

OPEN

Integrative analysis of blood and gut microbiota data suggests a non-alcoholic fatty liver disease (NAFLD)-related disorder in French SLA^{dd} minipigs

Marco Moroldo^{1*}, Peris Mumbi Munyaka², Jérôme Lecardonnel¹, Gaëtan Lemonnier¹, Eric Venturi³, Claire Chevalleyre⁴, Isabelle P. Oswald⁵, Jordi Estellé¹ & Claire Rogel-Gaillard¹

Minipigs are a group of small-sized swine lines, which show a broad range of phenotype variation and which often tend to be obese. The SLA^{dd} (DD) minipig line was created by the NIH and selected as homozygous at the SLA locus. It was brought to France more than 30 years ago and maintained inbred ever since. In this report, we characterized the physiological status of a herd of French DD pigs by measuring intermediate phenotypes from blood and faeces and by using Large White (LW) pigs as controls. Three datasets were produced, i.e. complete blood counts (CBCs), microarray-based blood transcriptome, and faecal microbiota obtained by 16S rRNA sequencing. CBCs and expression profiles suggested a non-alcoholic fatty liver disease (NAFLD)-related pathology associated to comorbid cardiac diseases. The characterization of 16S sequencing data was less straightforward, suggesting only a potential weak link to obesity. The integration of the datasets identified several fine-scale associations between CBCs, gene expression, and faecal microbiota composition. NAFLD is a common cause of chronic liver disease in Western countries and is linked to obesity, type 2 diabetes mellitus and cardiac pathologies. Here we show that the French DD herd is potentially affected by this syndrome.

Minipigs are a group of small-sized swine lines, which started to be developed in the 1940s for biomedical research. They present many advantages over conventional pigs, like size, ease of handling and decreased feed requirements¹. Today, a large variety of breeds is available for very different applications. Minipig lines are maintained using specific breeding strategies, which minimize inbreeding while keeping the genetic integrity of the population². Examples of minipigs include the MeLiM line, a model to study an inheritable form of melanoma undergoing spontaneous regression³ and the Yucatan strain, used for research on cardiovascular pathologies and diabetes⁴. Minipig lines homozygous at the major histocompatibility locus (MHC, also known as Swine Leukocyte Antigen or SLA in pigs) were also produced by the NIH, mostly for immunology studies and transplantation programs⁴.

Non-alcoholic fatty liver disease is the most common cause of chronic liver disease in developed countries^{5,6}. It spans a spectrum of pathologies ranging from hepatic steatosis to non-alcoholic steatohepatitis (NASH) and cirrhosis^{6–8}. While NAFLD was traditionally seen as a pathology affecting mainly the liver, it has recently been shown that its burden is not limited to this organ. There is increasing evidence that NAFLD is multisystemic, involving extra-hepatic organs and pathways. It is associated with obesity⁹, type 2 diabetes mellitus, insulin resistance, metabolic syndrome and cardiovascular diseases^{5–7,10–12}. Despite the complexity of this pathology, many NAFLD-related molecular pathways are known^{13–15}, and its development is usually modelled using the “two-hit hypothesis”^{12,13}. The “first hit” corresponds to insulin resistance-mediated fat accumulation in hepatocytes. The “second hit” leads to hepatocyte injury, inflammation and fibrosis and is triggered by oxidative stress and

¹Université Paris Saclay, INRAE, AgroParisTech, GABI, 78350, Jouy-en-Josas, France. ²Department of Agricultural, Food and Nutritional Science, University of Alberta, Edmonton, Canada. ³INRAE, PAO, 37380, Nouzilly, France. ⁴Université de Tours, INRAE, ISP, 37380, Nouzilly, France. ⁵Université de Toulouse, INRAE, ENVT, INP-Purpan, UPS, Toulalim, 31027, Toulouse, France. *email: marco.moroldo@inra.fr

proinflammatory cytokines¹². Subsequently, cytokines interact with their receptors to initiate various signalling cascades^{14–16}. Since NAFLD is linked to cardiac diseases, other processes such as platelet activation are involved¹⁰.

To date, rodents have been the pillar for research on obesity, NAFLD and NASH, but physiological differences with humans slow their use^{15,17}. Alternative models are being pursued, and the pig is acknowledged as a relevant one because of its anatomical, physiological, and metabolic similarities with humans^{15,17–19}. Minipigs show a distinct tendency towards obesity and related pathologies²⁰, and several lines are used for their study, such as Ossabaw, Göttingen, and Bama^{15,17,21}. In the aforementioned papers, the animals were fed specific high-calorie diets, but most minipigs develop obesity even when receiving *ad libitum* standard chow^{1,20,22}. Indeed, the recommended feed intake for minipigs corresponds to 40% *ad libitum* intake²³.

The inbred DD line was created and maintained homozygous at the SLA locus, harbouring the Hp-4.4 haplotype (IPD-MHC database, <https://www.ebi.ac.uk/ipd/mhc/>)²⁴. DD pigs were brought to France more than 30 years ago and maintained inbred thereafter. They do not seem especially susceptible to obesity as the DD pigs imported to Great Britain (Mick Bailey, personal communication), but they have not been characterized for any metabolic disease yet.

Our aim was to study the physiological status of the French herd of DD pigs by comparison to commercial Large White (LW) pigs. We measured three sets of intermediate phenotypes²⁵, i.e. complete blood counts (CBCs), whole blood transcriptome profiles and faecal DNA microbiota characterized by 16S rRNA gene sequencing. Integrative analysis suggested that the DD pigs were potentially affected by a disease of the NAFLD spectrum.

Results

We studied 12 DD and 12 LW contemporary pigs, measuring CBCs, blood transcriptome and faecal microbiota at 60 days of age. CBCs were also measured at four other time-points to provide a time-course that was used to confirm the data obtained at 60 days. The total number of animals used for the analyses ranged from 17 to 24 according to the measures and the time-points (Supplementary Table 1).

The LW and DD pigs differ by CBC parameters. CBCs were measured from blood sampled at 8, 20, 40, 60 and 100 days of age. Because some data were missing, the datasets comprised 23, 16, 24, 20 and 17 samples, respectively (Supplementary Table 1). Eighteen parameters were measured at each time-point (Supplementary Table 2, Supplementary Fig. 1).

Twelve parameters were differentially abundant between DD and LW pigs at least at one time-point. After quality control, five parameters were discarded as non-reliable. The other seven were split in two groups displaying opposite patterns (Supplementary Table 3A). The metrics of the first group showed higher values in DD pigs (Supplementary Table 3A). They were related to viscosity and platelet activation, and included hematocrit (hct), mean corpuscular volume (mcv), mean corpuscular hemoglobin (mch), hemoglobin (hgb), and platelet distribution width (pwd).

The hct and the derived metric mcv were significantly higher in DD pigs over the whole time-course of the experiment. In the case of mch, hgb and pwd, the minipigs also presented higher values during all the period, but some time-points lacked statistical support (Supplementary Table 3A).

The second group of differentially abundant CBC parameters was linked to inflammation and included metrics related to white blood cells. These parameters tended to be lower in DD pigs (Supplementary Table 3A). Monocyte absolute number (monum) and monocyte percentage (mopro) were decreased in DD pigs, but with some inconsistencies and low statistical support.

A principal component analysis (PCA) separated the two breeds, with the first component accounting for 47.9% of the total variability (Fig. 1A). Hierarchical clustering, instead, did not found a clear split (Supplementary Fig. 2A).

Blood transcriptomes of DD and LW pigs revealed a potential NAFLD syndrome in DD pigs. Blood transcriptome profiles were obtained from 60-day-old animals. After quality assessment, 21 samples were retained (Supplementary Table 1). A total number of 3,046 differentially expressed (DE) genes were identified (Supplementary Table 4), of which 1,405 were upregulated and 1,641 downregulated in DD pigs. The values of log₂ fold change (FC) ranged from 4.1 to –10.7, but were generally low. In fact, only 38 genes presented a log₂ FC higher than 1.5 or lower than –1.5. The 20 most upregulated and downregulated genes in DD pigs compared to LW pigs are listed in Table 1.

Several exploratory approaches were used to gather information about the structure of the dataset. A PCA was carried out using the expression values of all the genes. The first component accounted for 30.2% of the total variability and separated the two breeds (Fig. 1B). A clustering analysis confirmed this split, but three samples were misplaced (Supplementary Fig. 2B).

The functional analysis of the expression profiles was performed using several methods. First, the DE genes were processed using Ingenuity Pathway Analysis (IPA). Forty-six enriched pathways were found, 22 of which (or 47.8%) were related to NAFLD (Supplementary Table 5), like for example “adipogenesis”, “sirtuin signalling”, “mTOR signalling”, “PI3K/Akt signalling”, “B Cell receptor signalling”, and “hepatic fibrosis”^{13,14,26,27}.

A second analysis was performed with ClueGO 2.3.4²⁸. Fifty connected terms were found, 48 of which presented a group adjusted p-value < 10^{–10} (Fig. 2), and 13 of which (or 27.1%) fitted into the NAFLD spectrum, like “non-alcoholic fatty liver disease”, “actin cytoskeleton”, “PI3K/AKT signalling”, “natural killer cell-mediated cytotoxicity”, and “leukocyte transendothelial migration” (Supplementary Table 6).

As a third step, a Gene Set Enrichment Analysis (GSEA)²⁹ was carried out. Eleven MSigDB gene sets³⁰ were enriched at a FDR q-value < 0.05, 4 of which (or 36.4%) were related to NAFLD (Supplementary Table 7). The results were globally consistent with those of IPA and ClueGO, as in the case of the signatures “insulin resistance (pancreas β cells)” and “mTOR signalling”.

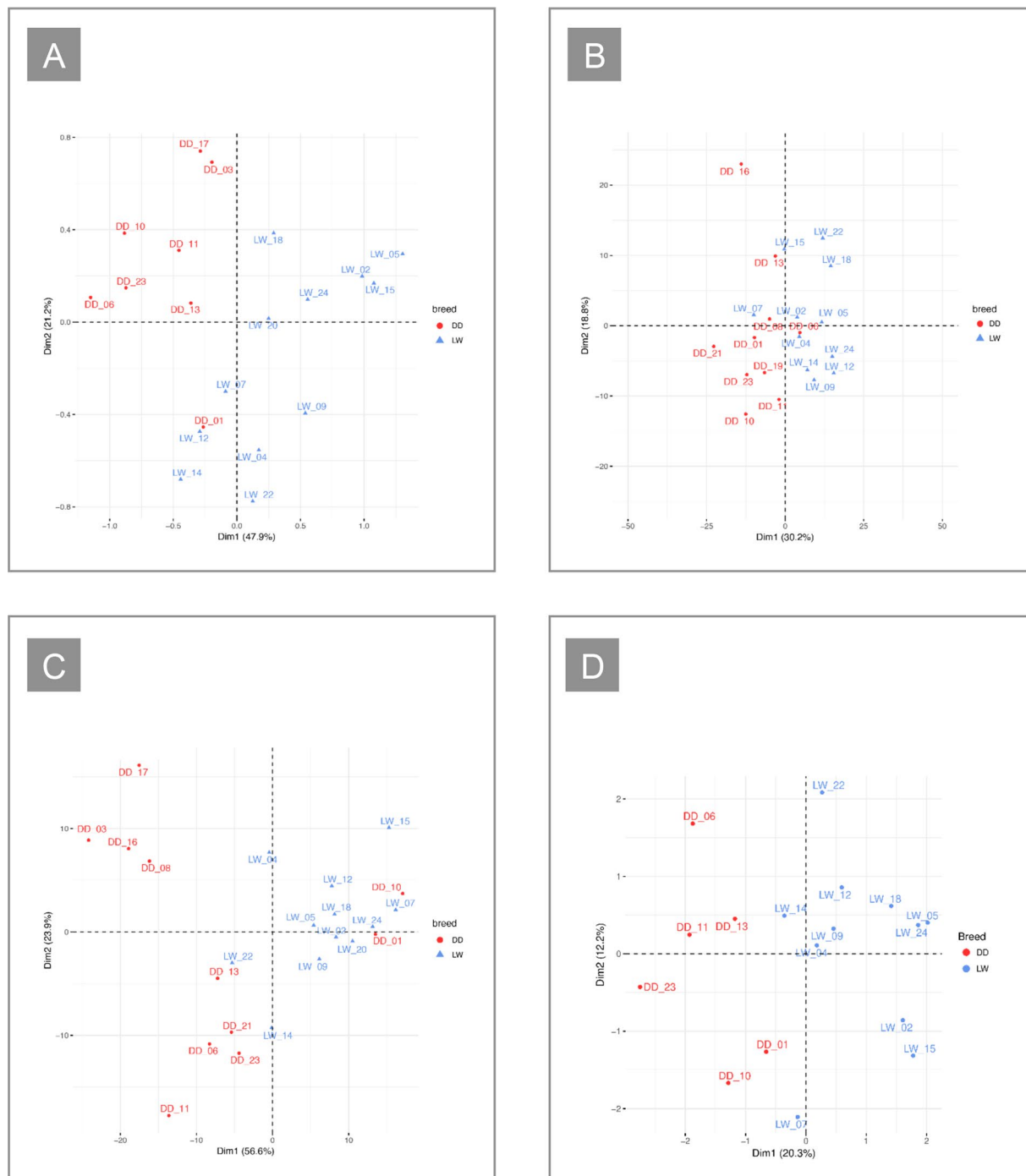


Figure 1. (A) Plot of the first two components of the PCA based on the CBC data. (B) Plot of the first two components of the PCA based on the expression values of all the genes expressed in the blood. (C) Plot of the first two components of the PCA based on the faecal microbiota data. (D) Plot of the first two components of the IFM obtained from the MFA analysis performed on CBCs, genes, and the microbiota data.

The pathways related to NAFLD were evenly distributed across all the steps of the two-hit model and were often confirmed by two or more methods of analysis.

Processes such as those related to insulin resistance (ClueGO, GSEA) corresponded to the first hit, while pathways such as “oxidative phosphorylation” (IPA, ClueGO, GSEA) and “cytokine-cytokine receptor interaction” (ClueGO) represented the mechanisms leading to the second hit¹². As found by others^{13,14,26}, a significant proportion of processes (i.e. 43.5% on average) was related to signalling mechanisms, such as “mTOR signalling” (IPA, GSEA) and “ERK/MAPK signalling” (IPA, ClueGO).

| Gene | Gene description | log ₂ (FC) | BH adjusted p-value |
|--------------|--|-----------------------|---------------------|
| IFITM1 | Interferon induced transmembrane protein 1 | 4.11 | 0.001878 |
| REXO2 | RNA exonuclease 2 homolog | 3.96 | 0.000044 |
| COL16A1 | Collagen type 16 alpha 1 chain | 2.87 | 0.000001 |
| PCP4 | Purkinje cell protein 4 | 2.58 | 0.000191 |
| GSG1 | Germ cell associated 1 | 2.57 | 0.000303 |
| TOR3A | Torsin family 3 member A | 2.45 | 0.000165 |
| LGALS7 | Lectin galactoside-binding soluble 7 | 2.29 | 0.000616 |
| SPINK4 | Serine peptidase inhibitor Kazal type 4 | 2.25 | 0.000001 |
| LY6D | Lymphocyte antigen 6 family member D | 2.15 | 0.000229 |
| SYT9 | Synaptotagmin 9 | 2.11 | 0.000314 |
| HEATR4 | HEAT repeat containing 4 | 2.11 | 0.000473 |
| KAZALD1 | Kazal-type serine peptidase inhibitor domain 1 | 2.07 | 0.000025 |
| OAS1 | 2'-5'-oligoadenylate synthetase 1 | 1.87 | 0.013548 |
| CXCL10 | Chemokine (C-X-C motif) ligand 10 | 1.86 | 0.017611 |
| BCL2L15 | BCL2 like 15 | 1.76 | 0.001981 |
| RSPH9 | Radial spoke head 9 homolog | 1.68 | 0.000935 |
| CREB3L3 | CAMP responsive element binding protein 3 like 3 | 1.56 | 0.000325 |
| DNAH9 | Dynein axonemal heavy chain 9 | 1.51 | 0.000094 |
| SRPK3 | SRSF protein kinase 3 | 1.50 | 0.000033 |
| LOC106505804 | Uncharacterized protein | 1.48 | 0.000128 |
| LOC100515340 | Uncharacterized protein | -1.51 | 0.030272 |
| VPREB1 | V-set pre-B cell surrogate light chain 1 | -1.52 | 0.000723 |
| LOC100510923 | Uncharacterized protein | -1.59 | 0.000180 |
| HPGD | 15-hydroxyprostaglandin dehydrogenase | -1.66 | 0.000153 |
| RGS18 | Regulator of G protein signaling 18 | -1.68 | 0.000723 |
| CRISPLD2 | Cysteine rich secretory protein LCCL domain containing 2 | -1.69 | 0.000002 |
| F2R | Coagulation factor II thrombin receptor | -1.76 | 0.000036 |
| ATOX1 | Antioxidant 1 copper chaperone | -1.85 | 0.006059 |
| FAM151B | Family with sequence similarity 151 member B | -1.89 | 0.000000 |
| ZSCAN25 | Zinc finger and SCAN domain containing 25 | -1.91 | 0.000000 |
| KCNQ3 | Potassium voltage-gated channel subfamily Q member 3 | -1.95 | 0.000005 |
| OAZ3 | Ornithine decarboxylase antizyme 3 | -1.98 | 0.001265 |
| MRPL42 | Mitochondrial ribosomal protein L42 | -2.09 | 0.000000 |
| TAGLN | Transgelin | -2.11 | 0.008512 |
| LAMB3 | Laminin subunit beta 3 | -2.15 | 0.000010 |
| TMEM98 | Transmembrane protein 98 | -2.27 | 0.000263 |
| TRDV3 | T cell receptor delta variable 3 | -2.32 | 0.000000 |
| TCRA | T cell receptor alpha locus | -2.34 | 0.000000 |
| ECHDC1 | Ethylmalonyl-CoA decarboxylase 1 | -3.46 | 0.000000 |
| IFITM3 | Interferon induced transmembrane protein 3 | -10.69 | 0.000000 |

Table 1. Table listing the 20 highest and the 20 lowest differentially expressed genes in whole blood in DD pigs respect to LW pigs. The second column shows gene descriptions, the third one the log₂ FC values and the fourth one the Benjamini-Hochberg-adjusted p-values.

The second hit was represented by pathways such as “hepatic fibrosis” (IPA) and “integrin signalling” (IPA), while the ontologies related to inflammation and to the immune response in obesity, like “leukocyte migration” (IPA, ClueGO), “phagocytosis” (IPA, ClueGO), and “cell adhesion” (IPA, ClueGO) were particularly abundant.

Because these three steps of analysis suggested the potential presence of an NAFLD-related pathology, we decided to explore the genes directly related to this disorder in more detail. For this, a literature-based meta-analytic approach was chosen. First, a non-redundant consensus list of 2,551 NAFLD-related genes was created (see methods). Among them, 1,679 were represented on the microarray and 386 were differentially expressed (Supplementary Table 4). The enrichment within the DE genes list was assessed using a Fisher’s exact test and a hypergeometric test, both of which were significant (respectively 0.0195 and 0.0104).

DD and LW pigs differ by their faecal microbiota. The V3-V4 16S rRNA region was sequenced on 24 60-day-old animals. After quality checks, one low quality sample with less than 300 reads was discarded. Thus, 23 samples were retained for the analysis (Supplementary Table 1), with an average number of 2,180 reads per

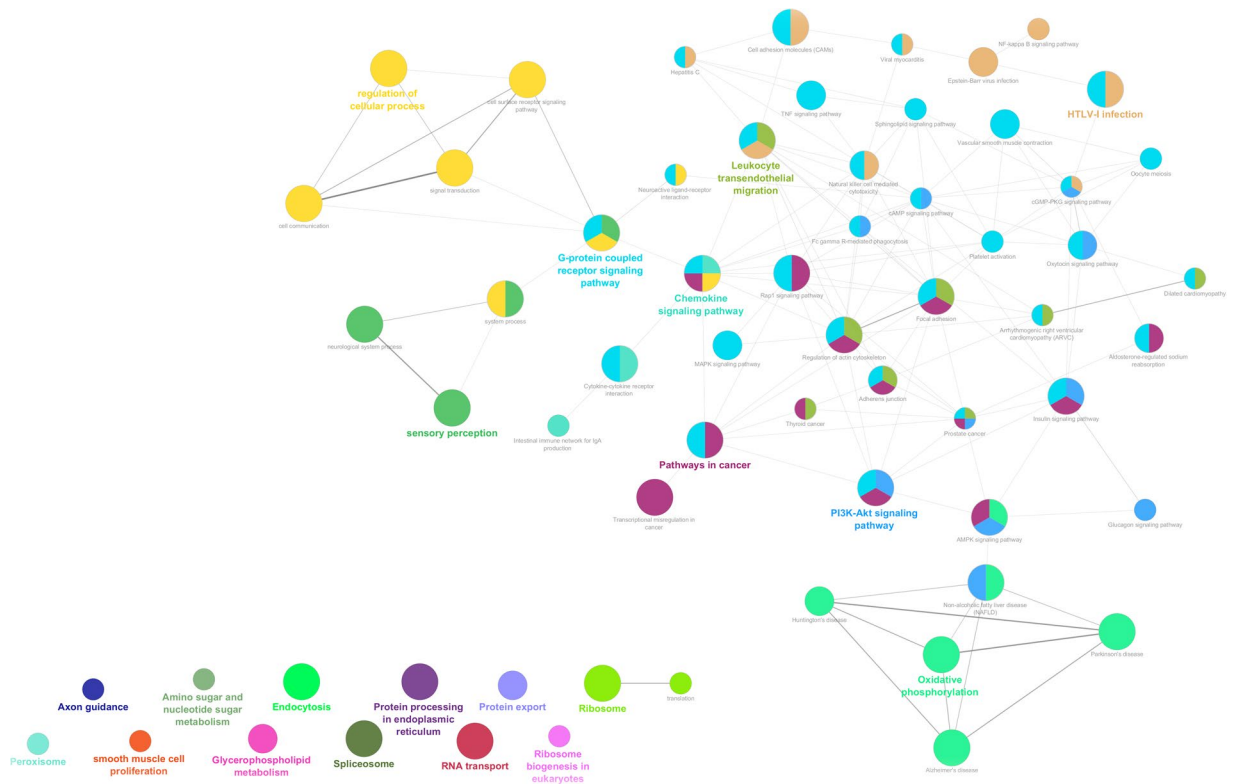


Figure 2. Representation of the functionally connected networks obtained using ClueGO 2.3.4²⁸. Each node corresponds to an enriched GO term, and each colour corresponds to a GO group. The same term can be included in several groups, and its size reflects its statistical significance (refer to Supplementary Table 6). The 13 nodes at the bottom left of the picture are not connected to the network. Line width is proportional to k-score values.

sample and a range of variation from 1,359 to 3,088 reads per sample. The sequences retained after pre-processing were used for the open reference-based operational taxonomic units (OTUs) picking using the Greengenes database (v. 13.8). A total of 1,224 OTUs and 49 genera were identified.

The values of α -diversity (OTU richness and Shannon diversity) did not differ between the two breeds, while β -diversity analyses suggested that taxonomic abundance profiles were significantly different between the groups, with p-values of 0.001 for both the PERMANOVA and the ANOSIM analyses (Supplementary Fig. 3). In agreement with the literature^{22,31}, more than 95% of the sequences in both breeds were represented by the phyla Firmicutes and Bacteroidetes (Fig. 3A), followed by Proteobacteria, Actinobacteria and Spirochaetes. Firmicutes were significantly more abundant in DD minipigs, while Bacteroidetes and Proteobacteria were significantly less abundant (Supplementary Table 8). The two most abundant genera in both breeds were *Prevotella* and *Lactobacillus* (Fig. 3B). *Peptococcus*, *Lactobacillus* and *Ruminococcus* were significantly more abundant in DD pigs, while *Anaerovibrio*, *Prevotella*, *Succinivibrio*, *Mitsuokella* and *Roseburia* were significantly less abundant (Supplementary Table 9).

To further characterize the composition of DD pigs microbiota, we compared our results to three available datasets, the first one obtained on NASH-affected Ossabaw minipigs³² and the other two obtained on obese Göttingen and obese Ossabaw minipigs, respectively²².

In the first comparison (i.e. the contrast between DD and LW pigs *versus* the contrast between NASH-affected Ossabaw and lean Ossabaw pigs), 14 taxa were common to the two studies, seven of which were significantly differentially abundant in both cases, and three of which showed the same direction of variation (Supplementary Information 2). This was the case with Firmicutes, which were more abundant in both the contrasts, and with Bacteroidetes and *Roseburia*, which were less abundant. In the second comparison (i.e. the contrast between DD and LW pigs *versus* the contrast between obese Göttingen and lean Göttingen pigs), seven taxa were shared by both studies. Only Firmicutes and Bacteroidetes showed statistically different abundances in the two studies, but with opposite directions of variation. The other five taxa were not significantly different in our study, and only three of them showed the same direction of variation. In the third comparison (i.e. the contrast between DD and LW pigs *versus* the contrast between obese Ossabaw and lean Ossabaw pigs) we found eight taxa common to both studies, six of which with the same direction of variation, i.e. Actinobacteria, Prevotellaceae, *Clostridium*, *Streptococcus*, *Bacteroides* and *Prevotella*. Of note, none of these contrasts was however statistically supported (Supplementary Information 2).

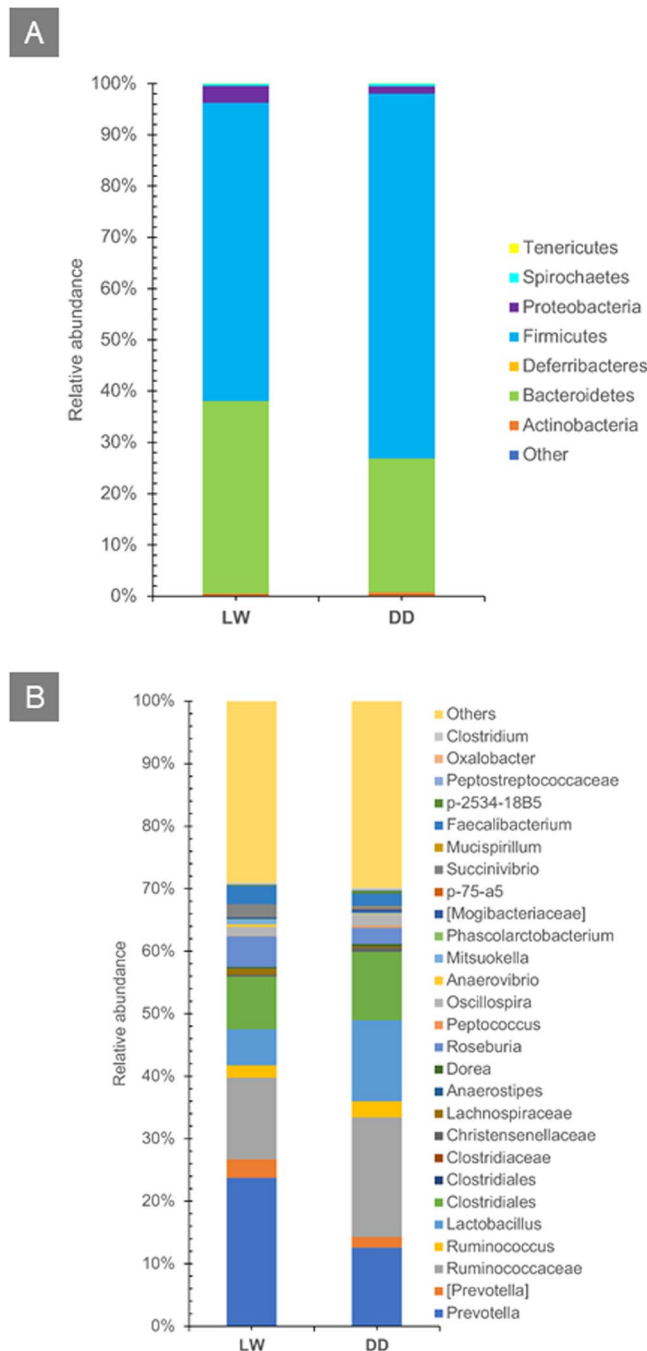


Figure 3. (A) Histogram representing the relative abundance (%) of phyla in DD (on the right) and LW (on the left) pigs. (B) Histogram representing the relative abundance (%) of microbial taxa in DD (on the right) and LW (on the left) pigs.

Both the PCA and the clustering analyses separated the two breeds, but several samples were misplaced (Fig. 1C, Supplementary Fig. 2C).

Integration of the different layers of phenotype information. First, a multiple factor analysis (MFA) was performed using CBCs, blood transcriptome, and faecal microbiota data. The individual factor map (IFM) revealed a split between the breeds, with the first component accounting for 20.3% of the total variability (Fig. 1D). The variable group plot (Supplementary Fig. 4) showed that CBCs gave the highest contribution to the first dimension of variability and the lowest contribution to the second dimension, while transcriptome data showed the opposite pattern. Microbiota data presented intermediate values.

Subsequently, the links between the different data layers were studied in a pairwise fashion with sPLS³³ and using the data obtained at 60 days. In the case of CBCs and transcriptome, 17 individuals were used. The results

were visualized producing a clustered image map (CIM)³⁴, showing that CBCs were split in two main groups, corresponding to the metrics which had already been found more abundant and less abundant in DD pigs (Fig. 4A).

The most relevant associations were represented by 20 genes and 12 CBCs. Some CBC metrics, which were not differentially abundant between DD and LW pigs, did show association with DE genes (Supplementary Table 3B). Indeed, eonum, eopro, grnum and wbc grouped with the parameters showing lower abundance in minipigs. The metric eopro was discarded as not reliable (See methods).

In the case of transcriptome and faecal microbiota, sPLS was performed on 20 individuals. A CIM (Fig. 4B) revealed that the most important associations concerned 20 genes and 18 taxa. The genes were split into three groups. The first included four genes with strong positive association with Firmicutes such as *Peptococcus* and *Ruminococcus* and strong negative association with *Prevotella* and *Roseburia*, while the second included six genes showing the opposite pattern. The third group showed weaker association with the taxa.

For CBCs and faecal microbiota, the sPLS was based on 20 individuals, but no significant association between the variables was found.

Discussion

The DD minipig herd maintained inbred in France does not seem especially prone to obesity, but to date there have been no attempts to characterize it in terms of metabolic diseases. Usually, these disorders are studied by feeding the animals with high-fat and high-sucrose diets. However, by a general point of view, minipigs can develop obesity even when they receive *ad libitum* standard chow²². This has been observed, for instance, in the DD minipig herd maintained in the United Kingdom (Mick Bailey, personal communication). In our study, the animals were fed *ad libitum* with classical pig diet and their physiological status was studied at 60 days by comparison to LW pigs. This latter commercial breed was chosen because of its lean meat percentage^{20,35}. We combined the results from three data layers, i.e. CBCs, transcriptome profiles, and 16S rDNA sequencing, and our findings suggested that DD pigs were potentially affected by a disease belonging to the NAFLD spectrum.

First, we studied the overall data structure. The MFA showed a clear split between DD and LW pigs, which was globally confirmed by the PCA analyses individually performed on each dataset. In the case of hierarchical clustering, the split between the breeds was clearly supported by the transcriptome data alone. The transcriptome was also the most informative dataset in terms of contribution to the total variation. The microbiota contributed to a similar, but slightly lower extent, while the role of CBCs was more limited. These findings were in agreement with those obtained by Mach and co-workers³¹ and was possibly related to the number of variables of each dataset, which was highest in the case of transcriptome and lowest in the case of CBCs.

Then, we studied each single dataset independently to understand the molecular mechanisms underlying our biological systems, and we used pairwise sPLS analyses to better characterize their interactions. Two important hemorheological parameters, i.e. hematocrit and hemoglobin, were significantly higher in DD pigs over almost the whole time-course, making the finding more robust. High levels of hematocrit and hemoglobin are associated with obesity and metabolic syndrome^{36,37}, respectively, and both metrics are related to insulin resistance, NAFLD and NASH in humans^{38–40}. Mean corpuscular volume was also higher in DD pigs; however, even though this parameter is increased in some liver pathologies, it does not seem to be linked to NAFLD⁴¹. The values of platelet distribution width and mean platelet volume were also consistently higher in DD pigs, indicating platelet activation. This process is linked to obesity and NAFLD in both pigs and humans^{23,42–44}. Platelets participate in the process of liver inflammation⁴⁵ and their activation is specifically predictive of cardiovascular diseases⁴⁶, which are strongly associated with NAFLD^{5,10}. Because these processes are deeply intertwined, it is difficult to state whether augmented values of pwd and mpv point to a NAFLD-related disorder, a cardiovascular disease, or a combination of both. However, in a multisystemic perspective, they might suggest the presence of cardiac diseases associated to NAFLD as comorbidities. In contrast, monocyte absolute number and monocyte percentage showed a trend towards decreased values in minipigs. Low levels of monocytes have been correlated to septic processes in patients affected by chronic liver diseases, possibly representing a response to an extensive and acute inflammatory process⁴⁷. In any case, the low levels of statistical support did not allow us to draw strong conclusions in this regard.

Transcriptome profiling also suggested a disorder belonging to the spectrum of NAFLD in DD pigs. This result was especially interesting because the enriched processes were found across each step of the two-hit model^{12,13} and because they were often confirmed by more than one analysis method. Furthermore, the pathways identified as enriched showed high levels of agreement with the literature available for humans^{13,14,48} and for Bama pigs¹⁵. These findings indicated a global shift in gene expression and suggested an impact of the disease on the physiological state of the minipigs. The pathways detected using traditional functional analysis corresponded to different molecular processes involved in NAFLD, but almost none of them represented an NAFLD-centered gene set. Indeed, a specific enrichment was obtained only with ClueGO, because both the IPA and the GSEA databases lacked appropriate ontologies. Since we wanted to study the potential presence of NAFLD in a more targeted way, we followed a meta-analytic approach based on the literature and then we obtained a significant enrichment of our DE genes list. This supported our characterization of the transcriptome in terms of NAFLD, with particular relevance because it was based on experimental expression studies (See methods). Moreover, looking at the 20 highest and the 20 lowest expressed genes, we found a strong overrepresentation of NAFLD-related genes, which added up to 25% (i.e. 5 out of 20 genes). Indeed, the same value calculated on the whole list was just 12.7% (i.e. 386 out of 3,042 genes). The genes belonging to the pathways used for the other methods of enrichment analysis, instead, accounted for only 5% of the total. This indicated that the deepest changes in the transcriptome affected genes that were directly involved in NAFLD. It was the case, for instance, with *IFITM1* (\log_2 FC = 4.1), *COL16A1* (\log_2 FC = 2.9), *TOR3A* (\log_2 FC = 2.4), *LY6D* (\log_2 FC = 2.2), *CXCL10* (\log_2 FC = 1.9), *F2R* (\log_2 FC = -1.8) and *TAGLN* (\log_2 FC = -2.1), which have key roles in many studies performed in pigs¹⁵, mice^{6,49,50} and humans^{14,51}.

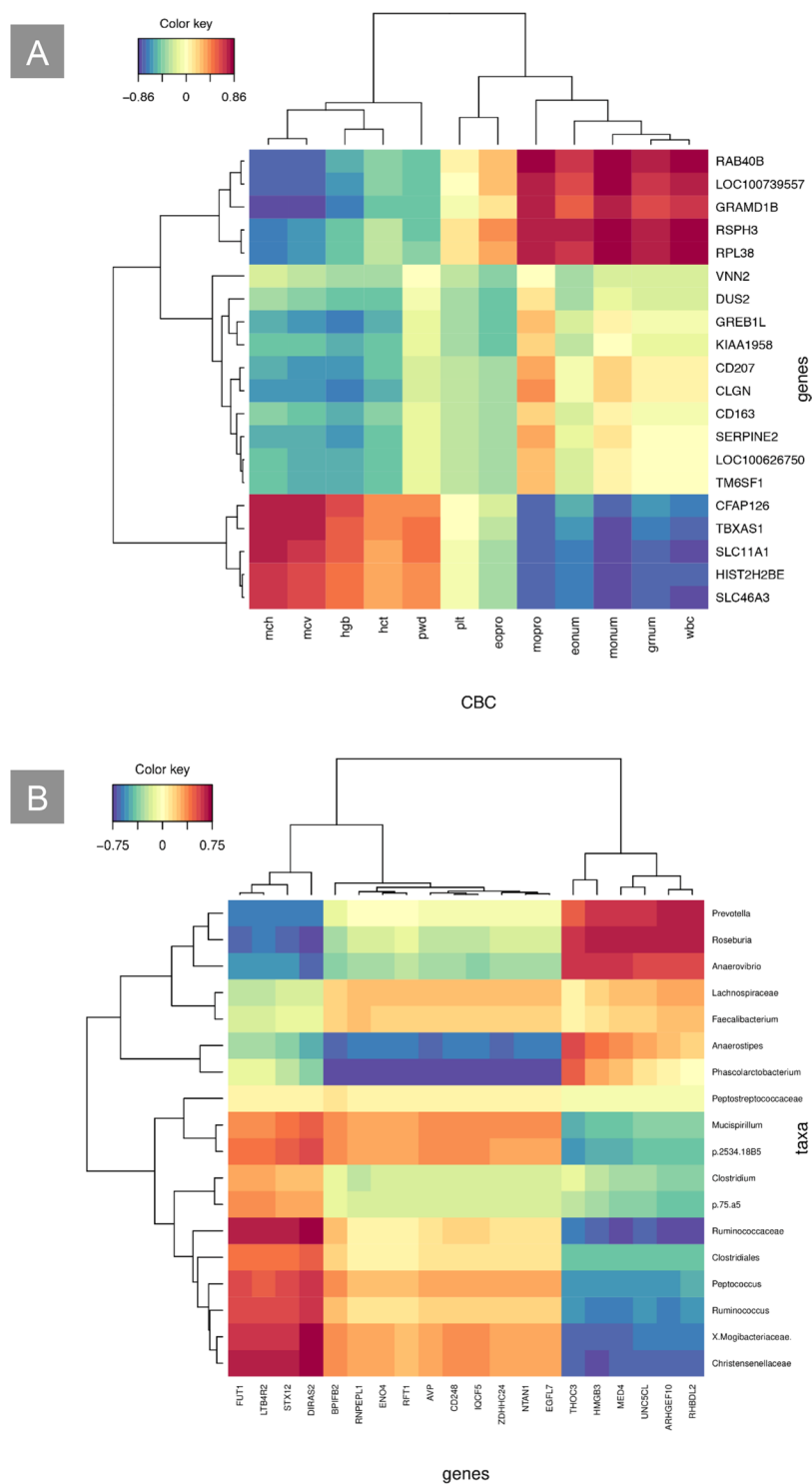


Figure 4. (A) A clustered image map (CIM) based on the association values obtained from the first two dimensions of the sPLS analysis integrating CBCs and DE genes. The CBCs parameters are plotted on the X-axis, and the genes are plotted on the Y-axis. The red and the blue coloured blocks indicate a positive or a negative association respectively, while the yellow colour indicates weak levels of association. (B) A clustered image map (CIM) based on the association values obtained from the first two dimensions of the sPLS analysis integrating DE genes in the blood and faecal microbial taxa. The DE genes are plotted on the X-axis, and the microbial taxa are plotted on the Y-axis. The red and the blue coloured blocks indicate a positive or a negative association respectively, while the yellow colour indicates weak levels of association.

In agreement with CBC data, a specific enrichment of pathways related to cardiac disorders was also found, like in the cases of “heme metabolism” (GSEA), “platelet activation” and “cardiomyopathies” (ClueGO). This finding is meaningful in a multisystemic perspective, which stresses the link between NAFLD and cardiovascular diseases⁵.

The sPLS provided further hints about the connections between the first two data layers (i.e. CBCs and blood transcriptome). On the one hand, it confirmed the importance of the differentially abundant CBCs in the definition of the NAFLD phenotype, showing strong association with some DE genes. Interestingly, the large majority of genes (18 out of 20) was not known to be related to NAFLD and could therefore represent new candidates. This was the case, for instance, with *CFAP126* and *HIST2HBE*, which showed strong positive association with hematological parameters and strong negative association with monocytes, and which grouped with the well characterized *SLC11A1*⁴⁹ and *TBXAS6* genes. On the other hand, the sPLS also indicated that some non-differentially abundant CBC parameters were strongly associated with DE genes, even though their biological meaning was not easy to interpret. This happened, for instance, with the metric eonum. A loss of eosinophils was reported in the adipose tissue of patients affected by NASH⁵², but it is not straightforward to compare these results to ours. Also, NAFLD is usually associated with high wbc⁵³ and mpv⁴⁴ values, which contrasts with our findings.

The characterization of microbiota was less clear-cut. On the one hand, the comparison of our data to those obtained in NASH-affected Ossabaw pigs³² and in obese Göttingen pigs²² highlighted only limited similarities. On the other hand, the microbiota of our DD pigs shared some characteristics with that of the Ossabaw breed which, according to the authors, displayed the features associated with obesity²². However, the lack of statistical support did not allow us to draw robust conclusions on this point. The Firmicutes to Bacteroidetes ratio was higher in DD (2.7) than in Large White pigs (1.6), which was also consistent with the pattern observed in obese Ossabaw minipigs²² and could provide some support to the presence of obesity^{32,54}. However, these pieces of evidence were not sufficient to make robust inferences about microbiota data.

The sPLS analysis provided more details about the relationship existing between microbiota and DE genes. The genus *Prevotella*, for instance, was negatively associated with obesity in Ossabaw minipigs²² and was similarly underrepresented in DD pigs. According to our results, the genus *Prevotella* exhibited negative association with the genes *FUT1*, *LTB4R2*, *STX12* and *DIRAS2*, which could be therefore related to this pathology. Except for *LTB4R2*, these genes have never been associated with obesity or NAFLD and could be therefore new candidates for these disorders. The genera *Roseburia* and *Anaerovibrio* grouped with *Prevotella*, pointing to a potential link to the NAFLD spectrum; however, no evidence for the involvement of these bacteria in obesity was found in Ossabaw pigs²².

Taken together, our results suggest that the DD minipigs could potentially present symptoms of a pathology belonging to the NAFLD spectrum. It is important to refer to NAFLD spectrum, and not to NAFLD *stricto sensu*, because of the inherent complexity of this disorder and because we could not produce additional physiological data to allow a specific diagnosis. In any case, NAFLD is increasingly being regarded as a multisystem disease⁵, which implies that it is constitutively linked to a large array of comorbidities.

Even if all the data layers supported our hypothesis, each one highlighted a slightly different combination of pathologies. In the case of CBCs, the results pointed to the presence of cardiac diseases and NAFLD, with a stronger representation of the former. In the case of transcriptome, NAFLD seemed to be the primary disorder, while cardiac diseases appeared as comorbidities. Microbiota data, instead, suggested the potential presence of obesity, but this evidence was not strong.

Transcriptome was the most informative data layer, and its integration with CBCs and microbiota produced consistent results. Therefore, we can speculate that our tentative characterization of the datasets in terms of NAFLD could potentially be the most accurate.

It must be considered that some limitations affect our results. Our data were obtained on young individuals, while NAFLD-related disorders are typically studied on 4- to 30-month-old animals. Therefore, many symptoms were likely not completely apparent. Moreover, transcriptome and microbiota data were only obtained at 60 days, while more time-points could provide a better characterization. Nevertheless, our work adds useful information to characterize NAFLD-related disorders in pigs, providing a robust workflow for the analysis of intermediate phenotypes.

Materials and Methods

Animals and diet. Twelve LW (i.e. 6 pairs of siblings) and twelve DD pigs (i.e. 6 pairs of siblings) of the same age were used (Supplementary Table 1). The suckling piglets had access to a pre-starter feed from day 3 (18% crude protein DM, 7.5% crude fat DM, 1.55% lysine DM) and were weaned at day 28. They were fed a starter diet from day 29 to day 40 (19% crude protein DM, 7% crude fat DM, 1.45% lysine DM) and were fed a grower diet *ad libitum* (16.5% crude protein DM, 2.4% crude fat DM, 1.17% lysine DM) from day 41 to day 150.

All the animals were raised and sampled at the INRA experimental farm of Nouzilly (France) in 2012 (farm agreement E 37-175-2). All the experiments were compliant to the French regulation for use of animals in research (articles R214-87 to R214-137 of the French rural code before the legal update enacted in 2013). No ethics approval was required for the collection of blood samples under the then current regulation.

Blood and stool sampling. The blood samples for CBCs were collected at 8, 20, 40, 60 and 100 days of age using BD Vacutainer EDTA tubes (Becton Dickinson, Franklin Lakes, NJ, USA). The blood samples for RNA extraction were collected from 60-day-old animals using PAXgene Blood RNA tubes (Qiagen, Hilden, Germany). Faecal aliquots (200 mg) were collected from the rectum of the animals at 60 days of age, snap-frozen in liquid nitrogen and stored at -80°C .

Complete blood counts: data production and analysis. CBCs were obtained using a scil Vet abc counter (scil animal care company, Altorf, France). Eighteen parameters were measured for each time-point: white blood cell count (wbc); red blood cell count (rbc); platelet count (plt); hemoglobin (hgb); hematocrit (hct); mean corpuscular volume (mcv); mean corpuscular hemoglobin (mch); mean corpuscular hemoglobin concentration (mchc); platelet distribution width (pwd); mean platelet volume (mpv); lymphocytes absolute number (lynum); monocytes absolute number (monum); granulocytes absolute number (grnum); eosinophils absolute number (eonum); lymphocytes percentage (lypro); monocytes percentage (mopro); granulocytes percentage (grpro); eosinophils percentage (eopro).

A linear-mixed model was fitted to each parameter using the lme4 R package⁵⁵. The breed was set as a fixed effect and the sow as a random effect. Data of each day were separately analysed.

The results obtained on 60-day-old individuals were used for the subsequent analyses, while the values obtained at the other time-points (i.e. 8, 20, 40 and 100 days) were used only as a means of validation. A metric was retained only if the direction of the change in the differential abundance at 60 days (i.e. positive or negative) was consistent with at least two out of five points.

The metrics retained were rescaled using the ‘rescale’ function of the scales R package (<https://CRAN.R-project.org/package=scales>) and analysed through a PCA and a hierarchical clustering using the FactoMineR R package and selecting the Pearson’s correlation coefficient and the Ward method⁵⁶.

RNA extraction and microarray processing. Total RNA from blood of 60-day-old individuals was isolated using the PAXgene Blood RNA Kit (Qiagen). RNA quality and integrity were determined using a NanoDrop 1000 (Thermo Scientific, Waltham, MA, USA) and a Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA).

Transcriptional profiling was performed using an Agilent 8 × 60K pig custom microarray (Agilent Technologies, AMADID 037880). Twenty-four samples were processed (i.e. three chips). All steps were performed by the @BRIDGE facility (INRA Jouy-en-Josas, France, <http://abridge.inra.fr/>), as described previously⁵⁷. Microarray data were submitted to GEO (accession number GSE111953).

Microarray data analysis. The annotation of the pig array was updated using the Sscrofa 11.1 version of the pig genome (see Supplementary Information 1).

Probe intensities were background corrected using the “normexp” method, log₂ scaled and quantile normalized using the Limma R package⁵⁸. The probes of the lowest quartile were filtered out, controls were discarded and the probes corresponding to genes were summarized.

The obtained expression matrix (“E1”) was processed with the arrayQualityMetrics R package⁵⁹ for quality assessment. Three outliers and a technical bias concerning the third chip were detected. After outlier removal, the raw data were pre-processed again to obtain the “E2” expression matrix.

The differential analysis was performed using the Limma R package. A linear model was fitted for each gene, setting the breed and the chip as fixed effects, and comparing DD to LW pigs. The sow was included as a random effect using the “duplicateCorrelation” function. The p-values were Benjamini-Hochberg corrected setting a threshold of 0.05.

An alternative linear model was tested, including also monocyte count as a fixed effect, but the small sample size (i.e. 17 individuals) determined a sharp decrease in the statistical power of detection.

The E2 expression matrix was further modified. In fact, this matrix was not corrected for the chip-related bias, which was treated as an effect by the Limma linear model. To perform other downstream analyses, this bias was corrected using the Limma “removeBatchEffect” function. The “E3” corrected matrix was used to perform a PCA and a clustering with FactoMineR⁵⁶.

Functional analysis of blood transcriptome. Transcriptome functional analyses were performed using four approaches. First, the DE genes were analysed with Ingenuity Pathways Analysis 01.08 (September 2017 Release, IPA, www.ingenuity.com) to identify canonical pathways. Only those with a -log(p-value) > 1.75 and including at least 10 genes were considered.

The second analysis was realized using ClueGO 2.3.4²⁸. A two-sided test was used to highlight enriched GO terms. Significance was set at an adjusted p-value of 0.05, the “GO fusion” option was used, the k-score was fixed at 0.2 and the global/local parameter at the sixth level. Three ontologies were selected: “Sus scrofa ontologies KEGG_06.07.2015”, “GO_ImmuneSystemProcess_03.07.2015_09h08” and “GO_BiologicalProcess_03.07.2015_09h08”.

A third analysis was performed using GSEA²⁹ as implemented by the Broad Institute (<http://software.broadinstitute.org/gsea/index.jsp>). The E3 matrix was analysed using the “h.all.v6.1” MsigDB²⁹, choosing the “gene_set” option for the permutations.

In all these three cases, the percentage of pathways putatively related to the NAFLD spectrum was estimated by performing a comparison with a pathway consensus list obtained using 19 papers describing transcriptional profiling^{6,8,14–16,49,50,60–69} or data meta-analysis^{7,51}. A further review paper was also added²⁷. More details are found in Supplementary Information 2.

The fourth approach involved a meta-analytical screening of the literature relative to NAFLD and other related diseases. The same already listed 19 papers were used, and the available DE gene lists were retrieved and merged to obtain a non-redundant set, which was intersected with our DE gene list. The enrichment was assessed using a Fisher’s exact test and a hypergeometric test.

Production and analysis of 16S data. DNA was extracted from frozen faecal samples as described by⁷⁰. Subsequently, the V3–V4 region of the 16S rRNA gene was sequenced using the 454 GS FLX Titanium platform (Roche, Pleasanton, CA, USA) as described by⁷¹. FastQC was used to perform quality checks on the raw data and the sequences were analysed with QIIME 1.9.1⁷² as described previously⁷¹.

Species richness estimates (Chao1, observed OTUs, and abundance-based coverage estimators or ACE) and evenness indexes (Shannon and Simpson), were calculated using the phyloseq R package after the reads were rarefied⁷³. For β -diversity, weighted UniFrac distances were first calculated on the OTUs table. Then, nonmetric multidimensional scaling (nMDS) and distances to centroid were calculated using the “betadisper” function of the vegan R package. Breed comparisons were assessed with PERMANOVA using the “adonis” function in the vegan R package⁷⁴, and p-values were obtained using Monte Carlo random draws. Also, an ANOSIM was performed using the vegan R package to test the significance of the difference between the breeds.

A linear-mixed model was fitted on the OTUs table using the lme4 R package and setting the breed as a fixed effect and the sow as a random effect. Relative abundance (%) differences at phylum and lower taxonomic levels between DD and LW pigs were assessed using MetagenomeSeq 1.20.1⁷⁵. A Kruskal-Wallis test was performed, and q-values were Benjamini-Hochberg adjusted. The differentially abundant taxa were compared to the results published in two recent papers^{22,32} by following the procedure described in Supplementary Information 3. A PCA and a clustering were also performed. 16S data were submitted to SRA (accession number SRP144235).

Data integration. An MFA⁷⁶ was used to integrate CBCs, blood transcriptome, and 16S data using the FactoMineR R package⁵⁶. The results were visualized using an individual factor map (IFM) and a variable groups plot.

The datasets were then studied in a pairwise fashion with sPLS³³ using the mixOmics R package (<https://cran.r-project.org/web/packages/mixOmics/>)⁷⁷. For each dataset pair, the analysis was performed using only the 60-day-old individuals for which the data belonging to both datasets were available. For CBCs and transcriptome and for transcriptome and 16S, the statistical models were optimized using the “perf” function and setting the “ncomp” parameter to two, and a CIM was used to visualize the data. No valid model was obtained for CBCs and 16S data.

Received: 13 August 2019; Accepted: 18 December 2019;

Published online: 14 January 2020

References

- McAnulty, P. *et al.* Minipig in Biomedical Research. Minipig in Biomedical Research, <https://doi.org/10.1201/b11356-38> (RC Press, Boca Raton, 2012).
- Larzul, C. Pig genetics: insight in minipigs. *Bilater. Symp. Miniat. Pigs Biomed. Res. Taiwan Fr.* 1–6 (2013).
- Bourneuf, E. *et al.* New susceptibility loci for cutaneous melanoma risk and progression revealed using a porcine model. *Oncotarget* **9**, 27682–27697 (2018).
- Sachs, D. H. The pig as a potential xenograft donor. *Vet. Immunol. Immunopathol.* **43**, 185–191 (1994).
- Byrne, C. D. & Targher, G. NAFLD: A multisystem disease. *J. Hepatol.* **62**, S47–S64 (2015).
- van Koppen, A. *et al.* Uncovering a Predictive Molecular Signature for the Onset of NASH-Related Fibrosis in a Translational NASH Mouse Model. *Cell. Mol. Gastroenterol. Hepatol.* **5**, 83–98 (2017).
- Ryaboshapkina, M. & Hammar, M. Human hepatic gene expression signature of non-alcoholic fatty liver disease progression, a meta-analysis. *Sci. Rep.* **7** (2017).
- Greco, D. *et al.* Gene expression in human NAFLD. *Am. J. Physiol. Gastrointest. Liver Physiol.* **294**, G1281–7 (2008).
- Fabbrini, E., Sullivan, S. & Klein, S. Obesity and nonalcoholic fatty liver disease: Biochemical, metabolic, and clinical implications. *Hepatology* **51**, 679–689 (2010).
- Patil, R. & Sood, G. K. Non-alcoholic fatty liver disease and cardiovascular risk. *World J. Gastrointest. Pathophysiol.* **8**, 51–58 (2017).
- Berlanga, A., Guiu-Jurado, E., Porras, J. A. & Auguet, T. Molecular pathways in non-alcoholic fatty liver disease. *Clin. Exp. Gastroenterol.* **7**, 221–239 (2014).
- Paschos, P. & Paletas, K. Non alcoholic fatty liver disease and metabolic syndrome. *Hippokratia* **13**, 9–19 (2009).
- Baciu, C. *et al.* Systematic integrative analysis of gene expression identifies HNF4A as the central gene in pathogenesis of non-alcoholic steatohepatitis. *PLoS One* **12**, 1–14 (2017).
- Zhu, R. *et al.* Systematic transcriptome analysis reveals elevated expression of alcohol-metabolizing genes in NAFLD livers. *J. Pathol.* **238**, 531–542 (2016).
- Xia, J. *et al.* Transcriptome analysis on the inflammatory cell infiltration of nonalcoholic steatohepatitis in Bama minipigs induced by a long-term high-fat, high-sucrose diet. *PLoS One* **9** (2014).
- Matz-Soja, M. *et al.* Hedgehog signaling is a potent regulator of liver lipid metabolism and reveals a GLI-code associated with steatosis. *Elife* **5**, 1–28 (2016).
- Spurlock, M. E. & Gabler, N. K. The development of porcine models of obesity and the metabolic syndrome. *J. Nutr.* **138**, 397–402 (2008).
- Kanuri, G. & Bergheim, I. *In vitro* and *in vivo* models of non-alcoholic fatty liver disease (NAFLD). *Int. J. Mol. Sci.* **14**, 11963–11980 (2013).
- Lunney, J. K. Advances in swine biomedical model genomics. *Int. J. Biol. Sci.* **3**, 179–184 (2007).
- Kogelman, L. J. A. & Kadarmideen, H. N. Applications of Systems Genetics and Biology for Obesity Using Pig Models. in *Systems Biology in Animal Production and Health*, Vol. 1 1, 25–43 (2016).
- Lee, L. *et al.* Nutritional model of steatohepatitis and metabolic syndrome in the Ossabaw miniature swine. *Hepatology* **50**, 56–67 (2009).
- Pedersen, R. *et al.* Characterisation of Gut Microbiota in Ossabaw and Göttingen Minipigs as Models of Obesity and Metabolic Syndrome. *PLoS One* **8**, 1–10 (2013).
- Bollen, P. J. A., Madsen, L. W., Meyer, O. & Ritskes-Hoitinga, J. Growth differences of male and female Göttingen minipigs during ad libitum feeding: a pilot study. *Lab. Anim.* **39**, 80–93 (2005).
- Ho, C. S. *et al.* Nomenclature for factors of the SLA system, update 2008. *Tissue Antigens* **73**, 307–315 (2009).
- Bovo, S. *et al.* Genome-wide association studies for 30 haematological and blood clinical-biochemical traits in Large White pigs reveal genomic regions affecting intermediate phenotypes. *Sci. Rep.* **9**, 1–17 (2019).
- Riordan, J. D. & Nadeau, J. H. Modeling progressive non-alcoholic fatty liver disease in the laboratory mouse. *Mamm. Genome* **25**, 473–486 (2014).
- Nassir, F. & Ibdah, J. A. Sirtuins and nonalcoholic fatty liver disease. *World J. Gastroenterol.* **22**, 10084–10092 (2016).
- Bindea, G. *et al.* ClueGO: A Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* **25**, 1091–1093 (2009).
- Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci.* **102**, 15545–15550 (2005).

30. Liberzon, A. *et al.* The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Syst.* **1**, 417–425 (2015).
31. Mach, N. *et al.* The Effects of Weaning Methods on Gut Microbiota Composition and Horse Physiology. *Front Physiol.* (2017).
32. Panasevich, M. R. *et al.* High-fat, high-fructose, high-cholesterol feeding causes severe NASH and cecal microbiota dysbiosis in juvenile Ossabaw swine. *Am J. Physiol Endocrinol Metab* **314** (2017).
33. Lê Cao, K. A., Martin, P. G. P., Robert-Granié, C. & Besse, P. Sparse canonical methods for biological data integration: Application to a cross-platform study. *BMC Bioinformatics* **10**, 1–17 (2009).
34. González, I., Cao, K. A. L., Davis, M. J. & Déjean, S. Visualising associations between paired ‘omics’ data sets. *BioData Min.* **5**, 1–23 (2012).
35. Alfonso, L., Mourot, J., Insausti, K., Mendizabal, J. A. & Arana, A. Comparative description of growth, fat deposition, carcass and meat quality characteristics of Basque and Large White pigs. *Anim. Res.* **54**, 33–42 (2005).
36. Rayis, D. A., Abdelmageed, R. E., Adam, I., Elmugabil, A. & Gasim, G. I. High level of hemoglobin, white blood cells and obesity among Sudanese women in early pregnancy: a cross-sectional study. *Futur. Sci. OA* **3**, FSO182 (2017).
37. Guiraudou, M., Varlet-Marie, E., Raynaud De Mauverger, E. & Brun, J. F. Obesity-related increase in whole blood viscosity includes different profiles according to fat localization. *Clin. Hemorheol. Microcirc.* **55**, 63–73 (2013).
38. Giorgio, V. *et al.* Elevated Hemoglobin Level Is Associated with Advanced Fibrosis in Pediatric Nonalcoholic Fatty Liver Disease. *J. Pediatr. Gastroenterol. Nutr.* **65**, 150–155 (2017).
39. Bai, C. H. *et al.* Relationship between hemoglobin levels and risk for suspected non-alcoholic fatty liver in taiwanese adults. *Chin. J. Physiol.* **57**, 286–294 (2014).
40. Vinci, P. *et al.* The Association between Hematological Parameters and Insulin Resistance Is Modified by Body Mass Index – Results from the North-East Italy MoMa Population Study. *PLoS One* **9**, e101590 (2014).
41. Das, S., Mukherjee, S., Vasudevan, D. & Balakrishnan, V. Comparison of haematological parameters in patients with non-alcoholic fatty liver disease and alcoholic liver disease. *Singapore Med. J.* **53**, 175–181 (2011).
42. Milovanovic Alempijevic, T. *et al.* Diagnostic Accuracy of Platelet Count and Platelet Indices in Noninvasive Assessment of Fibrosis in Nonalcoholic Fatty Liver Disease Patients. *Can. J. Gastroenterol. Hepatol.* **2017**, 1–5 (2017).
43. Ozhan, H. *et al.* Mean platelet volume in patients with non-alcoholic fatty liver disease. *Platelets* **21**, 29–32 (2010).
44. Celikbilek, M. Letter: increased platelet activation in chronic liver disease - hit two targets with a single shot. *Aliment. Pharmacol. Ther.* **43**, 1023–1023 (2016).
45. Chauhan, A., Adams, D. H., Watson, S. P. & Lalor, P. F. Platelets: No longer bystanders in liver disease. *Hepatology* **64**, 1774–1784 (2016).
46. Bekler, A. *et al.* Increased Platelet Distribution Width Is Associated with Severity of Coronary Artery Disease in Patients with Acute Coronary Syndrome. *Angiology* **66**, 638–643 (2015).
47. Antoniadis, C. G., Wendon, J. & Vergani, D. Paralyzed monocytes in acute on chronic liver disease. *J. Hepatol.* **42**, 163–165 (2005).
48. Harmon, R. C., Tiniakos, D. G. & Argo, C. K. Inflammation in nonalcoholic steatohepatitis. *Expert Rev. Gastroenterol. Hepatol.* **5**, 189–200 (2011).
49. Mitsumoto, K. *et al.* Time-course microarrays reveal early activation of the immune transcriptome in a choline-deficient mouse model of liver injury. *Life Sci.* **184**, 103–111 (2017).
50. Kirpich, I. A. *et al.* Integrated hepatic transcriptome and proteome analysis of mice with high-fat diet-induced nonalcoholic fatty liver disease. *J. Nutr. Biochem.* **22**, 38–45 (2011).
51. Qi, S., Wang, C., Li, C., Wang, P. & Liu, M. Candidate genes investigation for severe nonalcoholic fatty liver disease based on bioinformatics analysis. *Medicine (Baltimore)*. **96**, e7743 (2017).
52. Hart, K. M. *et al.* Type 2 immunity is protective in metabolic disease but exacerbates NAFLD collaboratively with TGF- β . *Sci. Transl. Med.* **9** (2017).
53. Chung, G. E. *et al.* Associations between White Blood Cell Count and the Development of Incidental Nonalcoholic Fatty Liver Disease. *Gastroenterol. Res. Pract.* **2016**, 1–6 (2016).
54. Ley, R., Turnbaugh, P., Klein, S. & Gordon, J. Microbial ecology: human gut microbes associated with obesity. *Nature* **444**, 1022–3 (2006).
55. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting Linear Mixed-Effects Models using lme4. *J. Stat. Softw.* **67** (2015).
56. Lê, S., Josse, J. & Husson, F. FactoMineR: An R Package for Multivariate Analysis. *J. Stat. Softw.* **25**, 1–18 (2008).
57. Jacquier, V. *et al.* Genome-wide immunity studies in the rabbit: Transcriptome variations in peripheral blood mononuclear cells after *in vitro* stimulation by LPS or PMA-Ionomycin. *BMC Genomics* **16** (2015).
58. Smyth, G. K. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* **3**, Article3 (2004).
59. Kauffmann, A., Gentleman, R. & Huber, W. arrayQualityMetrics - A bioconductor package for quality assessment of microarray data. *Bioinformatics* **25**, 415–416 (2009).
60. Younossi, Z. M. *et al.* Hepatic gene expression in patients with obesity-related non-alcoholic steatohepatitis. *Liver Int.* **25**, 760–771 (2005).
61. Arendt, B. M. *et al.* Altered hepatic gene expression in nonalcoholic fatty liver disease is associated with lower hepatic n-3 and n-6 polyunsaturated fatty acids. *Hepatology* **61**, 1565–78 (2015).
62. Almanza, D. RNA Seq Analysis of Non-Alcoholic Fatty Liver Disease (NAFLD) Induced by Metabolic Syndrome in a Mouse Model. Honor. Coll. Thesis Univ. Massachusetts Bost (2016).
63. Wang, R., Wang, X. & Zhuang, L. Gene expression profiling reveals key genes and pathways related to the development of non-alcoholic fatty liver disease. *Ann. Hepatol.* **15**, 190–199 (2016).
64. Hui, S. T. *et al.* The genetic architecture of NAFLD among inbred strains of mice. *Elife* **4**, 1–28 (2015).
65. Moylan, C. A. *et al.* Hepatic gene expression profiles differentiate presymptomatic patients with mild versus severe nonalcoholic fatty liver disease. *Hepatology* **59**, 471–482 (2014).
66. Wruck, W. *et al.* Multi-omic profiles of human non-alcoholic fatty liver disease tissue highlight heterogenic phenotypes. *Sci. Data* **2**, 1–10 (2015).
67. Gawrieh, S. *et al.* Hepatic gene networks in morbidly obese patients with nonalcoholic fatty liver disease. *Obes. Surg.* **20**, 1698–1709 (2010).
68. Carazo, A. *et al.* Hepatic expression of adiponectin receptors increases with non-alcoholic fatty liver disease progression in morbid obesity in correlation with glutathione peroxidase 1. *Obes. Surg.* **21**, 492–500 (2011).
69. Teufel, A. *et al.* Comparison of Gene Expression Patterns Between Mouse Models of Nonalcoholic Fatty Liver Disease and Liver Tissues From Patients. *Gastroenterology* **151**, 513–525.e0 (2016).
70. Lepage, P. *et al.* Biodiversity of the mucosa-associated microbiota is stable along the distal digestive tract in healthy individuals and patients with IBD. *Inflamm. Bowel Dis.* **11**, 473–480 (2005).
71. Mach, N. *et al.* Early-life establishment of the swine gut microbiome and impact on host phenotypes. *Environ. Microbiol. Rep.* **7**, 554–569 (2015).
72. Caporaso, J. G. *et al.* correspondence QIIME allows analysis of high-throughput community sequencing data Intensity normalization improves color calling in SOLiD sequencing. *Nat. Publ. Gr.* **7**, 335–336 (2010).
73. McMurdie, P. J. & Holmes, S. Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS One* **8** (2013).

74. Philip, D. Computer program review VEGAN, a package of R functions for community ecology. *J. Veg. Sci.* **14**, 927–930 (2003).
75. Paulson, J. N., Colin Stine, O., Bravo, H. C. & Pop, M. Differential abundance analysis for microbial marker-gene surveys. *Nat. Methods* **10**, 1200–1202 (2013).
76. Escofier, B. & Pagès, J. Multiple factor analysis (AFMULT package). *Comput. Stat. Data Anal.* **18**, 121–140 (1994).
77. Rohart, F., Gautier, B., Singh, A. & Lê Cao, K.-A. mixOmics: An R package for omics feature selection and multiple data integration. *PLOS Comput. Biol.* **13**, e1005752 (2017).

Acknowledgements

This work was supported by the FP7 Interplay grant (FP7-KBBE-2008-2B; grant agreement No. FP7-227549), and the FP7 NADIR grant (NADIR; grant agreement No. FP7-228394). Thanks are due to Frédéric Lantier (PAO, INRAE, Nouzilly, France) for his support for funding the 16S data thanks to the NADIR European project, to François Meurens for contributing to the organization of the pig production, to Catherine Denis (GABI, INRAE, Jouy-en-Josas, France) for contributing to blood and stool extractions, and to Mick Bailey (University of Bristol, UK) for providing information about pig physiology. The authors are grateful to Joan Lunney (Animal Parasitic Disease Laboratory, USDA, Beltsville, MD, USA) who proofread the manuscript.

Author contributions

M.M. performed the analysis of CBCs and transcriptome data and their integration, interpreted the results and wrote the article. P.M.M. analysed 16S data, carried out the integration of transcriptome and 16S data and the MFA and contributed to the final manuscript. J.L. produced the transcriptome data. G.L. performed blood and stool extractions. E.V. took the lead in performing the sampling. C.C. produced the CBCs. J.E. and I.P.O. contributed to the final manuscript. C.R.G. conceived the experiment, participated in the sampling and contributed to the critical interpretation of the data and to the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-019-57127-x>.

Correspondence and requests for materials should be addressed to M.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020