



**HAL**  
open science

## Data-Mining Approach on Transcriptomics and Methylo-mics Placental Analysis Highlights Genes in Fetal Growth Restriction

Floris Chabrun, Noemie Huetz, Xavier Dieu, Guillaume Rousseau, Guillaume Bouzillé, Juan Manuel Chao de La Barca, Vincent Procaccio, Guy Lenaers, Odile Blanchet, Guillaume Legendre, et al.

► **To cite this version:**

Floris Chabrun, Noemie Huetz, Xavier Dieu, Guillaume Rousseau, Guillaume Bouzillé, et al.. Data-Mining Approach on Transcriptomics and Methylo-mics Placental Analysis Highlights Genes in Fetal Growth Restriction. *Frontiers in Genetics*, 2020, 10, pp.1292. 10.3389/fgene.2019.01292 . hal-02469273

**HAL Id: hal-02469273**

<https://univ-rennes.hal.science/hal-02469273v1>

Submitted on 6 Feb 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



# Data-Mining Approach on Transcriptomics and Methyloomics Placental Analysis Highlights Genes in Fetal Growth Restriction

## OPEN ACCESS

### Edited by:

Mehdi Pirooznia,  
National Heart, Lung, and Blood  
Institute, United States

### Reviewed by:

Amit Kumar Yadav,  
Translational Health Science and  
Technology Institute, India  
Izabela Makalowska,  
Adam Mickiewicz University,  
Poland  
Amanda Vlahos,  
Murdoch Children's Research  
Institute, Australia

### \*Correspondence:

Floris Chabrun  
floris.chabrun@chu-angers.fr

### Specialty section:

This article was submitted to  
Bioinformatics and  
Computational Biology,  
a section of the journal  
Frontiers in Genetics

**Received:** 10 August 2019

**Accepted:** 25 November 2019

**Published:** 09 January 2020

### Citation:

Chabrun F, Huetz N, Dieu X, Rousseau G, Bouzillé G, Chao de la Barca JM, Procaccio V, Lenaers G, Blanchet O, Legendre G, Mirebeau-Prunier D, Cuggia M, Guardiola P, Reynier P and Gascoin G (2020) Data-Mining Approach on Transcriptomics and Methyloomics Placental Analysis Highlights Genes in Fetal Growth Restriction. *Front. Genet.* 10:1292. doi: 10.3389/fgene.2019.01292

Floris Chabrun<sup>1,2\*</sup>, Noémie Huetz<sup>2,3</sup>, Xavier Dieu<sup>1,2</sup>, Guillaume Rousseau<sup>1,2</sup>, Guillaume Bouzillé<sup>4,5</sup>, Juan Manuel Chao de la Barca<sup>1,2</sup>, Vincent Procaccio<sup>1,2</sup>, Guy Lenaers<sup>1,2</sup>, Odile Blanchet<sup>6</sup>, Guillaume Legendre<sup>7</sup>, Delphine Mirebeau-Prunier<sup>1,2</sup>, Marc Cuggia<sup>4,5</sup>, Philippe Guardiola<sup>8</sup>, Pascal Reynier<sup>1,2</sup> and Geraldine Gascoin<sup>2,3</sup>

<sup>1</sup> Département de Biochimie et Génétique, Centre Hospitalier Universitaire, Angers, France, <sup>2</sup> Unité Mixte de Recherche (UMR) MITOVASC, Équipe Mitolab, Centre National de la Recherche Scientifique (CNRS) 6015, Institut National de la Santé et de la Recherche Médicale (INSERM) U1083, Université d'Angers, Angers, France, <sup>3</sup> Réanimation et Médecine Néonatales, Centre Hospitalier Universitaire, Angers, France, <sup>4</sup> Laboratoire du Traitement de l'Image et du Signal, INSERM, UMR 1099, Université Rennes 1, Rennes, France, <sup>5</sup> Département d'Information médicale et dossiers médicaux, Centre Hospitalier Universitaire, Rennes, France, <sup>6</sup> Centre de Ressources Biologiques, Centre Hospitalier Universitaire, Angers, France, <sup>7</sup> Département de Gynécologie Obstétrique, Centre Hospitalier Universitaire, Angers, France, <sup>8</sup> Service de Génomique Onco-Hématologique, Centre Hospitalier Universitaire, Angers, France

Intrauterine Growth Restriction (IUGR) affects 8% of newborns and increases morbidity and mortality for the offspring even during later stages of life. Single omics studies have evidenced epigenetic, genetic, and metabolic alterations in IUGR, but pathogenic mechanisms as a whole are not being fully understood. An in-depth strategy combining methylomics and transcriptomics analyses was performed on 36 placenta samples in a case-control study. Data-mining algorithms were used to combine the analysis of more than 1,200 genes found to be significantly expressed and/or methylated. We used an automated text-mining approach, using the bulk textual gene annotations of the discriminant genes. Machine learning models were then used to explore the phenotypic subgroups (premature birth, birth weight, and head circumference) associated with IUGR. Gene annotation clustering highlighted the alteration of cell signaling and proliferation, cytoskeleton and cellular structures, oxidative stress, protein turnover, muscle development, energy, and lipid metabolism with insulin resistance. Machine learning models showed a high capacity for predicting the sub-phenotypes associated with IUGR, allowing a better description of the IUGR pathophysiology as well as key genes involved.

**Keywords:** data mining, methylomics, intrauterine growth restriction, multi-omics, text-mining, transcriptomics

**Abbreviations:** IUGR, intrauterine growth restriction; PE, pre-eclampsia; Se, sensitivity; Sp, specificity; tf, term frequency; idf, inverse document frequency; SVM, support vector machine; RMSE, root-mean-square deviation; C-section, caesarean section.

## INTRODUCTION

Intrauterine growth restriction (IUGR) is a frequent complication of pregnancy with a prevalence in up to 5% to 10% in the general population (Zhang et al., 2015). It is defined as a restriction of fetal growth during pregnancy, “a fetus that doesn’t reach its growth potential” (Vayssière et al., 2015). It can lead to a birth weight and/or length below the tenth percentile for a given gestational age in newborns, thus considered as “Small for Gestational Age” (Vayssière et al., 2015). IUGR represents a major public health problem, being one of the main causes of premature birth, perinatal mortality, and neurological and respiratory morbidities (Flamant and Gascoin, 2013). It is also suspected to be a determining factor in the development of cardiovascular diseases, obesity, and type 2 diabetes in adulthood (Gascoin and Flamant, 2013).

Fetal growth is a complex process that involves fetal genetics, nutrient and oxygen availability, and maternal nutrition, as well as growth factors and hormones from maternal, fetal, and placental origin (Murki, 2014). Fetal growth is inseparable from placental growth and requires a continuous supply of nutrients that is adapted to each period of pregnancy (Sharma et al., 2016).

IUGR remains a complex problem for the clinician. Placental dysfunction and vascular underperfusion are involved in the largest proportion of cases (Kaplan, 2007; Malhotra et al., 2019). It results from utero-placental insufficiency due to abnormal uterine artery remodeling in the first trimester of pregnancy and may or may not be associated with pre-eclampsia (PE). However, while many risk factors have been identified, placental insufficiency is still unexplained in up to 60% of cases (Malhotra et al., 2019).

Epigenetics (Xiao et al., 2016) and gene expression (Buffat et al., 2007; Madeleneau et al., 2015) reprogramming play a central role in IUGR. However, the pathophysiological connections between these two fields of high-throughput analyses have only recently begun to be studied (Ding and Cui, 2017). Although many tools have been developed to analyze and integrate multi-omics data, this task remains a challenge in medicine (Gomez-Cabrero et al., 2014). Many features originating from the variance between samples and the complexity of the statistical data processing require developing data-driven approaches rather than classical hypothesis-driven approaches (van Helden, 2013). The exploration of pathophysiological conditions with such data-driven approaches must integrate many processes from clinical and biological data collection, through complex data normalization and mathematical and bioinformatics modeling, to the final interpretation and data visualization.

When dealing with a short list of genes, the exploration of their roles and underlying patterns is usually carried out through “manual” interpretation, using both annotations and personal knowledge. This “manual” interpretation may be used to categorize the genes, or to seek patterns in roles, functions, or localizations, underpinning the pathology or context studied. When dealing with thousands of significant gene features (e.g.

expression levels or methylation levels), the interpretation becomes humanly untenable, due to time and memory limits. Rather than limiting our literature review to a small subset of the most significantly altered genes, we used text-mining algorithms to perform an unsupervised analysis of those genes. Those algorithms have already been used to categorize and summarize text corpora based on similarities in their content (Aggarwal and Zhai, 2012).

With the aim of having an extended vision of the pathophysiological processes at the origin of IUGR, while identifying the most predominant deregulated pathways that may be targeted for therapeutic purposes, we used machine-learning models to explore the relationship between placental transcriptomics and methylomics variations and IUGR. The highly predictive models obtained from IUGR and its sub-phenotypes were then used to highlight the genes with a high correlation with IUGR clinical severity, and thus with a high therapeutic potential.

## MATERIAL AND METHODS

The global workflow is summarized in **Figure 1**.

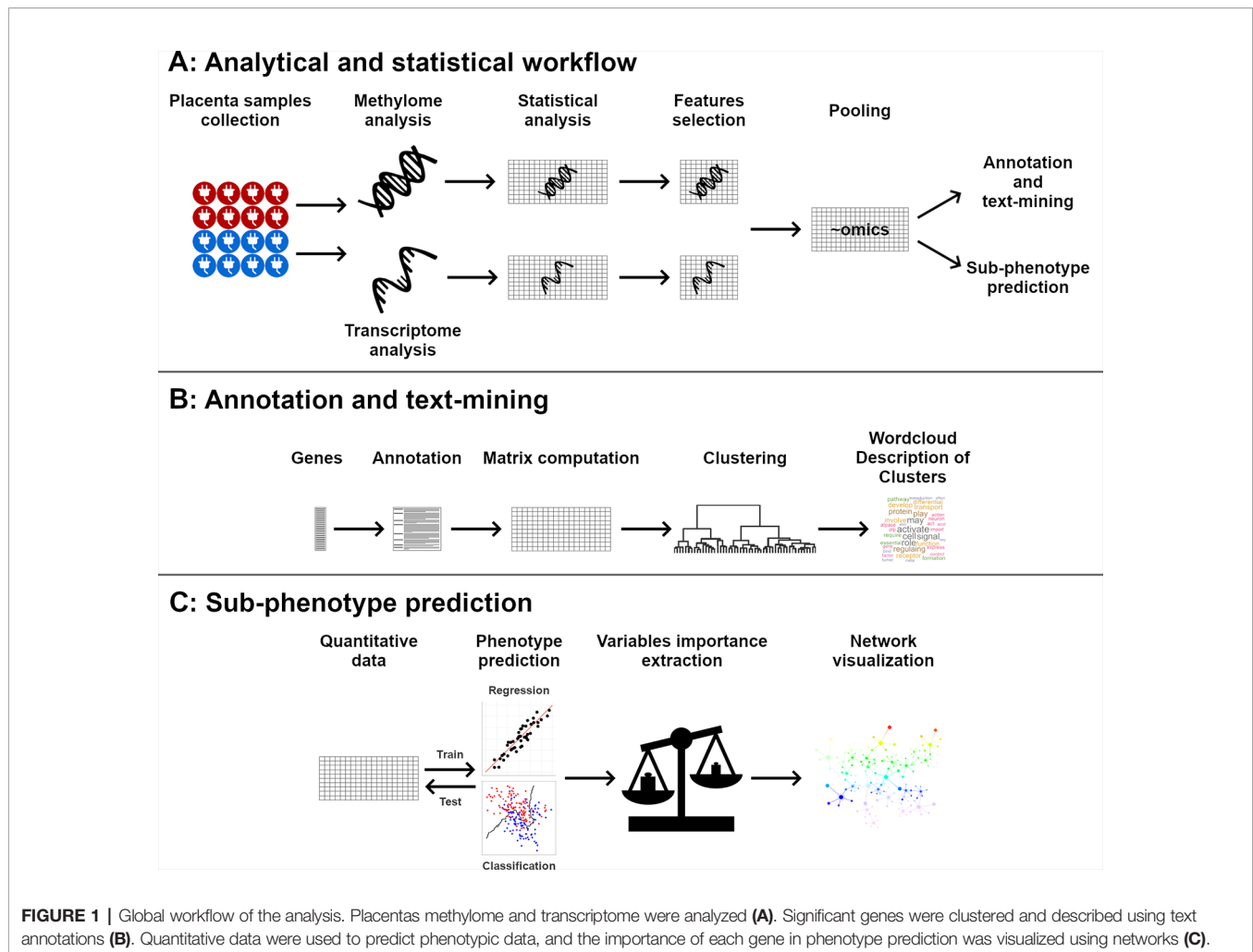
### Patients

All placentas were collected from Angers University Hospital. This study was approved by the Ethics Committee of Angers. All patients gave their informed consent for the use of their placenta. Clinical data related to the mother and the fetus, as well as neonatal data, were collected from the patients’ obstetric files. The cohort was registered at the French CNIL (*Commission Nationale de l’Informatique et des Libertés* no. pWP03752UL, ethics committee for the collection of clinical data from patient records). The study was validated by the French CPP (*Comité de Protection des Personnes*) and registered to the French Ministry of Research under number DC-2011-1467. The study was conducted in accordance with the declaration of Helsinki.

Placentas were obtained from caesarean sections before onset of labor or from vaginal delivery. For the analysis, patients were classified into two groups: IUGR and control group. The IUGR group was defined by a reduction of fetal growth during gestation, with a notch observed by Echo-Doppler in at least one uterine artery and with Doppler abnormalities on umbilical Doppler and/or cerebral Doppler and/or ductus venosus, and with a birth weight below the tenth percentile according to Audipog growth curves (American College of Obstetricians and Gynecologists, 2013) and confirmed by the anatomopathological analysis of the placenta after birth. The control group was defined by women with normal pregnancy and who underwent a planned caesarean section. All obstetrical and neonatal data were collected prospectively from medical records.

### Placental Samples

To avoid degradation, only placental tissues dissected within a time frame of 30 min after delivery were included. After removal



of the maternal decidua and amniotic membrane, sections of 1 cm<sup>3</sup> of placental villi were dissected from four different cotyledons between the basal and chorionic plates, as previously described (Gascoïn-Lachambre et al., 2010). After vigorous washing with PBS to remove maternal blood, tissues were immediately frozen in liquid nitrogen, before storage at -80 °C, to further extract DNA and RNA. Placentas were then sent for anatomopathological analysis or stored at the biological core facility at Angers University Hospital.

## DNA Preparation and Microarray Hybridization

Genomic DNA extraction was performed manually using a QIAamp DNA mini QIAcube Kit (Qiagen, Venlo, Netherlands), according to the manufacturer's protocol.

DNA was treated with bisulfite using an EZ-96 DNA Methylation Kit on a Zymo Spin I-96 column (Zymo Research, Irvine, CA, U.S.A.). Bisulfite-converted DNA was amplified, fragmented, and hybridized to Illumina Human Methylation 450k microarrays using an Illumina Hybridization Oven (Illumina, San Diego, CA, U.S.A.), according to the

manufacturer's protocol. Slides were analyzed by an Illumina I-Scan (Illumina, San Diego, CA, U.S.A.).

Raw iDAT files were directly imported in R software (R Development Core Team, 2008) and processed using the R minfi package (Aryee et al., 2014). Raw data were normalized using functional normalization (Fortin et al., 2014) before constructing the beta matrix for all 36 samples and 485,512 CpG sites (methylomics dataset).

## RNA Preparation and Microarray Hybridization

Total RNA was extracted after lysing samples with TRIzol reagent (Life Technologies, Carlsbad, CA, U.S.A.), using the RNeasy Micro kit (Qiagen, Venlo, Netherlands), according to the manufacturer's recommendations. Biotinylated, amplified cRNA was generated using the Illumina Total Prep RNA Amplification kit (Ambion, Life Technologies, Carlsbad, CA, U.S.A.), according to the manufacturer's recommendations. cRNA was hybridized on Illumina HumanHT-12 v4 Expression BeadChips, stained, and detected with the iScan system, according to the manufacturer's protocol (Illumina,

San Diego, CA, U.S.A.). A total of 47,323 marker probes were assessed, of which: 47,231 elements with sequences, with 46,841 with at least one genome alignment, including 34,627 elements mapped to at least one among 22,283 unique genes. GenomeStudio 2011 (version 1) and its Expression Analysis Module (version 1.9.0) were used for signal extraction and quantile normalization (Illumina, San Diego, CA, U.S.A.).

Normalized data for all 47,323 marker probes and 36 samples were imported into R software (R Development Core Team, 2008) and processed as described below (transcriptomics dataset).

## Omics Data Integration

Each omics dataset was processed independently. Levene's tests were used to assess the comparability of variances between control and IUGR groups. Significant features were determined using Student's *t*-tests. Alpha thresholds for *p*-value significance were set to  $\alpha = 0.05$ . For Student's *t*-tests, *p*-values were adjusted into *q*-values using the Benjamini-Hochberg method in order to control the false discovery rate. The  $\frac{IUGR}{control}$  fold-change was computed for all significant features. Only features with Levene's test *p*-value  $\geq 0.05$  and Benjamini-Hochberg adjusted Student's *t*-test *q*-value  $< 0.05$  were considered significant.

## Gene Annotation and Text-Mining

All genes showing a significant alteration in methylation or expression were annotated using abstracts available on PubMed, by automatic retrieval. Genes without available annotations were discarded. Abstracts were pre-processed by removing punctuation, short words (words of three characters or fewer) and stop words (i.e. common language non-specific words), and stemming (Willett, 2006). They were then analyzed by taking into account, in the same analytical process, unigrams, bigrams, and trigrams, commonly denoted as terms. A normalized term-frequency inverse-document-frequency (tf-idf) matrix (Aggarwal and Zhai, 2012) was then computed based on the frequency and specificity of each term in each gene summary, using the formula:

$$M_{i,j} = tf_i \times idf_i$$

With the inverse document frequency  $idf_i$  for the term *i*:

$$idf_i = \log_2 \left( \frac{|D|}{|\{d|t_i \in d\}|} \right)$$

where  $M_{i,j}$  is the value in the matrix for the term *i* and gene *j*,  $tf_i$  is the number of occurrences of the term *i* in the gene *j* summary divided by the total number of terms in the summary,  $|D|$  is the number of genes and  $|\{d|t_i \in d\}|$  is the number of gene summaries where the term *i* appears.

Due to the large dimension of the initial tf-idf matrix, a Latent Semantic Analysis (LSA) (Evangelopoulos, 2013) was performed in order to reduce its dimension and render further analyses possible. K-means was then used to perform clustering based on gene annotations similarity. Clusters were then summarized by terms closest to the cluster centers.

## Phenotype Prediction and Network Visualization

Support vector machines (SVM) are state-of-the-art machine-learning models that have already been successfully applied to several omics studies (Ben-Hur et al., 2008). They can successfully highlight non-linear correlations between genes and phenotypic traits, in order to highlight genes based on their links with several phenotypic traits (Altmann et al., 2010). Furthermore, SVM models are particularly suitable for high-dimensionality datasets, such as results of high-throughput analyses (Vanitha et al., 2015).

SVM models were trained using grid search cross-validation to predict four phenotypic traits as a function of omics data: control/IUGR group, premature birth (see below), birth weight, and head circumference at birth. These four phenotypic traits were chosen because of their known relevance in the IUGR pathophysiology. Term birth is defined by the International Classification of Diseases as between 37 (included) and 42 (excluded) weeks (Quinn et al., 2016), otherwise  $39.43 \pm 2.43$  weeks. To simplify, pregnancy term was expressed as a variable named premature birth, computed with the formula:

$$Premature\ birth = 39 - Gestational\ Age$$

Since gestational age and the newly-created variable, premature birth, are linearly correlated, this simplifies yet does not alter the interpretation of the results of the model's predictions. Values  $>2$  therefore indicate pre-term newborns, while values  $\leq -3$  indicate post-term newborns.

Both head circumference at birth and birth weight were expressed as Z-scores according to the gestational age and gender, based on Olsen growth curves (Olsen et al., 2010), to standardize values between infants born at different terms. Case-control classification is important to verify the integrity of the dimension-reduced dataset. Birth weight is a criterion of severity of the IUGR. Head circumference at birth is a criterion of high severity, due to the brain sparing effect (Cohen et al., 2015). Premature birth is indirectly linked to severity of these. Indeed, in most cases during IUGR pregnancies, a delivery is induced or carried out *via* caesarean section, to prevent either maternal or fetal damage. Exploring factors correlated with the premature birth may therefore allow exploring severity symptoms not directly and only linked to IUGR.

The dimensionality of the omics dataset had to be reduced before training the SVM, to reduce noise and achieve better model predictions (Keogh and Mueen, 2010). For this reason, only features with a significant difference between IUGR and control groups were used to train SVM models ( $q < 0.05$ , after Benjamini-Hochberg adjustment). Several methods may be used to reduce the dimensionality of a dataset (Guyon and Elisseeff, 2003). Features selection was preferred compared to other methods like Principal Components Analysis as it allows the use of the initial variables instead of computing new, abstract dimensions, making the final interpretation easier. Student's *t*-tests have already been evidenced as an effective method for features selection (Haury et al., 2011). By using Student's *t*-tests as the features selection method, this step could be applied



seamlessly to our omics analyses results, without modifying or altering the results.

The dataset was randomly partitioned into training and test sets, with a ratio of two-thirds/one-third, using stratified sampling in order to respect the original  $\frac{case}{control}$  ratio. Due to the low number of samples and the imbalance between IUGR and control samples, Synthetic Minority Over-sampling Technique (SMOTE) was used in order to synthetically increase the training set size (Chawla et al., 2002). Test sets were not modified to ensure unbiased results when measuring models' performances. Hyperparameters were fine-tuned with grid search cross-validation. Model results were assessed using accuracy for classification, and Pearson's correlation factor for regression.

The variable importance for predicting each phenotypic trait was computed for each feature by Permutation Importance (Breiman, 2001). These results were used to carry out a network visualization to assess the importance of each feature in the prediction of each phenotypic trait.

## Computational Tools

R software (version 3.4.1) and Python (version 3.6) were used to carry out all data processing and analysis, as well as to output all plots (van Rossum, 1995; R Development Core Team, 2008). Heat maps were created using the gplots package (Warnes et al., 2016). Gene functional annotation analysis was performed for both gene expression and gene methylation using the DAVID 6.8 online tool (Huang et al., 2009a, 2009b). Genes were annotated with abstracts available from PubMed (10/10/2019) using easyPubMed (Fantini, 2019). Text-mining and SVM

computing were processed using the python scikit-learn library (Pedregosa et al., 2011). Word clouds were created using the wordcloud R software package (Fellows, 2014). Hierarchical clustering was performed using the R software base package. Networks were constructed using Cytoscape (Shannon et al., 2003). The GIMP software was used to refine figures.

## RESULTS

### Cohort

Patient cohort is described in **Table 1**. It should be noted that while the control group is smaller, controls are much more homogeneous concerning clinically relevant phenotypic traits discussed below. F-tests show a significantly lower variance in this control group for gestational age at birth (in grams) ( $p = 4.48E-5$ ), head circumference at birth (in centimeters) ( $p = 1.08E-3$ ), and APGAR at 5 min ( $p = 3.48E-5$ ).

### Univariate Analyses

A total of 1651 features (1,072 DNA methylation sites, 579 transcripts) showed significantly different values between IUGR and control groups ( $q < 0.05$ ). The full list of significant features is available in **Supplementary Table 1**.

Since a significant difference in mean gestational age had been observed between IUGR and control groups, univariate analyses were re-run after excluding IUGR samples with a gestational age lower than 37 weeks. Kendall correlation tests were then performed to compare Student's t-tests results obtained for the whole cohort and for the high gestational age

**TABLE 1** | Description of the patient cohort.  $p$ -values were computed using Wilcoxon tests (quantitative values) or Fisher tests (percentages).

			Control group (n = 8)		IUGR group (n = 28)		p
<b>Maternal data</b>	Age (years)		35.4 ± 3.9	8	29.1 ± 5.9	28	0.006
	BMI before pregnancy	(kg/m <sup>2</sup> )	23.7 ± 7.0	8	25.1 ± 7.9	28	N.S.
	Tobacco consumption	Before pregnancy	0 (0.0%)	8	2 (7.1%)	28	N.S.
		During pregnancy	0 (0.0%)	8	9 (32.1%)	28	N.S.
	Ethnic group	European	7 (87.5%)	8	26 (92.9%)	28	N.S.
North African		1 (12.5%)	8	2 (7.1%)	28	N.S.	
<b>Obstetric data</b>	Gestivity		4.0 ± 2.1	8	2.5 ± 1.9	28	0.03
	Parity		2.6 ± 1.3	8	1.4 ± 0.9	28	0.005
	Weight gain (kg)		10.5 ± 10.5	8	9.1 ± 6.4	24	N.S.
	Type of delivery	Vaginal delivery	0 (0%)	8	5 (17.9%)	28	N.S.
		C-section	8 (100%)	8	23 (82.1%)	28	N.S.
	Pathology	IUGR	0 (0%)	8	16 (57.1%)	28	N/A
		IUGR + PE	0 (0%)	8	12 (42.9%)	28	N/A
	<b>Newborn data</b>	Gestational age (week)		38.7 ± 0.7	8	34.0 ± 3.9	28
Gender		Boy	4 (50.0%)	8	9 (32.1%)	28	N.S.
		Girl	4 (50.0%)	8	19 (67.9%)	28	N.S.
Birth weight		(Z-score)	-0.07 ± 0.89	8	-2.02 ± 0.75	28	<0.001
		(g)	3346 ± 444	8	1,524 ± 664	28	<0.001
Birth size		(Z-score)	-0.47 ± 0.74	7	-1.90 ± 0.80	26	<0.001
Birth size		(cm)	49.2 ± 1.8	7	39.2 ± 5.2	26	<0.001
Head circumference at birth		(Z-score)	0.22 ± 0.49	7	-1.30 ± 0.86	27	<0.001
Head circumference at birth		(cm)	34.6 ± 0.9	7	29.0 ± 3.4	27	<0.001
APGAR at 5 min			9.88 ± 0.35	8	9.11 ± 2.08	28	N.S.
Resuscitation at birth			0 (0%)	8	12 (42.9%)	28	0.03
NICU		0 (0%)	8	18 (64.3%)	28	0.003	

BMI, body mass index; PE, pre-eclampsia; NICU, neonatal intensive care unit; N.S., non-significant versus  $\alpha = 0.05$ ; N/A, not applicable.

restricted subset. Gene expression and gene methylation features were significantly correlated ( $p < 0.001$ ,  $\tau = 0.45$ ;  $p < 0.001$ ,  $\tau = 0.40$ , respectively).

Heat maps picturing all genes with significant expression (Figure 2) or methylation (Figure 2) alteration showed a global hypomethylation, as opposed to a balanced ratio between the number of overexpressed and underexpressed transcripts. While hierarchical clustering distinctly separated IUGR from control samples, IUGR samples appeared divided into two different clusters for both heatmaps, even though the exact distribution of IUGR samples is not exactly the same for epigenetic and expression alterations. In order to explain this behavior, gestational age at birth of IUGR samples according to clusters was plotted in Figure 3.

Gene functional annotation analysis, performed with DAVID, showed gene expression and/or methylation alterations significantly associated with several pathways ( $p < 0.05$ ), including: NAD-binding, histone acetylation, mTOR signaling pathway, lysosome, cell-cell adhesion and cell

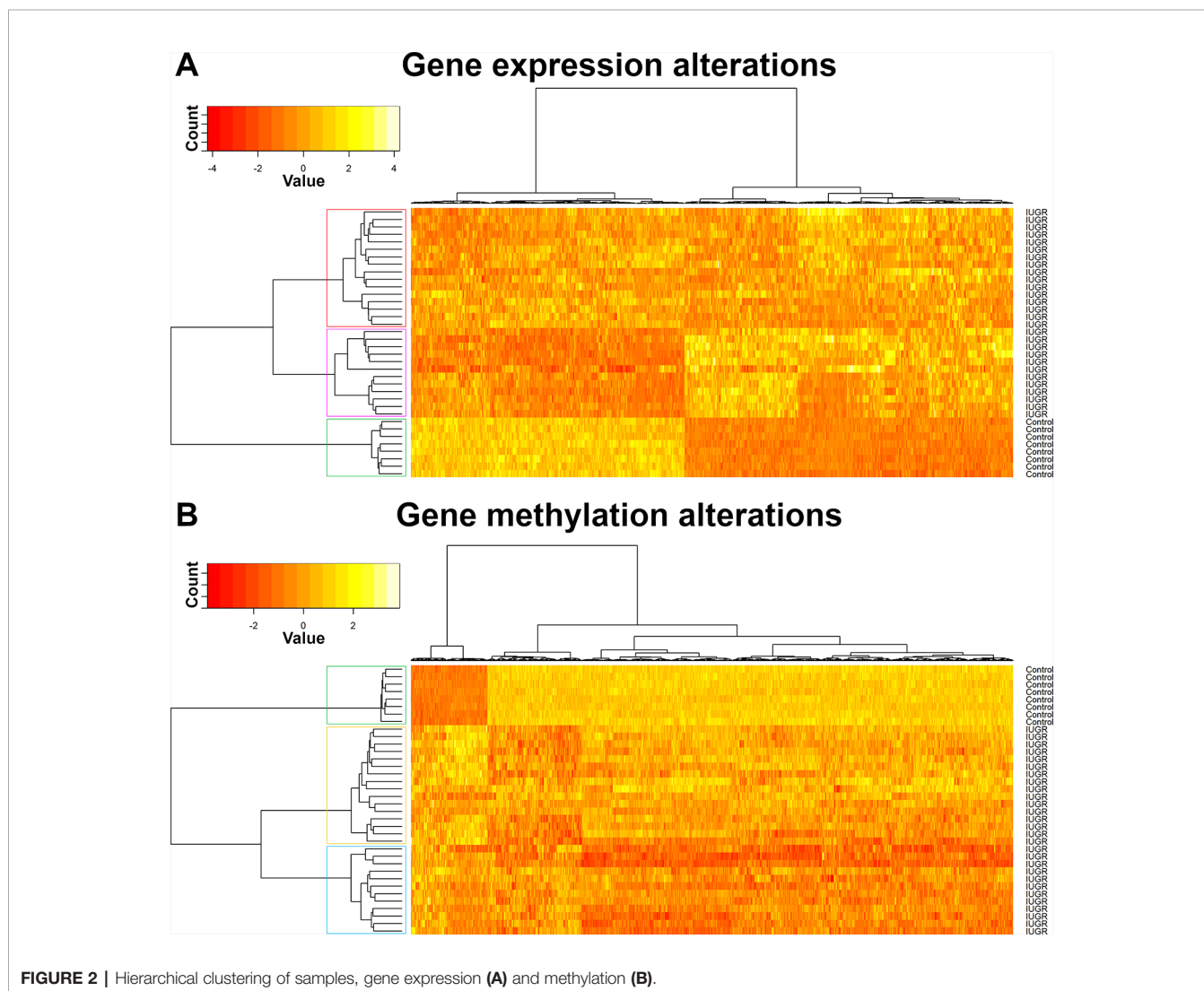
junction, calmodulin binding, and carbohydrates metabolism. The complete results are available in **Supplementary Table 2**.

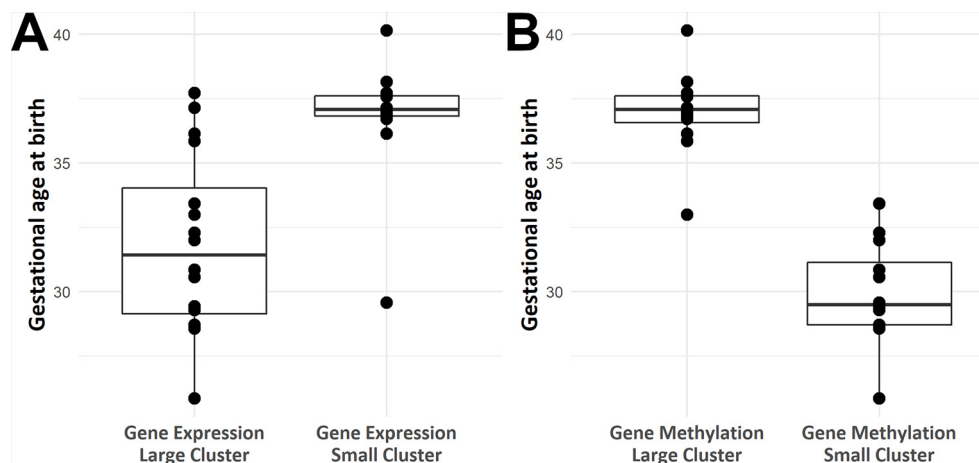
Only 25 genes were found to be altered both in methylome and transcriptome (Table 2). Among these 25 genes, eight show a significant linear correlation between methylation and expression.

### Textual Annotation and Text-Mining

Among these 1,651 features, 1,269 unique genes could be identified, and textual annotations were successfully retrieved for 1,259 of them. A total of 196,918 abstracts were retrieved (95% confidence interval: [146;167] abstracts per gene). LSA allowed reducing the dimension from 135,220 unique terms among all abstracts to 1,000 principal components, while retaining 97% of the initial tf-idf matrix variance. Genes were classified into 24 clusters. The cluster sizes ranged from 7 (0.6%) to 241 (19.1%) genes.

These clusters were summarized by word clouds picturing the most frequent and specific terms among the gene clusters,





**FIGURE 3 |** Box plots of gestational age at birth according to IUGR samples position in hierarchical clustering based on methylomics (A) and transcriptomics (B) data.

**TABLE 2 |** Genes found altered in both methylome and transcriptome. Numbers in brackets refer to the number of methylation sites (methylome) and transcripts (transcriptome) found significantly altered.

Gene symbol	Gene name	Epigenetics (sites count/total)	Gene expression (transcripts count/total)	r
<b>PAPPA2</b>	<i>Pregnancy-Associated Plasma Preproprotein-A2</i>	Hypomethylated (2/13)	Overexpressed (2/2)	-0.76
<b>AP2A1</b>	<i>Adaptor Related Protein Complex 2 Subunit Alpha 1</i>	Hypomethylated (1/26)	Underexpressed (2/3)	N.S.
<b>BCL6</b>	<i>B Cell CLL/Lymphoma 6</i>	Hypomethylated (2/55)	Overexpressed (1/2)	-0.65
<b>SLC2A1</b>	<i>Solute Carrier Family 2 Member 1</i>	Hypomethylated (2/34)	Overexpressed (1/1)	-0.42
<b>UNKL</b>	<i>Unkempt Family Like Zinc Finger</i>	Hypomethylated (2/74)	Underexpressed (1/3)	N.S.
<b>WSB1</b>	<i>WD Repeat and SOCS Box Containing 1</i>	Hypomethylated (1/19)	Underexpressed (2/3)	N.S.
<b>AFAP1</b>	<i>Actin Filament Associated Protein 1</i>	Hypomethylated (1/103)	Overexpressed (1/3)	N.S.
<b>ALDOA</b>	<i>Aldolase, Fructose-Bisphosphate A</i>	Hypomethylated (1/27)	Overexpressed (1/4)	-0.43
<b>ALKBH5</b>	<i>AlkB Homolog 5, RNA Demethylase</i>	Hypomethylated (1/23)	Overexpressed (1/1)	N.S.
<b>C1QTNF1</b>	<i>C1q And TNF Related 1</i>	Hypomethylated (1/40)	Underexpressed (1/3)	0.40
<b>CALM1</b>	<i>Calmodulin 1</i>	Hypermethylated (1/20)	Overexpressed (1/1)	N.S.
<b>DGKZ</b>	<i>Diacylglycerol Kinase Zeta</i>	Hypomethylated (1/62)	Overexpressed (1/3)	N.S.
<b>DLX5</b>	<i>Distal-Less Homeobox 5</i>	Hypomethylated (1/47)	Overexpressed (1/1)	N.S.
<b>FLNB</b>	<i>Filamin B</i>	Hypomethylated (1/40)	Overexpressed (1/1)	-0.58
<b>FOXP1</b>	<i>Forkhead Box K1</i>	Hypomethylated (1/175)	Underexpressed (1/2)	0.36
<b>LIMCH1</b>	<i>LIM and Calponin Homology Domains 1</i>	Hypomethylated (1/51)	Overexpressed (1/1)	-0.51
<b>PDP2</b>	<i>Pyruvate Dehydrogenase Phosphatase Catalytic Subunit 2</i>	Hypomethylated (1/13)	Underexpressed (1/2)	N.S.
<b>PDXK</b>	<i>Pyridoxal Kinase</i>	Hypomethylated (1/37)	Underexpressed (1/1)	N.S.
<b>PEA15</b>	<i>Proliferation and Apoptosis Adaptor Protein 15</i>	Hypomethylated (1/12)	Overexpressed (1/1)	N.S.
<b>PLEKHA2</b>	<i>Pleckstrin Homology Domain Containing A2</i>	Hypermethylated (1/22)	Overexpressed (1/4)	N.S.
<b>RALGPS1</b>	<i>Ral GEF With PH Domain and SH3 Binding Motif 1</i>	Hypomethylated (1/20)	Underexpressed (1/1)	N.S.
<b>RRAD</b>	<i>RRAD, Ras Related Glycolysis Inhibitor and Calcium Channel Regulator</i>	Hypomethylated (1/13)	Overexpressed (1/2)	N.S.
<b>SFRS8</b>	<i>Splicing Factor SWAP</i>	Hypomethylated (1/77)	Underexpressed (1/1)	N.S.
<b>UCKL1</b>	<i>Uridine-Cytidine Kinase 1 Like 1</i>	Hypomethylated (1/18)	Underexpressed (1/1)	N.S.
<b>USP5</b>	<i>Ubiquitin Specific Peptidase 5</i>	Hypomethylated (1/23)	Underexpressed (1/1)	N.S.

Pearson's correlation coefficient  $r$  is given for genes with a significant correlation between methylation and expression. N.S., Not significant.

allowing a quick and easy grasp and visualization of the global role of the clusters (Figure 4).

## Predicting Phenotypic Traits From Omics Data

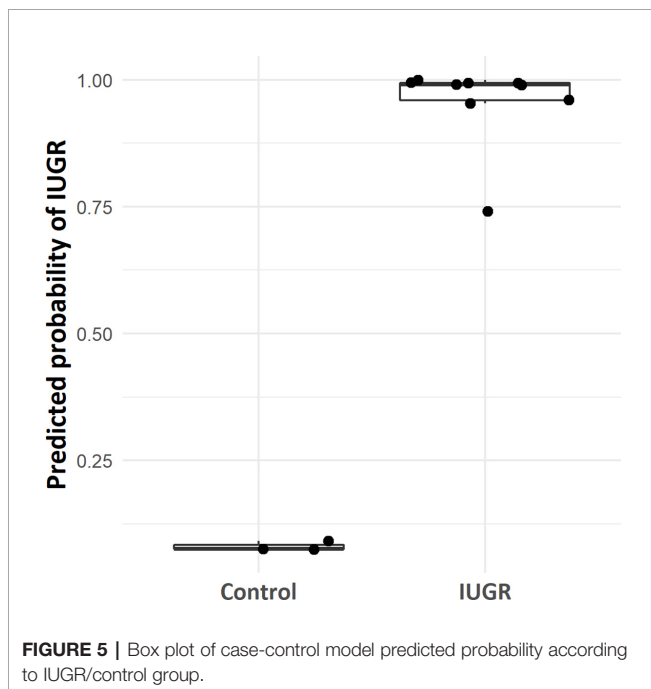
The 1,651 features were used as input data to predict the outcome for four phenotypic traits (IUGR, premature birth, birth weight, and head circumference), in order to measure the importance of each gene in sub-phenotypic prediction. Class-

control classification showed perfect predictions on the test set, with clearly distinct predicted probabilities between control and IUGR samples (Figure 5). This large gap of probabilities between IUGR and control samples confirmed the robustness of the model. These results were expected, as only features showing a significant difference between IUGR and control groups were selected for training the model. Furthermore, the previous unsupervised analysis (Figure 2) confirmed a clear distinction between IUGR and control samples.





**FIGURE 4 |** Word clouds summarizing the most frequent and specific terms among the 24 gene clusters (A–X).



Premature birth, birth weight, and head circumference scores predicted on test samples were linearly correlated with actual values ( $p < 0.01$ ) (Figure 6).

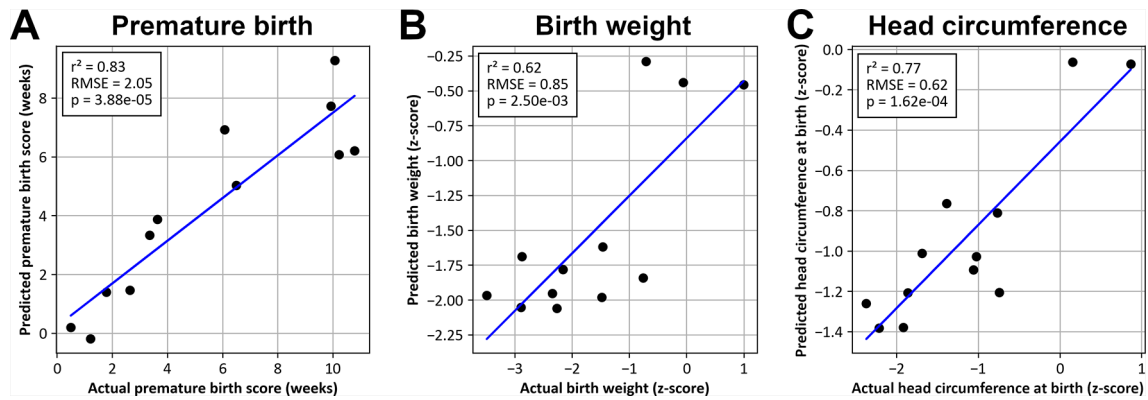
A network was created to represent all omics features with at least 10% importance for predicting at least one phenotypic trait (Figure 7). Among the nine genes with high importance (> 80%) in the prediction of at least one phenotypic trait, five (NMD3, ORC6L, MAPK8, PDCL, PLP1), in the center of the network share an importance in predicting most studied phenotypic traits.

The full list of methylomics and transcriptomics features with importance higher than 50% for phenotypic prediction is available in the Supplementary Table 3.

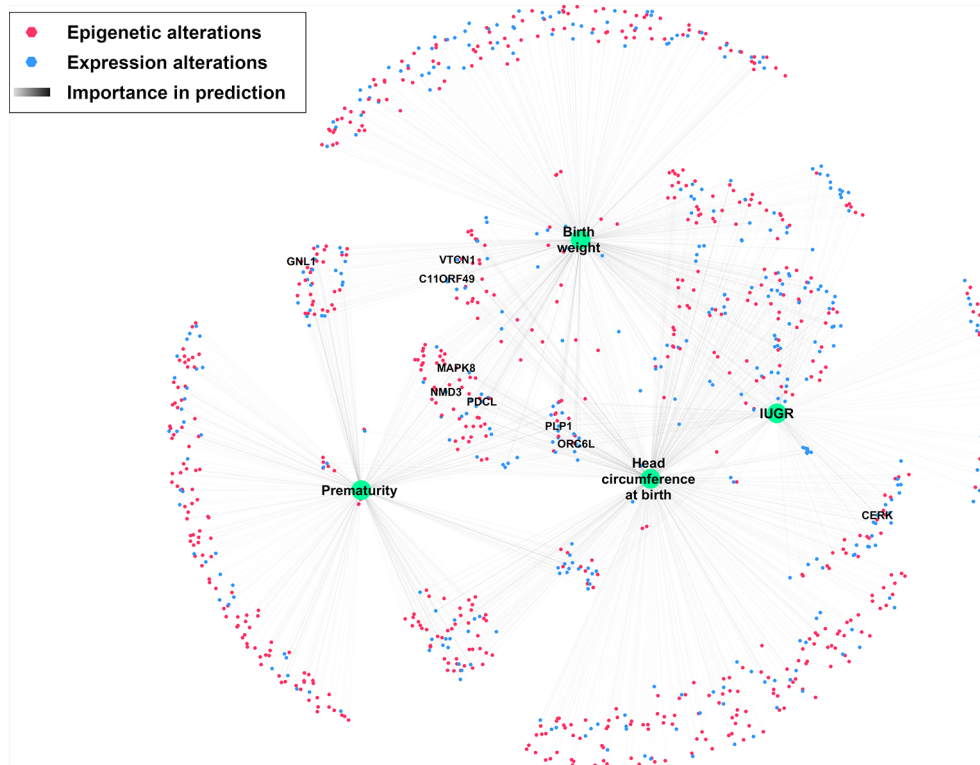
## DISCUSSION

### Text Annotation Clustering and Word Cloud Visualization

In most high-throughput gene studies, functional annotation analysis is a powerful tool, allowing the highlighting of pathways enriched in a particular pathophysiological context. However,



**FIGURE 6** | Values predicted by SVM models as a function of actual values for premature birth (A), birth weight (B), and head circumference at birth (C).



**FIGURE 7** | Network depicting significantly altered features and their importance in predicting IUGR phenotype. Nodes were positioned according to an Edge-weighted Spring Embedded Layout, based on feature importance for predicting each phenotypic trait. Only genes with at least 80% importance for predicting at least one phenotypic trait are labeled.

limiting gene annotation to categorical roles or pathways leads to a significant loss of knowledge in comparison with data available in literature.

Word clouds allowed a visual description of the main biological processes and pathways involved in the IUGR pathophysiology, in order to speed up and deepen the bibliographic work on genes significantly altered in IUGR.

### Cell Signaling and Proliferation

Many terms among the most frequent and specific refer to proto-oncogenes and cell proliferation and signaling and development mechanisms. This is confirmed by several genes isolated from both methylome and transcriptome (overexpression of *BCL6*, *CALM1*, *DLX5*, *PEA15*, *RRAD*, and underexpression of *FOXK1* and *UCKL1*).

## DNA, RNA Regulation, Transcription, Translation

Many gene clusters (C, D, F, G, I, L, respectively 5.4%, 5.5%, 2.1%, 1.8%, 2.0% and 2.9% of genes) refer to DNA methylation and repair, regulation of transcription, and RNA splicing and translation. Epigenetic and gene expression alterations in IUGR have been evidenced here as well as in literature (Kawai et al., 2015).

## Mitochondria and Oxidative Stress

Clusters H and T (2.0% and 1.8%, respectively) refer to mitochondria alterations, cell death and apoptosis, and redox reactions. Indeed, pregnancy increases ROS production and oxidative stress, causing damage to mitochondria and potentially leading to cell death, especially during pathological pregnancies like PE or IUGR (Myatt and Cui, 2004). These phenomena may have a role in the fetal programming of atherosclerosis (Leduc et al., 2010). *ALKBH5* (found hypomethylated, overexpressed) encodes a hypoxia-inducible factor playing a role in cell proliferation (Zhang et al., 2016).

## Intra- and Extra-Cellular Matrix

Several clusters (E, F, W, respectively 2.3%, 2.1% and 5.0%) suggest primarily cytoskeleton and cell-cell junction alterations. Furthermore, cluster N (1.0%) refers to intra-cellular trafficking and cell mechanisms relying heavily on the cytoskeleton. Riquelme and her colleagues (Riquelme et al., 2011) have already evidenced abnormalities in the lipid raft composition of the microvillous membrane of the placental syncytiotrophoblast, linked with alterations in the expression of several cytoskeletal proteins (actin, ezrin, and cytokeratin-7) in placentas from pathological pregnancies (PE and IUGR). They suggest that these cytoskeleton alterations might be responsible for alterations in the syncytiotrophoblast microvilli, which may play a major role in the IUGR pathophysiology. Among the genes found altered in both methylome and transcriptome, *AFAP1* is a major regulator of the cytoskeleton structure (Xiao et al., 2012). *FLNB* codes for an actin-binding protein crosslinking actin filaments and playing various roles including cell proliferation and angiogenesis through mechanotransduction (Xu et al., 2017). Clusters P and X (2.0% and 6.9%, respectively) refer to extracellular matrix alterations. Such alterations have already been evidenced in IUGR (Merchant et al., 2004; Swierczewski et al., 2012).

## Protein Degradation and Turnover

Cluster S (2.6%) refers to protein SUMOylation, ubiquitination, and degradation. It has been evidenced that protein ubiquitination is altered in IUGR and PE, particularly due to a modulation by oxidative stress, with an increased degradation of p53 and Mcl-1 proteins, contributing to the pathological mechanisms of the diseases (Rolfo et al., 2012). *WSB1* (underexpressed here) mediates ubiquitination and proteolytic degradation, and is also involved in cell and glucose metabolism, playing a role in hypoxia-related mechanisms (Haque et al., 2016). *USP5* (underexpressed here) codes for a deubiquitinating enzyme which has also been shown to play a role in cell cycle modulation.

## Heart and Skeletal Muscle Development

Heart and skeletal muscles are referred to in cluster P (2.0%). Wang et al. (Wang et al., 2013) and Yates et al. (Yates et al., 2012) already reported that hypoxemia and hypoglycaemia undergone during IUGR decrease muscle mass in offspring. *DGKZ* (found hypomethylated, overexpressed) is known to induce muscle fiber hypertrophy and plays a role in the adaptation to energy metabolism alterations (Benziane et al., 2017). *FOXK1* induces muscle progenitor cell proliferation and inhibits their differentiation (Shi et al., 2012). *FOXK1* was found here both hypomethylated and underexpressed. This underexpression might be due to another role of *FOXK1* in repressing starvation-induced atrophy and autophagy (Bowman et al., 2014).

## Energy Metabolism and Insulin Resistance

Major references are made to fat and lipid metabolism in cluster O (5.4%) and cluster Q (1.2%). These clusters support the hypothesis of an alteration of lipid and fat metabolism during IUGR, reflecting mechanisms of insulin resistance. Several genes found altered in both methylome and transcriptome support this pathway. Among these genes, *PAPPA-2* is the gene with the largest number of methylation sites significantly altered (hypomethylation), and with the largest number of transcripts significantly differently expressed (overexpression) in IUGR placentas. Its overexpression has already been reported in both maternal blood and the placenta in IUGR (Whitehead et al., 2013) and PE (Kramer et al., 2016). *PAPPA-2* encodes a protein cleaving the insulin-like growth factor 1 (IGF-1) from a ternary complex with IGF binding proteins (IGFBP-3) (Fujimoto et al., 2017). Via this regulation of the IGF-1 bioavailability, it plays a key role in both placenta development and fetal growth. Both low and high levels of IGF-1 have also been associated with insulin resistance (Friedrich et al., 2012). Interestingly, the *STC2* gene, encoding the *PAPPA2* inhibitor stanniocalcin-2, was found significantly hypomethylated here, but its expression was not significantly altered between IUGR and control groups.

*PEA15* encodes a phosphoprotein responsible for insulin resistance and diabetes. Higher levels of expression of *PEA15* have been reported in both patients with diabetes mellitus type 2 (Condorelli et al., 1998) and in euglycemic patients with impaired insulin sensitivity (Valentino et al., 2006). The *DGKZ* gene, already discussed above, has been proven to play a role in the protection against peripheral insulin resistance and in improving overall energy metabolism (Benziane et al., 2017). *SLC2A1*, also known as glucose transporter 1 (*GLUT1*), is the major glucose transporter in the human placenta and the rate-limiting step of glucose transport from the placenta to the fetus (Illsley, 2000). Its overexpression here might reflect mechanisms of adaptation to fetal nutrient restriction. *CIQTNF1*, also known as glucose-dependent insulinotropic polypeptide (GIP) is an adipokine, whose secretion by adipocytes is increased under hypoxia, partially under the control of HIF-1 $\alpha$ . It stimulates proinflammatory gene expression and impairs insulin sensitivity of adipocytes (Chen et al., 2015). However, *CIQTNF1* was found underexpressed in this study.

Two more genes supporting these mechanisms of insulin resistance were found here among the most overexpressed genes:



*HTRA4* (IGF binding domain containing protein, fold-change = 7.33) and *LEP* (leptin, fold-change = 4.89). This major overexpression had already been observed in both IUGR (Madeleneau et al., 2015) and PE (Brew et al., 2016).

### Sub-Phenotype Prediction

Unsupervised clustering (Figure 2) showed a clear distinction between IUGR and controls and suggested the existence of multiple sub-phenotypes in the IUGR group (Figure 3).

As expected, SVM models were able to accurately predict such phenotypic traits: gestational age at birth, birth weight, and head circumference, using only a small subset of the whole data, i.e. 1,651 (0.3%) methylome and transcriptome variables. These results confirmed the high predictive value of the genes highlighted in this study in the IUGR, as well as in several variables of severity and pathophysiology of the IUGR.

In particular, nine genes with high importance in the prediction of these phenotypic traits were observed. Network visualization (Figure 7) showed that most of these genes are correlated with most clinically relevant traits studied here.

Among these genes, *CERK*, *GNL1*, *PLP1*, and *MAPK8* are known to be altered or play a direct role in the pathophysiology of IUGR or PE in various pathways discussed above: differentiation and proliferation regulation, response to hypoxia and oxidative stress, and neurological maturation (Vaiman et al., 2011; Reid et al., 2012; Goyal et al., 2013; Chan et al., 2019). For the other genes (*VTCN1*, *C11ORF49*, *PDCL*, *ORC6L*, *NMD3*), no obvious link with IUGR was found in literature, creating a topic for future studies regarding their exact role in the IUGR pathophysiology.

### Limits

Our study was mainly limited by the imbalance between cases and controls and the relatively weak number of controls. However, as already stated, controls show a significantly lower variance for most phenotypic traits discussed in this study. Furthermore, oversampling methods were used in order to compensate this limit and prevent model overfitting, while assessing the importance of genes on unmodified test sets which were not previously used for training models.

### Conclusion

Many epigenetic and gene expression alterations in IUGR placentas have been observed here, some of them confirming previous mechanisms already published, and others being new findings. Several major pathways were highlighted by annotation text-mining analysis: cell cycle and proliferation, regulation of apoptosis, epigenetic modifications, transcription, translation, oxidative stress and hypoxia, cytoskeleton and cell structure, protein degradation and turnover, autophagy, muscle development, and glucose and lipid energy metabolism. The involvement of these pathways was supported by significant differences in both methylome and transcriptome. Finally, several key genes with high correlation with phenotypic traits clinically relevant for IUGR were observed and may constitute potential targets for future study.

### DATA AVAILABILITY STATEMENT

Array-based datasets for both genome methylation and expression have been deposited at the European Genome-phenome Archive (EGA), which is hosted by the EBI and the CRG, under accession number EGAS00001003467. Further information about EGA can be found on <https://ega-archive.org> (The European Genome-phenome Archive of human data consented for biomedical research, <http://www.nature.com/ng/journal/v47/n7/full/ng.3312.html>). Analysis output files are available in **Supplementary Material (Supplementary Tables 1–3)**.

### ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics committee of Angers. The patients/participants provided their written informed consent to participate in this study.

### AUTHOR CONTRIBUTIONS

FC: literature search, data analysis, data interpretation, figures, writing. NH: literature search, data collection, data analysis. XD: data interpretation. GR: data interpretation. GB proofreading, expertise in data analysis methods. JB: expertise in data analysis methods. VP: proofreading. GLen: proofreading. OB: data collection. GLeg: data collection. DM-P: proofreading. MC: proofreading, expertise in data analysis methods. PG: data analysis, data interpretation, expertise in data analysis methods. PR: co-director of the study, data interpretation, expertise, writing. GG: director of the study, literature search, data collection, data interpretation, expertise, writing.

### FUNDING

This study was funded by a grant from the Angers University Hospital, France.

### ACKNOWLEDGMENTS

We acknowledge support from the Institut National de la Santé et de la Recherche Médicale (INSERM), the Centre National de la Recherche Scientifique (CNRS), the University of Angers, and Angers University Hospital.

### SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.01292/full#supplementary-material>

## REFERENCES

- Aggarwal, C. C., and Zhai, C. (2012). "A Survey of Text Clustering Algorithms," in *Mining Text Data*. Eds. C. C. Aggarwal and C. Zhai (Boston, MA: Springer US), 77–128. doi: 10.1007/978-1-4614-3223-4\_4
- Altmann, A., Tološi, L., Sander, O., and Lengauer, T. (2010). Permutation importance: a corrected feature importance measure. *Bioinformatics* 26, 1340–1347. doi: 10.1093/bioinformatics/btq134
- American College of Obstetricians and Gynecologists (2013). ACOG Practice bulletin no. 134: fetal growth restriction. *Obstet. Gynecol.* 121, 1122–1133. doi: 10.1097/01.AOG.0000429658.85846.f9
- Aryee, M. J., Jaffe, A. E., Corrada-Bravo, H., Ladd-Acosta, C., Feinberg, A. P., Hansen, K. D., et al. (2014). Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* 30, 1363–1369. doi: 10.1093/bioinformatics/btu049
- Ben-Hur, A., Ong, C. S., Sonnenburg, S., Schölkopf, B., and Rätsch, G. (2008). Support Vector Machines and Kernels for Computational Biology. *PLoS Comput. Biol.* 4, e1000173. doi: 10.1371/journal.pcbi.1000173
- Benziane, B., Borg, M. L., Tom, R. Z., Riedl, I., Massart, J., Björnholm, M., et al. (2017). DGK $\zeta$  deficiency protects against peripheral insulin resistance and improves energy metabolism. *J. Lipid Res.* 58, 2324–2333. doi: 10.1194/jlr.M079723
- Bowman, C. J., Ayer, D. E., and Dynlacht, B. D. (2014). Foxk proteins repress the initiation of starvation-induced atrophy and autophagy programs. *Nat. Cell Biol.* 16, 1202–1214. doi: 10.1038/ncb3062
- Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi: 10.1023/A:1010933404324
- Brew, O., Sullivan, M. H. F., and Woodman, A. (2016). Comparison of normal and pre-eclamptic placental gene expression: a systematic review with meta-analysis. *PLoS One* 11, e0161504. doi: 10.1371/journal.pone.0161504
- Buffat, C., Mondon, F., Rigourd, V., Boubred, F., Bessières, B., Fayol, L., et al. (2007). A hierarchical analysis of transcriptome alterations in intrauterine growth restriction (IUGR) reveals common pathophysiological pathways in mammals. *J. Pathol.* 213, 337–346. doi: 10.1002/path.2233
- Chan, K. J., Swan, K. F., Narayanappa, A., Mersereau, C., and Intapad, S. (2019). Imbalance of Sphingolipids synthesis/degradation pathway in preeclamptic mouse placenta and kidney of intrauterine growth restricted mouse fetus. *FASEB J.* 33, 593.3–593.3. doi: 10.1096/fasebj.2019.33.1\_supplement.593.3
- Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *J. Artif. Intell. Res.* 16, 321–357. doi: 10.1613/jair.953
- Chen, S., Okahara, F., Osaki, N., and Shimotoyodome, A. (2015). Increased GIP signaling induces adipose inflammation via a HIF-1 $\alpha$ -dependent pathway and impairs insulin sensitivity in mice. *Am. J. Physiol. Endocrinol. Metab.* 308, E414–E425. doi: 10.1152/ajpendo.00418.2014
- Cohen, E., Baerts, W., and Bel, F. V. (2015). Brain-sparing in intrauterine growth restriction: considerations for the neonatologist. *Neonatology* 108, 269–276. doi: 10.1159/000438451
- Condorelli, G., Vigiotta, G., Iavarone, C., Caruso, M., Tocchetti, C. G., Andreozzi, F., et al. (1998). PED/PEA-15 gene controls glucose transport and is overexpressed in type 2 diabetes mellitus. *EMBO J.* 17, 3858–3866. doi: 10.1093/emboj/17.14.3858
- Ding, Y., and Cui, H. (2017). Integrated analysis of genome-wide DNA methylation and gene expression data provide a regulatory network in intrauterine growth restriction. *Life Sci.* 179, 60–65. doi: 10.1016/j.lfs.2017.04.020
- Evangelopoulos, N. E. (2013). Latent semantic analysis. *Wiley Interdiscip. Rev. Cogn. Sci.* 4, 683–692. doi: 10.1002/wcs.1254
- Fantini, D. (2019). *easyPubMed: Search and Retrieve Scientific Publication Records from PubMed*. Available at: <https://CRAN.R-project.org/package=easyPubMed> [Accessed October 10, 2019].
- Fellows, I. (2014). wordcloud: Word Clouds. Available at: <https://CRAN.R-project.org/package=wordcloud>.
- Flamant, C., and Gascoin, G. (2013). Short-term outcome and small for gestational age newborn management. *J. Gynecol. Obstet. Biol. Reprod. (Paris)* 42, 985–995. doi: 10.1016/j.jgyn.2013.09.020
- Fortin, J.-P., Labbe, A., Lemire, M., Zanke, B. W., Hudson, T. J., Fertig, E. J., et al. (2014). Functional normalization of 450k methylation array data improves replication in large cancer studies. *Genome Biol.* 15 (12), 503. doi: 10.1186/s13059-014-0503-2
- Friedrich, N., Thuesen, B., Jørgensen, T., Juul, A., Spielhagen, C., Wallaschofski, H., et al. (2012). The Association Between IGF-I and Insulin Resistance. *Diabetes Care* 35, 768–773. doi: 10.2337/dc11-1833
- Fujimoto, M., Hwa, V., and Dauber, A. (2017). Novel Modulators of the Growth Hormone - Insulin-Like Growth Factor Axis: Pregnancy-Associated Plasma Protein-A2 and Stanniocalcin-2. *J. Clin. Res. Pediatr. Endocrinol.* 9, 1–8. doi: 10.4274/jcrpe.2017.S001
- Gascoin, G., and Flamant, C. (2013). Long-term outcome in context of intra uterine growth restriction and/or small for gestational age newborns. *J. Gynecol. Obstet. Biol. Reprod. (Paris)* 42, 911–920. doi: 10.1016/j.jgyn.2013.09.014
- Gascoin-Lachambre, G., Buffat, C., Rebourcet, R., Chelbi, S. T., Rigourd, V., Mondon, F., et al. (2010). Cullins in human intra-uterine growth restriction: expression and epigenetic alterations. *Placenta* 31, 151–157. doi: 10.1016/j.placenta.2009.11.008
- Gomez-Cabrero, D., Abugessaisa, I., Maier, D., Teschendorff, A., Merckenschlager, M., Gisel, A., et al. (2014). Data integration in the era of omics: current and future challenges. *BMC Syst. Biol.* 8, I1. doi: 10.1186/1752-0509-8-S2-I1
- Goyal, R., Van Wickle, J., Goyal, D., Matei, N., and Longo, L. D. (2013). Antenatal maternal long-term hypoxia: acclimatization responses with altered gene expression in ovine fetal carotid arteries. *PLoS One* 8, e82200. doi: 10.1371/journal.pone.0082200
- Guyon, I., and Elisseeff, A. (2003). An introduction to variable and feature selection. *J. Mach. Learn. Res.* 3, 1157–1182.
- Haque, M., Kendal, J. K., MacIsaac, R. M., and Demetrick, D. J. (2016). WSB1: from homeostasis to hypoxia. *J. Biomed. Sci.* 23, 61. doi: 10.1186/s12929-016-0270-3
- Haury, A.-C., Gestraud, P., and Vert, J.-P. (2011). The influence of feature selection methods on accuracy, stability and interpretability of molecular signatures. *PLoS One* 6, e28210. doi: 10.1371/journal.pone.0028210
- Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 37, 1–13. doi: 10.1093/nar/gkn923
- Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4, 44–57. doi: 10.1038/nprot.2008.211
- Illsley, N. P. (2000). Glucose transporters in the human placenta. *Placenta* 21, 14–22. doi: 10.1053/plac.1999.0448
- Kaplan, C. G. (2007). Fetal and maternal vascular lesions. *Semin. Diagn. Pathol.* 24, 14–22. doi: 10.1053/j.semmp.2007.02.005
- Kawai, T., Yamada, T., Abe, K., Okamura, K., Kamura, H., Akaishi, R., et al. (2015). Increased epigenetic alterations at the promoters of transcriptional regulators following inadequate maternal gestational weight gain. *Sci. Rep.* 5, 14224. doi: 10.1038/srep14224
- Keogh, E., and Mueen, A. (2010). "Curse of Dimensionality," in *Encyclopedia of Machine Learning*. Eds. C. Sammut and G. I. Webb (Boston, MA: Springer US), 257–258. doi: 10.1007/978-0-387-30164-8\_192
- Kramer, A. W., Lamale-Smith, L. M., and Winn, V. D. (2016). Differential expression of human placental PAPP-A2 over gestation and in preeclampsia. *Placenta* 37, 19–25. doi: 10.1016/j.placenta.2015.11.004
- Leduc, L., Levy, E., Bouity-Voubou, M., and Delvin, E. (2010). Fetal programming of atherosclerosis: possible role of the mitochondria. *Eur. J. Obstet. Gynecol. Reprod. Biol.* 149, 127–130. doi: 10.1016/j.ejogrb.2009.12.005
- Madeleneau, D., Buffat, C., Mondon, F., Grimault, H., Rigourd, V., Tsatsaris, V., et al. (2015). Transcriptomic analysis of human placenta in intrauterine growth restriction. *Pediatr. Res.* 77, 799–807. doi: 10.1038/pr.2015.40
- Malhotra, A., Allison, B. J., Castillo-Melendez, M., Jenkin, G., Polglase, G. R., and Miller, S. L. (2019). Neonatal morbidities of fetal growth restriction: pathophysiology and impact. *Front. Endocrinol.* 10, 55. doi: 10.3389/fendo.2019.00055
- Merchant, S. J., Crocker, I. P., Baker, P. N., Tansinda, D., Davidge, S. T., and Guilbert, L. J. (2004). Matrix metalloproteinase release from placental explants of pregnancies complicated by intrauterine growth restriction. *J. Soc. Gynecol. Investig.* 11, 97–103. doi: 10.1016/j.jsg.2003.08.005
- Murki, S. (2014). Intrauterine growth retardation - a review article. *J. Neonatal Biol.* 3, 1–13. doi: 10.4172/2167-0897.1000135



- Myatt, L., and Cui, X. (2004). Oxidative stress in the placenta. *Histochem. Cell Biol.* 122, 369–382. doi: 10.1007/s00418-004-0677-x
- Olsen, I. E., Groveman, S. A., Lawson, M. L., Clark, R. H., and Zemel, B. S. (2010). New intrauterine growth curves based on United States data. *PEDIATRICS* 125, e214–e224. doi: 10.1542/peds.2009-0913
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830. doi: 10.1007/978-1-4842-0958-5\_8
- Quinn, J.-A., Munoz, F. M., Gonik, B., Frau, L., Cutland, C., Mallett-Moore, T., et al. (2016). Preterm birth: case definition & guidelines for data collection, analysis, and presentation of immunisation safety data. *Vaccine* 34, 6047–6056. doi: 10.1016/j.vaccine.2016.03.045
- R Development Core Team. (2008). “R: A language and environment for statistical computing,” in *R Foundation for Statistical Computing* (Vienna, Austria). Available at: <http://www.R-project.org>.
- Reid, M. V., Murray, K. A., Marsh, E. D., Golden, J. A., Simmons, R. A., and Grinspan, J. B. (2012). Delayed myelination in an intrauterine growth retardation model is mediated by oxidative stress upregulating bone morphogenetic protein 4. *J. Neuropathol. Exp. Neurol.* 71, 640–653. doi: 10.1097/NEN.0b013e31825cfa81
- Riquelme, G., Vallejos, C., de Gregorio, N., Morales, B., Godoy, V., Berrios, M., et al. (2011). Lipid rafts and cytoskeletal proteins in placental microvilli membranes from preeclamptic and IUGR pregnancies. *J. Membr. Biol.* 241, 127. doi: 10.1007/s00232-011-9369-3
- Rolfo, A., Garcia, J., Todros, T., Post, M., and Caniggia, I. (2012). The double life of MULE in preeclamptic and IUGR placentae. *Cell Death Dis.* 3, e305. doi: 10.1038/cddis.2012.44
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303
- Sharma, D., Shastri, S., and Sharma, P. (2016). Intrauterine Growth Restriction: Antenatal and Postnatal Aspects. *Clin. Med. Insights Pediatr.* 10, 67–83. doi: 10.4137/CMPed.S40070
- Shi, X., Wallis, A. M., Gerard, R. D., Voelker, K. A., Grange, R. W., DePinho, R. A., et al. (2012). Foxk1 promotes cell proliferation and represses myogenic differentiation by regulating Foxo4 and Mef2. *J. Cell Sci.* 125, 5329–5337. doi: 10.1242/jcs.105239
- Swierczewski, A., Kobos, J., Pasiński, J., Kowalska-Koprek, U., and Karowicz-Bilińska, A. (2012). Expression of metalloproteinase MMP-9 and tissue inhibitor of metalloproteinase TIMP-2 in placenta of pregnant women with intrauterine growth restriction. *Ginekol. Pol.* 83, 439–445.
- Vaiman, D., Gascoin-Lachambre, G., Boubred, F., Mondon, F., Feuerstein, J.-M., Ligi, I., et al. (2011). The intensity of IUGR-induced transcriptome deregulations is inversely correlated with the onset of organ function in a rat model. *PLoS One* 6, e21222. doi: 10.1371/journal.pone.0021222
- Valentino, R., Lupoli, G. A., Raciti, G. A., Oriente, F., Farinaro, E., Della Valle, E., et al. (2006). The PEA15 gene is overexpressed and related to insulin resistance in healthy first-degree relatives of patients with type 2 diabetes. *Diabetologia* 49, 3058–3066. doi: 10.1007/s00125-006-0455-5
- van Helden, P. (2013). Data-driven hypotheses. *EMBO Rep.* 14, 104. doi: 10.1038/embor.2012.207
- van Rossum, G. (1995). Python tutorial, Technical Report CS-R9526.
- Vanitha, C. D. A., Devaraj, D., and Venkatesulu, M. (2015). Gene expression data classification using support vector machine and mutual information-based gene selection. *Proc. Comput. Sci.* 47, 13–21. doi: 10.1016/j.procs.2015.03.178
- Vayssière, C., Sentilhes, L., Ego, A., Bernard, C., Cambourieu, D., Flamant, C., et al. (2015). Fetal growth restriction and intra-uterine growth restriction: guidelines for clinical practice from the French College of Gynaecologists and Obstetricians. *Eur. J. Obstet. Gynecol. Reprod. Biol.* 193, 10–18. doi: 10.1016/j.jejogrb.2015.06.021
- Wang, T., Liu, C., Feng, C., Wang, X., Lin, G., Zhu, Y., et al. (2013). IUGR alters muscle fiber development and proteome in fetal pigs. *Front. Biosci. Landmark Ed.* 18, 598–607. doi: 10.2741/4123
- Warnes, G. R., Bolker, B., Bonebakker, L., Gentleman, R., Liaw, W. H. A., Lumley, T., et al. (2016). *gplots: Various R Programming Tools for Plotting Data*. Available at: <https://CRAN.R-project.org/package=gplots> [Accessed October 15, 2018].
- Whitehead, C. L., Walker, S. P., Ye, L., Mendis, S., Kaitu'u-Lino, T. J., Lappas, M., et al. (2013). Placental specific mRNA in the maternal circulation are globally dysregulated in pregnancies complicated by fetal growth restriction. *J. Clin. Endocrinol. Metab.* 98, E429–E436. doi: 10.1210/jc.2012-2468
- Willett, P. (2006). The Porter stemming algorithm: then and now. *Program* 40, 219–223. doi: 10.1108/00330330610681295
- Xiao, H., Han, B., Lodyga, M., Bai, X.-H., Wang, Y., and Liu, M. (2012). The actin-binding domain of actin filament-associated protein (AFAP) is involved in the regulation of cytoskeletal structure. *Cell. Mol. Life Sci. CMLS* 69, 1137–1151. doi: 10.1007/s00018-011-0812-5
- Xiao, X., Zhao, Y., Jin, R., Chen, J., Wang, X., Baccarelli, A., et al. (2016). Fetal growth restriction and methylation of growth-related genes in the placenta. *Epigenomics* 8, 33–42. doi: 10.2217/epi.15.101
- Xu, Q., Wu, N., Cui, L., Wu, Z., and Qiu, G. (2017). Filamin B: The next hotspot in skeletal research? *J. Genet. Genomics Yi Chuan Xue Bao* 44, 335–342. doi: 10.1016/j.jgg.2017.04.007
- Yates, D. T., Macko, A. R., Nearing, M., Chen, X., Rhoads, R. P., and Limesand, S. W. (2012). Developmental programming in response to intrauterine growth restriction impairs myoblast function and skeletal muscle metabolism. *J. Pregnancy* 2012, 631038. doi: 10.1155/2012/631038
- Zhang, S., Regnault, T. R. H., Barker, P. L., Botting, K. J., McMillen, I. C., McMillan, C. M., et al. (2015). Placental adaptations in growth restriction. *Nutrients* 7, 360–389. doi: 10.3390/nu7010360
- Zhang, C., Zhi, W. I., Lu, H., Samanta, D., Chen, I., Gabrielson, E., et al. (2016). Hypoxia-inducible factors regulate pluripotency factor expression by ZNF217- and ALKBH5-mediated modulation of RNA methylation in breast cancer cells. *Oncotarget* 7, 64527–64542. doi: 10.18632/oncotarget.11743

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Chabrun, Huetz, Dieu, Rousseau, Bouzillé, Chao de la Barca, Procaccio, Lenaers, Blanchet, Legendre, Mirebeau-Prunier, Cuggia, Guardiola, Reynier and Gascoin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.