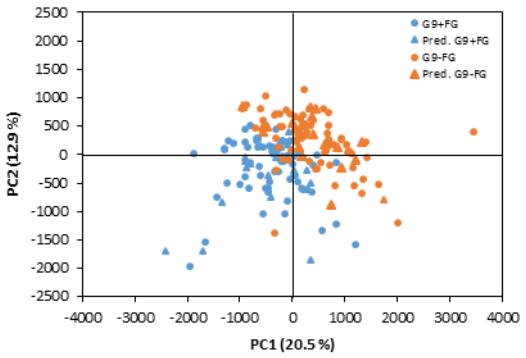

Figure 8. PLS-DA scores plot of plasma metabolites from the genetic lines G9+FG and G9-FG measured in negative mode.

Table 2. The peak area of the peaks discriminating the four genetic lines when analysing data using discriminant analysis of principal components.

Peak[1]	Genetic lines				*P*-value
	G2-FG	G2+FG	G9-FG	G9+FG	
4.74/226.018	420973	317422	223380	171887	<0.001
6.71/448.307	93067	102057	134637	171543	<0.001
0.72/215.033	816965	820429	846333	789546	0.001
3.26/198.033	1236133	1251571	1303988	1301261	<0.001
9.65/540.331	569557	600114	558407	496493	0.025
1.30/267.073	68629	60472	57078	41360	0.019
6.31/464.302	22802	24272	40942	76465	<0.001
3.65/212.002	209280	191942	161597	127065	<0.001
3.64/129.056	587402	548657	570505	530796	0.366
10.96/568.362	309720	297223	262081	267794	0.019
2.85/203.083	462044	437276	447713	355748	<0.001
1.16/180.067	308614	287804	271443	261653	0.001
4.25/187.007	260029	237703	150973	196074	0.001
9.24/564.331	406475	408898	399079	353756	0.108
0.71/92.928	876215	870982	880778	861762	0.553
3.28/129.056	278830	259316	276176	265251	0.405
0.71/268.801	260992	258929	262488	260394	0.957
6.66/465.305	3155753	3166220	3194250	3205678	0.075
7.28/453.286	73005	104782	122720	303518	<0.001
3.57/178.051	724943	647597	783216	678259	0.004
0.72/217.030	276866	277272	286364	266597	<0.001
6.71/449.310	25447	25373	32858	43263	<0.001
6.17/209.082	177982	148292	140061	207963	0.006
3.26/200.030	406380	411673	432423	429259	<0.001
0.71/94.925	607456	605987	613137	600419	0.628
2.71/117.056	66768	61499	81160	75124	0.005

[1]The peak is given as retention time/mz-value.
Metabolites marked in bold are common between Table 2 and Table 3.

Table 3. Regression coefficients of peaks in negative mode responsible for the separation between the genetic lines G9+FG and G9-FG.

Peak[1]	Scaled regression coefficients		Metabolite suggestions[2]
	PC#1	PC#2	
7.28/453.286	-0.0001	-0.0002	Glycerophosphocholine
7.82/437.291	-7.15E-05	-0.00016	No suggestions
6.17/209.082	-6.22E-05	-0.00019	Many suggestions
4.76/225.077	-5.81E-05	-0.00019	Many suggestions
6.71/448.307	-4.11E-05	-0.00011	Bile salt
4.25/187.007	-2.58E-05	-0.00021	p-Cresol sulfate
0.63/316.948	-2.00E-05	-1.55E-05	
5.49/264.088	-1.57E-05	-4.98E-05	
6.66/465.305	-1.41E-05	8.12E-05	Internal standard
1.05/128.035	-2.79E-06	-2.28E-05	A pyrroline
2.10/164.072	2.75E-06	-8.22E-05	Phenylalanine
10.96/568.362	8.52E-06	-9.49E-05	"Fat-like"
0.57/174.956	9.82E-06	4.59E-05	
3.26/198.033	1.21E-05	4.87E-05	Internal standard
9.79/265.148	1.59E-05	4.79E-05	
3.28/129.056	1.85E-05	-6.08E-05	Ketoleucine
10.88/309.174	1.96E-05	5.63E-05	
0.75/160.842	1.97E-05	6.72E-05	
0.72/146.046	3.82E-05	0.000134	Glutamate
3.64/129.056	4.79E-05	-5.02E-05	Ketoleucine
3.65/212.002	5.18E-05	3.43E-05	Indoxylsulfuric acid
9.27/588.330	7.25E-05	0.000154	"Fat-like"
4.74/226.018	7.42E-05	0.000106	No suggestions
3.94/192.067	7.42E-05	8.93E-05	Phenylacetylglycine
0.72/215.033	8.07E-05	0.000194	Many suggestions
10.00/566.346	8.33E-05	0.000145	Glycerophosphoserine
4.09/283.082	9.27E-05	0.000148	p-Cresol glucoronide
3.57/178.051	9.80E-05	0.000167	Hippuric acid
2.85/203.083	0.000106	0.000213	Tryptophan
0.99/89.024	0.00011	2.94E-05	Lactic acid

[1]The peak is given as retentiontime/mz-value.
[2]These are suggestions after search in databases. They need to be confirmed with MSMS and standards (when possible).
Metabolites marked in bold are common between Table 2 and Table 3.

Based on the results obtained using two different analytical approaches and the complexity of the experimental design, it was decided to involve a student in the analysis of the dataset. A student has been enrolled for these analyses, which will be performed during the first semester of 2019.

3.3 Metabolomic analysis of plasma from pigs with extreme feed efficiencies – Topigs Norsvin

To investigate further the use of metabolomics to find biomarkers for feed efficiency, 239 male pigs originating from a three-way cross (i.e., Synthetic boar x (Large White x Landrace)) and housed in pens equipped with individual feed intake registration stations were studied. The pigs were fed *ad libitum* according to a three-phase feeding program, in which all diets were commercial diets based on wheat, barley, and by-products, as typically fed in Europe. For the evaluation of the effect of metabolites on feed efficiency, the 40 pigs with the highest and the 40 pigs with the lowest individual feed efficiency (estimated from FCR) were used. Plasma samples were collected at the day before slaughter. Plasma metabolites were characterised by untargeted liquid chromatography–mass spectrometry (LC-MS), and results were subjected to a discriminant approach combined with principal component analysis to discriminate pigs based on feed efficiency extremes.

In the first measure of differences between the groups (i.e., by a principle component analysis), the high and low feed efficiency pigs did not seem to differ in their metabolite profiles (Figure 9). However, when using a more directed approach (DAPC), there was a clear difference between the high and low feed efficient pigs when measuring the metabolites in negative mode (Figure 10). In total, 16 metabolites were necessary to discriminate the groups. When comparing the mass information of the metabolites to the human metabolome database, the most likely metabolic candidates were, amongst others, L-leucine, phenol sulphate, L- or D-tryptophan, and p-cresol. There was also a difference measurable in the positive mode (Figure 10). Four metabolites were needed for discriminating the high feed efficiency pigs, with the most likely candidates being 2-piperidinone, tetracosahexaenoic acid, L-valine, and an unknown metabolite. The concentration of metabolites related to protein metabolism, amino acids L-leucine, L- or D-tryptophan, and L-valine, and the bacterial end product of protein fermentation (p-cresol) were all higher in the plasma of high feed efficiency pigs. This might indicate that the high feed efficiency pigs were fed a diet that did not fulfil their protein requirement.

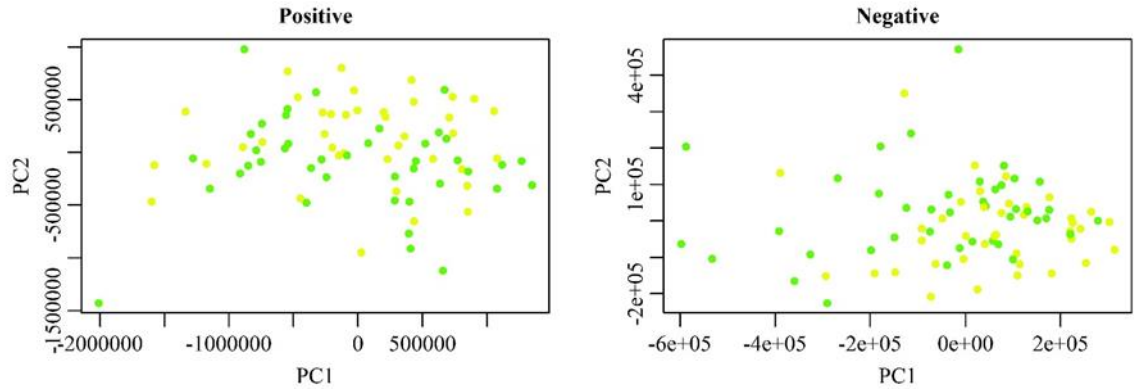

Figure 9. Principle component analyses scores plot of plasma metabolites in positive (left) and negative (right) mode testing differences between feed efficiency extremes in grower-finisher pigs. Yellow=high feed efficiency, Green=low feed efficiency.

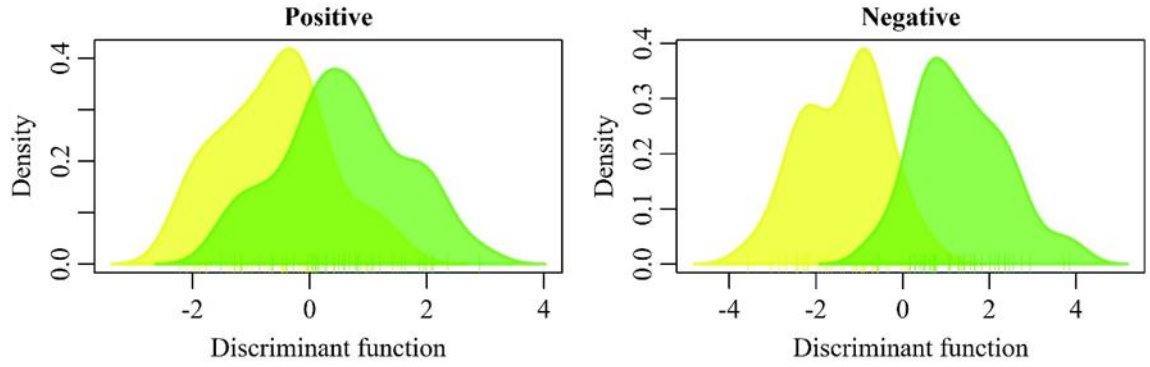

Figure 10. Gaussian kernel density plot of metabolites in positive (left) and negative (right) mode testing differences between feed efficiency extremes in grower-finisher pigs using discriminant analysis of principle components. Yellow=high feed efficiency, Green=low feed efficiency.

3.4 Molecular indicators of feed efficiency as proposed by a meta-analysis of transcriptomics data in tissues and fluids

Small molecules in biological samples, which can potentially predict animal variation in feed efficiency, were identified using different biological samples, analytical methodologies, and statistical approaches. The principle of 3R (refinement, reduction and replacement) was followed because the applied strategies involved not only pigs and biological samples newly collected in the Feed-a-Gene project, but also data available in open international repositories that were reanalysed with new computational strategies (machine learning) to identify important predictors of feed efficiency traits.

To identify molecular predictors of feed efficiency traits such as RFI (net efficiency), gain-to-feed ratio (FCR, overall feed efficiency), and gain-to-energy ratio (FCRe, energy efficiency), large sets of transcriptomics data were used that were acquired in muscle, one of the main tissues affected by differences in feed efficiency. The expression levels of 22,400 annotated molecular probes in loin muscles collected at slaughter were available (GEO subseries accession numbers GSE47769 and GSE84092) on 71 purebred Large White pigs divergently selected for RFI. Across experiments, muscles were sampled in both barrows (n=48) and females (n=23) weighing between 80 and 115 kg body weight. The pigs had free access to pelleted diets of standard composition (n=39) or to diets rich in fibre (n=24). A subset of pigs (n=8) were also feed-restricted (~10% below *ad libitum* intake) during the growing-finishing period. In the merged dataset, pigs were classified into low or high RFI categories according to their lines of origin (n=40 for the high RFI line and n=31 for the low RFI line). The genetic RFI values of these pigs were estimated by considering performance of their littermates in the selection farm. The FCR was calculated from individually measured daily feed intake and average daily gain during the test periods. Net energy-based FCRe was obtained by dividing the daily net energy feed intake by the average daily gain during these periods. The merged molecular data set included 20,405 annotated molecular probes that were expressed in muscle. Molecular data were normalised by mean centering (i.e., subtracting the mean value across all probes from all raw values for each sample) to obtain consolidated expression values. Machine learning procedures were applied on molecular probes for categorical (RFI group) and continuous (RFI value, FCR, and FCRe) traits. The random forest and Gradient TreeNet Boosting machine-learning procedures were used to propose the most important

molecular predictors of these traits. A good prediction (90.5% of success) of RFI categories was obtained by the random forest procedure with 30 molecular probes (out of the 22,406 probes) selected as very important variables to classify the pigs (Table 4).

Table 4. Classification of pigs between low or high RFI groups based on expressing genes studied in pig muscle (categorical values).

Actual class	Total class	Percent correct	Predicted categories	
			High RFI	Low RFI
High RFI	13	84.6%	11	2
Low RFI	8	100.0%	0	8
Total	21			
Average		92.3%		
Overall %Correct		90.5%		

The random forest procedure was applied on a microarray dataset (20,405 expressed annotated probes) in the *longissimus* muscle of 71 pigs divergently selected for RFI, with the aim to predict the separate pigs between low or high RFI classes. A randomly selected bootstrap sample (n = 50 pigs) was used for learning, whereas the remaining sample (n = 21 pigs) was used for validation. The best model (90.5% of good results) was obtained with 30 molecular probes declared as very important in prediction. Good prediction was obtained for 84.6% of the pigs of the high RFI line and 100% of the pigs of the low RFI line, respectively.

These probes corresponded to 20 unique genes, including PDZD2, MUM1, CD40, HLA-A, ST8SIA2, SERPINA1, and HBXIP. Good prediction was obtained for 100% for pigs of the low RFI line and 84.6% of the situations for pigs of the high RFI line, respectively. For continuous traits, the Gradient TreeNet Boosting procedure provided powerful models for the prediction of RFI, FCR, and FCRe. About 50 molecular probes corresponding to 27 unique genes can be considered as the most important predictors for RFI (R^2=0.63, RMSE=42.91). About 88 probes corresponding to 52 unique genes were proposed to predict FCR (R^2= 0.70; RMSE=0.22), and when comparing with the mean FCR in the evaluated pigs, the error of prediction was 8% of the trait. However, the prediction model was again clearly better for low than for high RFI lines (low RFI line: R^2=0.71; RMSEP=0.09; high RFI line: R^2=0.51; RMSEP=0.15). For FCRe, iterative steps led to propose seven probes corresponding to six identified genes as the most important predictors, but the accuracy of the model was lower (R^2=0.52; RMSE=0.002) as compared with the other analysed traits (Table 5).

Table 5. Number of molecular probes and unique genes identified as the most important predictors for feed efficiency traits (continuous values).

	Number of probes	Number of genes	R^2	RMSE
RFI	384	222	0.63	42.91
	280	161	0.64	39.6
	50	27	0.65	39.35
FCR	421	267	0.61	0.23
	88	52	0.70	0.22
	50	33	0.67	0.22
FCRe	318	218	0.49	0.002
	50	29	0.52	0.002
	7	6	0.52	0.002

RFI: residual feed intake; FCR: feed conversion ratio; FCRe: gain-to-energy intake.

Machine learning procedures were applied on a microarray dataset (20,405 expressed annotated probes) in *longissimus* muscle of 71 pigs with the aim to predict RFI, FCR, and FCRe values. A randomly selected bootstrap sample (n = 50 pigs) was used for learning, whereas the remaining (n = 21 pigs) was used for test validation. The best models were obtained with Boostrap procedures to select molecular probes as variables important in prediction. Iterative steps led to model reduction to identify a subset of very important predictors with increased accuracy of the prediction, evaluated by the root mean square error (RMSE) and the coefficient of determination (R^2).

Overall, linear combinations of 24 genes chosen among the different lists of very important predictors and for which expression levels in muscle were further studied by qPCR methodology (i.e., a cheaper and quicker method than microarrays) were significantly related to inter-individual variations in RFI (R^2=0.58), FCR (R^2=0.72) and FCRe (R^2=0.71). Top genes (Table 6) included FKBP5, MUM1, AKAP12, FYN, TMED3, PHKB, TGF, SOCS6, ILR4, and FRAS1, which are known to participate in a variety of biological processes (e.g., immunity, energy metabolism, cell growth, and signal transduction). In addition, FKBP5, SERINC3, IGF2, CSRNP3, EZR, and RPL16 were more intrinsically related to RFI.

Table 6. Most important genes involved in linear prediction of feed efficiency traits when studied by qPCR in loin muscle collected at the slaughterhouse.

RFI value		FCR		FCRe	
Gene name	***P*-value**	**Gene name**	***P*-value**	**Gene name**	***P*-value**
FKBP5	<0.001	FKBP5	<0.001	FKBP5	<0.001
SERINC3	0.02	MUM1	0.03	MUM1	0.04
IGF2	0.03	AKAP12	0.03	AKAP12	0.03
CSRNP3	0.03	FYN	0.03	PHKB	0.08
EZR	0.09	TMED3	0.08	SOCS6	0.07
RPL16	0.08	PHKB	0.08	FYN	0.08
		TGF	0.02	TGF	0.02
		SOCS6	0.07	TMED3	0.09
		ILR4	0.10	ILR4	0.10
		FRAS1	0.12	FRAS1	0.12
R^2 = 0.58		R^2 = 0.72		R^2 = 0.71	

RFI: residual feed intake; FCR: feed conversion ratio; FCRe: gain-to-energy intake.
A stepwise selection was used to retain the most significant variables in the regression model for feed efficiency traits.

Although working on muscle at the slaughterhouse may currently not be considered cost-effective, the results demonstrate that the applied methodological approaches are powerful to identify reliable predictors of feed efficiency traits. Similar approaches in fluids allowing early sampling in living animals have been considered. In blood, we again took advantage of two experiments for which transcriptomics datasets were available on low and high RFI pigs (Campos *et al.*, 2014, Le Floc'h *et al.*, 2016; n=36 pigs). We also performed a transcriptomics analysis in blood aliquots from a newly conducted experiment based on the same RFI lines but including pigs of a later generation of selection (n=79 pigs). In these three experiments, gene expression levels were monitored in tissue samples obtained at an early age (i.e., post-weaning and at the start of the growing period). It was assumed that it would be possible to predict FCR from early obtained transcriptomics data for the subsequent growing-finishing period. A machine-learning procedure was applied on the two first datasets and the selected molecular probes were validated on the last experiment dataset. About 154 molecular probes were identified by the Gradient TreeNet Boosting procedure as very important predictors to classify pigs as low or high RFI groups, and iterative steps can reduce the number of probes to the top 10 most important predictors with excellent performance (97.5% of success). The genes are known to participate to immunity, homeostatic processes, oxidation-reduction reactions, proteolysis, and response to hormone stimuli. For continuous values such as FCR, 766 molecular probes were first selected, and iterative steps reduced the number of molecular probes to 10 most important predictors (Table 7) with excellent performance (R^2=0.82, RMSE=0.195, error of prediction ~8%).

Table 7. Most important 10 expressing genes in blood sampled at early stage and able to predict FCR in next growing periods.

Gene name	Score in predictive model
Serine palmitoyltransferase, small subunit A	100.00
Vacuolar protein sorting 26, transcript Variant 1	83.91
Rho GTPase activating protein 22	80.78
PRP40 pre-mRNA processing factor 40 homolog A Selectin L	78.65
Selectin L	76.43
RNA guanylyltransferase and 5'-phosphatase	75.62
WASH complex subunit FAM21-like	73.90
DDB1 and CUL4 associated factor 6	73.55
Similar to centrosomal protein 290kDa	71.38
Eukaryotic translation initiation factor 2B, subunit 1 alpha, 26kDa	71.10
R^2	**0.58**

4. Conclusions

Studies on the urinary and plasma metabolome of high and low birth weight pigs showed that it was possible to obtain grouping according to dietary protein level in both the urinary and the plasma metabolome, whereas it was only possible to obtain separation between birth weight categories in the plasma metabolome. The urinary metabolites responsible for the separation according to dietary protein level belonged to the TCA-cycle and were amino acids and metabolites of microbial origin, likely absorbed from the gut. Furthermore, a number of glucuronidated metabolites were excreted when feeding adequate dietary protein. These were probably mainly xenobiotics. Using DAPC, it was possible to discriminate both birth weight categories and dietary protein levels in the plasma samples. Discriminating the birth weight categories required many metabolites (43 and 28 metabolites in positive and negative mode, respectively). However, when considering dietary protein supply level within a birth weight category, only two or three metabolites were required for separation of treatment groups. In the second experiment, where the breeding value for protein deposition was taken into consideration, it was possible to discriminate the groups based on breeding value for protein deposition and on dietary protein level in the urine and the plasma metabolome. Metabolites causing separation according to dietary protein level were similar to those identified in the first experiment. In plasma, cinnamoylglycine and an unknown metabolite were associated with a low breeding value for protein deposition. Piperidine and creatinine were associated with a high breeding value for protein deposition. Piperidine, a small metabolite of microbial origin, was associated with the discrimination of diets in the low birth weight piglet in the first experiment whereas it characterised a high breeding value for protein deposition piglets in the second experiment. The possible discrepancy between these findings needs to be investigated further. The results of experiment 1 and 2 showed that the breeding value for protein deposition gave a better separation of the metabolomes than using birth weight categories, which will make the selection of metabolites more reliable and thereby increase the probability of finding valid biomarkers.

Studying the plasma metabolome of pigs selected for cortisol level or feed efficiency showed that it was possible to obtain separation in the plasma metabolome according to feed efficiency line. Several metabolites, which differ between more and less efficient lines, were tentatively identified.

Applying the metabolomics approach to a larger population of pigs (239 animals) showed that it was possible to obtain separation between plasma metabolites based in information on feed efficiency when plasma from the 40 most efficient and the 40 least efficient pigs was analysed. Tentatively identified metabolites responsible for the discrimination were amino acids and microbial metabolites.

Overall, the metabolomics studies have shown that it is possible to identify metabolites associated with overall feed efficiency. The validation of these as predictive biomarkers of feed efficiency is in progress in Task 2.5 and will be continued during the coming months.

The assumption that it is possible to find important expressing genes in tissues and fluids that are able to predict feed efficiency traits was validated, and the variety of biological functions represented by the genes included in predictive models confirmed the integrative and complex nature of feed efficiency in growing pigs. The lists of biomarkers will be transferred to WP5 as possible traits to be included in next selection procedures.

5. Annexes

Presentations of results:

Verschuren L.M.G., Jansman A.J.M., Calus M.P.L., Bergsma R., Hedemann M.S. (2018). Plasma metabolites related to nitrogen efficiency in low and high birth weight pigs. 14th International Symposium Digestive Physiology in Pigs, August 21-24, 2018, Brisbane, Australia.

Gondret F., Koffi B., van Milgen J., Louveau I. (2017). Molecular indicators of feed efficiency as proposed by a meta-analysis of transcriptomics data. In: Book of Abstracts of the 68th Annual Meeting of the European Federation of Animal Science, p. 239. August 28 - September 1, 2017. Tallinn, Estonia. Wageningen Academic Publishers.

Messad F., Louveau I., Renaudeau D., Gondret F. (2018). Variability of feed efficiency in growing pigs: towards predictive biomarkers. Feed-a-Gene plenary meeting, Newcastle, 2018.

References

Campos P.H.R.F., Noblet J., Jaguelin-Peyraud Y., Gilbert H., Mormède P., de Oliveira Donzele R.F.M., Donzele J.L., Renaudeau D. (2014). Thermoregulatory responses during thermal acclimation in pigs divergently selected for residual feed intake. International Journal of Biometeorology 58: 1545-1557.

Kyriazakis I. (2011). Opportunities to improve nutrient efficiency in pigs and poultry through breeding. Animal 5: 821-832.

Le Floc'h N., Merlot E., Gilbert H. (2016). Metabolic response to an inflammatory challenge in pigs divergently selected for residual feed intake. Journal of Animal Science 94: 563-573.

Noblet J., Gilbert H., Jaguelin-Peyraud Y., Lebrun T. (2013). Evidence of genetic variability for digestive efficiency in the growing pig fed a fibrous diet. Animal 7: 1259-1264.

Sumner L.W., Amberg A., Barrett D., Beale M.H., Beger R., Daykin C.A., Fan T.W.M., Fiehn O., Goodacre R., Griffin J.L., Hankemeier T., Hardy N., Harnly J., Higashi R., Kopka J., Lane A.N., Lindon J.C., Marriott P., Nicholls A.W., Reily M.D., Thaden J.J., Viant M.R. (2007). Proposed minimum reporting standards for chemical analysis. Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). Metabolomics 3: 211-221.

van der Kloet F.M., Bobeldijk I., Verheij E.R., Jellema R.H. (2009). Analytical error reduction using single point calibration for accurate and precise metabolomic phenotyping. Journal of Proteome Research 8: 5132-5141.