This document is the Accepted Manuscript version of a Published Work that appeared in final form in Analyst, 2020, 145, 6859–6867, copyright © The Royal Society of Chemistry 2020. To access the final edited and published work see https://doi.org/10.1039/D0AN00254B

ARTICLE

Comparison of liver and plasma metabolic profiles in Piglets of different ages as animal models for paediatric population

Oihane E. Albóniga*[a], Oskar González[a], Rosa M. Alonso[a], Yun Xu[b], Royston Goodacre[b]

Liver plays an important role in drug metabolism, so studying the grade of maturation of this organ would help to develop more appropriate dosing regimens for paediatric populations. Nevertheless, considering the invasive nature of liver analyses there are obvious ethical boundaries, particularly in babies and children. In this work, we investigated the suitability of blood plasma as an alternative matrix to evaluate the biological age of liver. With this aim, we studied the correlation of plasma and liver metabolomic profiles obtained by HPLC-TOF-MS for piglets of different ages (newborns, neonates and infants). By means of Pearson correlation analysis we observed that 360 and 1784 pairs of metabolite features were significantly correlated in positive and negative ionization mode, respectively. Procrustes analysis was applied in order to assess the similarity of the clustering resulting from the data obtained from the two matrices and the two ionisation modes. The Procrustes distances were low for both ESI+ (0.3753) and ESI- (0.3673) and, hence, liver and plasma are expected to provide similar discriminatory information. Furthermore, we found that Multiblock Principal Component Analysis (MB-PCA) readily allowed us to combine the data obtained from both matrices and to better understand the clustering according to the three study groups. Considering all these results, we suggest that plasma can provide valuable insight into the maturation grade of liver in order to provide accurate dosing in paediatric population.

1. Introduction

Clinical studies involving paediatric population are scarce in part due to principal difficulties: ethical issues(1-4) and difficulties in obtaining a statistical meaningful population that can be used to estimate confidence intervals due to the distributions are heterogenic in general(5). As a consequence there is a lack of knowledge about drug metabolism in children and the amount of authorized drugs for paediatric use is limited. For these reasons, drug administration in children is normally made "off-label", which involves the usage of drugs approved for adults for the treatment of paediatric diseases. The drug dosage for children is normally calculated by an empirical approach from adults, based on bodyweight (mg_{drug}/kg), and assuming a linear relationship between them(6). This approximation can sometimes be inadequate, becoming inefficient or even toxic. Furthermore, it is well known that the pharmacokinetic response to a drug is substantially different in paediatric populations compared with adults because drug disposition is closely related to the organ maturation state(7,8). In some cases, the liver capacity for drug metabolism is insufficient in children (immature organ) and even lower doses than those calculated from adults are required(9). Organ maturation is a dynamic process that occurs not only during fetal life, but also during neonatal and childhood period. This maturation depends highly on the activation and repression of genes, or sets of genes, and it is related to the growth, which at the same time depends on environmental factors, endocrine regulation and nutrition(10). Therefore, all these factors lead to situations in which children with the same chronological age could have a different biological age. As a consequence, a thorough understanding of human developmental biology in neonates, infants and children is required to achieve an effective and safe drug therapy(8)and a better personalized medicine(11).

Metabolomics is a promising tool that has the potential to provide information about the maturation state of organs by means of investigating changes in organ metabolism. Metabolic profiles are commonly obtained from matrices such as urine, blood, plasma and serum because these are less invasive compared to tissue biopsies or other biofluids such as cerebrospinal fluid (CSF)(12). However, the most adequate matrix to study organ maturation state is the organ tissue itself, as this will contain organ-specific metabolites (13). Obviously, the inconvenience of tissue sample collection limits the direct analysis of tissue matrices(14,15).

This limitation leads to the use of animal models such as minipigs or piglets. Piglets are widely used as surrogates for clinical trials in paediatric populations due to the piglets size, similarity, physiology, organ development, and disease progression(16-20). The knowledge obtained from metabolic profiles in tissue samples from piglets can be extrapolated to paediatric population. These organ metabolic profiles can also be combined with plasma or urine profiles, obtained from the same individuals, to study the relationship between them. If a

[a] Department of Analytical Chemistry, Faculty of Science and Technology, University of the Basque Country (UPV/EHU), Barrio Sarriena s/n, 48940, Leioa, Spain.
[b] .Department of Biochemistry, Institute of Integrative Biology, University of Liverpool, Biosciences Building, Crown Street, Liverpool L69 7ZB, UK

meaningful relationship exists, these less invasive matrices could be used as surrogates of the organ samples, avoiding sample collection difficulties. The metabolomes of different biological samples or data sets can be explored together using data fusion techniques. In this aspect, multiblock statistical methods are widely used in metabolomics to integrate data from different experiments together, which distinguish for the type of sample, analytical platform, growth conditions, as well as other factors(21-25). Multiblock principal component analysis (MB-PCA) provides analogous information to classical Principal Component Analysis (PCA) and finds a correlation between several sets of possibly related data (the different data blocks) to reveal the common trend between those data blocks(26). Furthermore, the correlation between the data sets can be studied with Pearson correlation analysis to determine which variables are more in agreement with the underlying biology(22). It is important to highlight that an appropriate comparison between different data sets requires further analysis to determine if the distribution of two sets of points follow similar behaviors. For this purpose, Procrustes analysis is a useful tool since it measures the similarity in clustering and in information obtained from two different sets of data(27).

The aim of this work was to study liver tissue and plasma samples data from the same piglets as a whole in order to investigate if plasma could be an adequate surrogate matrix for liver. In order to investigate whether plasma metabolites can be used as liver maturation indicator or not, MB-PCA, Pearson correlation and Procrustes analysis were applied to the data acquired by High Performance Liquid Chromatography coupled to a Time-Of-Flight Mass Spectrometer (HPLC-TOF-MS) at positive and negative ionization modes. The key metabolites that significantly contributed to separation of three groups of piglets of different ages (newborns, neonates and infants) and may reflect organ maturation differences were then studied.

2. Materials and Methods

2.1. Reagent and solutions

Acetronitile (ACN) (LC-MS grade purity) and formic acid used in the mobile phases were purchased from Scharlau (Sentmenat, Spain) and Fisher Scientific (Pittsburgh, PA, USA), respectively. Methanol (MeOH) used for standards and sample preparation was also obtained from Scharlau. Ultra-high purity water was used in the preparation of mobile phase and reagent solutions and it was obtained from tap water pre-treated by Elix reverse osmosis, and subsequent filtration by a Milli-Q system from Millipore (Bedford, MA, USA).

Standard reagents used to assess the performance of the LC-MS system operation were supplied by different manufacturers: paracetamol, cholic acid, (±) verapamil hydrochloride, simvastatin, reserpine and leucine encephalin acetate salt hydrate were provided by Sigma-Aldrich (Steinheim, Germany), caffeine was purchased from Alfa Aesar (Karlsruhe, Germany) and salicylic acid from Fluka Analytical (Bucharest, Romania). Finally, sodium fluvastatin was kindly supplied by Novartis (Basel, Switzerland). A system suitability test (SST) was prepared with the nine compounds at a final concentration of 100 ng/mL in MeOH:H_2O 2:1 (v/v.

2.2. Animal model and experimental design

The piglets used in this study were *Topig F-1 Large White x Landrave* breed. Sample collection was performed by the Experimental Neonatal Physiology Unit of the BioCruces Health Research Institute (Cruces University Hospital, Basque Country, Spain), according to protocols approved by the Ethical Committee for Animal Welfare and were in compliance with the European and Spanish regulations for protection of experimental animals (86/609/EFC and RD 1201/2005). The 36 samples, 50% of each gender, were obtained from mechanically ventilated newborn piglets or group A (< 5 days, *n*=12), neonate piglets or group B (2 weeks, *n*=12) and infant piglets or group C (4 weeks, *n*=12). Plasma samples were obtained after the immediate centrifugation of blood, collected in EDTA tubes, at 950g for 10 min at room temperature. Liver tissue samples were immediate submerged in liquid nitrogen. Both types of samples were stored at -80 °C until analysis.

2.3. Preparation of metabolomic samples from piglets

Frozen plasma samples were thawed to room temperature and 50 µL were vortex mixed with 100 µL of cold MeOH during 2 min in a Signature Digital Vortex Mixer 945303 (VWR, Radnor, PA, USA). After centrifugation at 16110 x*g* for 15 min at 10 °C in a 5415R Eppendorf centrifuge (Hamburg, Germany) the supernatants were collected for HPLC-TOF-MS analysis.

Liver tissue samples were kept on liquid nitrogen and/or ice during the whole sample manipulation and treatment. 1 mL of MeOH:H_2O 2:1 (v/v) solution was added to 100 mg of tissue weighted in precellys tubes with 6 zirconium balls and extracted in a Precellys 24 Tissue Homogenizer coupled to a Cryolysis cooling system provided with N_2 stream, both from Bertin Instrument (Montigny-le-Bretonneus, France). After 3 cycles of 40 s (10 s between cycles) at 4500 rpm, the supernatants were centrifuged 3 times at 15866 g during 15 min at 10 °C to remove the suspension particles. Then the supernatants were ready for HPLC-TOF-MS analysis.

For the preparation of Quality Control (QC) samples. 5 µL of each plasma sample were taken and thoroughly mixed, reaching a total volume of 180 µL. From this pool, 50 µL were treated as previously described protocol for plasma samples. For the tissue analysis, 8 µL of the centrifuged supernatant of each liver sample were combined and mixed. The QC samples were injected at the beginning of the run to condition the LC system and then every six samples during plasma and liver analyses. These QCs were used to assess reproducibly of the metabolic profiles and when necessary for signal correction within the analytical sequence.

2.4. HPLC-TOF-MS analysis

The metabolomic profiles of plasma and liver supernatants were obtained separately, in positive and negative ionization modes, on a high performance liquid chromatography (HPLC) system (Agilent 1200 Series) coupled to a hybrid quadrupole-time-of-flight (Q-TOF) mass spectrometer (Agilent 6530) from Agilent Technologies (Santa Clara, CA, USA). The analysis order

was randomized to reduce any time-related effects. Chromatographic separation was performed by injecting 5 µL of sample on a reverse phase column (Zorbax SB-C18 (2.1 x 100 mm, 3.5 µm) equipped with a C8 guard column (2.1 x 12 mm, 5 µm), both from Agilent Technologies, at 35 °C and a flow rate of 0.4 mL/min. The mobile phase was 0.1% formic acid and 5% ACN in water (phase A), and 0.1% formic acid in ACN (phase B). The gradient increased in a linear way from 0 to 100% B in 10 min, it was kept at 100% B for 2.5 min and returned to the starting conditions in 1.5 min. Finally it was re-equilibrated for 5 min. An Agilent Jet Stream electrospray ionization source (ESI) operated in positive and negative ionization modes were employed at capillary voltages of +3800 V and -2500 V, respectively. The rest MS parameters were as follows: drying gas (nitrogen) temperature at 325 °C and a flow rate of 10 L/min, nebulizer gas pressure (nitrogen) at 30 psi, sheath gas at 350 °C and a flow rate of 11 L/min, skimmer at 65 V, fragmentor at 125 V and octopole RF peak at 750 V. Data acquisition was carried out in low mass range (< 1700 *m/z*), 2 GHz extended dynamic range and centroid mode at a full range from 50 to 1200 *m/z* (rate 2 spectra/s). A reference solution was directly infused into the source for internal calibration during the analysis and for accuracy and reproducibility controlling. For this purpose, *m/z* 121.0509 (purine, $[C_5H_4N_4+H]^+$) and *m/z* 922.0098 (HP-921, $[C_{18}H_{18}O_6N_3P_3F_{24}+H]^+$) for positive mode, and *m/z* 112.9855 (TFANH$_4$, $[C_2H_4O_2NF_3\text{-}NH_4]^-$) and *m/z* 966.0007 (HP-921COOH, $[C_{18}H_{18}O_6N_3P_3F_{24}\text{-}COOH]^-$) for negative mode, were used during the HPLC-TOF-MS run. Additionally, in order to control the analytical performance of the MS instrument and the LC system, the SST was injected in the HPLC-TOF-MS system at the beginning, in the middle and at the end of each sequence. The data were acquired using the Agilent MassHunter Workstation software (version B.05.01) and the raw data were processed with the Agilent MassHunter Qualitative (version B.07.00).

2.5. Data preprocessing

HPLC-TOF-MS raw data were converted to mzXML file format using msConverter (proteoWizard) from 0 to 13.5 min to avoid features coming from the cleaning step of the gradient. Data preprocessing was performed using XCMS 1.52.0 (Metlin, La Jolla, CA, USA) package (https://www.bioconductor.org/packages/release/bioc/html/xcms.html). The algorithms employed were centWave for peak picking, obiwarp for retention time correction and density for grouping. The parameters used in each algorithm were optimized with the Isotopologue Parameters Optimization (IPO) package (https://bioconductor.org/packages/release/bioc/html/IPO.html) and the conditions used are provided in Table 1. Finally, CAMERA 1.32.0 package (Bioconductor Open Source Software for Bioinformatics; https://bioconductor.org/packages/release/bioc/html/CAMERA.html) was used for isotopologues and adducts detection. The matrices obtained for each sample type and ionization modes were treated before the statistical analysis. The isotopes identified by CAMERA ([M+1], [M+2] and [M+3]), the features before the injection peak (less than 1 min) and the features with percentage of relative standard deviation (%RSD) in the QCs greater than 20 %(28,29) were removed.

2.6. Data analysis by Multiblock Principal Component Analysis (MB-PCA)

The resultant matrices of the two biological matrices (liver tissue and plasma) of the three different groups of piglets (newborn (A), neonates (B) and infants (C)), each analyzed in positive and negative ionization modes, , were further analyzed in MATLAB software (The MathWorks, Naticks, USA) using the in-house toolbox https://github.com/Biospec/cluster-toolbox-v2.0 to perform multivariate statistical analysis. In order to study the quality of the obtained data, classical exploratory principal component analysis (PCA) was performed for each data set (Liver ESI+, Plasma ESI+, Liver ESI- and Plasma ESI-). For this purpose, QC intensity drop correction and autoscaling were applied, and the resultant PCA scores plot was employed to assess the reproducibility of the data and to detect any outliers.

Table 1. XCMS parameter for liver and plasma samples at both ionization modes used in this study.

		LIVER		PLASMA	
Algorithm	**Parameter**	**ESI+**	**ESI-**	**ESI+**	**ESI-**
CentWave	ppm	29	20.75	31.68	31
	peakwidth	12, 39.5	9.76, 50	22.01, 81.26	20, 80
	mzdiff	0.0034	-0.0048	-0.0123	-0.012
Obiwarp	profStep	0.655	0.64	0.7324	1
	center	QC8	QC4	QC4	QC5
	gapInit	0.64	0.928	0.7552	0.928
	gapExtend	2.4	NULL	2.4	2.688
Density	bw	0.879	6	0.25	0.879
	mzwid	0.0265	0.025	0.027	0.0342

Multiblock modelling is designed to find the underlying relationship between several sets (data matrices or blocks) of possibly related data to reveal the "common trend" between them, and this was used to investigate correlations between plasma and liver matrices(26,30). For this purpose, unsupervised multiblock modelling (MB-PCA), which is the extension of commonly used PCA, was applied and two different MB-PCAs were carried out based on the two ionization modes of analysis. The first one (MB-PCA_ESI+) was constructed with the positive ionization mode data sets of liver (Liver ESI+ / Block 1a) and plasma (Plasma ESI+ / Block 2a) and the second one (MB-PCA_ESI-) with the negative ionization mode data sets for liver (Liver ESI- / Block 1b) and plasma (Plasma ESI- / Block 2b).

The algorithm CPCA-W proposed by Westerhuis *et al*. was used as the MB-PCA models(31). In this way, the model consisted of three main components. The super scores matrix, or $\boldsymbol{T_t}$, represents the common trend across all the data matrices incorporated into the model. *c* pairs of blocks scores, or $\boldsymbol{T_b}$, and loadings, or $\boldsymbol{P_b}$, matrices that represents the unique pattern of each block under the consensual view revealed in $\boldsymbol{T_t}$, where *c* is

the number of data sets (blocks), in this work two (plasma and liver) incorporated into the MB-PCA model. Finally, a block weights vector, or $\boldsymbol{W_t}$, that represents the relative contribution of each block to the common trend showed in $\boldsymbol{T_t}$ (25).

An equal-variance block weighting method was applied in order to solve such variance disparity. This consisted on autoscaling each data block, so every variable had a mean of 0 and a standard devition of 1, followed by the application of a scaling factor. The scaling factor was dependent of the number of variables and it was calculated as the inverse of the square root of number of variables in that certain block ($1/\sqrt{n}$, being n the number of variables within the block). After the weighting step, the MB-PCA was built with positive or negative ionization modes data matrices that hopefully have a common trend of interest.

Finally, the MB-PCA scores plots obtained from the combination of two data sets (positive or negative data sets) were compared with the classical PCA scores plot. That is, the PCA scores plots obtained for Liver ESI+ and Plasma ESI+ matrices were compared with the MB-PCA_ESI+ and the Liver ESI- and Plasma ESI- score plots were compared with the MB-PCA_ESI-. In this way, it was possible to study the inference improvements achieved by the *combination* of the two matrices.

2.7. Pearson correlation studies and significant metabolites

In addition to MB-PCA, correlation analysis was performed to find out which features in each block were highly correlated to each other. The features with a high correlation are likely to have a great biological interest as they might originate from the same biological pathway. For this purpose, a pair-wise correlation analysis between the liver and plasma data sets corresponding to the same ionization mode was performed using Pearson correlation coefficient(32). Then, a correlation coefficient heat map was generated for each ionization mode to have a global view of the connections of all the metabolites(25). In order to identify if the metabolite features showed significant differences among the study groups, non-parametric test (Kruskal-Wallis) followed by False Discovery Rate (FDR) was applied. Finally, a post-hoc Turkey HSD test (Honestly Significant Difference) was performed to obtain those significant features that discriminate the three groups of piglets. The criteria to select the features of interest in this study was as follows: (1) features had to be correlated between liver and plasma samples with a Pearson correlation coefficient value greater than 0.8.(25) and (2) features had to be significant using non-parametric test and after applying FDR, with a p-value lower than 0.001, as well as differentiate the three group of piglets (A≠B≠C).

2.8. Procrustes analysis of the significant features

In order to understand the similarities between liver and plasma better, Procrustes analysis was employed to perform a multivariate pattern comparison measurement of the clustering patterns generated from the ordination plots. This comparison shows if plasma and liver samples cluster in a similar way and if they offer similar discriminatory information(32).

Procrustes analysis is a method that analyses the distribution of two shapes or sets of points (or, in the present study, sets of samples) in high dimensional space. It determines a linear transformation (incorporating translation, scaling and rotation component) of the points in one matrix (e.g., Liver ESI+) to best conform them to the points in the other matrix (e.g., Plasma ESI+). The resulting goodness-of-fit value, the Procrustes dissimilarity (d), is the normalized sum of squared errors and the value of this dissimilarity ranges from 0 to 1. This value gives an indication of how similar the two shapes are. Thus, a value of 0 means that the shapes are identical and values close to 1 mean that the shapes have nothing in common(27). Similarity between liver and plasma matrices, obtained from the same newborn, neonate and infant piglets, were assessed to determine if plasma could be an appropriate biofluid to reflect liver behavior. The main limitation of the Procrustes dissimilarity is that it requires matrices with the same number of points (rows-samples) and small differences in dimensions (number of columns-variables). As a consequence, the matrices of Liver ESI+, Plasma ESI+, Liver ESI- and Plasma ESI- were not directly employed and the PCA scores of first z PCs were used instead. It is also important to point out that only significant features (p<0.001) that fulfilled non-parametric test were of interest. Thus, only significant features were considered to perform the Procrustes analysis. For these reasons, d was measured between the resulting z scores that included sufficient PCs to explain 75% of the variance of the data, considering only the significant features. The PCA scores were obtained from the separately Autoscaled data sets from both types of samples at same ionization modes(27,32). A permutation test (n = 1000000) was also performed in order to obtain p-values, which indicates probabilities that the similarity between matrices have occurred by chance(32).

3. Results and discussion

3.1. Liver and plasma data fusion by MB-PCA and study of correlations among features

Once the data matrices were pre-processed, Liver ESI+ generated 5603 features, Plasma ESI+ 2207, Liver ESI- 2242 and Plasma ESI- 1855. These resultant matrices were block weighted (autoscaling followed by factor scaling, $1/\sqrt{n}$) in order to generate new matrices to perform the MB-PCA models (MB-PCA_ESI+ and MB-PCA_ESI-). The MB-PCA super scores plots are shown in Fig. 1 where the separation between newborns (A), neonates (B) and infants (C) is clearly seen.

MB-PCA model also generated PCA block scores plots for each individual data set that can be observed in Fig. 2. It is worth noting that the separation within groups observed in the PCA blocks scores is very similar and could be compared to that in the related MB-PCA super scores plot. Furthermore, it is remarkable that on Fig. 2, in block scores of Liver ESI- (block 1b), the TEV% of PC2 is slightly higher than that of PC1. The reason is that the Liver ESI- data had lower signal-to-noise ratio and the separation between three classes were weaker than that in Plasma ESI+ data, particularly between group A and the other

two groups. Therefore, the MB-PCA is more dominated by Plasma ESI- block. The first PC in this MB-PCA model which were separation group A from the other two groups was not the largest variance in Liver ESI- data which resulted in TEV% of PC2 was slightly higher than that of PC1 for Liver ESI- block.

In addition to MB-PCA, classical PCA scores plots for each individual data set were generated (see ESI, Fig. S1). Even though classical PCAs (see ESI, Fig. S1) and MB-PCAs (Fig. 1) show a similar pattern within groups, the MB-PCA super scores plots demonstrate that using multiblock modelling improves the clustering and enhances the separation between newborns (A), neonates (B) and infants (C) piglets. Moreover, further comparisons were done between block scores PCAs (Fig. 2) and classical PCAs (see ESI, Fig. S1) to assess the clustering. In this context, Liver ESI+ and Liver ESI- show a clear separation of the three groups of piglets in all cases but the clustering within groups is improved when block PCAs were performed from the MB-PCA models. Plasma ESI+ is mainly separated in the PC1 when using classical PCA but, as it occurred with the liver data sets, the employment of MB-PCA improves the group clustering. Finally, the classical PCA of the Plasma ESI- data set shows a perfect separation between group A and groups B and C in the PC1, whereas no obvious separation could be observed between groups B and C. This trend occurs also in the Plasma ESI- block of the MB-PCA_ESI- model but it seems that the separation is slightly improved. Thus, the separation of groups is improved both in the MB-PCA super scores plot and in the respective block scores plots if compared to classical PCA scores plot.

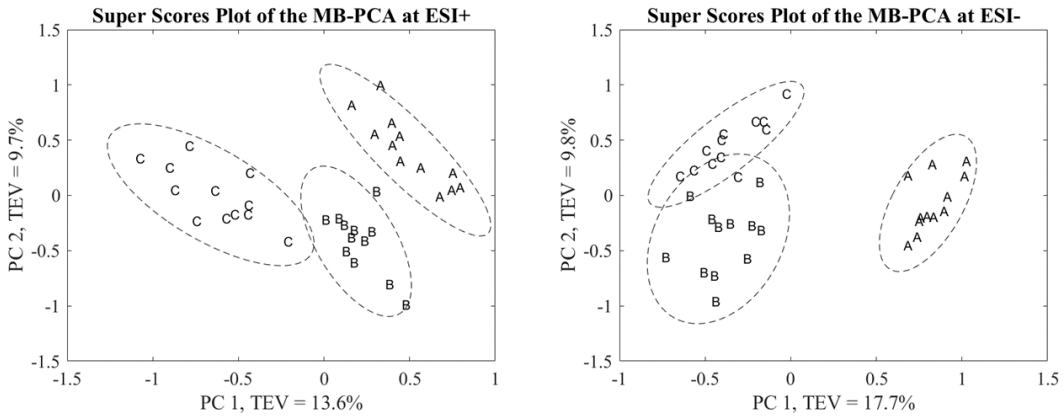

Fig. 1 MB-PCA super scores plots of combined data (liver and plasma) at positive (ESI+) and negative (ESI-) ionization modes. Coding for the piglet groups: newborns (A), neonates (B) and infants (C). TEV = total explained variance

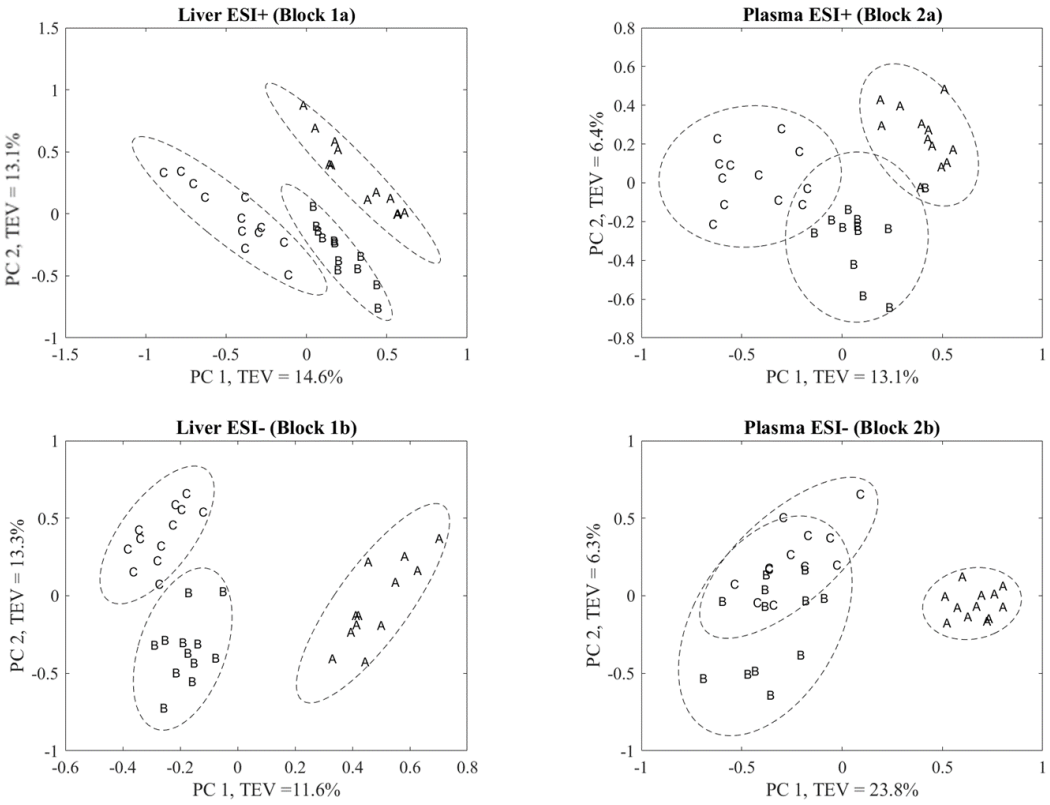

Fig. 2 MB-PCA block scores plots for liver and plasma samples at both ionization modes. Liver ESI+ or Block 1a and Plasma ESI+ or Block 2a belong to the MB-PCA_ESI+ model whereas Liver ESI- or Block 1b and Plasma ESI- or Block 2b belong to the MB-PCA_ESI-. Coding for the piglet groups: newborns (A), neonates (B) and infants (C). TEV = total explained variance

The next stage was to explore the correlation among the features from liver and plasma data sets. For that aim, Pearson correlation coefficients were calculated for all the features and the results were visualized as heat maps for each ionization mode (see ESI, Fig. S2). 360 pairs of metabolites in positive ionization mode and 1784 pairs in negative ionization mode showed an absolute correlation coefficient greater than 0.8. Among them, 284 unique metabolites for the positive ionization mode (167 from the Liver ESI+ block and 117 from Plasma ESI+ block) and 654 for the negative ionization mode (293 from Liver ESI- and 361 from Plasma ESI-) were involved.

The correlated features obtained after applying Pearson correlation were located and marked with red circles in the block loadings plot of the MB-PCA_ESI+ (Fig. 3) and MB-PCA_ESI- (Fig. 4) models in order to visualize how the variables were distributed. As it can be observed in the loadings blocks plot for Liver ESI+ and Plasma ESI+, the correlated features are distributed on a diagonal from the lower left corner to the upper right corner. This trend is coincident with the separation trend between newborns (A) neonates (B) and infants (C) piglets found in the block scores PCAs (Fig. 2). Similarly, the distribution of the correlated features along the PC1 in negative ionization mode fits with the separation between groups in the pertinent block scores PCAs (Fig. 2)

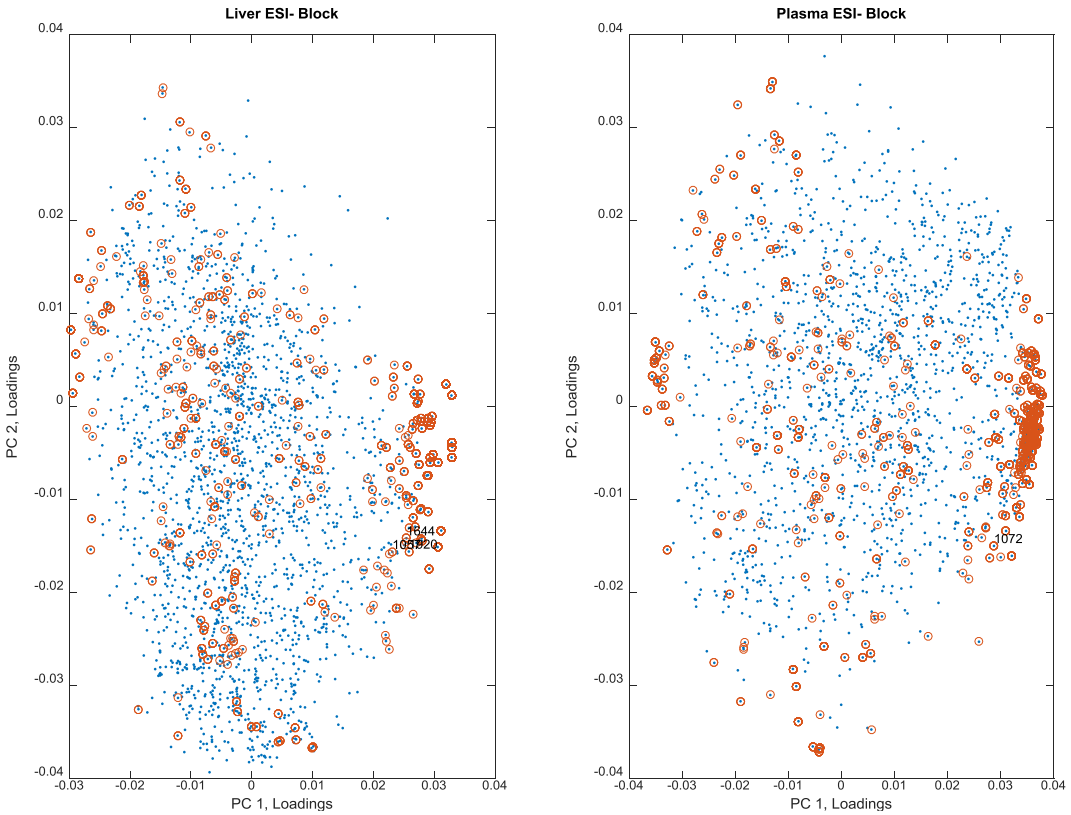

Fig. 3 MB-PCA_ESI- block loading plot for liver and plasma blocks. The metabolites circled in red are the ones with an absolute value of Pearson correlation coefficient greater than 0.8. The ID numbers correspond to the significant features that fulfill Kruskal-Wallis test (p-value <0.001) among the pairs of correlated features and distinguish the three groups (newborns, neonates and infants).

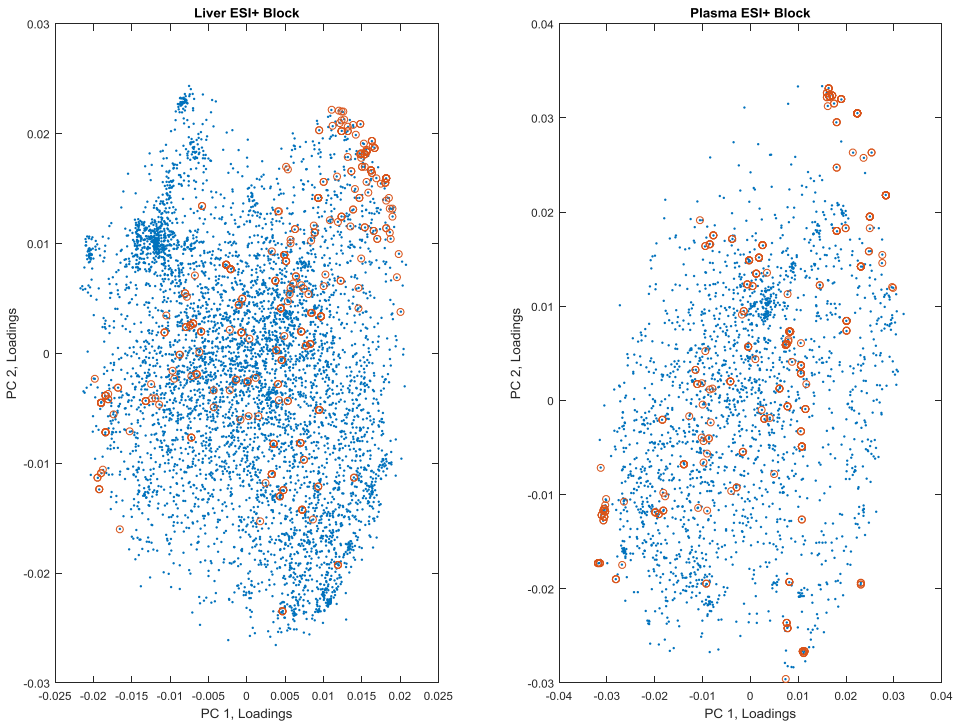

Fig. 4 MB-PCA_ESI+ block loading plot for liver and plasma blocks. The metabolites circled in red are the ones with an absolute value of Pearson correlation coefficient greater than 0.8.

Next, non-parametric test (Kruskal – Wallis) and FDR (p-value<0.001), followed by the post-hoc Tukey HSD tests, were applied. Among the 360 pairs of correlated features in positive ionization mode, there was no pair that fulfilled a p-value lower than 0.001 and differentiate the three groups under study. In negative ionization mode, 3 pairs out of 1784 fulfilled a p-value lower than 0.001 and distinguish newborns, neonates and infant piglets.

The three pairs obtained in negative ionization mode were combinations of a single feature from Plasma ESI- with three features from Liver ESI-. This means that this unique feature in Plasma ESI- (ID 1072) is correlated with the three features in Liver ESI- and that all of them distinguish the group of study in a significant way. Interestingly, the Plasma ESI- feature and the three Liver ESI- features are located in the same position in the block loadings plots (Fig. 4), which means that they have the same trend within groups. For this specific case, newborns (A) have greater amount of these features compared to infants (C) and a down-regulated trend was observed (Fig. 5). This fact is of special interest as correlated and significant features follow same tendency in both biological matrices, liver and plasma.

3.2.Procrustes analysis to assess the similarity of clustering in liver and plasma

Finally, Procrustes analysis was performed to determine the similarity between liver and plasma matrices. For this purpose, the *z* score matrices obtained by considering only the significant features of the Liver ESI+, Plasma ESI+, Liver ESI- and Plasma ESI- matrices were used to perform the Procrustes analysis. The number of PCs included in the *z* score matrices, which explained a TEV greater than 75 %, were four for Liver ESI+, Liver ESI- and Plasma ESI- and two for Plasma ESI+. The Procrustes dissimilarity values (*d*) for the pairs Liver ESI+/Plasma ESI+ was 0.3753 and for Liver ESI-/Plasma ESI- was 0.3673. The permutation tests indicated that similarity between liver and plasma at both ionization modes are very unlikely to have occurred by chance, with *p*-values below 1.0×10^{-6}. Therefore, liver and plasma samples cluster in a very similar way and, in the case studied, plasma could be considered a biofluid that provides vital information for liver maturation without need to analyse liver tissue directly.

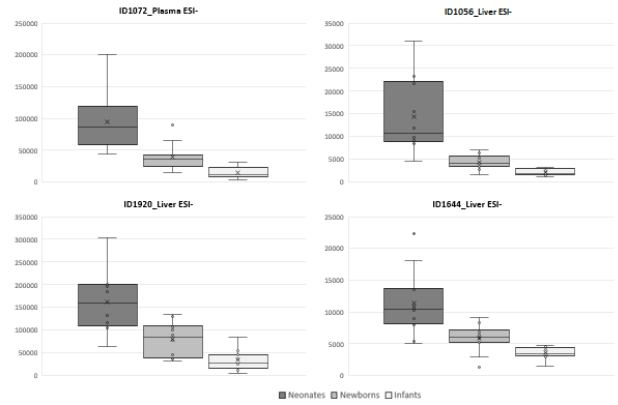

Fig. 5 Down-regulated tendency of correlated features in negative ionization mode within the three groups of piglets (newborns, neonates and infants).

Conclusions

In this study, HPLC-Q-TOF-MS system was used to obtain metabolic profiles from plasma and liver tissue samples. A thorough data treatment was carried out in order to ascertain whether the plasma is an appropriate biofluid to reflect liver maturation state.

Both the super scores and block scores plots obtained by MB-PCA modelling provided an improvement on clustering within groups compared to the classical PCA .Moreover, MB-PCA has the advantage of considering the relationship between liver and plasma samples as well as revealing the trend and correlation between plasma and liver.

Pearson correlation studies together with the MB-PCA was a successful tool to find and understand correlations between metabolites detected in liver and plasma. This approach, combined with the univariate analysis, allows to connect features from liver and plasma that show significant differences among groups and are strongly correlated. In other words, it can be used to find in one particular biological matrix metabolites that explain the changes in the other matrix. Furthermore, the feasibility of using plasma as a surrogate matrix of liver is also supported by the Procrustes analysis that indicates that both matrices are similar and comparable.

This is of special interest because metabolites detected in plasma could be used to investigate liver state.. In this way the limitation on tissue sample collection can be potentially avoided. With further studies in children this link between plasma and liver could be used to measure the maturation state accurately. If successful, this approach could be employed within clinical trials in children to improve the knowledge on organ maturation development which would lead to a better drug dosing in the paediatric populations.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

Authors thank the Experimental Neonatal Physiology Unit of the BioCruces Health Research Institute (Cruces University Hospital, Basque Country, Spain) for collecting and providing the samples. O.E.A. thanks the Ministry of Economy and Competitiveness for her predoctoral contract. Authors thank for technical and human support provided by SGIker of UPV/EHU and European funding (ERDF and ESF).

References

(1) Henschel AD, Rothenberger LG, Boos J. Randomized clinical trials in children - ethical and methodological issues. Curr Pharm Des 2010;16(22):2407-2415.

(2) Kshirsagar N, Swaminathan S, Jog P, Dalwai S, Mathur R, Shekhar C, et al. Regulatory and Ethical Issues in Pediatric Clinical Research: Recommendations From a Panel Discussion. J Clin Pharmacol 2017;57(8):943-946.

(3) Morales-Olivas FJ, Morales-Carpi C. Clinical trials in children. Rev Recent Clin Trials 2006;1(3):251-258.

(4) Pinxten W, Nys H, Dierickx K. Frontline ethical issues in pediatric clinical research: ethical and regulatory aspects of seven current bottlenecks in pediatric clinical research. Eur J Pediatr 2010;169(12):1541-1548.

(5) Kern SE. Challenges in conducting clinical trials in children: approaches for improving performance. Expert Rev Clin Pharmacol 2009 Nov 1;2(6):609-617.

(6) Leach RH, Wood BS. Drug dosage for children. Lancet 1967;2(7530):1350-1351.

(7) Kimland E, Odlind V. Off-label drug use in pediatric patients. Clin Pharmacol Ther 2012;91(5):796-801.

(8) de Wildt SN, Kearns GL, Leeder JS, Van den Anker JN. Cytochrome P450 3A: Ontogeny and drug disposition. Clin Pharmacokinet 1999;37(6):485-505.

(9) Saucedo AL, Paniagua-Vega D, Minsky NW, Saucedo AL, Paniagua-Vega D, Perales-Quintana MM, et al. Chronic Kidney Disease and the Search for New Biomarkers for Early Diagnosis. Curr Med Chem 2018;25(31):3719-3747.

(10) Bose K. Concept of human physical growth and development. ; 2007.

(11) Trivedi DK, Hollywood KA, Goodacre R. Metabolomics for the masses: The future of metabolomics in a personalized world. New horizons in translational medicine 2017;3(6):294-305.

(12) Atzori L, Antonucci R, Barberini L, Griffin JL, Fanos V. Metabolomics: A new tool for the neonatologist. J Matern -Fetal Neonat Med 2009;22(Suppl. 3):50-53.

(13) Wang X, Han W, Yang J, Westaway D, Li L. Development of chemical isotope labeling LC-MS for tissue metabolomics and its application for brain and liver metabolome profiling in Alzheimer's disease mouse model. Anal Chim Acta 2019;1050:95-104.

(14) Fanos V, Barberini L, Antonucci R, Atzori L. Metabolomics in neonatology and pediatrics. Clin Biochem 2011;44(7):452-454.

(15) Fanos V, Van den AJ, Noto A, Mussap M, Atzori L. Metabolomics in neonatology: fact or fiction? Semin Fetal Neonatal Med 2013;18(1):3-12.

(16) Odle J, Lin X, Jacobi SK, Kim SW, Stahl CH. The suckling piglet as an agrimedical model for the study of pediatric nutrition and metabolism. Annu Rev Anim Biosci 2014;2:419-444.

(17) Oosterloo BC, Stoll B, Premkumar M, Olutoye O, Thymann T, Sangild PT, et al. Dual purpose use of preterm piglets as a model of pediatric GI disease. Vet Immunol Immunopathol 2014;159(3-4):156-165.

(18) Sangild PT, Thymann T, Schmidt M, Stoll B, Burrin DG, Buddington RK. Invited review: the preterm pig as a model in pediatric gastroenterology. J Anim Sci (Champaign, IL, U S) 2013;91(10):4713-4729.

(19) Shulman RJ, Henning SJ, Nichols BL. The miniature pig as an animal model for the study of intestinal enzyme development. Pediatr Res 1988;23(3):311-315.

(20) Vodicka P, Smetana K,Jr, Dvorankova B, Emerick T, Xu YZ, Ourednik J, et al. The miniature pig as an animal model in biomedical research. Ann N Y Acad Sci 2005;1049:161-171.

(21) Ingerslev AK, Theil PK, Bach Knudsen KE, Hedemann MS, Karaman I, Bagcioglu M, et al. Whole Grain Consumption Increases Gastrointestinal Content of Sulfate-Conjugated Oxylipins in Pigs - A Multicompartmental Metabolomics Study. J Proteome Res 2015;14(8):3095-3110.

(22) Marshall DD, Lei S, Worley B, Huang Y, Garcia-Garcia A, Franco R, et al. Combining DI-ESI-MS and NMR datasets for metabolic profiling. Metabolomics 2015;11(2):391-402.

(23) Sayqal A, Xu Y, Trivedi DK, Al Masoud N, Ellis DI, Muhamadali H, et al. Metabolic analysis of the response of Pseudomonas putida DOT-T1E strains to toluene using Fourier transform infrared spectroscopy and gas chromatography mass spectrometry. Metabolomics 2016;12(7):1-12.

(24) Kassama Y, Xu Y, Dunn WB, Geukens N, Anne J, Goodacre R. Assessment of adaptive focused acoustics versus manual vortex/freeze-thaw for intracellular metabolite extraction from Streptomyces lividans producing recombinant proteins using GC-MS and multi-block principal component analysis. Analyst (Cambridge, U K) 2010;135(5):934-942.

(25) Xu Y, Correa E, Goodacre R. Integrating multiple analytical platforms and chemometrics for comprehensive metabolic profiling: application to meat spoilage detection. Anal Bioanal Chem 2013;405(15):5063-5074.

(26) Xu Y, Goodacre R. Multiblock principal component analysis: an efficient tool for analyzing metabolomics data which contain two influential factors. Metabolomics 2012;8(Suppl. 1):37-51.

(27) Vaughan AA, Dunn WB, Allwood JW, Wedge DC, Blackhall FH, Whetton AD, et al. Liquid Chromatography-Mass Spectrometry Calibration Transfer and Metabolomics Data Fusion. Anal Chem (Washington, DC, U S) 2012;84(22):9848-9857.

(28) Dunn WB, Broadhurst D, Begley P, Zelena E, Francis-McIntyre S, Anderson N, et al. Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid

chromatography coupled to mass spectrometry. Nat Protoc 2011;6(7):1060-1083.

(29) Dudzik D, Barbas-Bernardos C, Garcia A, Barbas C. Quality assurance procedures for mass spectrometry untargeted metabolomics. a review. J Pharm Biomed Anal 2018;147:149-173.

(30) Boccard J, Rudaz S. Harnessing the complexity of metabolomic data with chemometrics. J Chemom 2014;28(1):1-9.

(31) Westerhuis JA, Kourti T, Macgregor JF. Analysis of multiblock and hierarchical PCA and PLS models. J Chemom 1998;12(5):301-321.

(32) Wedge DC, Allwood JW, Dunn W, Vaughan AA, Simpson K, Brown M, et al. Is Serum or Plasma More Appropriate for Intersubject Comparisons in Metabolomic Studies? An Assessment in Patients with Small-Cell Lung Cancer. Anal Chem (Washington, DC, U S) 2011;83(17):6689-6697.