Metabolic prediction of maturity at birth in pigs

Elise Maigné
Mathematiques et Informatique Appliquées de Toulouse
Nathalie Marty-Gasset
Génétique Physiologie et Systèmes d'Elevage
Laure Gress
Génétique Physiologie et Systèmes d'Elevage
Cécile Canlet
Toxalim Research Centre in Food Toxicology
Agnès Bonnet
Génétique Physiologie et Systèmes d'Elevage
Pauline Brenaut
French Pork and Pig Institute
Laurence Liaubet
laurence.liaubet@inrae.fr

Génétique Physiologie et Systèmes d'Elevage

Article

Keywords:

Posted Date: April 8th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9150369/v1

Additional Declarations: No competing interests reported.

Metabolic prediction of maturity at birth in pigs

Elise Maigné[1], Nathalie Marty-Gasset[2], Laure Gress[2], Cécile Canlet[3], Agnes Bonnet[2], Pauline Brenaut[4], and Laurence Liaubet[2,*]

[1]UR MIAT, INRAE, Université de Toulouse, Castanet-Tolosan, 31320, France
[2]GenPhySE, Université de Toulouse, INRAE, INPT, ENVT, Castanet-Tolosan, 31326, France
[3]Toxalim (Research Centre in Food Toxicology, Université de Toulouse, INRAE, INPT, ENVT, MetaToul-AXIOM Platform, National Infrastructure of Metabolomics and Fluxomics: MetaboHUB, Toulouse, 31027, France
[4]IFIP-Institut du porc/Alliance R&D, Pacé, 35740, France
[*]laurence.liaubet@inrae.fr

ABSTRACT

Improving piglet survival is a key objective for breeders. Piglets that have not yet fully developed are more likely to die prematurely. Here, focus was to better characterize maturity at birth. Very immature piglets exhibit a distinctive head morphology with a reminiscent of a dolphin's, with prominent eyes. This study proposed integrating phenotyping data with blood sampling to develop a predictive metabolic signature of piglet maturity at birth.
Following analysis of the head morphology, the study categorized 278 newborns (99 Landrace, 87 Large White, 92 LR×LW) according to their maturity level. Furthermore, a metabolomic analysis was also performed by 1H-NMR on blood samples (serum) collected on piglets in the hours following birth. The raw spectra were analyzed using the R package ASICS. The following statistics were based on 55 metabolites with non-zero variance. A subset of 14 metabolites was selected to develop a predictive model based on random Forests and GLM methods. The two models accurately predict 100% of the severe immaturity status in both the training and test samples. Some piglets that are morphologically classified as mature may be metabolically immature. The 14-metabolite signature can qualify the maturity with a qualitative score as mature or not, and two quantitative scores, a mean predicted value and a stability of the prediction, which allow the confidence of the prediction to be assessed. The predictive model was applied to an independent dataset of blood collected on different farms and from piglets of different genetic origins. This allowed the relevance of the model to be evaluated, taking into account other phenotypes related to the status of birth piglets, such as birth weight, and body mass index.
Genetic selection for survival at birth and growth is primarily based on the measurement of birth weight. As these traits are correlated, it is important to unravel these correlations to understand the underlying molecular mechanisms. The identification of a molecular signature could facilitate future experiments aimed at deciphering the genetic architecture of complex traits, such as maturity. Therefore, we have developed a minimally invasive blood sample that allows for low-cost, user-friendly metabolic analysis of serum. While maturity is typically defined at the biometric level, we propose a novel approach to define this complex trait at the metabolic level.

Introduction

Genetic selection in pigs for productive and reproductive traits has resulted in an increase in carcass and litter size. Unfortunately, this genetic progress has been accompanied by a deterioration in certain traits, such as an increase in neonatal mortality. The number of piglets dying is estimated to be around 10 to 20% of the total birth losses before weaning[1]. Many factors influencing piglet mortality have been identified. They have been related to maternal effects (e.g., intrauterine effects, duration of farrowing, parity, health status), to piglet factors (e.g., genetic type, vitality at birth) or even to piglet characteristics that are partly under maternal control[1] (e.g. birth weight). The peak of preweaning mortality occurred in the first two days after birth and was mainly dependent on the maturity of the piglet[2]. Maturity is defined here as the state of development that allows the piglet to survive after birth. A delay in development influences the maturity status at birth and may penalize further growth[3,4] and survival. In previous research, we have studied fetal development during late gestation to identify delayed molecular processes in different tissues, including muscle[5–7], adipose tissue[8], and the intestinal tract[9]. We also previously examined the metabolic profiles of piglets during late gestation[10] and studied newborn piglets shortly after birth for their ability to thermoregulate[11] with an initial metabolic overview of these newborns[12]. The findings of these studies have identified specific metabolites, such as myo-inositol and proline, which appear to have the potential to distinguish the progression of perinatal development.

This new investigation aims to construct and evaluate a model that accurately predicts the maturity level at birth, which is a critical determinant of future health, growth, and survival. Extreme immaturity can be inferred from certain physical characteristics, such as a dolphin-shaped head shape[3]. This was used in this study as a proxy to train a predictive model to identify a metabolic signature that can be used to assess maturity at birth.

Material and methods

Animals and experimental design

This study was conducted following French legislation on experimentation and ethics. The experiments conducted on private farms were submitted to an ethics committee with the agreement number "SSA_2018_009" and "SSA_2020_006" from the SSA (for Animal Science and Health) Ethics Committee No. 115. The animals and samples were provided by Nucleus and Axiom breeding organizations of the ALLIANCE R&D. All information related to the animal identification and the farm was anonymized. Collected data were used solely for scientific purposes.

For samples collected in an INRAE experimental Unit, the French Ministry of Agriculture authorized the experiments on living animals in UE1372 GenESI facilities[13], with the agreement number APAFiS #13648-2018020417291866 v4). In these conditions, this study follows the ARRIVE guidelines (Animal Research: Reporting of In Vivo Experiments), and is committed to the 3Rs of laboratory animal research and consequently used the minimal number of animals to reach statistical significance.

Blood samples used for model development were obtained from 278 newborn piglets originating from three private farms (PicLet dataset). Three genotypes were used to develop a predictive model independent of genetic background. This group of animals is made up of 99 Landrace (LR), 87 Large White (LW), and 92 crossed LRLW. For these piglets, a morphometric phenotype was available. Three characteristics have been identified and defined. These characteristics were: 1) a domed skull similar in appearance to that of a dolphin's head, 2) bulging eyes, and 3) wrinkles around the mouth. The presence of all three traits is indicative of a severe form of immaturity, whereas the observation of only one or two traits is indicative of a lighter form of immaturity[3,4,14]. A trained person scored the head morphology using a grading system comprising normal mature (N, 78%), light immature (L, 15%), and severely immature (S, 7%).

An independant group of 91 piglets (SuBPig dataset) was used to assess the robustness and pertinence of the prediction. It was constituted of three groups of newborns. The first two groups were performed on 43 piglets from two divergent genetic lines for residual feed intake (RFI). The piglets corresponded to the 10th generation of the genetic lines and were born to two LRFI (low RFI, more efficient) and three HRFI (high RFI, less efficient) sows. Twenty-two piglets from the LRFI line and 21 from the HRFI line were included in the experiment. These first two sets of samples were obtained in an INRAE experimental unit. The third sets of samples was obtained from 48 newborn piglets of LW genotype collected in a private farm.

Blood samples were collected from all animals via the jugular vein, and newborns were phenotyped concurrently with the official weighing at the farms. This is in order to develop a predictive model adapted to real breeding conditions. Blood containing no anticoagulant was left on ice and serum was then prepared by low-speed centrifugation (2,000 g for 10 min at 4°C) and stored at -80°C until analysis.

Just after blood samplings, piglets were weighed (Birth_W) and their body length (crown-to-rump, LENGTH), shoulder width (WIDTH), and chest circumference (CIRC) were measured. Piglet body mass index (BMI, birth weight/crown-to-rump length^2) and ponderal index (PI, birth weight/crown-to-rump length^3) were calculated.

Metabolomic analysis

A metabolomics analysis was also performed by ^{1}H-NMR on serum samples collected on the 278 newborns (99 LR, 87 LW, 92 LRxLW). Each serum sample (200 µL) was diluted in a 500 µL phosphate-buffered solution prepared in deuterated water (0.2 M, pH 7.0) containing TSP (1.17 mM) as internal standard. After centrifugation (5000 g, 4°C, 15 min), 600 µL of supernatant was pipetted and transferred into 5 mm NMR tubes. All proton 1H spectra were acquired at 300 K on a Bruker Avance III HD 600 MHz Nuclear Magnetic Resonance (NMR) spectrometer (Bruker Biospin, Rheinstetten, Germany) operating at 600. 13 MHz for 1H resonance frequency using an inverse detection 5 mm 1H 13C 15N 31P cryoprobe attached to a cryoplatform (the preamplifier unit). During ^{1}H-NMR spectra acquisition, the presaturation of ^{1}H-NMR spectra for solvent signal suppression canceled the residual water contribution. Then, the one-dimensional (1D) Carr–Purcell–Meiboom–Gill (CPMG) pulse sequence attenuated the signal peaks from the macromolecules.

^{1}H-NMR spectra were processed using the R package ASICS, version 2.6.1[15]. Standard pre-processing of the spectra, including import of free induction decay signals, removal of water regions or baseline correction, was applied. All spectra were then aligned using the common alignment procedure[12,15]. Automated metabolite quantification was performed using the ASICS joint quantification procedure. Based on previous analyses on similar data[15], the parameters `noise.thres`, `add.noise` and `mult.noise` were set to 0.01, 0.04, and 0.09, respectively. In addition, the clean-up threshold (clean.thres) for the joint quantification procedure was set to 25%. This means that all metabolites identified in less than 25% of all samples were excluded during the quantification procedure. This choice was based on previous studies and validated by independent direct dosing of metabolites on the same samples[12,15]. Metabolites were identified with the R package `ASICS` and 55 with non-zero variance have been used.

Predictive algorithm

We used the PicLet dataset (278 piglets) to develop and test the model, and used the independent SuBPig dataset (91 piglets) in which the observed maturity was not available to assess the replicability and illustrate the possible usages of the algorithm. The PicLet dataset has been used to develop the predictive model since it came from the PicLet project in which the immature piglets were over-represented to study immaturity.

The PicLet dataset represents several characteristics: imbalanced data (the proportions of the observed maturity are 7% for severe (S), 15% for light and 78% for normal maturity) and small sample size which can lead to high variance when sampling as well as relative high number of feature (p) according to sample size (n).

To address these characteristics, we chose to use an ensemblist approach, bagging[16], by combining two predictive methods: one linear (LASSO or GLM) and one non linear (ExtraTrees). Each of these methods were run 100 times on different sub-samples of the training dataset (70% of the PicLet dataset) using bootstrap. We then aggregate all the 200 "base" learners into a robust meta-learner[16,17], which allowed us to reduce variance due to sub-sampling and increase accuracy. The imbalanced of the classes to predict were addressed by 1) grouping the severe/light maturity and predicting only the mature VS non mature response variable 2) using downsampling for linear method, after the sub-sampling procedure.

The PicLet dataset had 55 available metabolites used as features, as well as two categorical features: the sex (male and female) and the genotype (LW, LR, and LRLW). We first ran our algorithm (using LASSO and ExtraTrees) on training dataset using the 57 features.

From these models, we selected 14 features for their importance in 100 ExtraTrees models[18] (12 selected features) or their coefficient in 100 Lasso models[19] (10 selected features). This selection is presented in section Feature selection. The use of the two methods allowed to select features for their pertinence (higher importance or coefficient) but also their complementarity. Based on this feature selection, we re-performed the previous algorithm using the 14 selected features, but using GML instead of LASSO since $n < p$.

Here we briefly introduced the used methods:

Extremely Randomized Trees (ExtraTrees)[20], is an extension of the Random Forests algorithm, a collection of non-linear regression trees $\mathscr{T} = (T_1, \ldots, T_K)$, with $Y = f_{T_k}(X)$, partially grown at random using three sources of randomness: (i) each tree is grown using a random bootstrapped-with-replacement sample of the data, (ii) the variable used at each split node is selected exclusively from a random subset of all variables and (iii) each splitting value is randomly selected.

Least Absolute Shrinkage and Selection Operator (Lasso) method is used to solve the penalised linear regression problem $Y = X\theta^* + \varepsilon$, where Y is the immaturity score of an individual and the regressors X are the metabolites features while assuming Gaussian distributions of regressors X and Binomial noise distribution (ε) [21], under penalisation criterium $\sum_p \theta_p^* \leq \lambda, 0 \leq \lambda \leq 1$. We used the leave-one out cross-validation method for Lasso to determine an optimal value of the *lambda* parameter[22].

Generalized Linear Model (GLM) method is used to solve the linear regression problem $Y = X\theta + \varepsilon$, where Y is the immaturity score of an individual and the regressors $\boldsymbol{X}$ are the metabolites features while assuming Gaussian distributions of regressors X and Binomial noise distribution (ε), [23].

We used 3 metrics to analyse the results of the predictve algorithm.

For each model (base-learner) j ($j = 1 \ldots J$), the prediction (predicted class) for the individual i is determined as follows:

$$\hat{y}_{ij} = \begin{cases} LS \text{ (immature)} & \text{if } y_{ij} > 0.5 \\ N \text{ (mature)} & \text{otherwise,} \end{cases}$$

with y_{ij} the probability for the individual i to belong to class LS (immature).

From the algorithm several metrics were extracted for each individual i from $Y_i = y_{i1}, y_{i2}, \ldots, y_{iJ}$ the vector of probabilities for the J base models (here J=200).

- **MainPredicted**: the mode class from the J models, referred as $\hat{Y}_i$ for an individual i.

$$\hat{Y}_i = \begin{cases} LS & \text{if } \sum_{j=1}^{J} I(y_{ij} > 0.5) > \sum_{j=1}^{J} I(y_{ij} < 0.5) \\ N & \text{otherwise.} \end{cases}$$

- **MeanPredictions**: the mean accross J models of the predicted probability of belonging to the Immature class $\bar{Y}$ for an indivudual i.

$$\bar{Y}_i = \frac{\sum_{j=1}^{J} y_{ij}}{J}$$

- **Stability**: the stability of the prediction on all models, referred as S_i for an individual i :

$$S_i = \frac{\sum_{j=1}^{J} (\hat{y}_{ij} = \hat{Y}_i)}{J}$$

Technical specifications
This algorithm has been developed (choice of methods, feature selection, and tuning of hyperparameters) using a train dataset, on 195 individuals (70% of the PicLet dataset). The results shown in section Results use a test dataset of 83 individuals (30%), not used to adjust the parameters of the model.

The study was carried out with R and the R packages used for the analysis were `caret` for downsampling, `ranger`[24] for ExtraTrees, `glmnet`[22] for lasso and `stats` for glm.

The parameters used for these models were:

- For ExtraTrees: *num.trees* = 500 with replacement (*replace* = TRUE) and a fraction of replacement *sample.fraction* to 0.8. The importance has been computed by permutation.
- For ExtraTrees with the 57 features: *mtry*=20 and *num.random.split* = 10
- For ExtraTrees with the 14 metabolites features: *mtry*=3 and *num.random.split* = 5
- For Lasso : the loo (leave-one-out) algorithm has been used for the cross validation step then *lambda* set to equal to the largest value of lambda such that error is within 1 standard error of the minimum (*lambda.1se*).
- For both Lasso and glm, we used only 80% of the downsampled sub-dataset to avoid systematic representation of immature individuals.

We present the results on the model based on 14 selected features and discuss later the choice between two models (14 versus 57 features), see section Comparison of the metabolic signature with 55 and 14 metabolites. The set of 14 features gives as good results as the model with 57 features, and it is more rational in terms of number of features needed to run; the less data it requires, the more chance it has to be used in other experiments.

Results

Features selection
Using 57 features, including 55 metabolites and 2 categorical features (sex and genotype), we selected 14 features for the final model based on their importance in ExtraTrees models and the number of non-zero coefficients in LASSO models. All of the 14 features were metabolites. The features were ranked according to these criteria and thresholds were defined as the steps observed in Fig. 1. The thresholds were determined according to the observed gaps in the ranked indicators and a realistic number of metabolites to be obtained in future experiments.

We selected twelve metabolites from ExtraTrees models, and ten from the LASSO models. Eight metabolites were common to both models, six were unique to only one model. For both models, the first two metabolites identified were **myo-inositol** and **valine**. Seven metabolites are proteinogenic amino acids (**aspartate**, **glycine**, **leucine**, **lysine**, **methionine**, **proline**, **valine**). Four of these amino acids are considered essential (**leucine**, **lysine**, **methionine**, **valine**). Two other metabolites are methyl-branched fatty acids (**2-methylglutaric acid**, **isovaleric acid**). The other metabolites are three organic compounds (**creatine**, **isopropyl alcohol**, and **4-ethylphenol**), one inositol isoform (**myo-inositol**), and one choline derivate (**glycerophosphocholine**).

Models results
Table 1 display the results of predicting maturity using the metabolic signature in both the training and test samples. It is important to note that the head morphology was used as a proxy in this study while the signature was designed at the metabolic level. To develop the models, light and severe immature piglets were grouped, and three observed immature were predicted as mature among the 42 immature (light or severe) piglets in the training sample. Out of the 153 mature piglets observed with the proxy feature, 31 were predicted to be immature based on the metabolic signature in the training sample. In the test sample, five out of 20 (25%) observed immature piglets were predicted to be mature. In the test sample, 20 out of 63 (31.7%) observed mature piglets were predicted to be metabolically immature. All the 19 observed severe immature piglets (10 in the training sample and nine in the test sample) were predicted to be immature.

The prediction output consisted of two classes: mature or immature (Table 1). Additionally, two quantitative values are provided: the mean prediction and the stability (Fig. 2). The mean prediction is a quantitive value ranging from 0 to 1, with values below 0.5 indicating maturity and values above indicating immaturity.

Figure 1. Feature selection from **(a) ExtraTrees** and **(b) LASSO** models.

Table 1. Results, for train and test samples.

	Predicted maturity on **train sample**			Predicted maturity on **test sample**		
Observed maturity	Normal	Immature	Total	Normal	Immature	Total
Normal	122	31	153	43	20	63
Light	3	29	32	5	6	11
Severe	0	10	10	0	9	9
Total	125	70	195	48	35	83

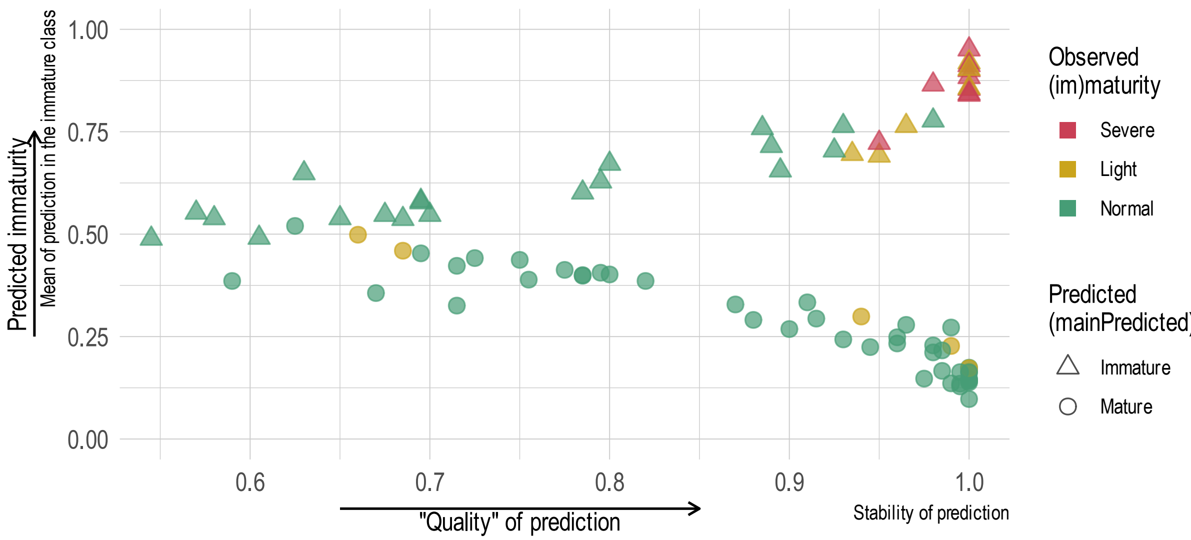

Figure 2. Results of the prediction model with 14 metabolites on test sample

We can see in Fig. 2 that a combination of the three criteriums (predicted maturity, mean prediction, and stability of the prediction) allows us to determine whether the individuals can be predicted as immature, but also the severity of immaturity (although that information was not included in the training models).

Fig. 3 illustrates the projection of the individuals from the test samples using the metabolic signature from the 14 selected variables. The principal component analysis (PCA) demonstrates a clear separation between the barycenters of the three observed maturity groups, ranging from normal on the right to severe immaturity on the left (Fig. 3-a). Fig. 3-b shows the projection of the mean predicted value of maturity for each piglet illustrating a similar projection to the observed maturity.

0.1 Comparison of the metabolic signature with 55 metabolites and the metabolic signature with 14 metabolites

To reduce the risk of missing data, it may be beneficial to decrease the number of required metabolites when predicting with a new dataset. The set of variables selected with non-zero variance consisted of 57 features (55 metabolites plus the sex and the genotype). To prevent missing data in a future independent ^{1}H-NMR study, a specific set of metabolites was chosen from both models. Fig. 4 shows a nearly perfect identical prediction between the two models. In addition to distinguishing piglets on the PCA (Fig. 3), a set of 14 metabolites allowed for accurate prediction of their metabolic status.

Investigation of the biological relevance of the molecular signature

Interest of the predictive model for genetics

In genetics, identifying sows and boars with a higher potential to produce mature offspring is of great interest. The experiment used to develop the molecular signature involved 51 sows and 24 boars. Fig. 5 presents a comparison between the observed and

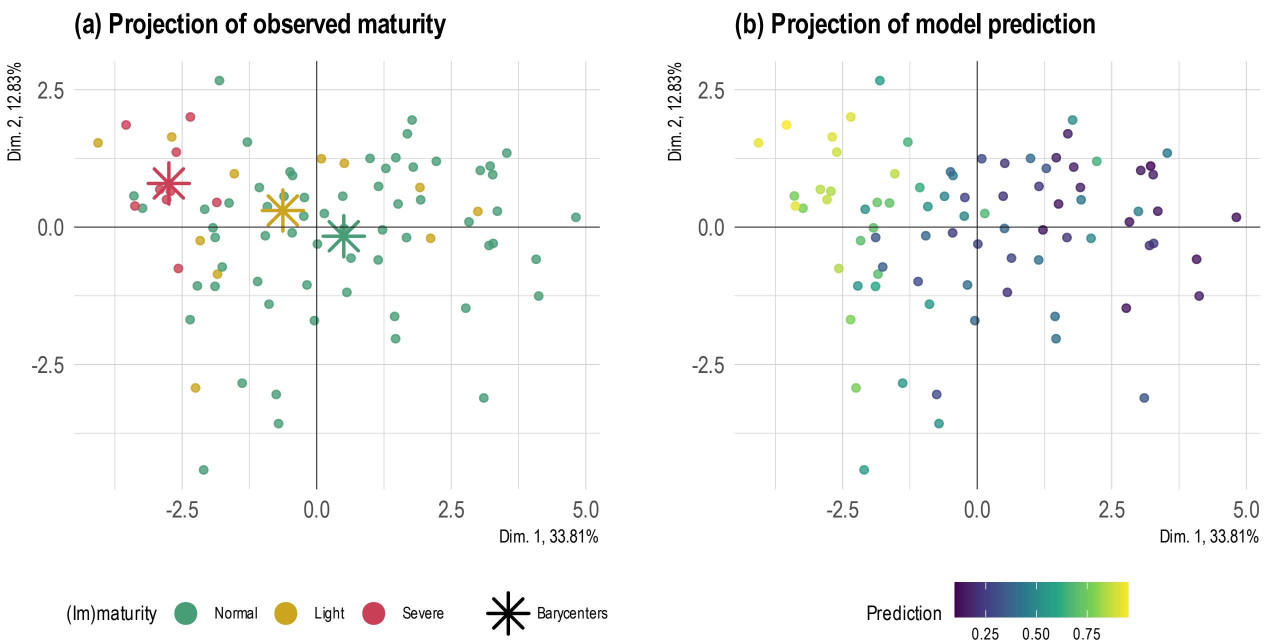

Figure 3. PCA on 14 metabolites, projection of individuals from test sample on PCA first 2 components. Comparison between (**a**) observed normal maturity or light/severe immaturity and (**b**) mean metabolic prediction of immaturity

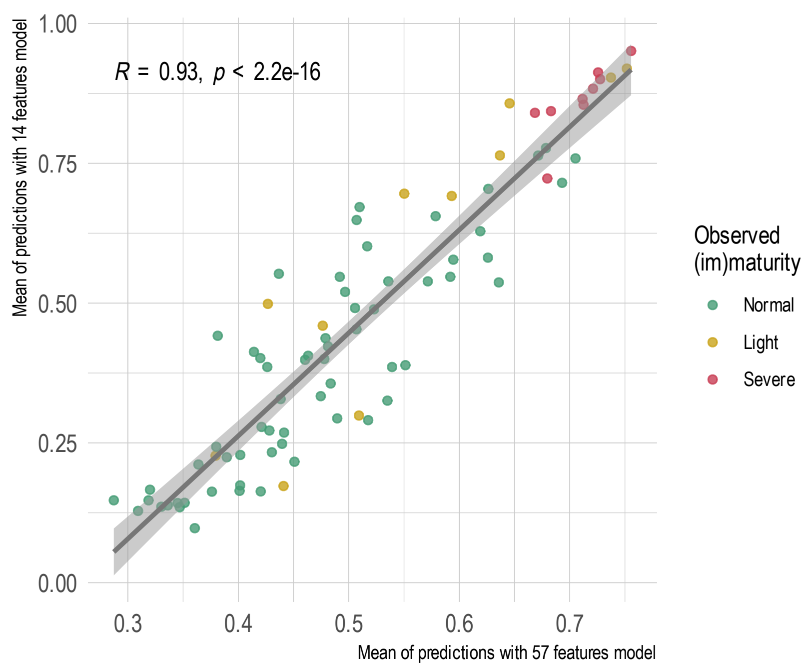

Figure 4. Comparison of mean predictions of immaturity by piglet, from the models with 14 or 57 variables, on test sample

Table 2. Results on **dead piglets**, for train and test samples.

	Predicted maturity on train sample			Predicted maturity on test sample		
Observed maturity	Normal	Immature	Total	Normal	Immature	Total
Normal	5	1	6	4	3	7
Light	0	7	7	0	3	3
Severe	0	4	4	0	4	4
Total	5	12	17	4	10	14

predicted percentages of mature piglets in the offspring of boars and sows. It is important to interpret this observation with caution as not all piglets from the litters were available in the PicLet project. However, we can observe that some sows and boars may have a main tendency to provide mature or immature piglets. Some sows (numbers 28 to 49) and boars (numbers 14 to 17) were identified with only mature offspring with the observed proxy (head shape), thirteen of them (10 sows and 3 boars) had offspring with a lower predicted metabolic maturity. More often, the percentages of observed mature piglets per litter were 60-100% and the percentages of predicted mature piglets were mostly lower. For three sows (numbers 5, 13, and 16) and two boars (numbers 2 and 4), the percentage of predicted metabolic maturity was better than with the observed proxy.

Furthermore, the most commonly employed proxy for maturity is the birth weight. Fig. 6-a illustrates the variation in birth weight in relation to the mean prediction measure of maturity. Despite a notable correlation (-0.68) was observed between the two variables, among the 35 piglets predicted as immature at the metabolic level, 43% (15/35) exhibited a birth weight exceeding 1 kg. Conversely, these 15 piglets, representing 17% of the total number with a weight above 1 kg, were predicted to be immature. Finally, nine piglets among the 28 piglets (32%) with a birth weight below 1kg were predicted to be mature.

Finally, a major concern for piglets at birth is the risk of mortality. In this experiment, there were too few piglets to statistically evaluate the relationship between mortality and the prediction of maturity. Nevertheless, in Table 2, we observed that all the 18 piglets (of the training and test samples) that died during the first three days after birth and were observed to be immature were indeed predicted to be immature (Light or Severe). Of the 13 piglets (of the training and test samples) that were observed to be mature but died, only four were predicted to be immature. Further investigation will be required to identify the underlying causes of mortality, i.e., whether or not this mortality could be related to the maturity status.

Assessing metabolic maturity in an independent experiment

A true independent experiment with piglets scored on head morphology, and the molecular signature was not available. To further assess the value of the metabolic signature, serum samples from independent experiments on piglets collected at birth were used to predict their maturity status using the prediction model. In Fig. 6-b, a small number of piglets were predicted to be immature (red dots). In this independent experiment, unlike the test sample (Fig. 6-a), piglets were not selected for severe immaturity. A negative correlation (-0.34) was observed between birth weight and the mean predicted value (as for the test samples in Fig. 6-a). Nevertheless, 12/16 of the predicted immature piglets in the independent data (Fig. 6-b) had a birth weight above 1 kg, suggesting that the prediction model can identify metabolic immaturity regardless of weight.

In this independent experiment, even when head morphology was not available, other biometric values were measured in addition to birth weight: crown-to-rump length, width between the shoulders, chest circumference, body mass index, and ponderal index. All of these measures are considered to be higher with an expected better maturity. A Partial Least Squares regression (PLS)[25] analysis between morphological values and the 14 metabolites enabled the projection of the piglets along the first axis, which was coloured according to the predicted maturity (Fig. 7-a). Fig. 7-b shows the variables projection from the two datasets with a correlation above 0.6 with axes. Higher biometric values are effectively more positively correlated with piglets being predicted to be mature.

The four metabolites more correlated (> 0.6) with the axes are 2-methylglutaric acid, valine, and glycine, which are more correlated with maturity, while myo-inositol is more correlated with immaturity. This was consistent with the expected projection according to the prediction model. In another way, the PLS projection of individuals showed that the morphological measures didn't fully explain immaturity status.

Discussion

The ability to predict the maturity of pigs at birth is of paramount importance in improving neonatal survival and reducing mortality[26], which may average 16 to 20 % in piglets[27] or 10 % to 20%[1]. Maturity is influenced by complex intra-uterine developmental processes and is a significant determinant of postnatal growth and health in mammalian species[28–30]. The comprehension of maturity can result in the improvement of breeding strategies and management practices. Predicting

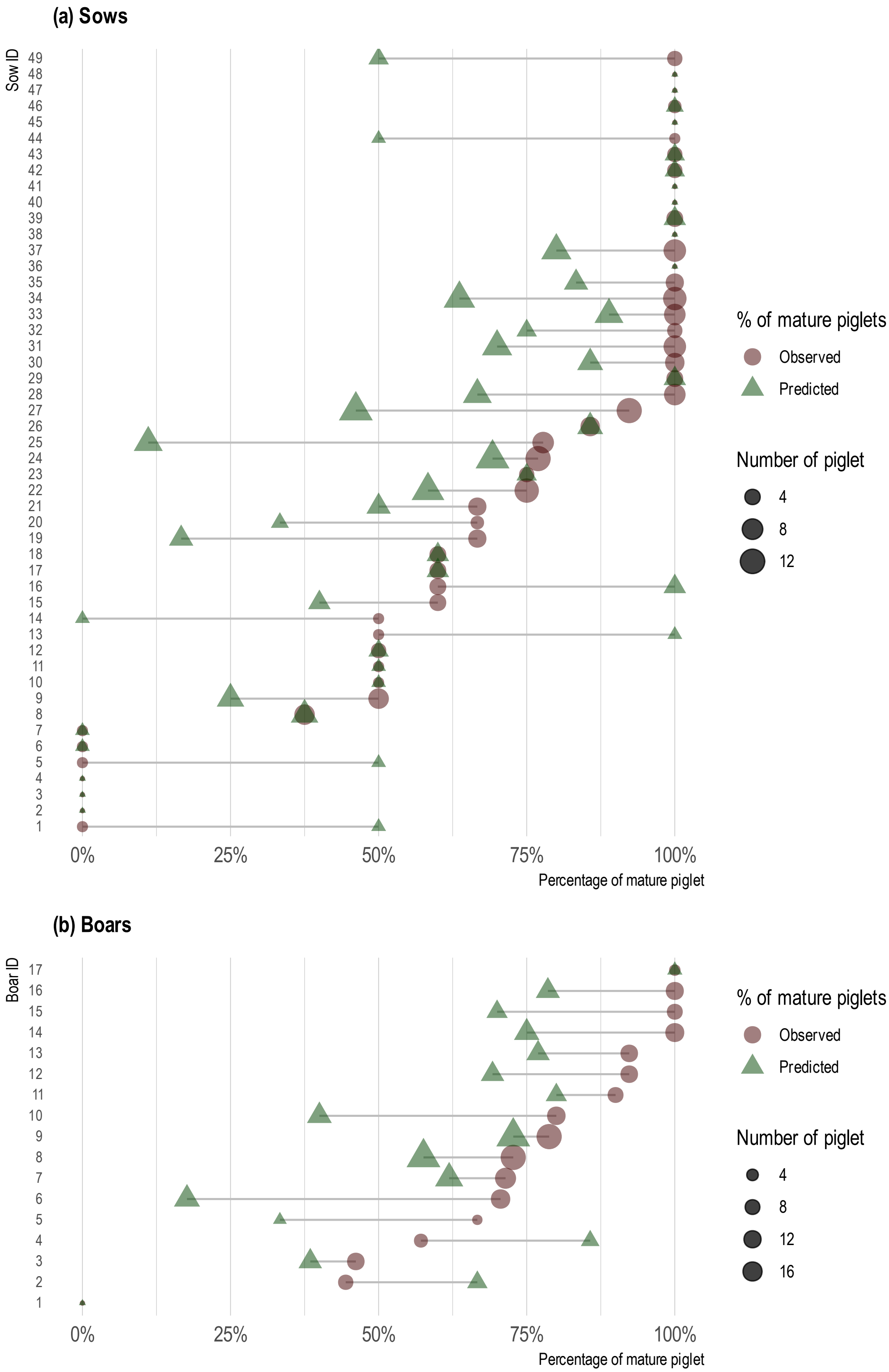

Figure 5. Observed and predicted maturity by **(a)** sow (278 piglets from the PicLet experiment), **(b)** boar (233 piglets from the PicLet experiment, with an identified boar)

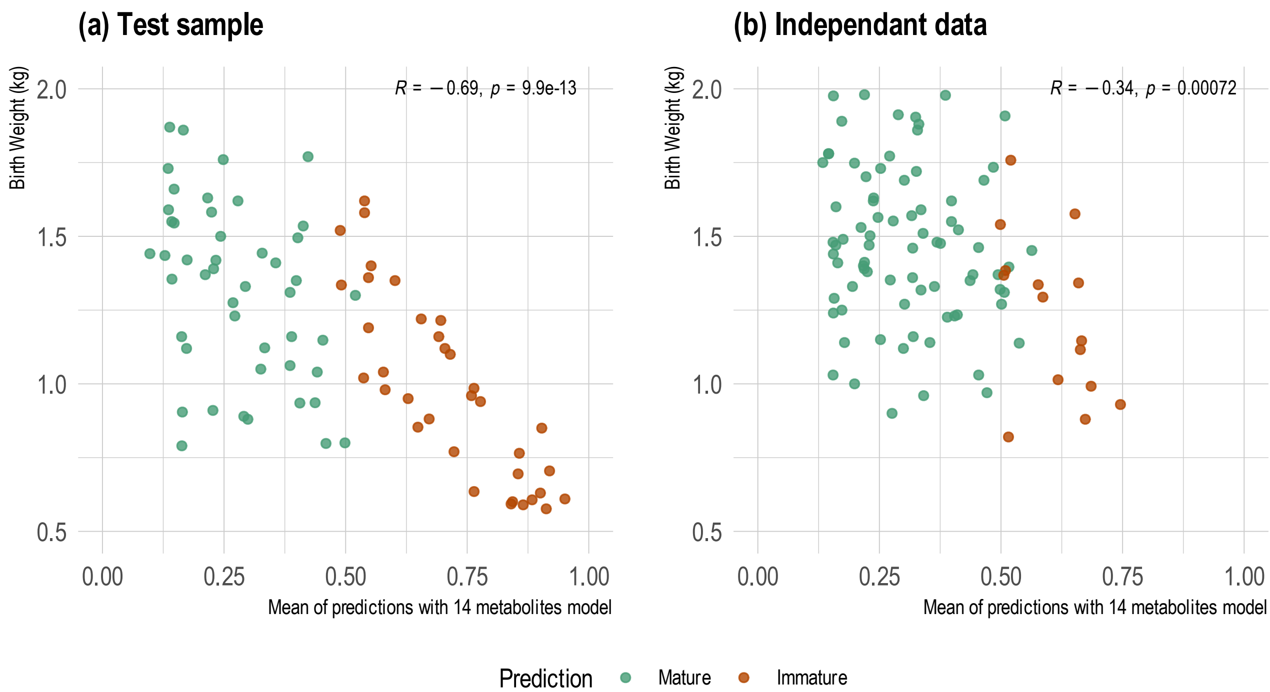

Figure 6. Relation between birth weight and model prediction on **(a)** test sample and **(b)** independant data

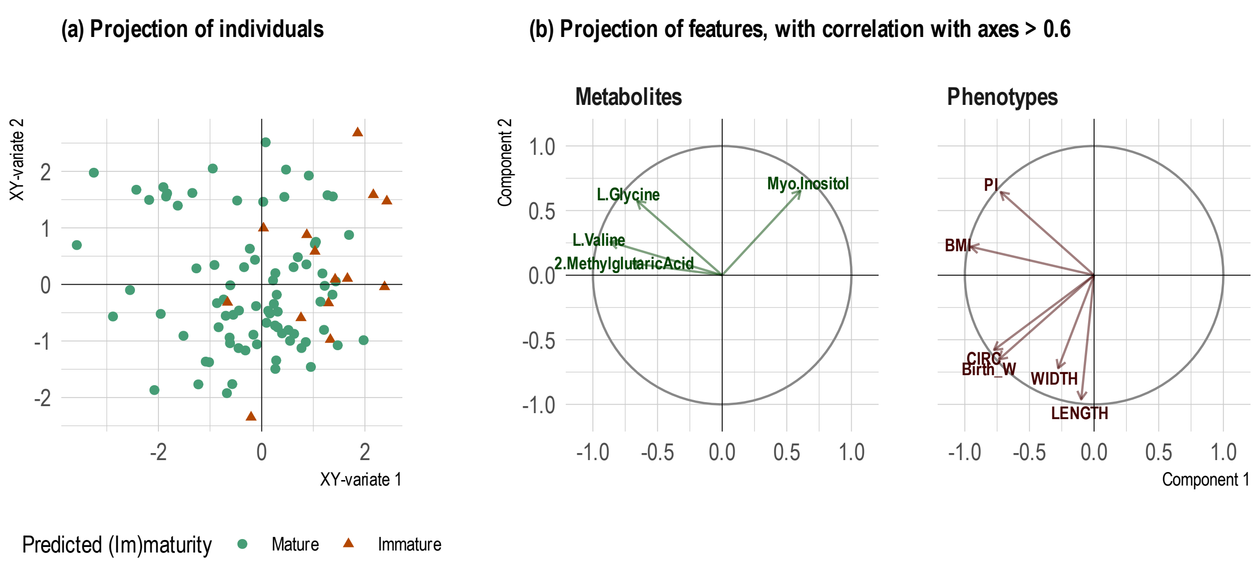

Figure 7. PLS (Partial Least Square) on independent data using the 14 selected metabolites and phenotypic features with (a) the projection of newborn piglets and (b) the projection of the features, one for the metabolites and the other one for the phenotypes. PI, ponderal index; BMI, body mass index; CIRC, chest circumference; Birth_W, birth weight; WIDTH, width between the shoulders; LENGTH, crown-to-rump length

immaturity in newborn piglets is therefore a challenging task[26]. To address this issue, we developed a predictive model based on serum metabolomics using animals with phenotyped head shape as a proxy for immaturity[3]. Despite the relatively small sample size of 278 piglets, our model combines the results of two different approaches (Random Forest and GLM Lasso) to provide accurate predictions. However, it is important to note that validation on a completely independent set of newborns using head shape as a proxy was not available. Therefore, a careful methodology was adopted. Rather than relying on a morphological proxy, maturity was assessed at the metabolic level. The shape of a newborn's head can indicate extreme immaturity, while intermediate immaturity can be difficult to detect. Being able to define the metabolic status of newborns would help in identifying those at risk more accurately. In this study, we developed a metabolic signature consisting of 14 metabolites that can predict all observed extremely immature piglets as metabolically immature. According to the metabolic signature, almost all piglets (90%) with mild morphological immaturity were classified as metabolically immature. Piglets with normal morphology were mainly classified as mature (74%), which means that a quarter of these piglets are classified as immature at the metabolic level. These results suggest that the morphology of the dolphin-shaped head is indeed characteristic of extreme immaturity, and that the molecular signature can predict even mild metabolic immaturity.

During the modelling phase, sex and genotype were included in the analysis in addition to metabolites. Many studies has demonstrated that gender is a determinant factor influencing survival rates[14]. Despite the fact that males have a higher birth weight, mortality seems to be more prevalent in male piglets[31]. In this study, gender was not retained in the final model. Genetics also plays a role in determining the survivability of animals, with breeds and crossbreeding showing notable effects. For example, a recent study[32] comparing Landrace, Duroc and their crosses showed that crossbreeding is a beneficial approach. Despite the low heritability of mortality at birth, Knap et al[33] demonstrated that genetics may improve this complex trait, particularly when considering underlying phenotypic traits, such as fetal maturity. Moreover, the specific trait under investigation was piglet maturity at birth, which may contribute to mortality and potentially influence subsequent growth and health. In our study, sex and genotype were not retained in the final molecular signature. This enabled the creation of a model that could be utilised irrespective of sex and genotype. Moreover, the availability of an experimental set of piglets with three genotypes (Landrace, Large White and crossed LWLR) enabled the development of a predictive model that would be more valuable for a larger variety of breeds. The 14-metabolites signature model was subsequently applied to an independent set of newborns of three genotypes: two INRAE experimental divergent LW breeds selected for higher (HRFI) and lower (LRFI) residual feed intake, and another LW breed from a private commercial farm. It is regrettable that no data was available regarding the shape of the head. However, alternative proxies for maturity at birth were employed, including the body mass index, ponderal index[26], chest circumference, shoulder width, and birth weight. Consequently, the metabolic signature was indirectly validated. Further experimental testing is required to provide a more rigorous validation of the predictive power of the model on other breeding systems. One potential avenue for further investigation could be the use of head shape as a proxy, in conjunction with other genotypes. It would also be of interest to test the model on an organic system, where mortality at birth is often higher than in a conventional system[32]. It would also be beneficial to further clarify the rationale for attempting to qualify the parental propensity to produce litters and offspring with a higher level of maturity on a global scale. Our data suggest that this is a viable proposition.

It is of greater interest to note that the variable of birth weight was not included in the modelling process, despite the high correlation between this trait and survival. This was due to the fact that one of the objectives was to identify factors other than weight. In humans, birth weight, while important, is not considered to be completely sufficient to truly capture the degree of immaturity of a newborn baby[29]. Previous studies in pigs have highlighted the need to distinguish between weight, maturity and survival in order to identify physiological traits that may explain the observed decline in survival during the last decade of genetic selection for litter size and piglet birth weight[26,34]. In essence, the proposed metabolite signature is capable of predicting a maturity status independently of weight. It is anticipated that this will facilitate the prediction of a metabolic status at birth, which is not directly associated with a morphological trait. The experimental design included an insufficient number of newborns that died to enable a valid evaluation of the predictive power of the mortality prediction model. However, it is noteworthy that a dead piglet in the first three days after birth was predicted to be immature more often than with the observed maturity status.

In humans, a recent review summarised the results of metabolomic studies in maternal, fetal and neonatal samples associated with fetal growth restriction and small for gestational age[35]. In this review, the authors didn't discuss maturity, but traits that are somewhat related to it (growth restriction, gestational age). All 14 metabolites of the predictive signature (except two) were found to be regulated in at least one of these 48 neonatal studies.

Myo-inositol (or inositol phosphates IP6) was identified as the primary ranked metabolite among the fourteen predictive metabolites. Myo-inositol belongs to the class of lipids named phosphatidylinositol, with a central role in cell signal transduction and energy metabolism[36]. We previously identified myo-inositol in plasma, urine, and amniotic fluid at the end of gestation[10] with decreased quantification at day 110 compared to day 90 of gestation (birth at day 114). The concentration of myo-inositol was found to be higher in Large white than in Meishan fetuses while Meishan newborns are known to be more robust, while the mortality rate of Large White fetuses is higher at birth[37]. Myo-inositol has been identified in the urine of newborns by means of ^{1}H-NMR and proposed as a biomarker for human neonates with intrauterine growth retardation[38].

It is notable that seven of the 14 predictive metabolites are proteinogenic amino acids (aspartate, glycine, leucine, lysine, methionine, proline, valine). Amino acids have numerous biological functions in addition to their role in protein synthesis. One such amino acid is valine, a branched-chain amino acid (BCAA) that is the second most important predictive metabolite. In the present study, valine was found to be more associated with higher maturity, and associated with higher BMI and PI in the independent samples. Another branched-chain amino acid (BCAA), leucine, was identified as a predictive metabolite. An essay with supplementation of lactating sows with BCAA resulted in a reduction in preweaning mortality in another study[39]. Luise et al.[40] confirmed this observation only if sows were supplemented with BCAA together with arginine. However, the two publications suggested a positive impact on the growth of piglets if sows were supplemented with BCAA. In another way, five predictive metabolites of the 14, four amino acids (valine, leucine, glycine, proline) and creatine, were differentially detected by ^{1}H-NMR in the plasma of newborn calves during sepsis, another high risk for neonatal death or impaired health[41].

Another metabolite of the predictive signature is glycerophosphocholine, which is a choline derivative that plays a role in brain development and myelination during the neonatal period[42]. Its presence in the metabolic signatures underscores the significance of choline metabolism in achieving maturity at birth. This metabolism is a key component of the one-carbon metabolic pathway and the methionine cycle. In addition, methionine is closely linked to choline metabolism, affecting its availability and utilisation during fetal growth[43]. The brain of piglets exposed perinatally to lower choline status was found smaller than the brain of those exposed to sufficient choline quantity[44].

The way in which the quantification of these metabolites has changed compared to healthy neonates and compared between immediately after birth and in the hours after birth is somewhat difficult to interpret. This finding suggests that quantifying isolated metabolites may present a challenge in terms of interpretation, and that a more complex signature may facilitate overcoming the physiological variability underlying the trait of interest, in this case, maturity.

Conclusions

Genetic selection for survival at birth and growth is mainly limited to the measurement of birth weight. As these traits are correlated, it is essential to find a way to unravel these correlations to reveal the underlying molecular regulation. One of our aims was to develop a simple method of collecting a minimally invasive sample, blood, to perform a low-cost and easy-to-use metabolic analysis on serum, with the acquisition of ^{1}H-NMR spectra coupled with the automatic identification and quantification of metabolites using the ASICS R package. Consequently, identifying a molecular signature could help future experiments to decipher the genetic architecture of a complex trait, in this case, maturity. Further experiments are needed to investigate the biological significance of each of the 14 metabolites in more detail.

Acknowledgements

The authors would like to thank the Pig breeding companies for allowing access to the farms of their selection scheme. The authors are very grateful to the breeders and technicians from the commercial farms who allowed access and helped with the sampling and phenotyping of the piglets. Many thanks to the veterinarians who carried out the blood sampling and to Léa André for the piglet head phenotyping. The authors also thank the staff of the experimental pig facilities (INRAE, 2018. Pig phenotyping and innovative breeding facility, https://doi.org/10.15454/1.5572415481185847E12). They also thank all the participants of the GenPhySE, GenESI (INRAE) laboratories for their help in collecting samples and data.

Author contributions statement

EM and LL developed the statistical models and wrote the manuscript. EM performed the data analyses. NMG performed all the NMR studies on the MetaToul-AXIOM plateform under the supervision of CC. NMG performed all the raw spectra analyses to the identification and quantification of the metabolites of the both studies (PicLet and SuBPig projects). LG carried out the piglet phenotyping, samples collection, samples and data management. AB participated in the conceptualization of the study, carried out the piglet phenotyping and samples collection. PB supervised the PicLet project, performed and supervised the phenotyping of the piglets. LL conceived and designed the both molecular studies of the two projects (PicLet and SuBPig). LL managed and supervised the SuBPig project. All authors reviewed and approved the manuscript.

Additional information

Competing interests

The authors declare that they have no competing interests.

Funding
This work is part of the PicLet project funded by CASDAR-RT 2019. We thank the animal and sample providers (Nucleus and Axiom breeding organisations of the ALLIANCE R&D). Additional samples for validation (independent data) were collected in the framework of the SuBPig project "Enhancing survival at birth in piglets" funded by the metaprogram INRAE GISA (Integrated Animal Health Management) 2018-2019, Grant n° DI2017. The SuBPig animals were raised at the French Experimental Unit GenESI, INRAE, UE1372, Surgères, France, doi: 10.15454/1.5572415481185847E12.

Ethics approval
The data from the PicLet project was conducted in accordance with the French legislation on experimentation and ethics. Authors obtained the authorization from the "Comité d'Ethique en expérimentation animale, Sciences et Santé animales N°115" delivered by the Veterinary school of Toulouse with the n° SSA_2020_006. The independant data from the SuBPig project was conducted in accordance with the French legislation on experimentation and ethics. The French Ministry of Research and Innovation authorized this experiment on living animals at the INRAE facilities UE1372 GenESI[13] with the agreement number APAFiS 13648-2018020417291866 v4. In these conditions, this study follows the ARRIVE guidelines (Animal Research: Reporting of In Vivo Experiments), and is committed to the 3Rs of laboratory animal research and consequently used the minimal number of animals to reach statistical significance.

Data availability
All raw and treated data are available, on INRAE Omics Dataverse repository, Recherche Data Gouv.

The ^{1}H-NMR datasets, the one with the 55 metabolites quantified on the serum of 278 newborns (PicLet project) used to develop the predicitive model, and the one with the 53 metabolites quantified on the serum of 95 newborns from the SuBPig project are available on https://doi.org/10.57745/8OZUKK together with a file for the metadata for the 373 piglets (278 PicLet and the 95 SuBPig) used in this study.

Raw spectra data for the ^{1}H-NMR spectra of the PicLet project are also available on https://doi.org/10.57745/8OZUKK.

Raw ^{1}H-NMR spectra of the SuBPig project are available on the INRAE Omics Dataverse repository from https://doi.org/10.57745/QPVLNS.

Code availability
The code developped for this paper is available at https://forge.inrae.fr/piclet/metabolicprediction.

References

1. Muns, R., Nuntapaitoon, M. & Tummaruk, P. Non-infectious causes of pre-weaning mortality in piglets. *Livest. Sci.* **184**, 46–57, DOI: 10.1016/j.livsci.2015.11.025 (2016).
2. Canario, L. *et al.* Estimation of genetic trends from 1977 to 1998 of body composition and physiological state of large white pigs at birth. *Animal* **1**, 1409–1413, DOI: 10.1017/s1751731107000766 (2007).
3. Amdi, C. *et al.* Intrauterine growth restricted piglets defined by their head shape ingest insufficient amounts of colostrum. *J. Animal Sci.* **91**, 5605–5613 (2013).
4. Hansen, C., Hales, J., Amdi, C. & Moustsen, V. Intrauterine growth-restricted piglets defined by their head shape have impaired survival and growth during the suckling period. *Animal Prod. Sci.* **59**, 1056–1062 (2018).
5. Voillet, V. *et al.* Muscle transcriptomic investigation of late fetal development identifies candidate genes for piglet maturity. *BMC genomics* **15**, 1–17 (2014).
6. Voillet, V. *et al.* Integrated analysis of proteomic and transcriptomic data highlights late fetal muscle maturation process*. *Mol. & Cell. Proteomics* **17**, 672–693, DOI: 10.1074/mcp.M116.066357 (2018).
7. Marti-Marimon, M. *et al.* A new approach of gene co-expression network inference reveals significant biological processes involved in porcine muscle development in late gestation. *Sci. reports* **8**, 10150 (2018).
8. Gondret, F. *et al.* Proteomic analysis of adipose tissue during the last weeks of gestation in pure and crossbred large white or meishan fetuses gestated by sows of either breed. *J. Animal Sci. Biotechnol.* **9**, 1–16 (2018).
9. Yao, Y. *et al.* Comparing the intestinal transcriptome of meishan and large white piglets during late fetal development reveals genes involved in glucose and lipid metabolism and immunity as valuable clues of intestinal maturity. *BMC genomics* **18**, 1–17 (2017).

10. Lefort, G. *et al.* The maturity in fetal pigs using a multi-fluid metabolomic approach. *Sci. Reports* **10**, 19912, DOI: 10.1038/s41598-020-76709-8 (2020).

11. Schmitt, O. *et al.* Thermoregulation at birth differs between piglets from two genetic lines divergent for residual feed intake. *Animal* **15**, 100069, DOI: 10.1016/j.animal.2020.100069 (2021).

12. Liaubet, L. *et al.* Plasma 1h-nmr metabolic and amino acid profiles of newborn piglets from two lines divergently selected for residual feed intake. *Sci. Reports* **13**, 7127, DOI: 10.1038/s41598-023-34279-5 (2023).

13. GenESI. Pig Innovative Breeding Experimental Facility, Surgères, France, DOI: 10.15454/1.5572415481185847E12 (2018).

14. Hales, J., Moustsen, V. A., Nielsen, M. B. F. & Hansen, C. F. Individual physical characteristics of neonatal piglets affect preweaning survival of piglets born in a noncrated system1. *J. Animal Sci.* **91**, 4991–5003, DOI: 10.2527/jas.2012-5740 (2013).

15. Lefort, G. *et al.* Joint automatic metabolite identification and quantification of a set of 1h nmr spectra. *Anal. Chem.* **93**, 2861–2870, DOI: 10.1021/acs.analchem.0c04232 (2021).

16. Breiman, L. Bagging predictors. *Mach. learning* **24**, 123–140, DOI: 10.1007/BF00058655 (1996).

17. Boehmke, B. & Greenwell, B. M. *Hands-on machine learning with R* (CRC press, 2019).

18. Genuer, R., Poggi, J.-M. & Tuleau-Malot, C. Variable selection using random forests. *Pattern recognition letters* **31**, 2225–2236, DOI: 10.1016/j.patrec.2010.03.014 (2010).

19. Lima, E., Davies, P., Kaler, J., Lovatt, F. & Green, M. Variable selection for inferential models with relatively high-dimensional data: Between method heterogeneity and covariate stability as adjuncts to robust selection. *Sci. Reports* **10**, 8002, DOI: 10.1038/s41598-020-64829-0 (2020).

20. Geurts, P., Ernst, D. & Wehenkel, L. Extremely randomized trees. *Mach. learning* **63**, 3–42 (2006).

21. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. Royal Stat. Soc. Ser. B (Methodological)* **58**, 267–288 (1996).

22. Friedman, J. *et al.* Package 'glmnet'. *CRAN R Repositary* DOI: 10.32614/CRAN.package.glmnet (2021).

23. Nelder, J. A. & Wedderburn, R. W. Generalized linear models. *J. Royal Stat. Soc. Ser. A: Stat. Soc.* **135**, 370–384 (1972).

24. Wright, M. N., Wager, S., Probst, P. & Wright, M. M. N. Package 'ranger'. *Version 0.11* **2**, DOI: 10.32614/CRAN.package.ranger (2019).

25. Tenenhaus, M. *La régression PLS: théorie et pratique* (Editions technip, 1998).

26. Baxter, E. *et al.* Investigating the behavioural and physiological indicators of neonatal survival in pigs. *Theriogenology* **69**, 773–783, DOI: 10.1016/j.theriogenology.2007.12.007 (2008).

27. Edwards, S. & Baxter, E. Piglet mortality: causes and prevention. In *The gestating and lactating sow*, 253–278, DOI: 10.3920/9789086868032_012 (Wageningen Academic, 2015).

28. Leenhouwers, J., Knol, E., De Groot, P., Vos, H. & Van der Lende, T. Fetal development in the pig in relation to genetic merit for piglet survival. *J. Animal Sci.* **80**, 1759–1770, DOI: 10.2527/2002.8071759x (2002).

29. Behrman, R. E., Butler, A. S. *et al.* *Preterm birth: causes, consequences, and prevention* (National academies press Washington, DC, 2007).

30. Basso, O. & Wilcox, A. Mortality risk among preterm babies: immaturity versus underlying pathology. *Epidemiology* **21**, 521–527, DOI: 10.1097/EDE.0b013e3181debe5e (2010).

31. Baxter, E. M., Jarvis, S., Palarea-Albaladejo, J. & Edwards, S. A. The weaker sex? the propensity for male-biased piglet mortality. *PLoS One* **7**, e30318 (2012).

32. Chu, T. T., Zaalberg, R. M., Bovbjerg, H., Jensen, J. & Villumsen, T. M. Genetic variation in piglet mortality in outdoor organic production systems. *animal* **16**, 100529 (2022).

33. Knap, P. W. *et al.* Genetic and phenotypic time trends of litter size, piglet mortality, and birth weight in pigs. *Front. Animal Sci.* **4**, 1218175 (2023).

34. Knol, E. F., van der Spek, D. & Zak, L. J. Genetic aspects of piglet survival and related traits: a review. *J. Animal Sci.* **100**, skac190, DOI: 10.1093/jas/skac190 (2022).

35. Conde-Agudelo, A. *et al.* Metabolomic signatures associated with fetal growth restriction and small for gestational age: a systematic review. *Nat. Commun.* **15**, 9752 (2024).

36. Chatree, S., Thongmaen, N., Tantivejkul, K., Sitticharoon, C. & Vucenik, I. Role of inositols and inositol phosphates in energy metabolism. *Molecules* **25**, 5079 (2020).

37. Canario, L. *et al.* Between-breed variability of stillbirth and its relationship with sow and piglet characteristics. *J. animal science* **84**, 3185–3196 (2006).

38. Dessì, A. *et al.* Metabolomics in newborns with intrauterine growth retardation (iugr): urine reveals markers of metabolic syndrome. *The J. Matern. & Neonatal Medicine* **24**, 35–39 (2011).

39. Rezaei, R., Gabriel, A. S. & Wu, G. Dietary supplementation with branched-chain amino acids enhances milk production by lactating sows and the growth of suckling piglets. *J. Animal Sci. Biotechnol.* **13**, 65 (2022).

40. Luise, D. *et al.* Productive and physiological implications of top-dress addition of branched-chain amino acids and arginine on lactating sows and offspring. *J. animal science biotechnology* **14**, 40 (2023).

41. Basoglu, A., Sen, I., Meoni, G., Tenori, L. & Naseri, A. Nmr-based plasma metabolomics at set intervals in newborn dairy calves with severe sepsis. *Mediat. Inflamm.* **2018**, 8016510 (2018).

42. Derbyshire, E. & Obeid, R. Choline, neurological development and brain function: a systematic review focusing on the first 1000 days. *Nutrients* **12**, 1731 (2020).

43. Mudd, A. T., Getty, C. M., Sutton, B. P. & Dilger, R. N. Perinatal choline deficiency delays brain development and alters metabolite concentrations in the young pig. *Nutr. neuroscience* **19**, 425–433 (2016).

44. Getty, C. M. & Dilger, R. N. Moderate perinatal choline deficiency elicits altered physiology and metabolomic profiles in the piglet. *PLoS One* **10**, e0133500 (2015).