Role of PPARγ in normal and pathological placental development

Fulin Liu

Université de Paris

Inserm
La science pour la santé
From science to health

Université de Paris

Ecole doctorale ED 562

Physiopathologie et pharmacotoxicologie placentaire humaine -

Microbiote pré & post natal, INSERM UMR-S 1139

Role of PPARγ in Normal and Pathological Placental Development

Fulin LIU

Thèse de doctorat du **Développement**

Dirigée par Thierry FOURNIER

Présentée et soutenue publiquement le 20/09/2021

Devant un jury composé de :

Vincent SAPIN	PU-PH	Université d'Auvergne	Rapporteur
Nadia ALFAIDY	DR	Université de Grenoble	Rapporteure
Nabila BOUATIA-NAJI	DR	Université de Paris	Examinatrice
Véronique FERCHAUD-ROUCHER	PHD HDR	Université de Nantes	Examinatrice
Thierry FOURNIER	DR	Université de Paris	Directeur de thèse

Abstract

In the first part, we mined the data sets of gene expression that combined the human placenta genome across pregnancy from 4 to 40 gestational weeks. Our results showed a total of 5173 genes involved in different periods of placentation with peroxisome proliferator-activated receptor (PPAR) signaling pathway confirmed to mediate the constant decrease of placental lipids throughout pregnancy. **In the second part,** we investigated one of the PPARs, PPARγ, in the trophoblasts, the cells that play a major role in the placenta. The research was conducted on extravillous cytotrophoblasts (EVCTs) and villous cytotrophoblasts (VCTs), where PPARγ is strongly expressed. We explored the genome-wide effects of activated PPARγ on EVCTs and VCTs. From our microarray data, we provided a broad perspective of PPARγ-activated biological processes in human trophoblasts. **In the third part,** we attempted to figure out the relationship between PPARγ and hypoxia-inducible factor (HIF) targets in the human first-trimester placenta transcriptome in response to physiological increased in oxygen levels. By comparing the transcriptomes of human placentas at 8-10 gestational weeks (2-3% O2) and 12-14 gestational weeks (8% O2), we characterized the similarities and differences between the enrichment patterns, as well as those associated with HIF targets. However, an intersection between the targets of PPARγ and HIF showed no overlapped element. **In the fourth part,** we explored the relationship between the polymorphisms of PPARγ and the susceptibility to preeclampsia, a pregnancy disease. With clinical characteristics and genotyped polymorphisms of PPARγ from a total of 1648 women, we applied eight machine learning algorithms to optimize predictive models. The decisive tree with the highest performance of accuracy and the area under receiver operating characteristic curve (AUC) was selected and displayed to show the procedure of preeclampsia prediction. **To sum up,** our findings support that PPARγ mediates the constant decrease of placental lipids throughout pregnancy through PPAR signaling pathway, provide a broad perspective of PPARγ-activated biological processes in EVCTs and VCTs through the analysis of transcriptomic signatures, and develop a pragmatic model for clinical practice with the polymorphisms of PPARγ.

Keywords: Placenta; Trophoblasts; PPARγ; Sequencing; Bioinformatics; Machine learning; Preeclampsia

Résumé

Le placenta est un organe transitoire indispensable au maintien de la grossesse et à la croissance du fœtus. La barrière placentaire permet les échanges en nutriments et en gaz entre le sang maternel et fœtal. La placentation commence dès l'implantation du blastocyste dans l'endomètre. Les aberrations de la structure et de la fonction placentaires ont un effet immédiat sur l'issue de grossesse. Le récepteur nucléaire PPARg est essentiel au cours des premières étapes de la placentation et est fortement exprimée dans les trophoblastes. Il joue un rôle important dans la différenciation des trophoblastes extravilleux et villeux et la mise en place de la vascularisation utéro-placentaie. La létalité embryonnaire observée chez les souris PPARγ-/- peut être contrée par la transfection de PPARγ dans le trophoblaste.

Dans le présent travail, nous avons apporté de nouvelles informations sur le rôle de PPARγ dans le développement placentaire humain d'un point de vue général à l'aide de technologies de séquençage, suivie de l'analyse en aval axée sur l'enrichissement des ensembles de gènes identifiés.

Partie I

Il est admis que PPARγ est indispensable pour la placentation et que sa délétion entraîne la mortalité de l' embryon. Mais sa régulation est-elle un processus statique dans une durée déterminée, comme le premier trimestre, ou un agencement dynamique et ordonné tout au long de la gestation ? Et si nous mettions PPARγ en arrière-plan de toute la période de gestation, son importance peut-elle encore être détectée plutôt que d'être noyée par d'autres facteurs plus importants ? C'est-à-dire quelle est la vue d'ensemble de l'expression à l'échelle du génome placentaire humain depuis le tout début de l'âge gestationnel jusqu'au terme.

Notre objectif est d'étudier les changements dynamiques dans l'expression des gènes tout au long de la placentation. Dans notre étude, les profils d'expression génique des placentas humains de 4 à 40 semaines de gestation ont été collectés. Une régression linéaire et une analyse de réseau de corrélation pondérée ont été appliquées pour le filtre génétique. L'analyse d'enrichissement génétique, y compris l'ontologie des gènes et les termes de la voie de l'Encyclopédie de Kyoto des gènes et des génomes, a été effectuée par clusterProfiler. Un graphique linéaire dessiné avec une expression génique mise à l'échelle et ajustée a été appliquée pour afficher les changements dynamiques. Nos résultats ont montré un total de 5173 gènes impliqués dans différentes périodes de la placentation. L'annotation en aval de ces gènes a révélé les processus et voies biologiques impliqués, parmi lesquels nous avons sélectionné la «voie de signalisation PPAR». Cette carte des voies montre les gènes impliqués dans le stockage et le métabolisme des lipides, y compris les membres de la famille FABP et LPL. De plus, l'expérience de coloration des lipides neutres sur des coupes placentaires a montré une diminution significative de la teneur en gouttelettes lipidiques dans les

placentas du premier trimestre par rapport aux placentas à terme. Notre étude fournit plus d'informations sur les processus biologiques et les voies à travers la placentation humaine. Ces résultats nous donnent de nouveaux indices pour déchiffrer les fonctions normales de la placentation et leurs dérèglements peuvent être liés à des maladies liées à la grossesse. À titre d'exemple, nos résultats montrent que la voie de signalisation PPAR médie une diminution constante des lipides placentaires tout au long de la grossesse.

Partie II

Le placenta est composé de divers types cellulaires, tels que des trophoblastes, des cellules mésenchymateuses, cellules de Hofbauer cellules endothéliales etc.

Les trophoblastes, en tant que cellules qui constituent la partie principale du placenta, subissent des processus de différenciation cellulaire tels que l'invasion, la migration et la fusion. Des anomalies dans ces processus peuvent conduire à une série de maladies gestationnelles dont les mécanismes sous-jacents ne sont pas encore clairs. Une protéine qui s'est avérée essentielle à la placentation est le récepteur activé par les proliférateurs de peroxysomes γ (PPARγ), qui est fortement exprimée dans les noyaux des cytotrophoblastes extravilleux (EVCT) au cours du premier trimestre et des cytotrophoblastes villeux (VCT) tout au long de la grossesse. Ici, nous avons cherché à explorer les effets à l'échelle du génome de PPARγ sur les EVCT et les VCT via un traitement avec la rosiglitazone, un agoniste de PPARγ. Les EVCT et les VCT ont été purifiés à partir de villosités choriales humaines, cultivées in vitro et traitées à la rosiglitazone. Les transcriptomes des deux types de cellules ont ensuite été quantifiés à l'aide d'un profilage par microarray. Les gènes différentiellement exprimés (DEG) ont été filtrés et soumis à l'annotation de l'ontologie génique (GO) et à l'analyse des voies avec ClueGO. L'outil en ligne STRING a été utilisé pour prédire les interactions entre les protéines PPARγ et DEG, tandis que iRegulon a été utilisé pour prédire les sites de liaison des promoteurs PPARγ et DEG. Les termes GO et chemin ont été comparés entre les EVCT et les VCT avec ClusterProfiler. Les visualisations ont été préparées dans Cytoscape. À partir de nos données sur les puces à ADN, 139 DEG ont été détectés dans les EVCT traités à la rosiglitazone (RT-EVCT) et 197 DEG dans les VCT traités à la rosiglitazone (RT-VCT). L'analyse d'annotation en aval a révélé les similitudes et les différences entre les RT-EVCT et les RT-VCT en ce qui concerne les processus biologiques, les fonctions moléculaires, les composants cellulaires et les voies KEGG affectées par le traitement, ainsi que des sites de liaison prédits pour les interactions protéine-protéine et les interactions facteur de transcription-gène cible. Ces résultats fournissent une large perspective des processus activés par PPARγ dans les trophoblastes ; une analyse plus approfondie des signatures transcriptomiques des RT-EVCT et des RT-VCT devrait ouvrir de nouvelles voies pour la recherche future et contribuer à la découverte de gènes ou de voies possibles ciblés par les médicaments dans le placenta humain.

Partie III

Notre intérêt s'est concentré sur le rôle de PPARγ dans deux types de trophoblastes au cours du premier trimestre, mais n'oubliez pas que le premier trimestre comprend une période entre 8 et 10 semaines de grossesse où les niveaux d'oxygène dans l'environnement placentaire augmentent considérablement. Il peut être intéressant d'étudier s'il existe un lien entre PPARγ ou ses cibles et les gènes ou voies liées à l'oxygène.

La tension physiologique en oxygène augmente considérablement dans l'environnement placentaire entre 8 et 14 semaines de gestation, entraînant des altérations significatives des processus biologiques dépendants de l'oxygène. Les anomalies de cette période peuvent conduire à une série de maladies gestationnelles, dont les mécanismes sous-jacents restent flous. Pendant que cette augmentation spectaculaire de l'oxygène se produit, les cytotrophoblastes villeux, les cellules qui constituent la partie principale du placenta, subissent une série de transformations fondamentales. Nous avons exploré ces changements au niveau de l'ARNm en comparant les transcriptomes des placentas humains à 8-10 semaines et 12-14 semaines de grossesse. Au total, 20 échantillons ont été collectés et divisés également en quatre groupes en fonction du sexe et du terme. Les cytotrophoblastes ont été isolés et séquencés à l'aide de RNAseq. Dans l'ensemble de données traité, les gènes clés ont été identifiés à l'aide de deux méthodes différentes : DESeq2 et l'analyse de réseau de co-expression génétique pondérée (WGCNA), basée sur les packages R DESeq2 et WGCNA, respectivement. Nous avons également construit une base de données locale de cibles connues des sous-unités alpha et beta du facteur inductible par l'hypoxie (HIF) afin d'étudier spécifiquement les modèles d'expression susceptibles d'être liés à des changements en oxygène. Les modèles d'enrichissement génique dans et parmi les quatre groupes ont été analysés sur la base des annotations de l'ontologie génique (GO) et des voies KEGG à l'aide de clusterProfiler. À partir de nos résultats DESeq2, nous avons identifié 457 gènes clés qui étaient corrélés à l'âge gestationnel et 15 qui étaient corrélés au sexe, tandis qu'avec WGCNA, nous avons identifié 2015 et 233 gènes clés associés à l'âge et au sexe, respectivement. Le croisement de ces résultats avec la base de données locale des cibles HIF a généré un sous-ensemble de 466 gènes, dont 463 étaient associés à l'âge et 3 étaient associés au sexe. Nous avons caractérisé les similitudes et les différences entre les schémas d'enrichissement révélés par les deux méthodes pour les deux conditions (âge et sexe), ainsi que ceux associés aux cibles HIF. Nos résultats offrent une large perspective des processus actifs dans les cytotrophoblastes lors de l'augmentation physiologique de la teneur en oxygène.

Partie IV

La recherche biologique fondamentale est intéressante, mais elle serait plus attrayante si nous pouvions appliquer les résultats à la pratique clinique, par exemple, la prédiction de maladies. Les polymorphismes (SNP), connus pour leur contribution à

la susceptibilité aux maladies, nous viennent à l'esprit puisque PPARγ est si critique pour la placentation et l'issu de grossesse. Ses SNP pourraient être utiles pour le diagnostic de certaines complications de la grossesse.

Certains polymorphismes de PPARg augmentent la susceptibilité à certaines maladies métaboliques et complications de la grossesse. Par exemple, le dysfonctionnement placentaire associé à la prééclampsie a été lié à une perturbation de PPARγ dans laquelle le polymorphisme génétique pourrait jouer un rôle. Notre objectif était d'étudier les facteurs de risque génétiques de cette maladie et de construire un modèle pragmatique pour la prédiction de la prééclampsie. Les données ont été recueillies auprès d'un total de 1648 femmes de l'étude de cohorte mère-enfant EDEN. De nombreuses caractéristiques cliniques ont été enregistrées, ainsi que des données de génotype pour trois polymorphismes PPARγ : Pro12Ala, C1431T et C681G. Une analyse univariée a été réalisée pour comparer les 35 patientes prééclamptiques aux 1613 femmes témoins. Des caractéristiques d'intérêt fortement corrélées ont été identifiées en utilisant trois méthodes de sélection de caractéristiques et de curation manuelle ; huit algorithmes d'apprentissage automatique différents ont ensuite été appliqués pour créer des modèles prédictifs. Les performances du modèle ont été évaluées sur la base de métriques de précision et de l'aire sous la courbe caractéristique de fonctionnement (AUC) du récepteur. Le polymorphisme C1431T de PPARγ était le seul facteur significativement associé à la prééclampsie ($p < 0,05$) dans les analyses univariées, avec un rapport de cotes allant de 4,90 à 8,75. Le processus de sélection des caractéristiques et de curation manuelle a également suggéré l'inclusion des variantes maternelles C1431T et C681G en tant que facteurs, ainsi que les caractéristiques cliniques associées à la grossesse ou aux délais d'accouchement, l'indice de masse corporelle, l'éducation et la consommation de cigarettes. Parmi les algorithmes d'apprentissage automatique testés, le modèle basé sur l'arbre boost a donné les meilleurs résultats, avec des valeurs de précision et d'AUC dans l'ensemble d'entraînement de 0,971 ± 0,002 et 0,991 ± 0,001, respectivement, et dans l'ensemble de test de 0,951 et 0,701, respectivement. Un organigramme de l'arbre final a été construit pour décrire la procédure de prédiction de la prééclampsie. Nos résultats montrent pour la première fois que la variante C1431T de PPARγ peut jouer un rôle dans la détermination de la susceptibilité à la prééclampsie. L'arbre de décision créé ici, basé sur de multiples facteurs prédictifs, y compris les variantes C1431T et C681G de PPARγ, les délais de grossesse ou d'accouchement, l'indice de masse corporelle, l'éducation et la consommation de cigarettes, pourrait avoir des applications dans la prédiction clinique de la prééclampsie aux stades très précoces de la grossesse.

Pour résumer, nos résultats soutiennent que PPARγ médie la diminution constante des lipides placentaires tout au long de la grossesse via la voie de signalisation PPAR, fournit une large perspective des processus biologiques activés par PPARγ dans les EVCT et les VCT grâce à l'analyse des signatures transcriptomiques, et développe un modèle pragmatique pour la clinique avec les polymorphismes de PPARγ.

Mots-clés : Placenta ; Trophoblaste; PPARγ; Séquençage ; Bioinformatique; Apprentissage automatique ; Prééclampsie

Publications during candidature

1. **Liu F**, Rouault C, Clément K, Zhu W, Degrelle SA, Charles MA, Heude B, Fournier T. C1431T variant of PPARγ is associated with preeclampsia in pregnant women. Life. 2021; 11(10):1052.
2. **Liu F**, Zhou J, Guo J, Zhang W, Wang H. Prenatal ethanol exposure increases maternal bile acids through placental transport pathway. Toxicology. 2021, June 30; 458, 152848.
3. **Liu F**, Simasotchi C, Vibert F, Zhu W, Gil S, Degrelle SA, Fournier T. Age and Sex-Related Changes in Human First-Trimester Placenta Transcriptome and Insights into Adaptative Responses to Increased Oxygen. Int J Mol Sci. 2021 Mar 12;22(6):2901.
4. **Liu F**, Zhu W, Shoaito H, Chissey A, Degrelle SA, Fournier T. Mining of combined human placental gene expression data across pregnancy, applied to PPAR signaling pathway. Placenta. 2020 Sep 15; 99: 157-165.
5. **Liu F**, Rouault C, Guesnon M, Zhu W, Clément K, Degrelle SA, Fournier T. Comparative Study of PPARγ Targets in Human Extravillous and Villous Cytotrophoblasts. PPAR Res. 2020 Apr 1; 2020: 9210748.
6. Wu K, **Liu F**, Wu W, Chen Y, Zhang W. Bioinformatics approach reveals the critical role of TGF-β signaling pathway in pre-eclampsia development. Eur J Obstet Gynecol Reprod Biol. 2019 Sep; 240: 130-138. (**Co-first author**)

Remerciements

La réalisation de ce thèse a été possible grâce au concours de plusieurs personnes à qui je voudrais témoigner toute ma reconnaissance. Je voudrais tout d'abord adresser toute ma gratitude au directeur de ce thèse, mon superviseur, Dr. Thierry FOURNIER, pour sa patience, sa disponibilité et surtout ses judicieux conseils, qui ont contribué à alimenter mon projet. Je désire aussi remercier Dr. Séverine A. DEGRELLE., qui m'a fourni des aides nécessaires à la réussite de mes recherches. Je tiens à remercier spécialement Jasmina ROGOZARSKI, Isabelle HERNANDEZ et Jessica DALMASSO, qui m'ont soutenu beaucoup sur la vie dans la laboratoire. Je voudrais exprimer ma reconnaissance envers les collègues, Audrey CHISSEY, Camille GOSSEAUME, Margaux NEDDER, Ioana FERECATU, Josette BADET, Françoise VIBERT, Amal ZERRAD-SAADI, Giovana VINCI, Hussein SHOAITO, *etc.*, qui m'ont apporté leur support moral et intellectuel tout au long de ma démarche. Un grand merci aussi à tous ceux qui ont contribué à la réalisation de ce travail.

I would like to thank Dr. Petri Auvinen, University of Helsinki, who accepted me to visit their lab, which helped me a lot in understanding bioinformatics and coding. Thanks for Icem Duru's favor in my life during Helsinki. I would also like to thank Dr. Nabila Bouatia-Naji, INSERM U-970, who accepted me to visit their lab to study Genetics and promised to be one of the members of the jury of my thesis defense. Thanks to all that helped me a lot in my visiting period.

终于进入学生时代的终章。毕业的时候应该听一上曲《凤凰花开的路口》，于是在这里插入一段虚拟的播放器。

--I< II >I □ 00:00----------凤凰花开的路口--------04:02

首先感谢我的父母，刘西汉先生和罗秀姑女士，是他们的耐心教导让我常常满怀着好奇心去探求新鲜的知识，是他们不辞辛苦地劳作给我创造了一直求学到现在的条件。他们的勤劳与坚韧，以身作则地影响着我的成长和对待事物的态度。他们对我无条件地支持和理解，让我得以轻装上阵地去追寻自己想要去追寻的，无论何时，也无论何地。感谢我的姐姐和哥哥，刘琴和刘改，虽然我们并不常常联系，也各自忙于奔赴自己的前途，但血浓于水的亲情让我们无时无刻都作着为家人提供任何帮助的准备。他们对父母的守护与照顾也得以让我安心地求学于外。

感谢思瑜，因为她我才有了出国的契机与勇气，可惜相逢和离别都有定时。感谢雯雯，在巴黎还有赫尔辛基的访学期间陪伴我最久，听过我说过最多的话，陪我度过最艰难时光，安慰和鼓舞我最多的好朋友。感谢周瑾，从硕士时期起就和我无所不谈的好朋友，时时刻刻站在我的角度考虑问题，一直以来在科研上以及生活上给予了我很多有益的建议。感谢文灿，在数学统计以及 R 语言的编程上给予的帮助和指导。感谢 Isabelle，在我初到法国时候，陪我练习英语。感谢 Jasmina，成为我早期学习法语几乎唯一的朋友和动力。感谢 Barabora，布拉格的小妹妹，上法语课时总是耐心听我说法语和纠正我的错误。感谢 Alina，Tatarstan 的姐姐，陪我一起练拳以及英语。说到练拳，感谢陈杰，带我接触到了梅花拳，终于找到了能长期坚持锻炼并且有趣的运动。感谢在我求学期间出现过的朋友，广州法语班的朋友们，和他们维系了最长久和稳定的情谊，还有 CROUS 的朋友们，在生活的琐事上给予的帮助，还有那些让我变得更好的朋友们，感谢你们的出现。

回顾三年多的生活，从刚来时的新奇，到独自一人的孤单，再到学业繁忙的烦扰困苦与目不暇接，以及最后平静生活的令人生厌。如果说异国他乡的生活能带给人什么的话，那一定是波澜起伏的情绪变化。有的人在其中成长，有的人在其中迷失。我自己呢？我不知道。然而，时光如果能够重来的话，我想我是不会选择来到巴黎的。现在的巴黎已经不再是当年海鸣威口中的巴黎了，巴黎的流动盛宴在长达百年的时光里终于也垂垂老矣，而上世纪还有上上世纪的香屁在 21 世纪的今天已经烟消云散了。但这恐怕才是历史的必然，万物繁荣有时，消弭有时，终究不过是笼罩在地球这个大生态下的内部流转。日光之下的事，复归转行原道。而人活着的意义，终究只能从人心内部去寻找，而不是那些转瞬即逝的外物。

王安石说，世之奇伟、瑰怪、非常之观，常在于险远，而人之所罕至焉，故非有志者不能至也。我更愿意把它理解为一种内心的精神追求。如果一个人要进步，要获得持久的激情与动力，就应该主动去拥抱痛苦，痛苦是进步的尺度，痛苦的结局也是快乐的源泉。而平静是留给一滩死水的。

以此自勉。

10/09/2021 À Paris

刘福林

Fulin LIU

Liste des principales abréviations

AF	activation function
AMPK	AMP-activated protein kinase
ARNT	aryl hydrocarbon receptor nuclear translocator
ATP	Adenosine triphosphate
AUC	area under receiver operating characteristic curve
BMI	body mass index
BP	basal plate
BT	boost tree
CDS	coding sequences
CP	chorionic plate
DBD	DNA binding domain
DEG	differentially expressed gene
DHA	docosahexaenoic acid
DNA	deoxyribonucleic acid
DT	decision tree
ENR	elastic net regression
ER	endoplasmic reticulum
EVCT	extravillous cytotrophoblast
FATP	fatty acid transport protein
FCS	fetal calf serum
FDR	false discovery rate
FFA	free fatty acid
FGF	fibroblast growth factor
FN	false negative
FP	false positive
GEO	genome expression omnibus
GO	gene ontology
GS	gene significance
GW	gestational week
hCG	human chorionic gonadotropin
HDL	high-density lipoprotein
HIF	hypoxia-inducible factor
HIPEC	human extravillous cytotrophoblast cell line
IDE	integrated development environment
IFN	interferon
IGF	insulin-like growth factor
IVS	intervillous space
KEGG	kyoto encyclopedia of genes and genomes
KNN	k-nearest neighbor
LBD	ligand-binding domain
LDL	low-density lipoprotein

MAPK	mitogen-activated protein kinase
MLP	multilayer perceptron
MM	module membership
MMP	matrix metalloproteinase
NB	naïve bayes
NGS	next-generation sequencing
ORO	Oil red O
p.c.	post-coitus
PAI	plasminogen activator inhibitor
PAPP-A	pregnancy-associated plasma protein-A
PCA	principal component analysis
PCR	polymerase chain reaction
PGE_2	prostaglandin E2
PIGF	placental growth factor
PO_2	arterial oxygen tension
PPAR	peroxisome proliferator-activated receptor
PPI	protein-protein interaction
PPRE	peroxisome proliferator response element
PUFA	polyunsaturated fatty acid
RF	random forest
RNA	ribonucleic acid
ROC	receiver operating characteristic
ROS	reactive oxygen species
RT-EVCTs	rosiglitazone-treated EVCTs
RT-VCTs	rosiglitazone-treated VCTs
RXR	nuclear receptor retinoid X receptor
SEG	significantly expressed genes
sEng	soluble endoglin
sFIt-1	soluble fms-like tyrosine kinase-1
SNP	single nucleotide polymorphism
SVM	support vector machine
TN	true negative
TOM	topological overlap matrix
TP	rrue positive
t-SNE	t-distributed stochastic neighbor embedding
TZD	thiazolidinedione
uPA	urokinase plasminogen activator
UPR	unfolded protein response
UTR	untranslated region
VCT	villous cytotrophoblast
VEGF	vascular endothelial growth factor
WGCNA	weighted correlation network analysis

CONTENTS

Liste des figures

Liste des tableaux

1 Introduction

1.1 Human placenta

1.1.1 General description of the human placenta

All viviparous vertebrates develop a placenta, a system surrounding the fetus to facilitate some functions. The placenta acts as a substitute of the still immature embryonic and fetal organs in some functions partially or completely, which includes the aspects:

1. Gas transfer (complete substitute of the lung)
2. Excretory functions, water balance, pH regulation (complete substitute of the kidney)
3. Catabolic and absorptive functions (complete substitute of the gut)
4. Synthetic and secretory hormones of most endocrine glands (partially substitute of endocrine glands)
5. Numerous metabolic and secretory substances of the liver (partially substitute of the liver)
6. Hematopoiesis of the bone marrow (complete substitute during early stages of pregnancy)
7. Heat transfer of the skin (partially substitute of the skin)
8. Immunological interactions and protection (partially substitute of the immunity)

The placental shapes vary in some species according to the comparative placentology from structural characteristics, including diffuse placenta (placenta membranacea), cotyledonary placenta, zonary placenta, discoid placenta (Fig. 1). While from the functional characteristics, the placental types have been delicately classified as epitheliochorial, endotheliochorial, and hemochorial (Fig. 2).

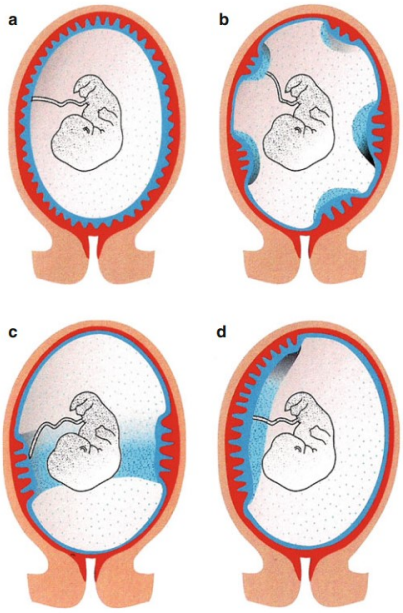


Figure 1 Placental types based on structural classification

Fig. 1 Placental types based on structural classification. According to the types of interdigitations between the mother (red) and the fetus (blue), the placental shapes are classified as: (a) Diffuse placenta, (b) cotyledonary placenta, (c) zonary placenta, and (d) discoid placenta [1].

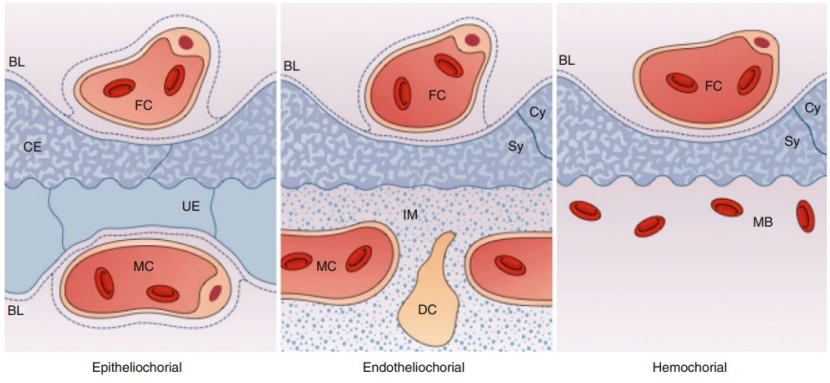


Figure 2 Placental types based on functional classification

Fig. 2 Placental types based on functional classification. According to the tissue layers of the maternal-fetal interface, the placental shapes are classified as: epitheliochorial placenta, characterized by no invasion of the maternal tissues where chorionic epithelium (CE) is located in the uterine epithelium (UE); endotheliochorial placentas:

characterized by loss of the uterine epithelium and direct contact between the syncytiotrophoblast (Sy) and the interstitial matrix (IM) where situates the maternal capillaries (MC) and decidual cells (DC); and hemochorial placentas, characterized by eroded maternal endothelial cells and the substitute maternal blood (MB). BL: basal lamina; Cy: cytotrophoblast; FC: fetal capillaries. [1]

According to the features of these classification systems, the mature human placenta is characterized by the hemochorial and discoidal type. But this typical performance doesn't last for the whole duration of placentation due to the dramatic changes in the interface along with the gestational age. For example, the intervillous space is full of a clear fluid that is only secreted by the endometrial glands in the uterus, while the intra-placental circulation is not yet established unless the end of the first trimester after which the placenta can be considered as hemochorial [1]. The variation in this duration might give rise to many of the complications of pregnancy [2] such as the oxygen concentration rises three-fold at the end of the first trimester.

As common sense, the generally realized human placenta is the term form, a local and disk-like thickening of the membranous sac that is surrounded by two split membranes, the chorionic plate and the basal plate (Fig. 3). Inside the membranes is the intervillous space where is filled with maternal blood with fetal villi floating. The villi are complex tree-like projections sprouting from the chorionic plate, with the trophoblast covered on the surface and fetal vessels located internally whereby the connection between the fetal circulatory system and maternal blood is built. At the placental margin, the basal plate and the chorionic plate fuse with each other to form the chorion laeve leaving no space for villi.

Figure 3 The normal and mature human placenta

Fig. 3 The normal and mature human placenta is surrounded by the chorionic plate (CP, fetal side) and the basal plate (BP, maternal side) with the intervillous space (IVS) inside. The villous trees sprouting from the CP are submerged by the maternal blood full of the intervillous space where exchanges happen between the maternal spiral vessels and fetal villous trees. Abbreviations from left to right: M: myometrium; P: placental bed; BP: basal plate; S: septum; IVS: intervillous space; *: cell island; MZ: marginal zone; CL: chorion laeve; CP: chorionic plate; A: amnion; UC: umbilical cord [3].

1.1.2 Trophoblast

The Origin of Trophoblast

After fertilization, the zygote continues to divide mitotically, which is known as cleavage. It afterward goes with the 8-cell stage on day 3 post-coitus (p.c.), and the 16-cell stage which is known as the morula. After undergoing a transformation of compaction, the morula is compassed by a smoother spherical outline while intercellular clefts are surrounded by tight junctions across the apical regions. The morula cells are polarized with the favor of tight junctions that help cytoskeletal and other proteins to redistribute from the apical to basal domains or *vice versa*. As a consequence, ionic gradients can go into the center of the mass and expand it to form a blastocyst by day 4.5 p.c. Inside the blastocyst, asymmetrically division goes on in cells that two distinct populations with different cell components are formed with the outer trophoblast cells giving rise to the wall of the blastocyst, and the inner cell mass to the embryo [4]. The embryonic stage of development thus begins as soon as the blastocyst hatches from the zona pellucida and implants in the endometrium. In this process, the trophoblast is the first cell lineage to differentiate, followed by cytotrophoblast, syncytiotrophoblast, extravillous, and villous trophoblasts, *etc.* (Fig. 4).

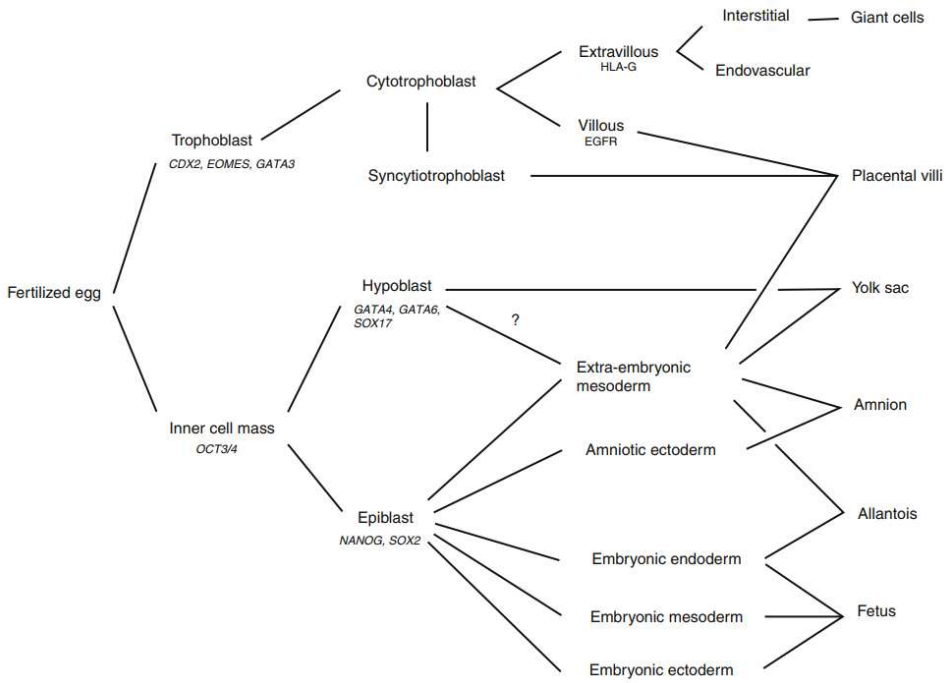


Figure 4 A schema of differentiation of fertilized egg

Fig. 4 A schema of differentiation of fertilized egg [4].

Prelacunar Stage

This first stage is defined after the implantation of the blastocyst when the blastocyst establishes close and stable contacts with the endometrium. The first step of implantation is called apposition, taking place around day 6–7 p.c., when it comes out of an outer layer of trophoblast from the wall of the blastocyst and of an inner layer of extra-embryonic mesoderm from the inner cell mass [5]. The inner cell mass gradually evolves into the embryo, umbilical cord, yolk sac, amnion, and part of the placenta. Following apposition, invasion happens in trophoblasts toward the endometrium with established cell-cell contacts between trophoblasts and uterine epithelial cells, in which the trophoblast cells migrate toward the endometrium and displace the endometrial cells [6].

During this process, part of the trophoblast cells undergoes further differentiation from mononucleated cells into multinucleated masses of syncytiotrophoblast, while the rest

remaining mononucleated are called cytotrophoblast (Fig. 5a). In the following days, the blastocyst is gradually encompassed by the endometrium with the rebuild occurring nearby the site of implantation. In between the external blastocyst and the endometrium, it is filled with the rapidly increased mass of syncytiotrophoblast, which functions as a complete mantle over the blastocyst. The mantle will achieve such a substantial thickness of the syncytiotrophoblast that finger-like extensions are subsequently developed from the implantation pole and deeply invade into the endometrium. Notably, the syncytiotrophoblast has lost its generative potency after the differentiation, which is replenished continuously by the stem cell-like cytotrophoblast underneath owing to its proliferation and fusion. This duration, lasting from day 7 to day 8 p.c., when the syncytiotrophoblast is a rather solid mass, is defined as the prelacunar stage [7].

Lacunar Stage

Around day 8 p.c., there appear small vacuoles within the mass of syncytiotrophoblast at the implantation pole, which form a system of lacunae (Figs. 5b&5c). The lacuna formation subsequently spreads over the whole surface of the chorionic sac. In the interior of the syncytiotrophoblastic mass, there exist the separating lamellae and pillars of syncytiotrophoblast, called the trabeculae. Their appearance defines the onset of the lacunar or trabecular stage of placentation, lasting from day 8 to day 13 p.c. By day 12 p.c., when the blastocyst is deeply implanted into the uterine epithelium, the cytotrophoblast cells underneath the syncytiotrophoblastic mantle enlarge and spread through the trabeculae as cellular columns [8]. Approximately 2 days later, they penetrate the tips of the trabeculae and amalgamate with each other to form a new layer called the cytotrophoblastic shell, which is interposed between the mantle and the endometrium. Therefore, the exterior of the blastocyst contains three layers in the

lacunae stage (Fig. 5c&5d): (1) the primary chorionic plate; (2) the lacunar system along with the trabeculae; and (3) the cytotrophoblastic shell, *i.e.*, the precursor of the basal plate.

Especially, in the lacunae stage, maternal erythrocytes can be observed in the lacunae despite few numbers. The dilation of venules containing maternal erythrocytes within superficial endometrium near the conceptus form sinusoids, which simplify the subsequent erosion of the syncytiotrophoblast into the sinusoids [8], resulting in the observation of erythrocytes. However, the connection to the maternal artery is not yet established unless the end of the first trimester [9, 10]. In this duration, the glandular secretion also contributes to the flourish of the blastocyst thanks to the erosion of the syncytiotrophoblast into the nearby necks of the endometrial glands [11].

Around day 14 p.c., the new layer of the cytotrophoblastic shell enables the interaction between cytotrophoblast cells and the endometrial tissues. On one hand, the shell contains large amounts of glycogen; on the other hand, the endometrium accumulates a population of differentiated pleomorphic cells, which are considered as extravillous trophoblast. Taken together, these preconditions make possible the adaptation of the maternal vessels to pregnancy conditions and the anchorage of the developing placenta during implantation and placentation [12].

Early Villous Stages

On about day 13 p.c., along with the increased cytotrophoblastic proliferation, side branches protrude from the trabeculae and into the lacunae (Fig. 5d&5e). These side branches are primary villi, composed of internal cytotrophoblast cells and its external derivatives, syncytiotrophoblast cells. The presence of the primary villi is taken as the onset of the villous stages. Subsequently, primary villi develop into primitive villous trees alongside its active proliferation, while the former trabeculae act as the stems

(Fig. 5e). The anchors of the primitive villous trees would be those villi keeping in touch with the trophoblastic shell, which is called anchoring villi. These villi are surrounded by a transformed lacunar system, which is known as the intervillous space.

Two days later, the primary villi begin to transform into secondary villi because of the invasion from mesenchymal cells stemming from the extra-embryonic mesenchyme layer of the chorionic plate. The invasion spreads soon towards the villous tips in several days but spares the anchoring villi as the trophoblastic shell is not reachable [9]. Instead, these intact villi remain as cytotrophoblast cell columns until the late stages of pregnancy, and function as the source of extravillous trophoblast. However, not all the villi will end in that case. Part of the intact villi probably peel off the cytotrophoblast shell and become free-floating villi at the tips, which are called trophoblastic cell islands.

Between days 18 and 20 p.c., the fetal capillaries appear for the first time in the mesenchyme of the villi as well as the primary chorionic plate. The endothelium of the fetal capillaries derives from hemangioblastic progenitor cells in the mesenchyme that are also the precursors of hematopoietic stem cells [13-15]. The appearance of capillaries in the villous stroma and its later derivatives represents the first tertiary villi, while the primary or secondary villi correspond to cell columns, cell islands, or newly formed villus from trophoblastic and villous sprouts. However, the placental and fetal circulations are established independently except for limited communications at the start of the fourth-week p.c.. Even though the beginning of the sixth-week p.c., the fetoplacental circulation still has a high resistance in between owing to the nucleated fetal erythrocytes [16]. This causes a relatively low oxygen concentration (<20 mmHg) with the fetoplacental during the first trimester [17, 18]. The low concentration of oxygen may benefit the development of the placenta by maintaining the pluripotent

state of cells and the embryo by minimizing the risk of oxygen-free radical-mediated teratogenesis [19, 20].

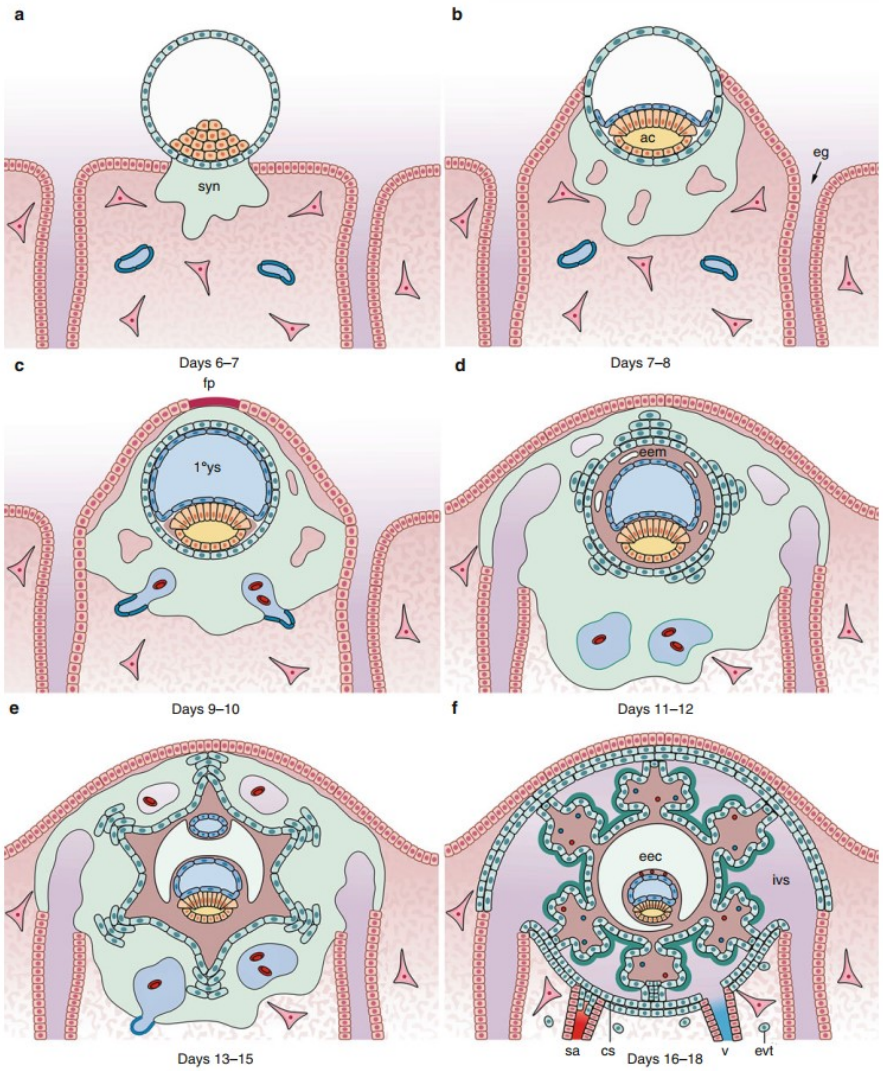


Figure 5 Schematic drawing of the typical stages of early placental development

Fig. 5 Schematic drawing of the typical stages of early placental development. (a, b) Prelacunar stages. (c) Lacunar stage. (d) Transition from lacunar to the primary villous stage. (e) Secondary villous stage. (f) Tertiary villous stage. ac amniotic cavity, fp: fibrin plug, eec: extraembryonic coelom, eem: extraembryonic mesoderm, eg: endometrial gland, evt: extravillous trophoblast, cs: cytotrophoblastic shell, ivs: intervillous space, sa: spiral artery, syn: syncytiotrophoblast, v: uterine vein [9].

Villus Regression and Formation of the Definitive Placenta

Villi initially and continuously develop over the entire surface of the chorionic sac until the eighth-week p.c. [9]. Then the superficial pole of the villous mass connected to decidua capsularis initiates the formation of the smooth chorion or chorion leave, which is called villus regression, while the remainder at the deep pole the formation of the definitive placenta. Villus regression is resulted from the pump of the maternal arterial circulation into the placenta, especially in the center of the implantation site [21, 22], allowing for the flow in the early beginning (Fig. 6). The flow thus brings the intervillous space a threefold rise of the oxygen concentration [23], which will lead to oxidative stress in the peripheral villi [24]. Meanwhile, the formation of the definitive placenta results from the following steps, separately or collaboratively [25]: firstly, the cytotrophoblast and its syncytial fusion cause the syncytial sprouts which are taken as the primary villi containing only the trophoblast; secondly, these sprouts, owing to the invasion of villous mesenchyme, evolve into villous sprouts which are taken as the secondary villi; thirdly, the components of the definitive placenta, mesenchymal villi, are formed because of the formation of fetal vessels with the stroma. The steps repeat and thus expand the sprouts of the villi throughout the placentation. Notably, due to the appearance of capillaries within the villous stroma, the step of the tertiary villi formation is a complex process associated with the differentiation of various villous types. However, its complicated differentiation results in an enormous rise in villous surface area and wide spread of the villous membrane dividing the fetal and placental circulation [26, 27], which effectively advances placental maturation.

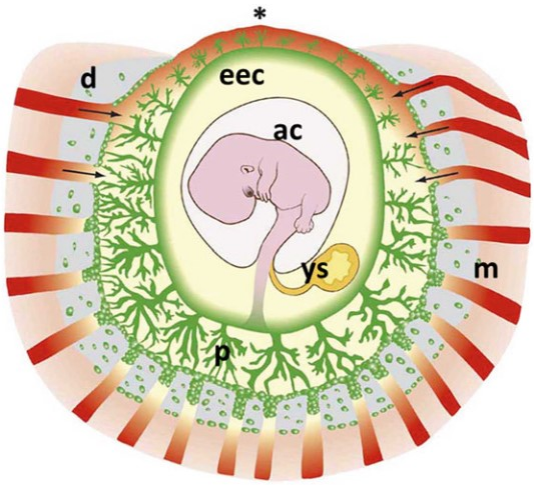


Figure 6 Schematic drawing of a gestational sac at the end of the second month

Fig. 6 Schematic drawing of a gestational sac at the end of the second month (8–9 weeks) showing the yolk sac (ys), amniotic cavity (ac), extraembryonic coelom (eec), placenta (p), decidua (d), and myometrium (m). Maternal blood flow (arrow) pumps into the periphery of the developing placenta, where the invasion of trophoblast and plugging of spiral arteries is least extensive. Due to the blood flow, bringing high levels of oxidative stress, the villi regress over the superficial pole of the sac (asterisk) and form the chorion laeve [28].

Basic Structure of the Villous Trees

During placental development, villi types vary in functions but their basic structures are similar (Fig. 7). Generally, the villi are covered by the syncytiotrophoblast, beneath which is the cytotrophoblast, while the stromal core of the villi is separated by the trophoblastic basement membrane. The stroma contains connective tissue cells, connective tissue fibers, villous macrophages (Hofbauer cells), and ground substance, as well as the fetal vessels.

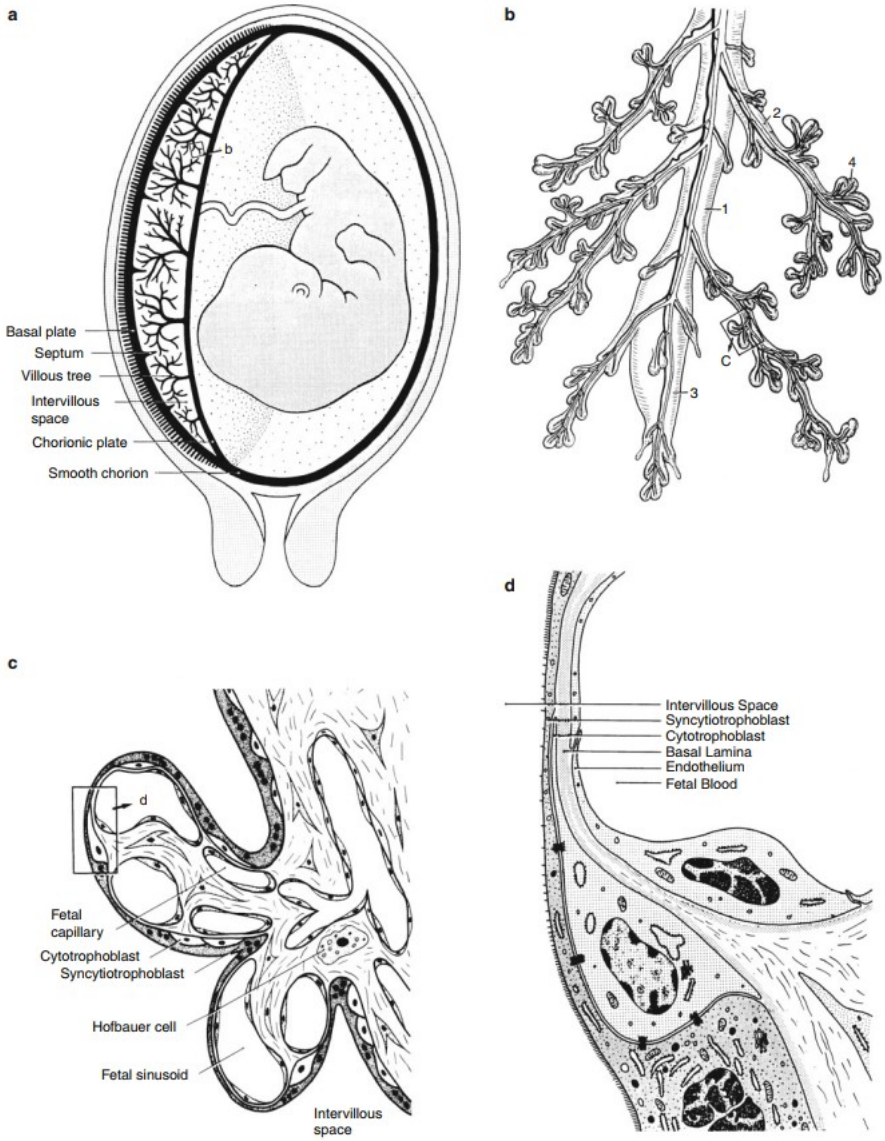


Figure 7 Basic structure of human placental villi

Fig. 7 Basic structure of human placental villi. (a) Simplified sagittal plane of the uterus, placenta, and membranes in the human [29]. (b) Typical mature villous tree, consisting of a stem villus (1) and its extending immature intermediate villus (3), the side branches (2), and terminal villi (4). (c) Highly simplified light microscopic structure of terminal villi. (d) Schematic electron microscopic structure of the placental barrier (Source: Modified after Kaufmann (1983)).

Syncytiotrophoblast

Microvillous Surface

The syncytiotrophoblast is a continuous and highly polarized epithelial layer that extends over the surfaces of villi, the apical of which carries an abundant covering of microvilli. The existence of microvilli profoundly amplifies the surface area to function by a factor of five to sevenfold [30, 31]. The microvilli are supported by cytoskeletons including actin filaments, α-actinin, and ezrin [32, 33], whose abnormality is related to placental diseases, such as preeclampsia [34].

Functions

The syncytiotrophoblast can secret the glycocalyx to cover the surface of the microvilli, which may prevent the intervillous space from thrombosis and function partially as immune barriers. The microvilli act as a semipermeable membrane, which can affect the placental transfer. As a transporting epithelium, a high density of transporter proteins exists in the apical membrane. For example, a non-ATP-dependent process, involving the GLUT family of proteins, can help to transfer glucose through the placenta, while an active process, involving amino acid transporters, mediates the amino acid exchange between the intra- and extra-cellular compartments. Herein also contains ionic and solute transporters. Except for transporters, several enzymes, phospholipids, and junctional complexes also can be found in the apical surface [35-37]. As to the basal surface, which contacts directly with the cytotrophoblast or the trophoblastic basement membrane, contains only a lower number of transporter proteins than that of the apical surface [38].

Vasculosyncytial membranes

Vasculosyncytial membranes exist on the surface of terminal villi, as is indicated by the name, representing the connection of fetal capillaries and the syncytiotrophoblast. It becomes ubiquitous after 32 weeks of gestation, the formation of whom is associated with a progressive reduction in the thickness of the villous membrane [26]. The

membranes are often sinusoidally dilated in the capillary, which results in its bulge against the villous surface, followed by the stretch of the syncytiotrophoblast and displacement of nuclei and organelles laterally. Therefore, the combined two ways, the reduction of thickness and the lack of organelles, aids gaseous exchange without many oxygen consuming.

Villous Cytotrophoblast

Villous cytotrophoblast, beneath the syncytiotrophoblast, can be identified by their position and also the appearance of their nucleus, which is larger generally rounder, and larger than that of the syncytiotrophoblast. Despite the general appearance, the different stages of differentiation of villous cytotrophoblast cells can be also distinguished from a range of morphological features. The shape of cells can be cuboidal, closely juxtaposed to form a complete layer in the first trimester, while flattened and separated in the later pregnancy [8, 39].

Villous cytotrophoblast cells constitute the germinative trophoblastic layer [40], which can proliferate and fuse to form the syncytiotrophoblast, the fusion of the two type of cells involves a rehearsal of events, including disintegration of cell membranes, and merge of cytoplasm, organelles, and nucleus into the syncytium. The nuclei number between the cytotrophoblast and the syncytiotrophoblast keeps a constant according to the stereological estimates, which suggests the fixed fusion rate of these two cell types across pregnancy [8]. The progress has been diagrammed below to indicate the critical cellular events during the fusion process (Fig. 8).

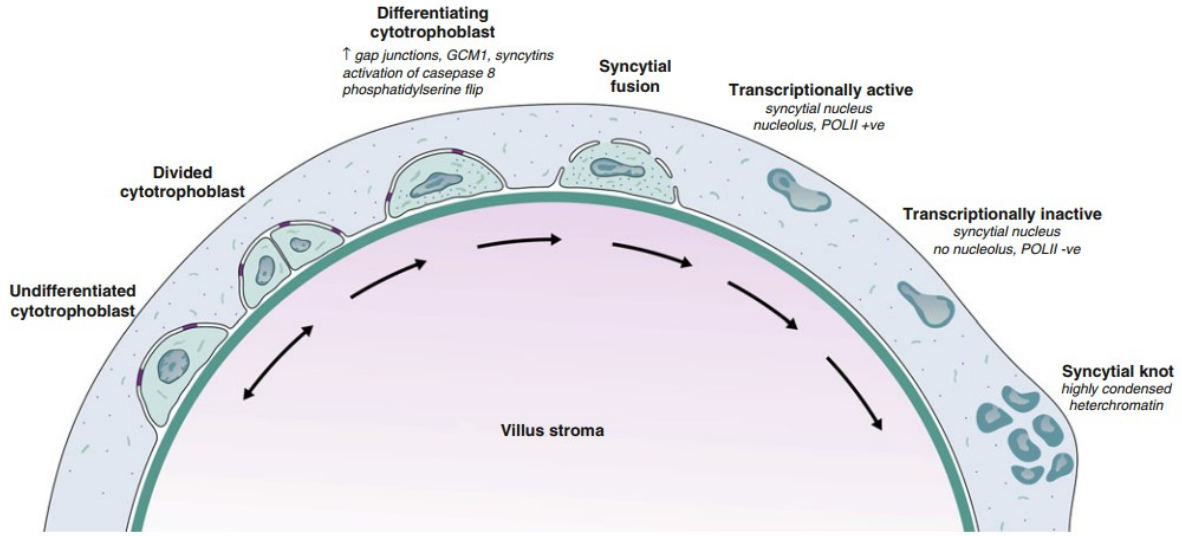


Figure 8 Schematic graph representing the differentiation of cytotrophoblast cells and the fusion process

Fig. 8 Schematic graph representing the differentiation of cytotrophoblasts and the fusion process, and the formation of a syncytial knot [40].

Villous Macrophages

Villous macrophages are a component of the stromal core, characterized by the morphology of being large and round. Their cytoplasm is highly vacuolated with many vacuoles during the first trimester. Villous macrophages are first observed on day 18 p.c. in placenta villi, which are thought to be the offspring of hemangioblastic cells from the mesenchymal cells [14]. Its population can be augmented because of the macrophage proliferation during early pregnancy but rapidly increase with newly bone marrow-derived monocytes around 8 weeks of pregnancy when the establish of fetal circulation [41]. This owes to the close location near the trophoblastic basement membrane and the fetal capillaries, which indicates its association with regulating trophoblast differentiation and angiogenesis. The cells are easily identified with the loose morphology in early pregnancy but less evident with the over-condensed stromal core as well as the decreasing numbers in the later of pregnancy [42, 43]. As the free

connective tissue cells, villous macrophages account for the most, but there still exists a few other cells without macrophage character, including mast cells and plasma cells. As to their functions, immunological defense of the fetus and the prevention of vertical transmission of pathogens are the most important, as expected.

1.1.3 Vasculogenesis and Angiogenesis

Vasculature

The umbilical cord is inserted into the chorionic plate with the umbilical arteries connected by Hytrl's anastomosis. The branches of the umbilical arteries, with the concomitant veins, equally and widely spread over the chorionic plate, aiming to balance flow to the different placental territories. The penetration of the arteries in the chorionic plate terminates in the stromal core of stem villi. Along with the chorionic arteries, the fetal vascular tree is formed. The key component of arteries, smooth muscles, increases in accordance with its thickness as gestation advances [44], while the existence of outer muscle fibers, the myofibroblasts, and the elastic lamina is not clear. The endothelial cells of these vessels, displaying large numbers of caveolae, can be observed and strongly immunoreactive for nitric oxide synthase [45-47].

The lateral intercellular spaces between adjacent endothelial cells are connected by intercellular junctions that consist of tight and adherent junctions [8, 48], where the leaf-like cell membranes can fold into the lumen of the capillary. These junctional complexes thus serve as intercellular mechanical links and reconcile the permeability for paracellular transport. However, the capillaries are not initially surrounded by a basement membrane. Until the first trimester, a bundle of intermediate filaments begins to show up in the form of fibronectin close to the endothelial cells, while it shifts to type IV collagen, laminin, and fibronectin by the third trimester [37]. Additionally,

apart from the membrane, the endothelial cells are also surrounded from the very early stage by the pericytes, which share the common hemangioblastic progenitor cells as the endothelial cells.

Notably, the endothelial cells in different origins show different properties. For instance, in the stem villous arteries and the villous capillaries, the latter owns lower immunoreactivity for nitric oxide synthase than the former [47]. Yet they can also share the same properties, like the susceptibility to oxidative stress. The endothelium of both exhibits high levels of nitro-tyrosine residues ensuing the period of hypoxia-reoxygenation *in vitro*, which is highly related to the cases of intrauterine growth restriction and preeclampsia [49]. Meanwhile, following ischemia-reperfusion during vaginal delivery, high levels of stress can also lead to fluctuating levels of placental oxygenation secondary to mal-perfusion, thus affecting the development of the villous vasculature [50].

General Description of Placental Vasculogenesis and Angiogenesis

Vasculogenesis and angiogenesis are the common steps happening in organs for vessel formation, to the placental vessels (Fig. 9). Vasculogenesis occurs from the *de novo* formation of blood vessels originating from mesodermal precursor cells. It begins within the mesenchymal layer of the secondary yolk sac in the conceptus. Later, vasculogenesis is restricted in mesenchymal villi, during their formation from immature intermediate villi. Angiogenesis involves the expansion of preexisting vessel beds and the creation of new vessel branches, which thus form the vessels in immature intermediate villi, stem villi, mature intermediate villi, and terminal villi. Placental angiogenesis can be further classified into two different types, branching angiogenesis and non-branching angiogenesis, regarding the mechanisms and geometry. Branching angiogenesis is a complex, multiply branched capillary formation, either through the

process of capillary sprouting or through intersusception. By contrast, nonbranching angiogenesis relates to elongating the existing capillary loops.

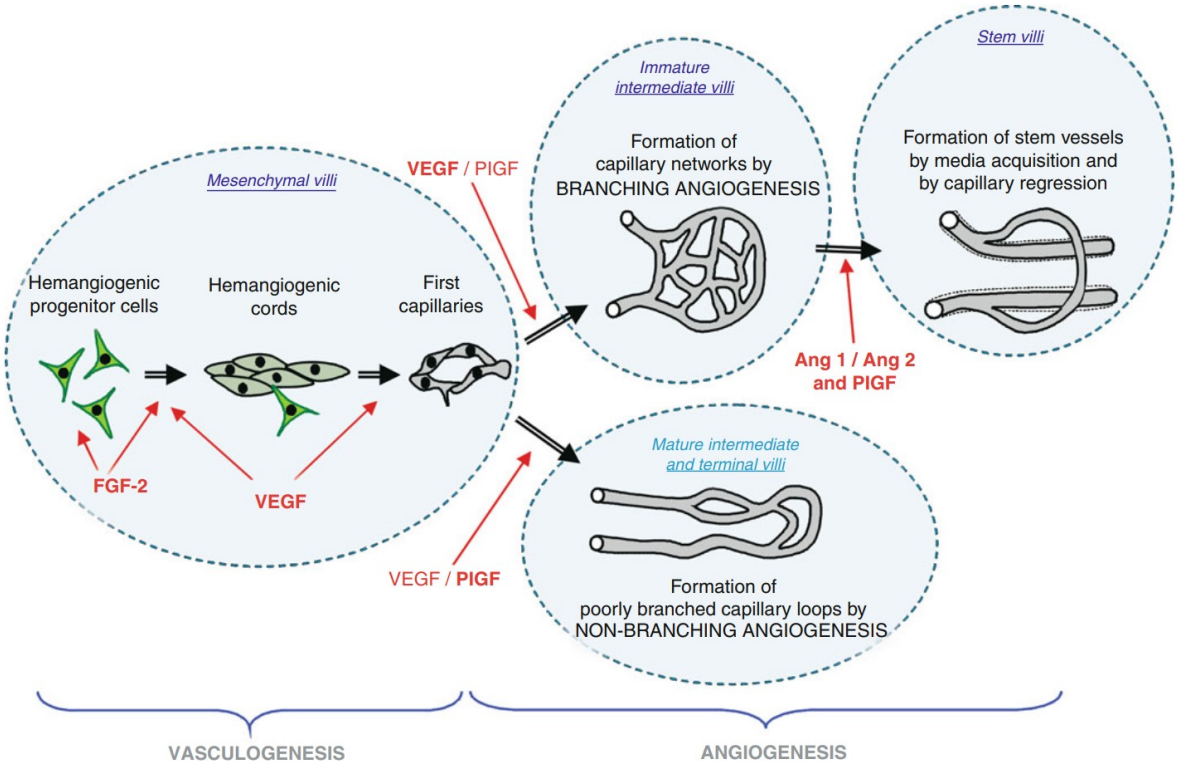


Figure 9 Schematic graph representing of general mechanisms of vasculogenesis and angiogenesis

Fig. 9 Schematic graph representing of general mechanisms of vasculogenesis and angiogenesis, with the regulating molecules. Ang: angiopoietins, FGF: fibroblast growth factor, PIGF: placental growth factor, VEGF: vascular endothelial growth factor [52].

Vasculogenesis (Day 15–32)

Vasculogenesis can be observed firstly through morphological evidence of the cores of mesenchymal villi at 18-20 days p.c. [14, 15]. Hemangiogenic progenitor cells, associated with the trophoblastic basement membrane, differentiate in the villi from the fetal-derived mesenchymal cells [51]. These cells shape string-like polygonal cells tied by desmosomes and primitive tight junctions. The lumen formation is firstly

observed around day 23 p.c. while by day 28 p.c., clearly defined, long, polygonal capillary lumens can be shown in most villi with endothelial cells nearby becoming flattened. As soon as capillary lumens have been built, the first hematopoietic stem cells begin to delaminate from the primitive vessel walls and differentiate into other types. Even though the formation of these cells *in situ*, neither the circulation is possible under the condition of the isolated segments of most of the endothelial tubes, nor yet the anatomical connection through the cord into the embryonic circulation. As to allantoic vessels, they are structured also in forms of vasculogenesis, within the allantois [52], then spread from the embryonic to placental directions, and finally, connect the intraembryonic with placental vascular beds.

Angiogenesis and Vascular Remodeling (Day 32 to Week 40)

At around day 32 p.c., a primitive fetoplacental circulatory network has been established with the villous endothelial tubes in contact with each other as well as the fetal allantoic vessels (Fig. 10). Hereafter, an extensive and continuous remodeling of vessels goes on in the placental vascular network during pregnancy. In the first and early second trimesters, the vessels increase gradually in the number, volume, and surface area with the villi, with continuously sprouting blind-ending capillaries [27, 53, 54], while the proliferation and elongation of endothelial cells will continue to increase the total surface area until term [55]. The construction is partially sustained by the angiopoietins secreted by pericytes around the endothelial cells and capillaries, which indicates the degree of maturity and stability of the network [56]. The pericytes and endothelial tubes also begin to form the basal lamina material around from about 6 weeks p.c., while the complete encirclement of the capillary is not to show until the last 10 weeks of pregnancy (Fig. 11) [14]. Besides, the differentiation of stem villi commences from the third month out of immature intermediate villi, along with the

expression of the cytoskeletal protein to form thin sheaths around endothelial tubes [57, 58]. In parallel, the superficial capillary net rarefies into a few largely unbranched para-vascular capillaries (Fig. 10).

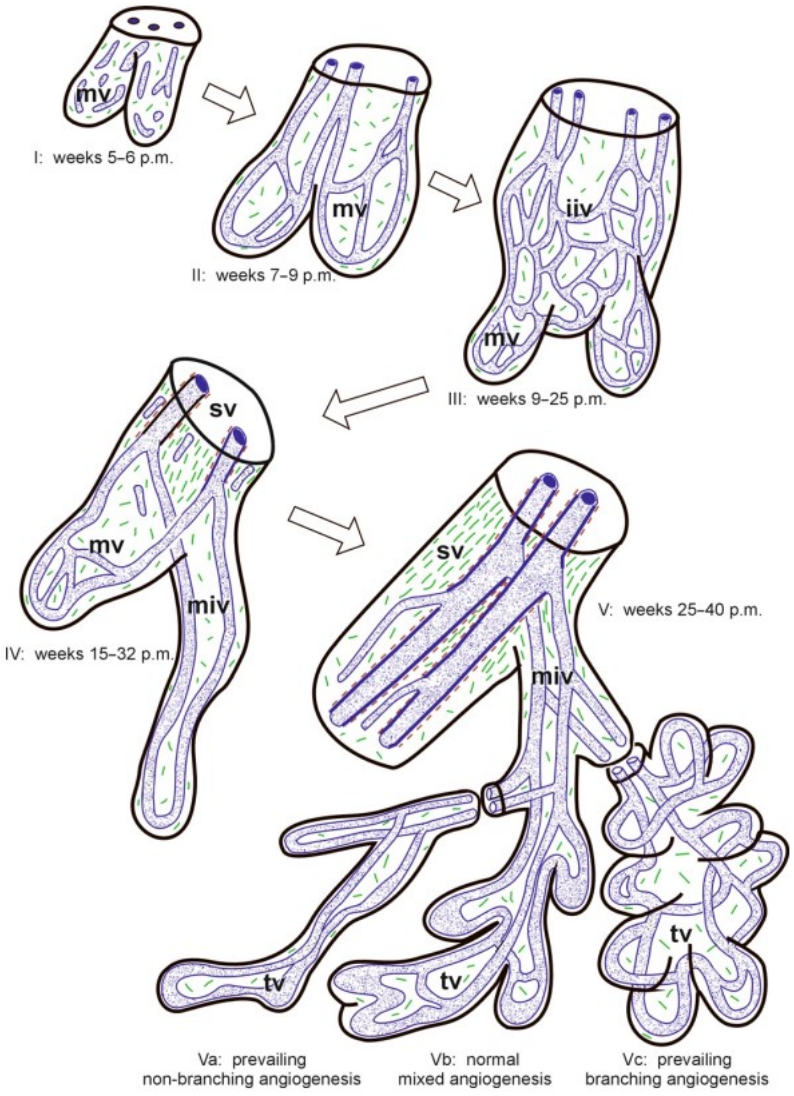


Figure 10 Showing angiogenesis and vascular remodeling of fetal vascular during villous development

Fig. 10 Showing angiogenesis and vascular remodeling of fetal vascular during villous development. I: vasculogenesis of fetal capillary segments. II: A simple netlike capillary bed fused by the primitive vessel segments. III: Development of immature intermediate villi (iiv) from mesenchymal villi (mv) by branching angiogenesis from the preexisting capillary bed. IV: Transformation of the centrally located capillaries

into stem vessels (arteries and veins) and the regression of the peripheral capillaries. V: Further growth and elongation of capillary loops with three different types of terminal villi because of imbalanced branching or nonbranching angiogenesis. *Green:* collagen fibers, *brown:* vascular smooth muscle cell, *blue:* endothelial tubes, mv: mesenchymal villi, tv: terminal villous [14].

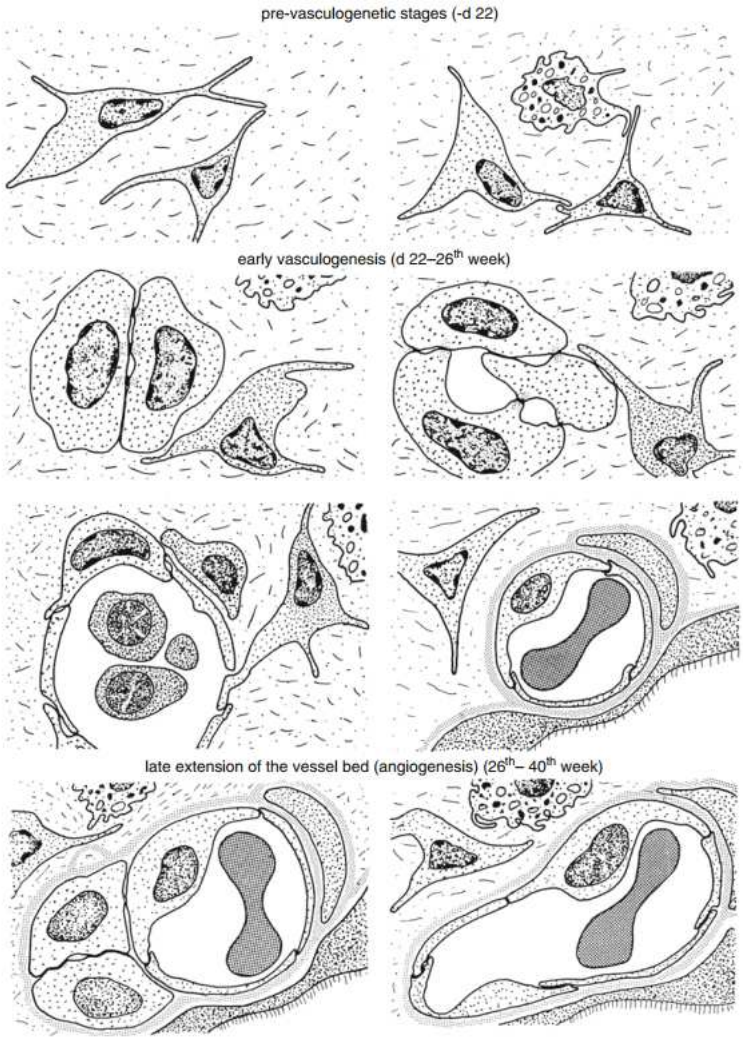


Figure 11 Vasculogenesis and angiogenesis in early and late placental villi

Fig. 11 Vasculogenesis and angiogenesis in early and late placental villi [14]. For further details, see the text.

1.1.4 Oxygen

Oxygen and Oxygen-Controlled Growth Factors as Regulators of Villous and vascular Development

The oxygen concentration plays an important role in villous development since it is heavily dependent on angiogenesis. Research has indicated that oxygen, along with its allied oxidative stress, has powerful positive and negative influences on villous growth [59]. Placental development requires a huge amount of oxygen which accounts for at least 30% of the total consumed by the uteroplacental unit, with one-third for protein synthesis, one-third for transport, and the rest for other functions. Meanwhile, it's important not to consume too much oxygen in the intervillous space for the sake of fetal development. This is controlled by minimizing the interposed trophoblast between the fetal capillaries and intervillous space with vasculosyncytial membranes, which are regulated by hypoxia.

"Hypoxia" in the mother-placenta-fetus unit refers to delicate conditions of which compartment reference is being made. Because hypoxia is a dynamic regulating state among these three, for instance, when the mother is exposed to hypoxia, the placenta will adapt to guarantee sufficient oxygen in the fetus. The metabolic requirements thus become the only standard to define the hypoxia state, but not a low oxygen environment. The types of hypoxia thus can be defined into three classes depending on the compartments (Fig. 12) [60]: 1) Pre-placental hypoxia, with all involved, the mother, the placenta, and the fetus, which can be caused by maternal anemia [61-63], and cyanotic maternal cardiac diseases; 2) Uteroplacental hypoxia, with the placenta and the fetus involved and the mother in normal, which can be caused by preeclampsia with the preserved umbilical end-diastolic flow; 3) Post-placental hypoxia, with only

the fetus involved, which can be caused by intrauterine growth restriction with the absent umbilical end-diastolic flow.

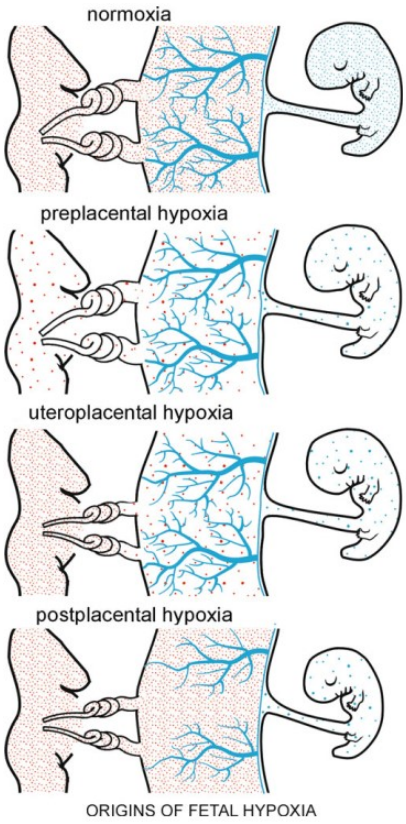


Figure 12 Defined types of hypoxia depending on the compartments

Fig. 12 Defined types of hypoxia depending on the compartments. Preplacental hypoxia: the hypoxia of the mother, the placenta, and the fetus. Uteroplacental hypoxia: the hypoxia of the placenta and the fetus but not the mother. Post-placental hypoxia: the hypoxia of the fetus but not the mother and the placenta. Red point: oxygenation of maternal blood, blue point: oxygenation of fetal blood. The dense of points represents the oxygen concentration referencing to normal or abnormal [60].

While the classification provides a useful framework to consider the impact of oxygen on the placenta and the fetus, it spares further revisions to be more accurate in the following aspects. Firstly, the physiological adaptations of maternal circulation should be considered in the models since changes in maternal and placental blood flows can

be sufficient to guarantee an adequate exchange of substances for fetal development [64-66]. Secondly, the different types of hypoxia should be addressed according to the placental responses [67]: 1) Hypoxic or hypobaric hypoxia due to insufficiency of oxygen transfer across the mother's lungs or living at high altitude, characterized by a low maternal arterial oxygen tension (PO_2) and oxygen content; 2) Anemic hypoxia due to a reduced carrying capacity of the blood, characterized by a normal PO_2 but low oxygen content; 3) Ischemic hypoxia due to reduced glucose and nutrient supply, characterized by injury of highly reactive oxygen species (ROS) resulted from the ischemia-reperfusion [49].

Oxygen Sensing in Placental Tissues

There are at least five ways for the placenta to sense hypoxia and respond to the prevailing conditions at different physiological forms of hypoxia, including oxygen-sensitive transcription pathways, ROS signaling pathways, oxygen-sensitive ion channels, ATP metabolites, and the unfolded protein response.

- Oxygen-sensitive transcription pathways. It focuses on the hypoxia-inducible factors (HIFs), containing HIF-1 and HIF-2. HIF-1 is a heterodimer composed of the hypoxia-dependent HIF-1a subunit and the HIF-1b subunit (also referred to as ARNT, arylhydrocarbon receptor nuclear translocator). Functioning as a transcription factor, it targets genes ranging from erythropoietin, VEGF, glucose transporters, to glycolytic enzymes, *etc.*, which protect the placenta from hypoxic situations [68, 69]. Notably, except for hypoxia, HIF-1 itself can also be regulated by a range of nonhypoxic factors such as nitric oxide, steroid hormones [70, 71], indicating consideration of the specific situation of the samples, as well as the location of HIF activity that happens. While HIF-1a may mediate acute responses, HIF-2a may contribute to longer-term adaptations to

milder oxidative stress [72, 73].

- ROS signaling pathways. It regulates placental development and function through a change in redox potential whose sensitive transcription factors include AP-1, SP-1, CREB, p53, Mash2, NF-kB, and STAT3, responding to the homeostatic balance of ROS [74-76]. The transcription factors can either modulate stress responses or regulate cell proliferation and differentiation. Activation of the ROS signaling pathways can activate the NF-kB, p38, and stress-activated protein kinase mitogen-activated protein kinase (SAPK MAPK) [77, 78], increasing secretion of pro-inflammatory cytokines.
- Ion channels. Ion transport can be an active and oxygen-consuming process, which is thus oxygen-sensitive and energy-demanding. The membrane where ion channels situate will become more impermeable under the hypoxia condition. For instance, inhibition of K+ channels because of hypoxia leads to membrane depolarization, activating voltage-gated calcium channels [79].
- ATP metabolites. ADP and AMP are the metabolites of ATP. The increased concentration of the latter act as a stimulus for the cytosolic enzyme AMP-activated protein kinase (AMPK), which regulates a variety of metabolic pathways, such as glycogen synthesis, glycolysis, and fatty acid oxidation, and inhibits protein synthesis via the mTOR pathway.
- The unfolded protein response. Oxidative stress and endoplasmic reticulum stress have a mutual influence on each other [80, 81]. The endoplasmic reticulum (ER) is a critical site for responses to various stress, including hypoxia, which is mediated mainly through the unfolded protein response [82-84]. It plays an important role in the homeostasis within the ER through suppressing mRNA translation and new peptides synthesis, upregulating ER chaperone proteins, synthesizing ER cisternae to increase the capacity, and stimulating the

proteasomal degradation pathway.

Oxygen and Villous cells

Villous cytotrophoblast was reported to be sensitive to oxygen. The number of villous cytotrophoblasts increases under the intrauterine hypoxia condition [61, 85]. The increased cytotrophoblast contributes to the boosting proliferation under 2% oxygen concentration, compared with 8% or 21% [86], and the shrinking fusion into the syncytiotrophoblast [87]. Apart from the influence from cytotrophoblast, villous syncytiotrophoblast itself is also sensitive to elevated or fluctuating oxygen levels, considering its booming and abundant syncytial sprouts under the low oxygen conditions [88] while the performance of oxidative stress under high oxygen conditions at the end of the first trimester [23, 24]. By contrast, the villous connective tissue reacts to the oxygen concentration in a different performance. The villous fibrosis is low, consistent with the low intervillous PO_2 across the first trimester, and high in the second trimester with steeply increased oxygen levels [75].

Oxygen and Intervillous Circulation

The flow of the intervillous circulation is determined largely by the flow inside the spiral arteries, which has been demonstrated by a computational model showing both the velocity and the volume of the blood flow [89]. Meanwhile, the trophoblast also functions to connect with the maternal blood, which is highly influenced by the degree of trophoblast invasion. Then it comes the influence of oxygen level. The results related to the effect of oxygen on the invasion are conflicting since the majority found the inhibition of either outgrowth or invasion of the trophoblast under reduced oxygen levels [90-94] while the other suggested hypoxia increases invasiveness [65, 95, 96].

Therefore, the situation between oxygen and the intervillous circulation should be carefully addressed on one side. The low oxygen levels during early pregnancy can promote the proliferation of cytotrophoblast cells and facilitate the extravillous trophoblast to invade into the maternofetal interface as the cytotrophoblastic shell. The success of the invasion will bring a plentiful supply of oxygen for the trophoblast cells, which again provides positive feedback for the further proliferation and invasion. However, on another side, the role of oxygen in trophoblast invasion maybe not the only factor since the other factors such as growth factors and cytokines can also affect the process, which is perhaps the reason for the paradox.

1.2 PPARG

1.2.1 PPARγ gene and protein structure

PPARγ gene structure

Eukaryotic genes are composed of intron and exon. In the human, there are more than 30,000 genes. Genes can be translated to mRNA and then translated to protein. In the contrast, with the fair conditions, we can retro-translate the mRNA according to amino acid sequences. Meanwhile, we can also retro-translate the cDNA as we have known the mRNAs. This is called genetic central dogma. This character was determined by the structure of the gene. Eukaryotic genes include not only the nucleotides for coding proteins but also the non-coding nucleotides which might for transcript regulation. Generally, these protein-coding genes will be regulated by a different type of protein-binding DNA sequence, which is called transcription control regions. In this area, many types of regulation elements were involved.

Peroxisome proliferator-activated receptor-gamma, with gene synonyms NR1C3, PPARG, PPARG1, PPARG2, PPARγ, locates in chromosome 3: 12,287,368-12,434,356

forward strands in human (Fig. 13). This gene has 17 transcripts (splice variants), 339 orthologues, 18 paralogues, is a member of 1 Ensembl protein family, and is associated with 45 phenotypes [97]. As to the 17 transcripts, 14 of them are of strength evidenced to transcribe and code protein, but for the rest 3 (PPARG-8, PPARG-213, PPARG-214) are still processed transcripts. In addition, except for PPARG-213 and PPARG-214, all the other transcripts can code proteins. Taking PPARG-205 for example, the box represents exon and line represents intron; hollowed box represents an untranslated region (UTR), while solid box represents coding regions of the transcript (also known as the coding sequences and CDS). There are 2 UTRs, 6 CDS, and 7 introns in the gene of PPARG. However, these components vary according to the variants of PPARγ. For example, PPARG-212 has only 1 UTR, 1 intron, and 1 CDS. These components of the variants have been proved by the human cDNAs detection as showing below in the figure. What should be emphasized is that the most involved PPARγ in the previous research are of 2 UTRs, 6 CDS, and 7 introns [98].

For the regulatory regions, the location of the first promoter is in chromosome 3: 12,287,200-12,294,201. According to the location of PPARγ, this nearest promoter is located at the upstream site ≈of 200 base pairs. According to the figure, we can see that there are promoters, CTCF (CCCTC-binding factor) binding regions, enhancers, promoter flank regions, and transcription factor binding sites.

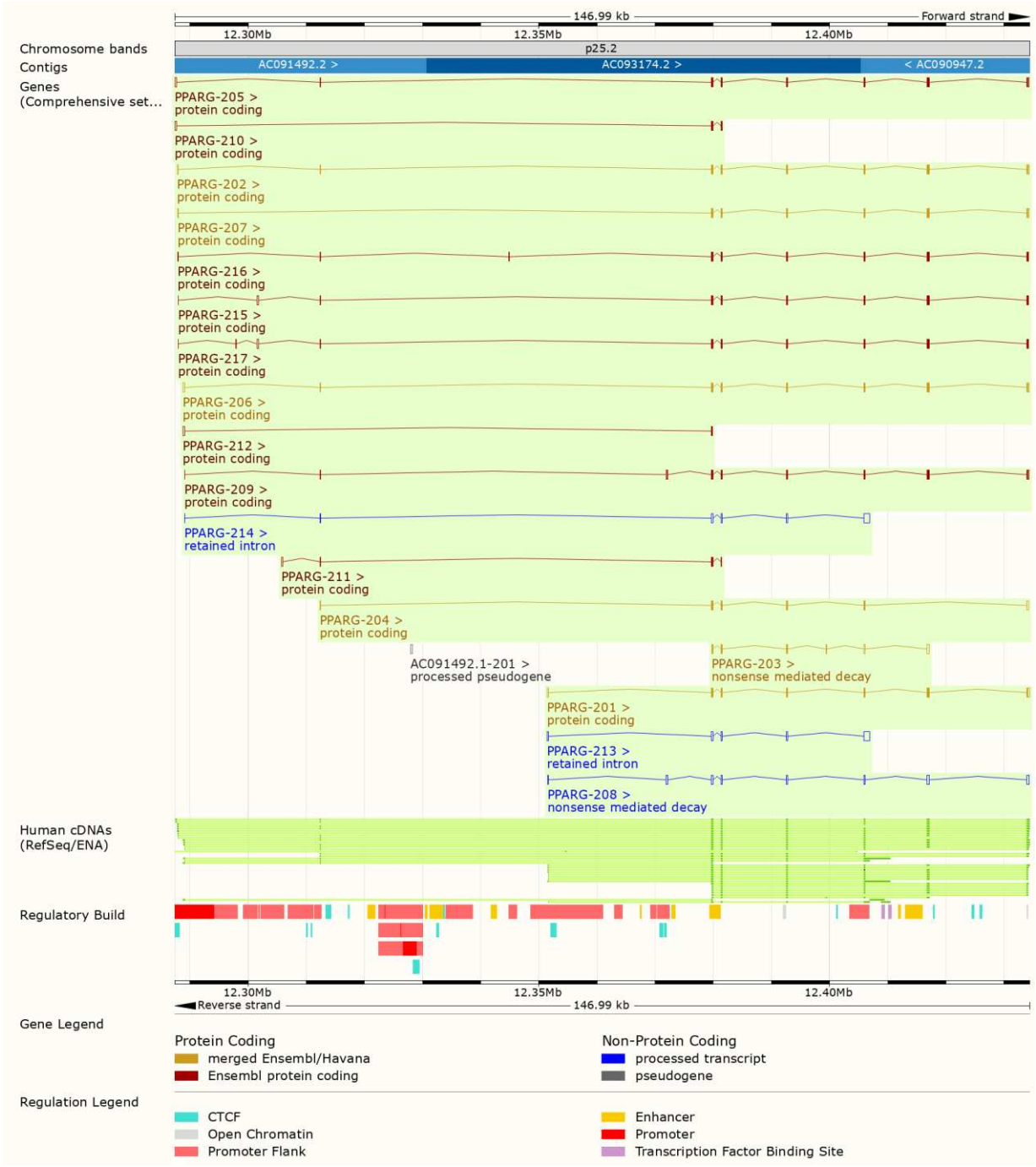


Figure 13 PPARγ Gene diagram from ENSEMBL

Fig. 13 PPARγ Gene diagram from ENSEMBL

In humans, the PPARγ gene will transcribe to RNA and then translated to PPARγ protein. Transcription is the first step. At the beginning of transcription, PPARγ gene double helix unwinds and is opened up for transcription. The mRNA of PPARγ is from

the forward strand of the PPARγ gene. The first RNA nucleotide is transcribed from the start site on the forward strand. The RNA polymerases are involved in the biological process, including initiation, promoter escape, elongation, and termination. For the initiation, RNA polymerase combined with transcription factors forms the RNA polymerase-promoter closed complex and binds to the PPARγ promoter. DNA double-strands are opened up and transcribed to the first bond, then RNA polymerase will escape from the promoter, following the release of the RNA transcript. However, these transcripts are generally truncated, which need to be capped with RNA polymerase combined with transcription factors. Then the RNA polymerase will traverse according to the template DNA of PPARγ to elongate the RNA sequences with the addition of RNA nucleotides. Finally, the transcription will be terminated and 3' -poly end will be added at the end of the RNA sequences. In addition, to form a mature mRNA of PPARγ, the pre-mRNA will need to be modified, such as splicing. With the existence of the mature mRNA of PPARγ, PPARγ protein will be translated.

The genetic code mRNA recognizes the amino acid according to the three-nucleotide sequences of the mRNA. Then amino acids of PPARγ protein were transferred by ribosome which is made up of 40S subunit and 60S subunit. Each amino acid is added at one time to the end of the chain of the polypeptide. In detail, the 40S subunit firstly binds to the 5'- end of mRNA with the assistance of initiation factors, and this subunit combines with tRNA will bring the recognized amino acids moving to the large subunit of 60S, which leads to the translation elongation. The amino acid sequences form the primary structure of PPARγ protein and then are modified to form a more complicated spatial structure. However, as the existence of the PPARγ variants, the number of amino acids of PPARγ varies according to the nucleotide sequences. As the base pairs PPARγ vary from 428bp to 2029bp, the amino acids of PPARγ vary from 40aa to 477aa.

PPARγ protein structure

Besides the primary structure of PPARγ protein, through further modification of the primary structure, PPARγ will constitute 4 main domains, A/B, C, D, and E/F domains. A/B domain, containing the N-terminal residue, is the least conserved domain, where the intrinsically disordered activation function 1 (AF1) was contained. The following domain is the most conserved C domain where the DNA binding domain (DBD) is included. DBD contains two C4 zinc fingers which can recognize the AGGTCA hexanucleotide motif. The D domain, also called the hinge domain, is located between the former domain and the latter one, which is related to nuclear localization, phosphorylation, and regulation. Then it comes to the E/F domain, also called ligand-binding domain (LBD). In this domain, C-terminal residue is contained, as well as a large hydrophobic pocket and a ligand-dependent transactivation function.

In detail, for the domain A/B, AF-1 is the site involved in ligand-independent coregulator binding, such as Map-kinase phosphorylation serine sites. This region greatly differs in different nuclear receptors according to its poor conserved characteristic. For the DBD in domain C, its conservation originates from the primary and tertiary structure of the protein. The folded two C4 zinc fingers are related to the recognition of specific DNA half-sites termed peroxisome proliferator response elements (PPRE). A C4 zinc finger is one type of zinc finger, because of its four cysteines in contact with the zinc ion. The existence of the zinc fingers of PPARγ distinguishes it from other DNA binding proteins, different from the intracellular receptors for non-steroid hormones. PPARγ binds to nuclear receptors usually combining with RXR heterodimer. The D domain, as a flexible hinge allowing for rotation, links the DBD and LBD. For the LBD, the largest domain of PPARγ protein is the second most conserved domain as its secondary structure is more conserved than the primary amino

acid sequences. Ligand binding can stabilize the ligand-binding domain structure and facility the interaction with cofactors for chromatin remodeling and transcriptional machinery recruitment. Therefore, LBD is involved in four main functions: a second dimerization interface, a coregulator binding surface, the ligand-binding pocket, and activation function 2(AF2). The scale of the ligand-binding pocket varies from classic receptors, adopted orphan receptors, and true orphan receptors. PPARγ is a typically adopted orphan receptor for its larger ligand-binding pocket than the classic receptors. As a member of the ligand-dependent nuclear receptor superfamily, PPARγ regulates downstream target genes by forming a heterodimer with the nuclear receptor retinoid X receptor α (RXRα) and then binding to the PPRE of target genes. PPRE is composed of two copies of the core motif organized by one nucleotide (DR-1) and a 5'-extension A(A/T)CT as a direct repeat space. Thus PPARγ/RXR heterodimer can recognize PPRE and exhibit the following consensus sequences 5'-A(A/T)CT(A/G)GGNCAAAG(G/T)TCA-3', which can be partially recognized by DBD. The D or hinge domain links the DBD to LBD (E/F), which is related to nuclear localization and interaction with regulatory proteins. PKC phosphorylation sites are also included in this domain. Meanwhile, the C-terminal E/F domain or LBD carries a large hydrophobic pocket that functions to bind lipophilic ligands and achieves a ligand-dependent transactivation function (AF-2).

1.2.2 Post-transcriptional modifications of PPARγ

PPARγ protein is a transcription factor, a DNA-binding protein that can bind to a particular promoter to control the transcription of downstream genes. The DNA control elements bound by the transcription factor are often located not close to the promoters of the target genes, which in some cases can be tens of thousands of base pairs away from the upstream or downstream from the promoters. This also leads to a

possibility of non-single control of the elements by multiple factors, which allows complex and elaborate regulation in the gene expression. The regulators for the DNA-binding domains can be activators or repressors according to the final effects. Activators connecting to a sequence-specific DNA-binding domain can contribute to transcription activation while repressors inhibit the transcription of genes, both through flexible protein domains. In contrast, mutations in these positions can alternate the final effects. That is, mutation of an activator-binding site decreases gene expression while mutation of a repressor-binding site is the opposite.

Either activators or repressors conform to the general mechanisms to regulate associated protein-coding genes, such as histone acetylation [99], phosphorylation [100], sumoylation [101], and ubiquitination [102]. These post-transcriptional modifications function through the reversibly modified histones in order to regulate either the condensation of chromatin or the exposure of residues. The regulators function to modulate chromatin structure to open up the regions or close off them so that binding to promoters will be influenced, whose flexibility is consistent with the condensation of chromatin.

For the acetylation, these processes happen often in the protein residues of the N-terminal region and the C-terminal region of histone, called histone tails, especially for the histone H3 and H4 tails. Acetylation of these tails decreases the chromatin condensation, increasing the accessibility of proteins with the transcription initiation position while deacetylation performs oppositely. To be specific, the reversibly modified histone tails in nucleosomes locate in the TATA box and promoter-proximal region of the genes. Evidence shows that the N-terminal lysine of unacetylated histones are positively regulated and interact with DNA phosphates, as well as neighboring histone octamers, thus favoring the folding of chromatin into condensed, while hyperacetylated tails fail to assemble into a preinitiation complex on a promoter

because of the elimination of neutralized and electrostatic interactions with DNA phosphates in the lysine. For the phosphorylation, it usually occurs on the residues in the location of serine, threonine, and tyrosine via phosphor-ester bond formation or on histidine, lysine, and arginine via phosphor-amidate bonds, or aspartic acid and glutamic acid via mixed anhydride linkages, such as the histidine at 1 and 3 N-atoms of the imidazole ring [103, 104]. For the sumoylation, it alters the interactions between molecules related to targets by masking or adding interaction surfaces. The target proteins can cause different outcomes, including changes in localization, activity, and protein stability [105]. For ubiquitination, it mainly affects the degradation of proteins through the proteasome and lysosome, with which it coordinates the cellular localization of proteins, activates and inactivates proteins, and modulates protein-protein interactions [106-108]. Notably, just as deacetylation, the modifications of phosphorylation, sumoylation, and ubiquitination can also be reversed, called de-phosphorylation, de-sumoylation, and de-ubiquitination, which leads to the opposite effects on gene expression.

Therefore, as a transcription factor, PPARγ shares as well these post-transcriptional modifications, which are intrigued by various substances. For example, phosphorylation of its serine 112 and 273 inhibits the transcriptional activity; sumoylation of its lysine 107 in the AF1 region and lysine 395 in the AF2 region stimulates PPARγ by restricting the interaction between the nuclear receptor co-repressor of HDAC3 and PPARγ; ubiquitination of it leads to protein degradation following by treatment of TZDs, PPARγ agonist (Anbalagan et al. 2012; Christianson et al. 2008; Floyd and Stephens 2002; Hauser et al. 2000); acetylation of it leads to PPARγ activation following the treatment of pioglitazone [109].

1.2.3 PPARγ and its ligands

Natural agonists of PPARγ

Selected fatty acids are confirmed to be natural modulators of PPARγ, even though they do not always activate PPARγ as well as the target gen transcription. Polyunsaturated fatty acids (PUFAs), mainly docosahexaenoic acid (DHA) and eicosapentaenoic acid, as natural agonists, can activate PPARγ and intrigue functional responses. DHA was reported to mediate the inhibition of tumor growth in human lung cancer cells. Meanwhile, delivery of DHA through albumin or enriched LDL with n-3 PUFAs to breast cancer cells, the proliferation of tumor cells will be diminished, combined with increased apoptosis [110-112]. Long-chain monounsaturated fatty acids (LC-MUFAs) with lengths more than 18 (*i.e.*, C20:1 and C22:1 isomer combined) increases PPARγ expression to enhance obesity-related metabolism and decreases inflammation in white adipose tissue [113]. Apart from PUFAs, phytanic acid, common in the human diet, can also activate PPARγ in a way similar to omega-3 PUFA by increasing glucose uptake and insulin sensitivity [114].

PPARγ pharmacological agonists

PPARγ mainly regulates lipid and glucose metabolism. As to the classical thiazolidinediones, including troglitazone, rosiglitazone, and pioglitazone, they can decrease levels of free fatty acid (FFA) and increase lipid storage in adipose tissue. Pioglitazone and rosiglitazone are used to treat patients with type 2 diabetes because of decreased hepatic glucose production and prolonged pancreatic β-cell function, which can prevent β-cells from apoptosis [115, 116]. While pioglitazone owns a positive effect in reducing cardiovascular complications by 16% in the main secondary endpoint compared with placebo treatment [117], rosiglitazone can controversially increase

myocardial infarction and cardiovascular-caused death with even only a short-term exposure [118]. The differences may result from the discrepant effects on lipid sub-fractions [115]. To be specific, pioglitazone can decrease triglycerides and fasting plasma free fatty acids and increase HDL cholesterol without changing total cholesterol and LDL cholesterol, while rosiglitazone does augment HDL levels and total cholesterol and LDL fraction [119-121]. Notably, among the thiazolidinediones, troglitazone is the only one to be revealed of properties in tumor-promoting and pro-angiogenic, which showed its positive effects on hepatic carcinogenesis and liposarcomas [122, 123].

Glitazones, as one of the synthetic ligands of thiazolidinediones, are of similar characters to synthetic ligands of thiazolidinediones and can function in the same way to improve insulin sensitivity by enhancing insulin and glucose parameters. Furthermore, activation of PPARγ by glitazones can attenuate systemic inflammation [124, 125] and reduce the growth of tumor cells and inhibit angiogenesis. For example, agonist RS5444 of PPARγ was reported to inhibit the growth of anaplastic thyroid cancer [126]. Despite its beneficial features in metabolic and anti-arteriosclerotic activity, *etc.*, glitazones also show side effects, including weight gain, bone fractures, heart failure, edema, and increased risk of myocardial infarctions, which should be considered carefully for use, especially in diabetic patients with high lipid levels [127]. Meanwhile, new selective PPARγ modulators, such as S26948 [128] and INT131 [129] are currently under research and development.

PPARα/γ dual agonists

The new synthetic agonists - PPARα/γ dual agonists – have multi-functions lipid and glucose metabolism. They not only have a capacity of anti-diabetic but also can reduce the development of arteriosclerosis by inhibiting inflammatory and anticoagulant

action, improving endothelial function, decreasing plasma free fatty acids, and lowering blood pressure to benefit the vasculature. However, as same as the side effects with glitazones, PPARα/γ dual agonists show aftereffects of weight gain and edema [130, 131]. What's more, the clinical use was also strictly limited because of the increased risk of bladder cancer and hyperplasia (ragaglitazar and naveglitazar) [132], cardiovascular risk (muraglitazar) [133], and renal dysfunction (tesaglitazar). And the promising new product aleglitazar, which is shown to decrease HbA1c and reduce triglyceride and LDL, and increases HDL cholesterol, has been withdrawn, owing to its toxicity and lack of efficacy [134].

PPARγ antagonists

Apart from the classical GW9662, which can prevent rosiglitazone-mediated PPARγ activation, and enhance rather than reverse rosiglitazone-induced growth inhibition [135], to satisfy the criteria of therapeutic efficacy with decreased side effects, new classes of compounds of PPARγ antagonists is proposed, including 13–16 (bexarotene, 2-phenylamino pyrimidine, and N-biphenylmethylindole derivatives), which performed well antidiabetic activity in rodent models of diabetes [136, 137]. There is also another alternative to targeting PPARγ for the therapeutic intervention in insulin resistance and type-2 diabetes, for example, Diospyros bipindensis, some of whose secondary metabolites have been purified and identified to function, including plumbagin, betulinic acid, caniculatin, 4-hydroxy-5-methyl-coumarin and ismailin22 [138].

Besides, novel ligands of PPARγ have also been synthesized, but their functions remain to decipher in the future. Here we list the compounds: PPARγ ligands: Synthetic compounds (thiazolidinedione salts) (US9126959B2), 5-hydroxy-4-phenyl-butenolide and derivatives (US9943501B2), Benzoate and phenylacetate (US20190000790A1),

and (E)-2-(5-((4-methoxy-2-(trifluoromethyl)quinolin-6-yl)methoxy)-2-((4-(trifluoromethyl)benzyl)oxy)-benzylidene)-hexanoic acid (MTTB) and its derivatives (PPARγ antagonist) (US20170210711A1); PPARα/γ dual agonist: US20170121268A1 [139].

Table 1. Natural peroxisome proliferator-activated receptors-γ ligands and synthetic agonists and antagonists.

Table 1 Natural peroxisome proliferator-activated receptors-γ ligands and synthetic agonists and antagonists

Natural Ligands	Synthetic agonists	Antagonists
• fatty acids	• rosiglitazone	• GW9662
• oxidized low-density lipoprotein	• thiazolidinedione	• T0070907
• 15-deoxy-12,14 prostaglandin J2 (15dPGJ2),	• ciglitazone	• BADGE
• prostaglandin D2	• troglitazone	• G3335
•9-and13-hydroxyoctadecadienoic acid (HODE)	• pioglitizone	• Fmoc-Leu
	• GW1929	• betulinic acid
• Unsaturated fatty acids	• farglitazar	
• 15- hydroxy-eicosatetraenoic acid	• S26948	
	• INT131	
• prostaglandin PGJ2		

1.2.4 PPARG and trophoblast cells

Placentation is a complicated process involving a complex process of cell cooperation. Trophoblast cells, as the main part of the placenta, function mostly important in placentation, which contains secretion, maturation, fusion, proliferation, migration,

and invasion from the first trimester to the term. Here we introduce the processes that have been taken part in by PPARG and its heterodimers RXRα.

Secretion

PPARγ signaling pathway in trophoblast cells took part in regulating visfatin through the secretion of interleukin (IL)-6 in BeWo, which indicated that the secretion of inflammatory cytokines promoted by PPARγ activation might thus promote the energy metabolism in trophoblast cells and their growth [140]. Similarly, the treatment of PPARγ agonist rosiglitazone in the medium where trophoblast cells were cultured could enhance the secretion of the cytokines interferon (IFN)-γ and prostaglandin E2 (PGE2) in trophoblast cells which were mediated by the MAPKs pathway [141]. Moreover, activation of PPARγ seemed to also affect the inflammation in HTR-8/SVneo cells via the NF-κB pathway [142]. Apart from the inflammation cytokines, PPARγ can also promote hCG expression and secretion in trophoblast cells, leading to villous trophoblast differentiation [143].

Fusion, differentiation, and Maturation

In the placenta, the fusion happens in the process when villous cytotrophoblast cells transform into the multinuclear syncytiotrophoblast cells. The formation of syncytiotrophoblast cells can be regulated by PPARγ/RXRα signaling directly through targeting syncytin-1 along with the MAPK or cAMP/PKA pathway [144]. In terms of villous cytotrophoblasts, activation of PPARγ by the agonist troglitazone induces the development of syncytiotrophoblasts *in vitro* [145]. Activation of PPARγ promotes the process that mononucleated villous cytotrophoblasts fuse into syncytiotrophoblast *in vitro* with an accumulation of neutral lipids, examined by oil red O staining [143]. Despite common sense that PPARγ and its heterodimeric nuclear receptor partner

RXRα can regulate fatty acid uptake [146], PPARγ can also promote villous trophoblast differentiation through promoting hCG expression and secretion in trophoblast cells [143]. Meanwhile, to knock down the PPARγ in mice embryos, the maturation of labyrinthine trilaminar trophoblast could be inhibited and occurred the deficiency of the vascular development in the placenta [147], while treating with PPARγ agonist rosiglitazone, it occurred a disorganization of the placental layers and an altered placental microvasculature [148].

Proliferation, Migration, and Invasion

Notably, proliferation, migration, and invasion of trophoblast cells are not strictly separative steps. These steps can happen simultaneously and also are promoted by one single molecule, for example, ANGPTL4, as a direct transcription target of PPARγ, mediated the proliferation, migration, and invasion in HTR-8/SVneo cells [149]. Generally, these steps are influenced by different PPARγ ligands. Pioglitazone, the PPARγ agonist, could promote the migration of extravillous trophoblast cells by up-regulating insulin-like growth factor (IGF) signaling pathway [150]. Synthetic and natural ligands that activate PPARγ could inhibit the invasion of the HIPEC 65 cell line without affecting proliferation [151-154]. However, the cell migration and invasion process might be decreased in a concentration-dependent manner concurrently in the extravillous cytotrophoblast cell line (HIPEC) as well as the primary extravillous cytotrophoblasts extracted from the first-trimester chorionic villi by either natural (15deoxy-Prostaglandin J2, oxidized lipids, *etc.*) or rosiglitazone [152, 155].

The invasion of trophoblast cells can also be affected by lots of factors. Activation of RXRα by 9-cis retinoic acid increased PPARγ-induced inhibition of trophoblast invasiveness [155]. Treatment with PPARγ or pan-RXR antagonists in extravillous cytotrophoblast cells promoted cell invasion. Inhibition of pregnancy-associated

plasma protein-A (PAPP-A) and the secretion of insulin-like growth factor (IFG), as the regulators of PPARγ, modulated trophoblast invasion [156]. Lysyl oxidase 1/2 negatively regulated PPARγ target genes and thus affected trophoblast invasion or itself could also be served as a target of PPARγ to control cell invasion [157], as well as the target genes matrix metalloproteinase (MMP)-2 and MMP-9 [157, 158].

Particularly, the effects of PPARγ on trophoblast cells are cell type-depending since the effects varied in cell types and even in the ligand concentration. The different PPARγ ligands might influence the different part roles of transcriptional or post-transcriptional levels, leading to various effects of interference in trophoblast invasion. Besides, our study also showed that PPARγ mediated the mono ethylhexyl phthalate (MEHP)-inhibited trophoblast invasion by disturbing the balance of MMP-9 and tissue inhibitors of metalloproteinase (TIMP)-1 expression in early pregnancy loss[159].

PPARγ and Energy Metabolism in the Trophoblast

The placenta mediates the transport of nutrients from the mother to the fetus, which happens specifically in the villous trophoblast. The nutrients, including lipid from lipoproteins, glucose, and amino acids, are consumed to supply energy, synthesize hormones and promote fetus growth. PPARγ plays important role in the metabolism of fat, glucose, *etc.* in trophoblasts.

PPARγ modulates fat storage, fat transport, and fat metabolism in trophoblasts that contain abundant lipid droplets. Associated with protein adipophilin, lipid droplets can be upregulated by PPARγ/RXR that promotes the synthesis of protein adipophilin [160]. The level of protein adipophilin increased due to the treatment of PPARγ agonists troglitazone, owing to the activation of PPARγ in the trophoblast [161]. The underlying mechanism might involve the enhanced uptake of free fatty acids, increased neutral lipids, and promoted expression of fatty acid transport protein 4 (FATP4) that

could be activated by the p38 MAPK pathway, as well as the decrease of FATP2 that affects RXR activation in trophoblasts [146, 162]. Besides, the lipids can conversely stimulate PPARγ activity by increased hCG in human trophoblasts treated with oxidized lipids such as 9S-hydroxy-10E,12Z-octadecadienoic acid (9-HODE), 13S-hydroxy-9Z,11E-octadecadienoic acid (13-HODE), or 15S-hydroxy-5Z,8Z,11Z,13E-eicosatetraenoic acid (15-HETE) [163].

PPARγ regulates glucose homeostasis in a way of multiple mechanisms. The energy metabolism of glucose can be verified by the fact that PPARγ agonists thiazolidinedione and pioglitazone promoted the expression of visfatin by IL-6 [140]. Meanwhile, Hyperglycemia could induce the apoptosis of human cytotrophoblast cells and the anti-angiogenesis of vessel branches by upregulating PPARγ and p38 MAPK phosphorylation [164]. By contrast, under the condition of hyperglycemia, the invasion of cytotrophoblast cells decreased due to the activated PPARγ pathways, including inhibiting urokinase plasminogen activator (uPA) and plasminogen activator inhibitor 1 (PAI-1) and enhancing the expressions of IL-6 and soluble fms-like tyrosine kinase-1 (sFIt-1) [165]. Besides, PPARγ was necessary for normal insulin sensitivity and adipogenesis, whose absence leads death of embryo due to placental dysfunction [166] and whose presence in trophoblast cells could rescue embryonic lethality [167]. The expression of PPARγ increased with the treatment of insulin sensitizer in primary extravillous trophoblast [150], while adiponectin inhibited insulin-mediated amino acid uptake in the cultured cells [168].

1.3 Sequencing technique

The rapid development of sequencing technologies over the past four decades has advanced a lot in the ability to detect genomics in individuals. DNA sequencing

gradually evolved from low throughput DNA fragment sequencing to high throughput next-generation sequencing (NGS) and third-generation sequencing techniques including rapid ways for genome-wide characterization and profiling of mRNAs, small RNAs, transcription factor regions, the structure of chromatin, and DNA methylation patterns [169]. Even single-cell sequencing technologies have also been rapidly developed for observing the multilayered status of single cells in different tissues. Sequencing technologies are widely used in molecular biology to study wide-genomes from different aspects. Information obtained using sequencing different technologies allows researchers to detect the expression of global genes under specific conditions, which indicates associations with diseases and phenotypes.

The general strategy for the downstream analysis focuses on the enrichment of identified gene sets. With the detection of the expression of genes, the enrichment of the genes is conducted to figure out the critical biological functions. These functions are generally provided via matching to the Gene Ontology (GO) knowledgebase, which is the world's canonical and largest source of information on the functions of genes [170]. The GO knowledge base is widely regarded as a reference to guide further research as soon as the differentially expressed genes (DEGs) were enriched. GO terms represent the comprehensive aspect of the function of the genes and gene products. A series of complicated biological processes would be involved in the disease occurrence, owing to the outcome happening inside or between cells [171]. The processes were summed up in the GO knowledgebase which contains 44,085 terms, 7,931,218 annotations, and 1,564,454 gene products to 4,743 different biological organisms as of February 2021. Among the terms, it contains 28748 biological process terms, 11153 molecular function terms, and 4184 cellular component terms. The abundance and diversity of GO terms indicate the possibility of disease characterization, which means a more general way, at the biological process level, should not be underestimated in

clinical practice. Specifically, instead of linking a single gene or genes with diseases, it is potentially reasonable to use GO terms to predict diseases. Further research will be mostly focused on the study of specific mechanisms in order to discover valuable drug targets or seeking links between the DEGs and diseases for diagnosis and prognosis in clinical practice [172].

1.4 Prospects

The human placenta plays a pivotal role in development by regulating the exchange of nutrients, gas, and waste between the mother and the fetus. Placentation involves a complex interaction between growth, rates of blood flow, transporter protein expression, trans-membrane concentration gradients, and metabolic demands. An overview of the normal development of the human placenta at transcriptome level needs more details to illustrate the complicated biological processes during the three trimesters. Among the processes, the first trimester lays the foundation for all subsequent processes, which requires more studies, especially for the development of trophoblasts and the oxygen-related process. For the former, different trophoblast cell lineages that constitute the main part of the placenta begin to differentiate in this period, where PPARγ, located in the nuclei of EVCTs and VCTs, plays an important role. For the latter, the evolution of this process is complex and intricately regulated by O_2 tension, especially in the period from 8 to 12 gestational weeks (GW). Oxygen, and the oxidative stress that accompanies it, play an important role in the positive or negative development and growth of chorionic villi, as well as the differentiation of villous cytotrophoblast. In our work, we applied various strategies to tackle the inadequate dataset, technology, and methodology.

1.5 Overview of the thesis

Here we give the overview of our study which is composed of four parts. In the first part, we mined the gene expression across the human pregnancy, which allowed us to discover the importance of PPAR signaling pathway and confirmed the role of PPARγ in the whole pregnancy. Based on the results from the first parts, here come the ideas of part two and three. In part two, we selected PPARγ to activate in the human placental cells, the cytotrophoblast and extravillous cytotrophoblast, followed by the comparison of the enriched terms in the two types of cells. In part three, we were intrigued by the simultaneous occurrence of lipid metabolism and the increased oxygen level in the first trimester. We hypothesized that the lipid metabolism was probably related to the oxygen level increase. In part four, since we have confirmed the importance of PPARγ, we were wondering if we could use it to predict the pregnant disease. So, we collected the clinical information national wide and the variants of PPARγ in the cohort study, and used eight machine learning methods to build models and performed prediction.

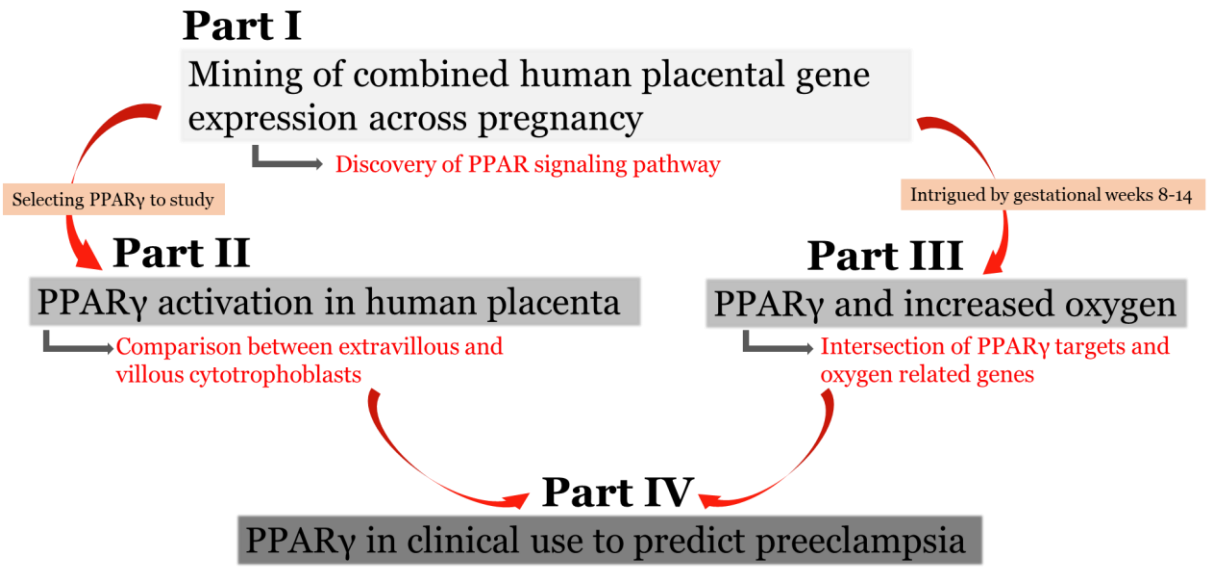


Figure 14 Overview of the thesis

Fig. 14 Overview of the thesis.

Part I

2 Part One

Lead-in

The pregnancy will last about 40 gestational weeks in the human. In the other research, it's common to study only a specific period, for example, the first trimester or the term of the placenta. However, the development of the human placenta is a dynamic process, to study a fixed period is not enough to get an overview of its development. In our proposal, we aimed to explore more information across the whole pregnancy by collecting the gene expression all over the gestational weeks. Instead of studying the partial biological process in a specific period, we choose to study the integral biological processes that happen throughout the whole period. Otherwise, the partial biological processes may play important role in that specific period but not the whole period. The idea to discover the process that can function throughout the pregnancy will be meaningful, which would pave a way for the treatment from the beginning to the end of the pregnancy. For example, we already know the importance of PPARγ in the human placenta, whose deficiency can lead to the mortality of an embryo. But is its regulation a static process in a specific duration, like in the first trimester, or a dynamic and ordered arrangement across the whole gestation period? What if we put PPARγ into the background of the whole gestation period, can its importance still be detected rather than be overwhelmed by the other more important stuff? Based on the proposal, we mined the dataset of gene expression across the pregnancy so as to discover the interesting and important processes.

2.1 Mining of combined human placenta genome across pregnancy, applied to PPAR signaling pathway

Fulin Liu [a], Wencan Zhu [b], Hussein Shoaito [a], Audrey Chissey [a], Séverine A. Degrelle [a, c, d, #], Thierry Fournier [a, c, #]

[a] Université de Paris, INSERM, «Pathophysiology & Pharmacotoxicology of the Human Placenta, pre & postnatal Microbiota», 3PHM, Paris F-75006, France

[b] UMR Applied Mathematics & Informatics, AgroParisTech-Université Paris-Saclay, 75005, Paris, France

[c] Fondation PremUp, Paris, F-75006, France

[d] Inovarion, Paris F-75005, France

[#] These authors contributed equally to this work.

Correspondence should be addressed to Thierry Fournier; thierry.fournier@parisdescartes.fr

2.1.1 Abstract:

Introduction: Overview of the human placental genome-wide expression from the very beginning gestational age to the term is lacked. Our aim is to investigate the dynamic changes in gene expression throughout placentation.

Methods: In our study, gene expression profiles of human placentas from 4 to 40 gestational weeks were collected. Linear regression and weighted correlation network analysis were applied for gene filter. Gene enrichment analysis including gene ontology and Kyoto Encyclopedia of Genes and Genomes pathway terms were performed by clusterProfiler. Line graph drawn with scaled and adjusted gene expression was applied to display the dynamic changes.

Results: Our results showed a total of 5173 genes involved in different period of placentation. Downstream annotation of these genes revealed the biological processes and pathways involved, among the intersection of which we selected "PPAR signaling pathway". This pathway map shows the genes involved in lipid storage/metabolism, including FABP family members, LPL. Moreover, lipid staining experiment on placental sections showed a significant decrease in lipid droplets content in first trimester placentas compared to term placentas.

Conclusion: Our study provides more information on biological processes and pathways across human placentation. These findings give us new clues for deciphering the normal functions of placentation and their mis-regulations may be linked to pregnancy-related diseases. As an example, our results show PPAR signaling pathway mediates constant decrease of placental lipids throughout pregnancy.

Highlights:

- This study shows human placental genome-wide expression from 4 to 40 gestational week.
- Linear regression and weighted correlation network analysis are combined to identify significant genes.

- Scaled and adjusted gene expression drawn in line graph displays the gene dynamic changes across placentation.
- PPAR signaling pathway mediates lipids decrease during placentation.

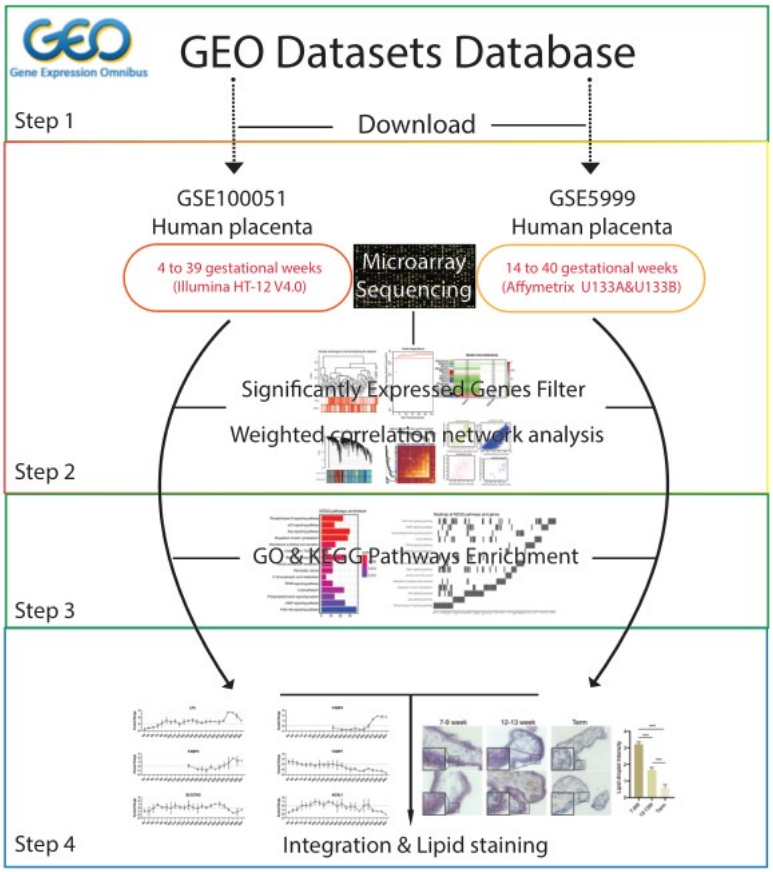


Figure 15 Graphical abstract

Graphical abstract: Weighted correlation network analysis reveals dynamic changes of biological processes involved in human placentation. PPAR signaling pathway has been selected to show its role in mediating lipids decrease through placentation.

Key words: WGCNA; Placenta; Bioinformatics; Microarray; Peroxisome Proliferator Activated Receptor (PPAR); Lipids

2.1.2 Introduction

The placenta acts as a bridge between the developing fetus and the mother, functioning as an exchanger in gas, water, nutrients exchange. The placentation starts out upon the implantation of the blastocyst into the endometrium. The outer layer of the blastocyst, known as the trophoblast, forms the out layer of the placenta. This trophoblast further differentiates into two subtypes, known as villous cytotrophoblast (inner layer) who forms by fusion the syncytiotrophoblast (outer layer) and the extravillous cytotrophoblast who has the property to invade the maternal endometrium[173]. The syncytiotrophoblast is renewed by fusion of underlying villous cytotrophoblast throughout pregnancy [174].

Aberrations of placental structure and function lead to an immediate effect on the outcome of a pregnancy as well as an influence in the life-long health of the offspring [175]. For example, the abnormal invasion of extravillous trophoblast results in the dysfunction of uterine spiral artery remodeling, associating to the pregnancy-related disease pre-eclampsia[176]. Moreover, the placentation, owning to its complicated process, involves a complex interaction between growth, rates of blood flow, transporter protein expression, trans-membrane concentration gradients and the metabolic demands[176]. Therefore, it is necessary to have an overview of the normal development of human placenta. The findings may thus, provide more details of the alterations of placental structure and function in pregnancy-related diseases.

In human placental transcriptome studies using microarray technology, groups comparison among first trimester, second trimester or term, is commonly used to uncover the alterations underneath. For example, Soncin et al [177] has conducted a

genome-wide expression profiling of human placental specimen from gestational week 4 to 16. As well, Winn et al. [178] has provided the expression profiling of the human placenta from gestational week 14 to the term of pregnancy. These two expressions data were sequenced on Illumina and Affymetrix chips, respectively. With the gene expression profiles, moderated t statistic and adjusted P value were generally applied. In our study, we furtherly applied linear regression and weighted correlation network analysis (WGCNA) to explore the dynamic changes in gene expression during normal placentation. Reasonably, we combined the samples from the two studies mentioned above which indicated the normal placentation ranging from 4 to 40 weeks. Finally, we examined lipid content in placental tissues corresponding to the enrichment results.

2.1.3 Materials and Methods

Datasets

The microarray profiling of GSE100051 and GSE5999 were retrieved from the Genome Expression Omnibus (GEO) database (ncbi.nlm.nih.gov/geo/query/acc.cgi?acc= GSE100051/GSE5999). GSE100051 profiled gene expression of human placenta from 4 to 39 gestational weeks on Illumina HumanHT-12 V4.0 [177], while GSE5999 profiled gene expression of human placenta from 14 to 40 weeks on Affymetrix Human Genome U133A&U133B [178]. To be convenient, in our study, we labeled microarray data in GSE100051 as “dataset1”, and U133A as “dataset2”, and U133B as “dataset3”. Additionally, we only retained the samples in first trimester and second trimester in dataset1, to be successive with dataset2&3.

Clustering analysis and construction of co-expression modules of human placenta microarray data

Dataset1 includes the gene expression from first trimester, second trimester and term human placenta samples. A total of 11405 genes were detected in the microarray data. While 22217 genes were detected for dataset2 and dataset3. To clarify the sample combination methods, we chose the samples with same gestational weeks (14, 16, 39) from these datasets to exam the compatibility. Specifically, distance clustering, principal component analysis (PCA), t-distributed stochastic neighbor embedding (t-SNE) were used. After the classification of samples, linear regression method [179] was used to filter the significantly expressed genes (SEGs) with adjusted P value less than 0.05. The WGCNA algorithm[180] was subsequently used to evaluate SEGs expression with the relevant WGCNA package (version: 1.68) in R (version: 3.6), accompanied with the clustering analysis of the human placenta samples in appropriate threshold values. To be specific, clustering analysis is the first step to detect outliers of samples. Secondly, the soft thresholding power value will be determined through a range of power value set (from 1 to 30) to reduce the background noise of the correlations in the adjacency matrix. The optimum power value will be selected according to the measurement of the scale independence which acts as a criterium under threshold 0.9. The accurate construction of co-expression modules depends on the adjacency matrix which shows the correlations of eigengenes from default unsigned network. The background noise of the adjacency matrix can be reduced with optimum soft thresholding power value. The minimum module size was set as 30 to give access to high reliability results. The modules were consequently constructed according to the power value provided by WGCNA package.

Analysis of co-expression modules for human placenta microarray data

With the constructed modules, the clustering dendrogram was plotted corresponding to the genes. Heatmap of the gene co-expression values was performed to show the strength of the associations. Likewise, module-trait associations of the module eigengenes with the clinical trait (gestational weeks) were estimated by defining Gene Significance (GS) and module membership (MM). The former represents the absolute value of the correlation coefficient between the genes and the gestational weeks, while the latter was defined to calculate the correlation coefficient of the gene expression profile and the module eigengene. The selection threshold for significant modules: correlation value of modules was set as 0.6 combined with P value less than 0.05.

Functional enrichment analysis of the co-expression modules and line graph drawing

The significant modules for further inspection were selected with the criteria: the correlation coefficient exceeds 0.6 and P value is less than 0.05. Genes contained by the significant modules were submitted afterward for functional enrichment analysis. The clusterProfiler (version 3.9) was applied for gene enrichment analysis including gene ontology and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway terms [181]. P value less than 0.05 was set as the threshold and top 15 terms were kept for visualization. Subsequently, we extracted the genes from selected pathway and performed the log2-transformation, followed by standardization (Z-scores transformation). That is, each expression of gene was adjusted by the mean and

standard deviation ($Z = \frac{x-\bar{x}}{S}$, where x is the raw expression value, $\bar{x}$ is the sample mean and S is the sample standard deviation). For the sake of combination and continuity of dataset1 and dataset2&3, the scaled gene expression in the overlapped gestational weeks (14w&16w) was set as a baseline value Δ ($\Delta = Mean_1 - Mean_{2\&3}$, where $Mean_1$ is the mean value from dataset1 in gestational weeks 14&16, $Mean_{2\&3}$ is from dataset2&3). Baseline value was then added to each scaled gene expression in dataset2&3. The scaled and adjusted gene expression was hence drawn in line graph corresponding to gestational weeks. The Y coordinate of the line graph presents the relative scaled range instead of the absolute expression of genes.

Lipid (Oil red O) staining

First term (7-13wk of pregnancy) and term placental tissues were obtained with the patients' written informed consent from Cochin Port-Royal, Antony, and Montsouris maternity units (Paris, France). Our protocol was approved by the local ethics committee (CPP 2015-mai-13909). Oil-red O staining was performed to detect the lipid droplet accumulation in the first trimester (early, n = 9; late, n = 9) and term placental (n = 5) tissues. Briefly, placental villi were frozen with cryomatrix gel (Thermoscientific, Runcorn, UK) under liquid nitrogen vapor. Frozen tissues were sectioned with 10 μm thickness. ORO working solution (150 mg O-red oil powder + 50 ml 100% isopropanol + 80 ml distilled water) was added to the slides to cover the sections which subsequently incubated at room temperature for 7 min, followed by the counter stain of hematoxylin at room temperature for 30 s. After 2 h rinse under tap water flow, the sections were covered by glass slides with the mounting medium (Dako North America, Inc., CA, USA) and then examined under a light invert microscope (Olympus BX60,

Tokyo, Japan) at ×40 amplification. The intensity and number of lipid droplets were evaluated blindly by two persons. Ordered logistic regression method was applied to select the effective factors from gestational age, sex and smoking.

Statistics

Data are presented as means ± standard error. Statistical analysis was performed using one-way ANOVA combined with scheffe post-hoc for groups and ordered logistic regression for multivariate data. P value less than 0.05 was considered a statistically significant difference.

2.1.4 Results

1. Datasets evaluation and SEGs filter and soft thresholding power value selection

To evaluate the possibility of combination analysis of dataset1&2&3, we performed the clustering analysis. Our sample compatibility test shows that these samples from different sequencing platforms or experiments are clustered in different groups in distance clustering (Figure 1A&1B), PCA analysis (Figure 1C) and t-SNE analysis (Figure 1D) and hence not suitable to combine directly. Therefore, we detected the gene set modules for these microarrays separately. For dataset1, all the gene expression from 4 to 16 gestational weeks were processed by WGCNA. While 2583 SEGs in dataset2 and 1536 SEGs in dataset3 were precedingly filtered using the linear regression analysis. These two sorts of SEGs were integrated, with the matched expression matrix (36 samples) in datasets, and then submitted to WGCNA for clustering and selection. Our result shows no outlier samples in dataset1 and the optimum soft thresholding power

value is 4 according to the plot (Figure 2A). In dataset2, no outlier sample was shown and the optimum power value is 24 (Figure 2B) while no outliers in dataset3 and the optimum power value is 18 (Figure 2C).

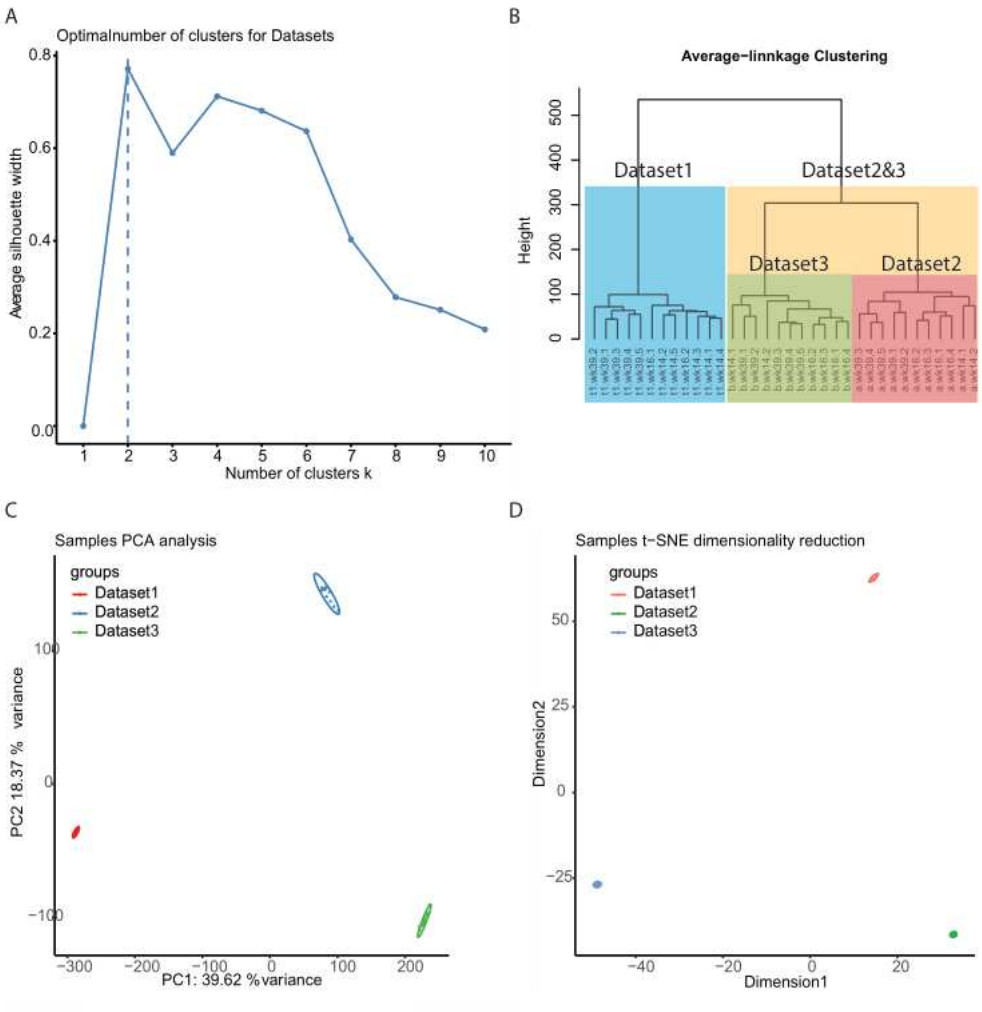


Figure 16 The results of sample compatibility test from different sequencing platforms or experiments

Fig. 1. The results of sample compatibility test from different sequencing platforms or experiments. (A) Optimal number of clusters for samples in combined dataset 1&2&3 according to K-means clustering algorithm. (B) K-means clustering for samples in dataset1, dataset2 and dataset3. (C) Samples PCA analysis for dataset1, dataset2 and dataset3. (D) Samples t-SNE dimensionality reduction for dataset1, dataset2 and

dataset3. PCA: principal components analysis; t-SNE: t-distributed stochastic neighbor embedding.

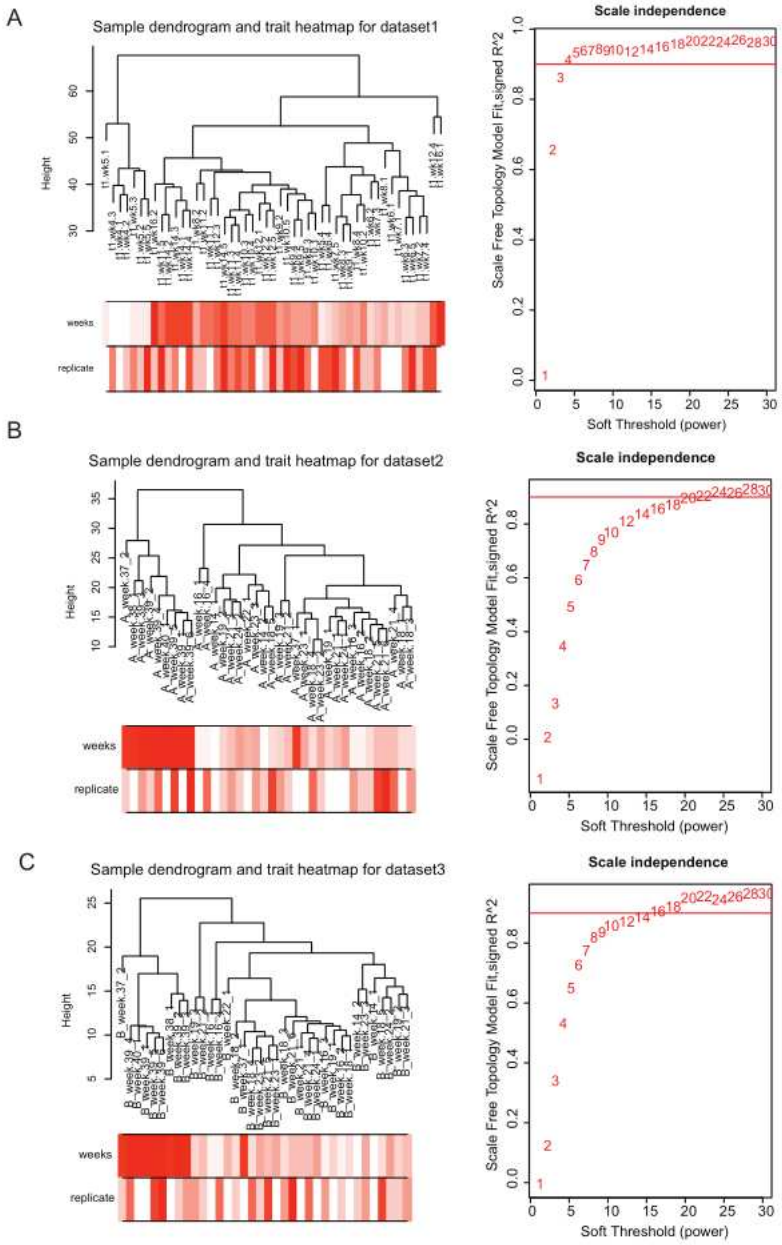


Figure 17 Sample clustering and soft-thresholding power selection for microarray datasets

Fig. 2. Sample clustering and soft-thresholding power selection for microarray datasets. Sample clustering is based on the Euclidean distance and the trait heatmap of samples includes gestational weeks and replication. Soft-thresholding power selection was determined through a range of power value set from 1 to 30. Graphs A, B

and C represent dataset1, dataset2 and dataset3, respectively. For each graph, the left panel shows the sample dendrogram and trait heatmap while the right panel shows the scale independence in which the threshold is set as 0.9 (red line) for value filter.

2. Construction and analysis of co-expression modules of human placenta microarray data

The original co-expression modules in R environment were constructed under the dynamic tree cut for branch cutting method. The number of genes contained in the original modules were shown in Table S1. After the construction of original modules, the modules with similarity were further merged to form the merged modules, along with the corresponding overview of the Topological Overlap Matrix (TOM) heatmap between genes (Figure S1). With the merged modules, interaction analysis between the module eigengene and gestational weeks was performed, as well as the scatterplots of GS vs. MM for merged modules. The results show that, under the selection threshold, the correlation coefficient of greenyellow, blue, pink and royalblue modules in dataset1 (Figure 3A) and grey modules in dataset2 (Figure 3B) and grey modules in dataset3 (Figure 3C) satisfy the criteria of correlation coefficient more than 0.6 and P value less than 0.05.

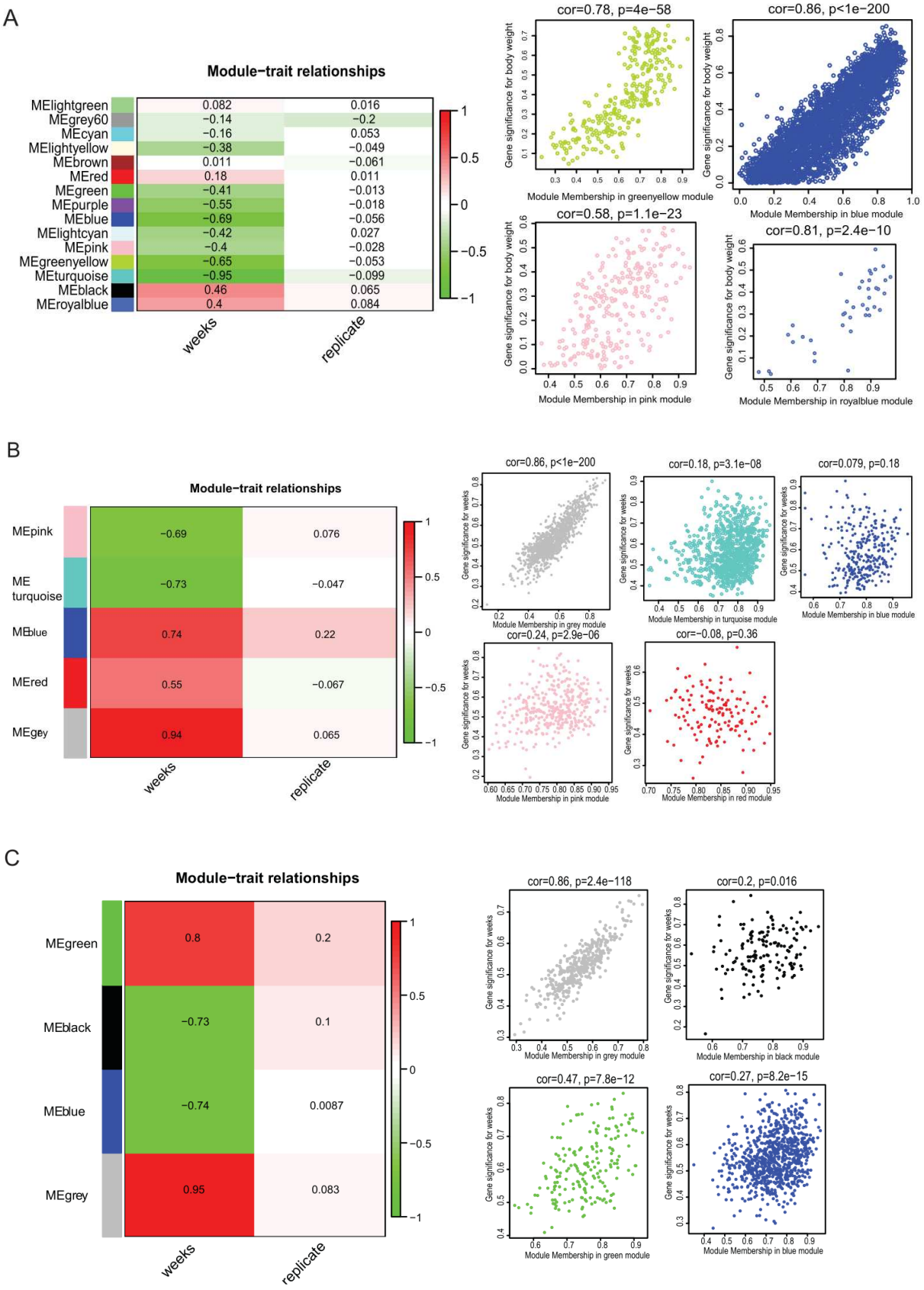


Figure 18 Heatmap of module-trait associations and scatterplots for merged modules

Fig. 3. Heatmap of module-trait associations and scatterplots for merged modules. Module-trait association heatmap for dataset1-3 in the left panels of graph A&B&C is composed of traits in columns and color modules in rows. The legend on left side of the heatmap represents the color module types and the legend on right side represents the value scale. According to the module-trait value, trait of gestational weeks along with the corresponding merged modules were selected to plot. Scatterplots based on gene significance score and module membership value were subsequently performed for dataset1-3, which were shown in the right side of the graph A&B&C, respectively. Correlation coefficient more than 0.6 and P value less than 0.05 were set as the inclusion criteria.

3. Functional enrichment analysis of the co-expression modules

Next, we integrated the genes within the significant modules. Before the submission, we combined the significant genes in dataset2 and dataset3 since they originated from the same samples. A total of 3651 genes in dataset1 and 1522 genes in dataset2&3 (1122 in dataset2 and 400 in dataset3) were submitted separately to clusterProfiler for functional enrichment analysis. Top 15 biological process GO terms for dataset1 and dataset2&3 in bar plot are shown in Figure 4A&4B, respectively, while top 15 KEGG pathways for them in Figure 4C&4D, respectively. Furthermore, the intersection of GO terms and KEGG pathways between dataset1 and dataset2&3 are shown in upset plot (Figure 4E). GO terms and KEGG pathways in details are given in supplementary material (Table S2).

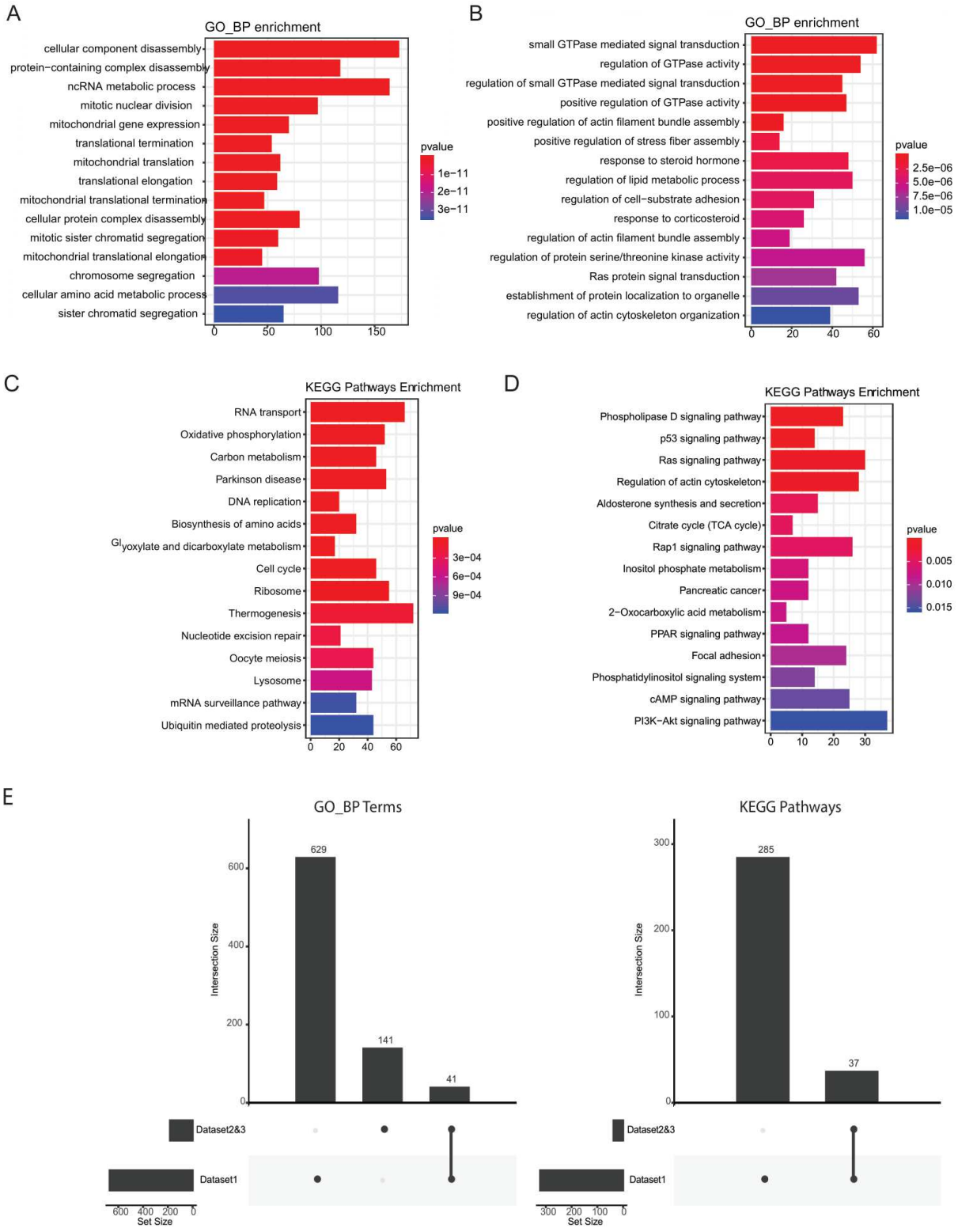


Figure 19 Functional enrichment analysis of the significant modules

Fig. 4. Functional enrichment analysis of the significant modules. In the functional enrichment analysis, 3651 genes in dataset1 and 1522 genes in dataset2&3 (1122 in dataset2 and 400 in dataset3) were submitted to plot using clusterProfiler. A&B: Top

biological process GO terms in bar plot for dataset1 and dataset2&3, respectively. C&D: Top KEGG pathways in bar plot with its corresponding involved genes in heatmap for dataset1 and dataset2&3, respectively. E: The intersection of GO terms and KEGG pathways between dataset1 and dataset2&3 in upset plot. GO: gene ontology; KEGG: Kyoto encyclopedia genes and genomes pathway.

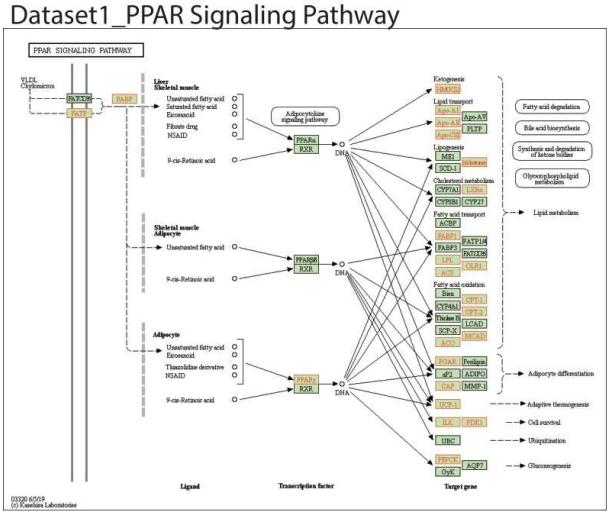
4. PPAR signaling pathway and lipid droplets detection

According to the intersection, we were interested in "lipid metabolic process" and "PPAR signaling pathway", genes involved in PPAR signaling pathway were partly extracted and mapped to the pathway for the datasets (Figure 5A), which included FABP family members, PCK2, LPL, SLC27A2, ACSL1, PPARA, and PPARγ and its heterodimer RXRα. The scaled and adjusted gene expression of all these genes corresponding to gestational weeks were drawn in line graph, agreeing with the map (Figure 5B). According to previous analysis, these genes vary significantly in expression during gestational age. Among them, FABP family members and LPL show apparent up- and down- regulation in their expression (Figure 5B). To have an overview on human placenta lipid metabolism during pregnancy, we detect the lipid content on human placental sections. To do so, an Oil-red O staining experiment was performed on early (7-9w) and late (12-13w) first trimester and term tissue sections (Figure 5C). The results show an abundant of lipid droplets in first trimester placentas compared to a deficiency in term. To go further, a semi-quantitative analysis was performed to reveal the difference of lipid droplets in first trimester (7-9w, 12-13w) and term, and explore the correlation of the lipid droplet content with gestational age, sex and smoking. One-way ANOVA test combined with scheffe post-hoc shows significant difference between groups ($p < 0.0001$, Figure 5C). A negative correlation between first

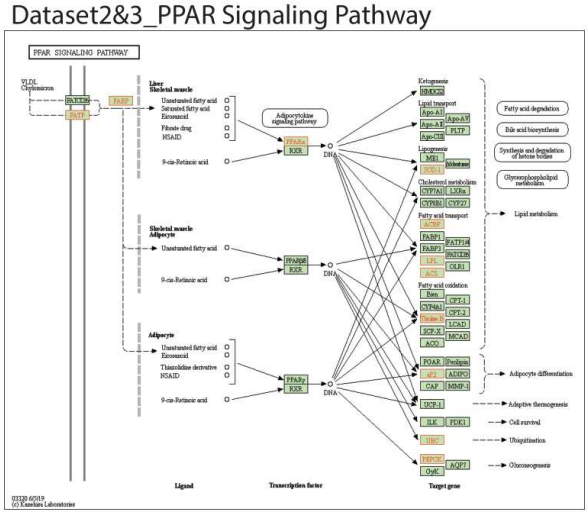
trimester gestational age and the lipid droplets intensity was observed ($p < 0.0001$, Table 1). Data for ordered logistic regression analysis were provided in details in supplementary material (Table S3).

A

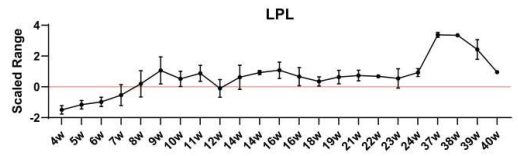
Dataset1_PPAR Signaling Pathway

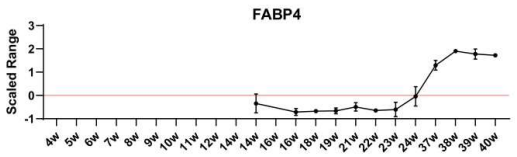


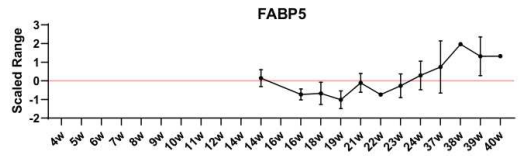
Dataset2&3_PPAR Signaling Pathway

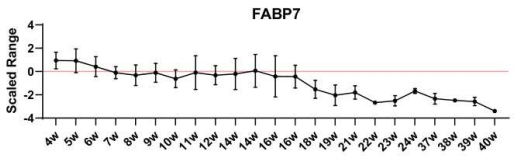


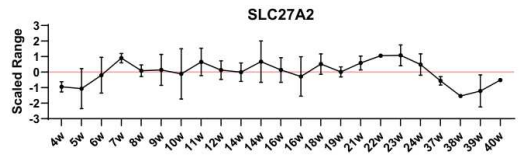
B

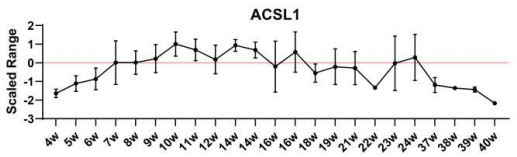


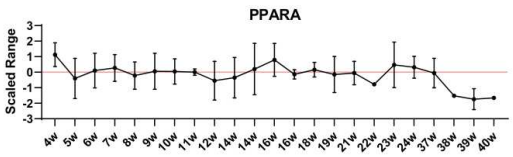


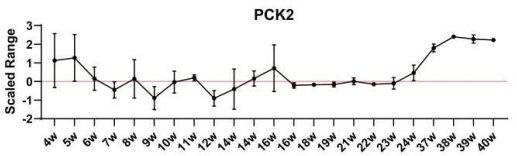


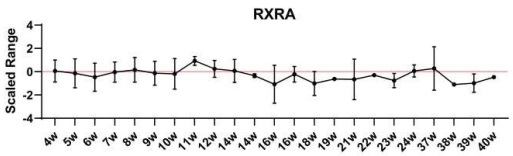


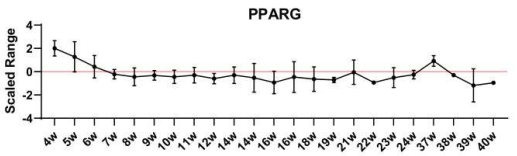












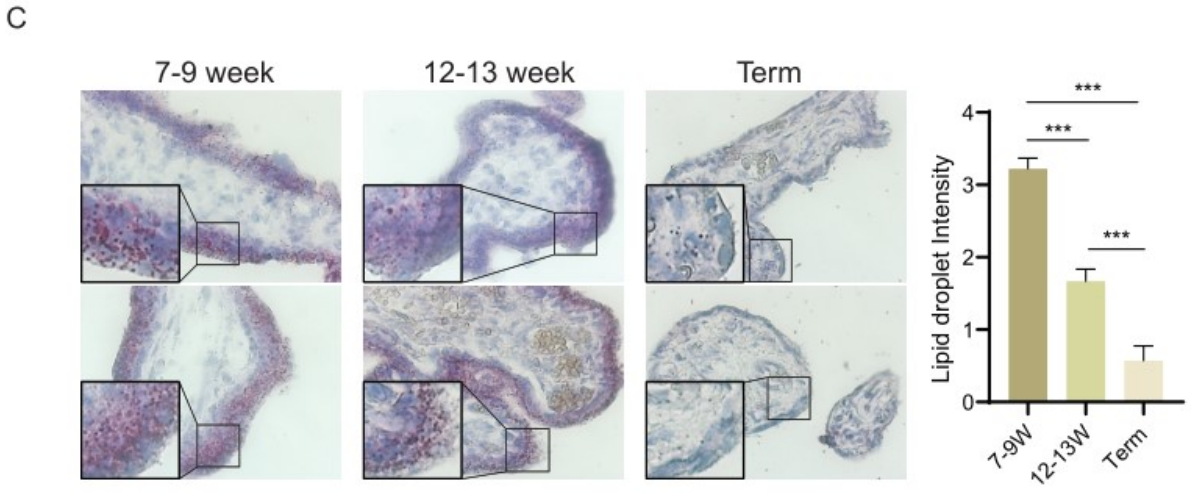


Figure 20 PPAR signaling pathway mapping and lipid droplets detection

Fig. 5. PPAR signaling pathway mapping and lipid droplets detection. Genes involved in PPAR signaling pathway map in dataset1 and dataset2&3 showing in red text and rectangle (A) using clusterProfiler. The scaled and adjusted expression of involved genes responding to gestational weeks are presented in line graph (B). Lipid droplets were detected with Oil-red O staining in early (7-9w, n = 9) and late (12-13w, n = 9) first trimester and term human placenta (n = 5) tissues and graphed under 40× magnification, with the comparison aside(C). PPAR: peroxisome proliferator-activated receptor; RXR: retinoid X receptor. ***: $p < 0.0001$.

Table 1. Result of ordered logistic regression analysis

Table 2 Result of ordered logistic regression analysis

Factor	Coefficient	Standard Error	t	P>|t|	[95% Conf. Interval]	
Gestational Age	**-576.0346**	**104.1833**	**-5.53**	**0.000**	**-803.0306**	**-349.0386**

Sex	-292.2495	439.55	-0.66	0.519	-1249.947	665.4477
Smoking	-279.8577	426.1765	-0.66	0.524	-1208.417	648.7012
Constant Term	8220.114	1109.687	7.41	0.000	5802.314	10637.91

2.1.5 Discussion

In the past decades, the DNA microarray experiments were generally performed based on different types of tissues or cells, as well as for human placenta. However, most of the research solely performed the sequencing on a specific but short period, such as pre-eclampsia development in the first trimester [182], dose response of placenta to maternal choline intake in the third trimester[183], effects of PPARγ-agonist rosiglitazone on cytotrophoblasts in the first trimester and term [157, 184], comparison of gene expression profiles of the first (45-59 days) and second trimester (109-115 days) and term [185]. Different to the preceding studies, our research gave an overview of the human placental genome-wide expression from the very beginning gestational week 4 to the term 40 by integrating the microarray data published by Soncin et .al [177] and Winn et al. [178].

Moreover, other than comparing different static stages with t-test, as the overall design of Winn et al.[178] in mid-trimester and term, we considered the dynamic change of gene expression along with the gestational weeks. Dynamic data, as Bar-Joseph [186] mentioned, indicates a strong autocorrelation between successive points, rather than

independent identically distributed samples in static data. Comparison from a static period to another might distort the actual role of a biological process serving different specialized functions in different stage [187]. For example, the surge of the oxygen tension in the placental villi only occurs in a narrow span during 10-12 weeks' gestation [59]. Therefore, the generalized linear regression was implemented, and the adjusted P-value associated with each gene was used as an indicator for is significance related to the response variable (gestational age). This statistic method helped to select more than 4000 significant probe sets for dataset2&3, far more than 505 in Winn's study. It seems reasonable to have such huge number of SEGs during pregnancy since the structure and functions of human placenta should have adapted to a lot of dramatical modifications.

Additionally, we considered the pairwise correlations between genes and co-expression gene sets, as well as the network topology of different networks as reported frequently [188-190]. Therefore, we used WGCNA with the default unsigned network to explore the whole genes in dataset1 and those selected genes in dataset2&3, which are correlated to gestational weeks. WGCNA is a method for construct correlation network, especially for analyzing large, high-dimensional data sets like DNA microarray[180]. As the WGCNA package noted, its application to differential expression is not recommended since it will invalidate the scale-free topology assumption. In our case, we applied it successfully in filtered SEGs which possibly because of the way they were defined by linear regression rather than direct comparison between groups. After all, with this combined method, we narrowed forward the scale of SEGs to 5173 in different period of placentation, compared to hundreds of differentially expression genes identified previously. Subsequently, these SEGs were submitted for further enrichment and we compared our top terms of the biological process, molecular function and

pathway with the original research [178], which indicates more detailed and abundant terms (Supplementary Material Figure S2).

With respect to the terms, our results indicated the involvement of "PI3K-Akt signaling pathway", which has reported to participate in the decidualization of trophoblast during early pregnancy[191]. As well, "Ras signaling pathway" has been reported to control the trophoblast stem cell survival by regulating the phosphorylation and destabilization of proapoptotic proteins[192]; "Rap1 signaling pathway" and "cAMP signaling pathway" regulate the placental cell fusion[193, 194]. Among the intersection of the top enriched terms in biological process and pathway, we selected "PPAR signaling pathway" for the further investigation since our team is devoted to studying the role of PPARs in human placental development.

According to the PPAR signaling pathway map, the up- and down-stream genes of PPARs nuclear receptors, such as LPL, SLC27A2, ACSL1, FABPs, were well matched. We therefore enclosed these genes and PPARs and its heterodimer RXRα for graphing the time-guided lines. These line graphs drawn with the scaled and adjusted gene expression dynamically displayed the relative changes across the placentation. Our results show a continuous increase in LPL, FABP4, FABP5, and a continuous decrease in FABP7, ACSL1. As it reports, PPARs, as nuclear receptors, regulate gene expression through binding to PPAR response element, with which the target gene transcription will be promoted or inhibited [195]. In response to the knowledge, our previous study investigated the effects of rosiglitazone on trophoblast. Trophoblast treated with PPARγ agonist rosiglitazone have a significant up-regulation in FABP4, FABP5, and down-regulation in LPL, but no significance was shown in the rest [157], which is consistent with the performance of the line graph, indicating the critical role of PPARγ

in lipid metabolism during placental development. Indeed, in mouse, the PPARγ deletion leads to the absence of adipose tissue resulting in a series of metabolic phenotype related to dysfunctional lipid metabolism like hypermetabolism, hyperphagia *etc.* [196]. Previous enrichment analysis also shows the susceptibility of PPAR signaling pathway in first trimester placenta, related to lipid metabolism and other complex biological functions[197]. Therefore, we detected the neutral lipid droplets by ORO staining in first trimester and term placenta tissues, whose quantity commonly applied to indicate the expression of LPL, FABPs and PPARs [162, 198, 199]. Our results show the apparent decrease of lipid droplets from first trimester to term. Meanwhile, the negative correlation of first trimester gestational age and lipid droplets intensity was revealed by the ordered logistic regression statistic method. The role of "PPAR signaling pathway" in lipid metabolism throughout placentation hence suggests the reliability of our analysis.

Last but not least, in our study, the decrease of lipid droplets throughout the gestation age seems own to up-regulation of FABPs which may further affect the activation of PPARγ. The activation of PPARγ up-regulates the expression of lipid metabolic genes such as LPL which can increase the metabolism of lipid droplets in cytotrophoblasts. However, other studies report the increase of FABPs expression resulted from the activation of PPARγ in *in vitro* cultured cytotrophoblasts, which results in lipids uptake and accumulation [200, 201]. Here, the paradoxical effect observed in our study and others might be explained by the fact that FABP4 could also downregulate PPARγ activity, as observed in adipocyte through ubiquitination and subsequent proteasomal degradation [202]. Therefore, we could hypothesize that there might be an interestingly dynamic regulation between FABPs and PPARγ activation which might lead to the similar function of FABP4 acting in term placenta as in adipocyte. Taking

for example, FABP4 is upregulated in case of preeclampsia, a major pregnancy-related disease [203].

To conclude, using new methods for the analysis of the global gene expression on human placenta from 4 to 40 gestational weeks, our study provides more significant genes, more information on biological processes and pathways, giving us new clues for deciphering the normal functions of placentation. Their mis-regulations may be linked to pregnancy-related diseases. Finally, we give an example that PPAR signaling pathway mediates constant decrease of placental lipids throughout pregnancy.

Declaration of competing interest

The authors declare that they have no competing interests.

Acknowledgments

No

Supplementary Materials

The number of genes from unsigned and signed network contained in the original color modules were shown in Table S1. Completed results of biological process and pathway analysis for differentially expressed genes were deposited in Table S2 with four sheets inside. Heatmap of the gene co-expression values was performed to show the strength of the associations in Figure S1. A comparison of the top terms of biological process,

molecular function and pathway between the original research and our present study was provided in Figure S2. Characteristics of placenta used for logistic regression were described as Table S3.

Funding Statement

This work was carried out with the funding support of the China Scholarship Council (CSC) in Chegongzhuang Avenue, Beijing 100044, P.R China, as well as Inserm and Université de Paris.

Part II

3 Part Two

Lead-in

In the last part, we have explored the important biological processes happening across the whole pregnancy, in which the PPAR signaling pathway was discovered. In the matched pathway map, PPARγ was found to be the key gene to regulate the downstream genes in the PPAR signaling pathway. Since PPARγ is the key gene for lipid metabolism and PPAR signaling pathway was the only pathway involved in the lipid metabolism in our enrichment, we detected the lipid droplet in the human placenta to verify the existence of PPARγ. The result confirmed the role of PPARγ in our study, but we found that the gene expression of PPARγ remained unchanged across the pregnancy. If it's not the changing expression of genes or pathways, we proposed that PPARγ, as a nuclear receptor, must function in the form of protein, via activity change, to regulate the target genes, instead of changing its gene expression. In this part, we aimed to study the activity of PPARγ protein in the human placenta. The placenta is composed of various materials, such as trophoblasts, stroma, *etc.*, while trophoblasts are the main element. The trophoblasts can be divided into villous cytotrophoblast and extravillous cytotrophoblast that are the most important cells in the human placenta. If the PPARγ is going to function in the human placenta, we believe the two cell types would play a key role in the process. We, therefore, decided to study the role of PPARγ protein in these two cell types and to find out the main genes or pathways that would have been regulated by PPARγ in the long period.

3.1 Comparative Study of PPARγ Targets in Human Extravillous and Villous Cytotrophoblasts

Fulin Liu [1], Christine Rouault [2, 3], Mickael Guesnon [1], Wencan Zhu [5], Karine Clément [2, 3, 4], Séverine A. Degrelle [1, 2, 6, #], Thierry Fournier [1, 2, #]

[1] Université de Paris, INSERM, UMR-S1139 "Pathophysiology & Pharmacotoxicology of the Human Placenta, pre & postnatal Microbiota" (3PHM), Paris F-75006, France

[2] Fondation PremUp, Paris, F-75006, France

[3] Sorbonne Université, INSERM, « Nutrition et Obesités : approches systémiques research unit », Paris, F-75013, France

[4] Assistance Publique-Hôpitaux de Paris, Nutrition department, Pitié-Salpêtrière hospital, Paris, F-75013 France

[5] UMR Applied Mathematics & Informatics, AgroParisTech-Université Paris-Saclay, 75005, Paris, France

[6] Inovarion, Paris F-75005, France

[#] These authors contributed equally to this work.

Correspondence should be addressed to Thierry Fournier; thierry.fournier@parisdescartes.fr

3.1.1 Abstract

Trophoblasts, as the cells that make up the main part of the placenta, undergo cell differentiation processes such as invasion, migration, and fusion. Abnormalities in these processes can lead to a series of gestational diseases whose underlying mechanisms are still unclear. One protein that has proven to be essential in placentation is the peroxisome proliferator-activated receptor γ (PPARγ), which is expressed in the nuclei of extravillous cytotrophoblasts (EVCTs) in the first trimester and villous cytotrophoblasts (VCTs) throughout pregnancy. Here, we aimed to explore the genome-wide effects of PPARγ on EVCTs and VCTs via treatment with the PPARγ-agonist rosiglitazone. EVCTs and VCTs were purified from human chorionic villi, cultured *in vitro*, and treated with rosiglitazone. The transcriptomes of both types of cells were then quantified using microarray profiling. Differentially expressed genes (DEGs) were filtered and submitted for gene ontology (GO) annotation and pathway analysis with ClueGO. The online tool STRING was used to predict PPARγ and DEG protein interactions, while iRegulon was used to predict the binding sites for PPARγ and DEG promoters. GO and pathway terms were compared between EVCTs and VCTs with ClusterProfiler. Visualizations were prepared in Cytoscape. From our microarray data, 139 DEGs were detected in rosiglitazone-treated EVCTs (RT-EVCTs) and 197 DEGs in rosiglitazone-treated VCTs (RT-VCTs). Downstream annotation analysis revealed the similarities and differences between RT-EVCTs and RT-VCTs with respect to the biological processes, molecular functions, cellular components, and KEGG pathways affected by the treatment, as well as predicted binding sites for both protein-protein interactions and transcription factor–target gene interactions. These results provide a broad perspective of PPARγ-activated processes in trophoblasts; further analysis of the transcriptomic signatures of RT-EVCTs and RT-VCTs should open new

avenues for future research and contribute to the discovery of possible drug-targeted genes or pathways in the human placenta.

Key Words: Placenta; Cytotrophoblast; Extravillous; Peroxisome proliferator-activated receptor-γ; Rosiglitazone; Microarray

3.1.2 Introduction

The human placenta serves as a critical bridge between mother and fetus, and thus plays a crucial role in maternal and fetal physiology. The placenta is composed mainly of trophoblast cells, which derive from the outer layer of the blastocyst. Certain trophoblasts can be further distinguished as villous cytotrophoblasts (VCTs), whose development progresses along with that of the placenta. In the process of embryo implantation and placenta formation, VCTs that invade the maternal uterus are known as extravillous cytotrophoblasts (EVCTs); these anchor the chorionic villi. Other VCTs differentiate and fuse to form the syncytiotrophoblast layer, which has critical functions in gas- and nutrient-exchange between the fetus and the mother. Defects in EVCT invasion and VCT differentiation and fusion contribute to a series of gestational diseases, such as fetus-related miscarriage [204], preterm birth [205], and pre-eclampsia [206]. The causes of and mechanisms behind these diseases have been the focus of much research, but as yet remain unclear.

As a member of the ligand-dependant nuclear receptor superfamily, PPARγ regulates many downstream target genes involved in lipid metabolism, cell differentiation, and tumorigenesis. PPARγ functions by forming a heterodimer with the nuclear receptor retinoid X receptor α (RXRα) and then binding to the PPAR response element (PPRE) of target genes [207]. It has been reported that a lack of PPARγ leads to defects in trophoblast differentiation and abnormal vasculogenesis in mice [147, 166], and PPARγ$^{-/-}$ embryonic lethality can be rescued via PPARγ

transfection in the trophoblast [167]. PPARγ thus appears to play a crucial but poorly understood role in placental development.

To explore the role of PPARγ in biological processes, the PPARγ-agonist rosiglitazone has been widely applied to various tissues. In human placenta, rosiglitazone has been used for the study of placental metabolism [208, 209], inflammation [210, 211], antioxidant response [212, 213], and pre-eclampsia [149]. *In vitro* treatment with rosiglitazone has been shown to reverse inflammation of the placenta that is mediated by the PPARγ-NF-κB pathway [210]. Similarly, rosiglitazone can improve the survival rate of trophoblasts under oxidative stress via its effects on the PPARγ pathway [212]. Other investigations into the activity of this drug have identified new potential target genes of PPARγ [149, 214]. Taken together, these studies show the enormous potential and benefit of rosiglitazone use in studies of the placenta.

In the human placenta, PPARγ is exclusively located in the nuclei of EVCTs during the first trimester and of VCTs throughout pregnancy [195, 215, 216]. To date, there is a lack of systematic research on the effects of PPARγ in these tissues and during these developmental periods. Our purpose here was to investigate the performance of PPARγ-activated trophoblasts by analyzing the transcriptomic signatures of rosiglitazone-treated EVCTs (RT-EVCTs) and VCTs (RT-VCTs), with the ultimate goal of identifying potential target genes or pathways and providing a broad knowledge base for future research. Specifically, we isolated EVCTs and VCTs from first-trimester and term human chorionic villi, respectively, cultured these cells with rosiglitazone, and quantified the transcriptome of each type of cell using microarray analysis. The procedure for this study is shown in Figure 1.

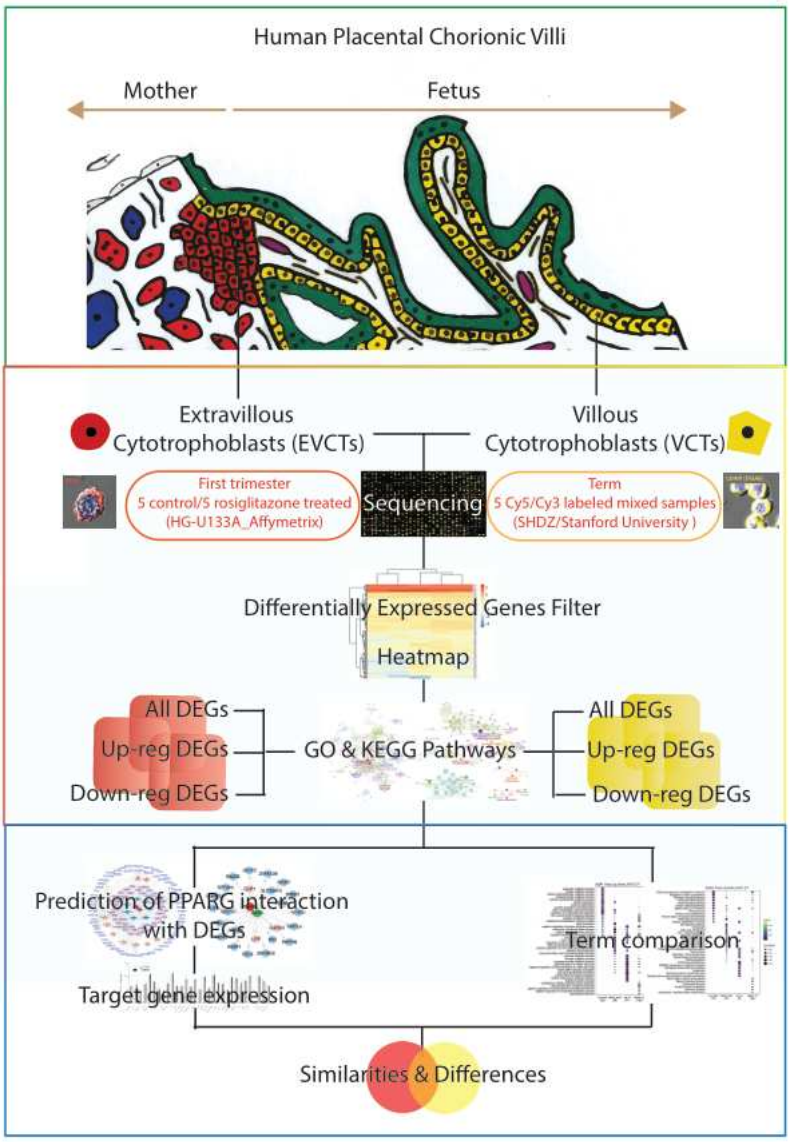


Figure 21 Summary of procedures

Fig. 1. Summary of procedures. Extravillous cytotrophoblasts (EVCTs) and villous cytotrophoblasts (VCTs) were isolated from first-trimester and term placentas, respectively, treated with rosiglitazone, and analyzed using microarrays. Differentially expressed genes (DEGs) were filtered for quality control and submitted for annotation. Terms associated with DEGs and predictions of PPARγ-target genes were compared between the rosiglitazone-treated EVCTs and VCTs. PPARγ: peroxisome proliferator-activated receptor γ.

3.1.3 Materials and Methods

Ethics statement

Placenta samples in this study were collected with patients' written informed consent, in compliance with the Declaration of Helsinki. Placenta tissues were collected from women with normal pregnancies during the 8-9th gestational weeks and at term (39 gestational weeks). Our ethics committee (CCPRB Paris Cochin n° 18-05) approved the collection of placentas from legal and voluntary terminations of pregnancy in the first trimester as well as of the normal term placentas.

Cell isolation and culture

As previously described [157], five effective first-trimester placentas were obtained for EVCT isolation. Villous tissues were rinsed and minced in Ca^{2+}-, Mg^{2+}-free Hanks' balanced salt solution for membrane removal. Mononucleated VCTs were isolated using digestion with trypsin-DNase and fractionation on a discontinuous Percoll gradient according to the protocol of Kliman et al. [217] and Alsat et al. [218]. In brief, villous tissues were digested in Hanks' balanced salt solution, containing 5 IU of DNase I per ml, 4.2 mM $MgSO_4$, 0.25% (wt/vol) trypsin powder (Difco), 100 IU/mL penicillin, 25 mM HEPES, and 100 μg/mL streptomycin (Biochemical Industry), and monitored under invert microscopy. The initial digested solution (consisting mostly of red blood cells) was discarded while the subsequent digested solution (clearly consisting of EVCTs) was retained for stratification. A discontinuous Percoll gradient (5–70% in 5% steps) was used to stratify the digested solutions; the middle layer (which included EVCTs) was retained for further analysis. The purified EVCTs were diluted with Dulbecco's modified Eagle's medium (DMEM), with 2 mM glutamine, 100 IU/mL penicillin, 100 mg/mL streptomycin and 10% decomplemented fetal calf serum

(FCS), to a final density of 0.9 × 10^6 cells/mL in 60-mm-diameter plastic tissue culture dishes (TPP). In preparation for culturing, culture plates (Techno Plastic Products, Switzerland) were coated with Matrigel™ (7 μg/cm^2; Collaborative Biomedical Products, Le Pont de Claix, France), then seeded with EVCTs at a density of 5×10^4 cells/cm^2. To maintain continuous culture conditions, DMEM-F12 medium was used that contained 10% heat-inactivated fetal calf serum (FCS), Glutamax, 100 μg/mL streptomycin, and 100 IU/mL penicillin (Invitrogen). Plates were incubated for 2 h at 37°C and 5% CO_2, then non-adherent EVCTs were rinsed off. At this point, fresh medium that contained either 1 μM rosiglitazone (Cayman) dissolved at 1 mM in ethanol (treatment) or 0.1% ethanol (vehicle) was added for another 24 h of incubation.

VCTs were isolated from five term placentas using the following procedure. Placentas were oriented with the maternal side facing upwards, and tissues were sampled at a depth of 1.5 cm, half the distance from the edge to the centre. Villous tissues were rinsed, minced, digested, and purified using the steps described above. Culture dishes containing 0.9×10^6 cells/mL were placed in a humidified incubator at 37°C under 5% CO_2 for 3 h. Non-adherent VCTs were rinsed off, fresh medium was added that contained either 1 μM rosiglitazone (Cayman) dissolved at 1 mM in ethanol (treatment) or 0.1% ethanol (vehicle), and dishes were incubated for another 24 h.

Microarray experiments

After 24 h of incubation, RT-EVCTs and control EVCTs were harvested for microarray experiments. Cell RNA was extracted using TRIzol® reagents (Invitrogen) and purified using RNeasy® Mini Kits (Qiagen). RNA integrity and purity were examined with a 2100 Bioanalyzer with the RNA 6000 LabChip kit (Agilent Technologies). The U133A 2.0 GeneChip (Affymetrix, Inc.) was used for gene

expression detection according to the method outlined in the manufacturer's manual. From the 22 000 probe sets on the gene chip, 14 500 genes were detected.

RT-VCTs and control VCTs were likewise harvested after 24 h of incubation for microarray experiments. VCT RNA extraction, purification, and quality control were performed as described above. The SHDZ gene chip (Stanford University) was used for gene detection as described in [219]: for each sample, the MessageAmp RNA kit (Ambion) was used, with 1 μg total RNA, for RNA amplification, and 3 μg amplified RNA were then labeled with Cy-dye using the 26 CyScribe first-strand cDNA labeling kit (Amersham Biosciences). Amplified RNA from rosiglitazone-treated VCTs was labeled with Cy5, and amplified RNA from control VCTs was labeled with Cy3. A Microcon YM 30 column (Millipore) was used to purify and concentrate the labeled mixture (Cy5 and Cy3) after additional modifications with human cot-1, yeast tRNA, and poly A. The probes were denatured and the mixture was hybridized at 65°C overnight in a sealed humidified hybridization chamber, then rinsed with 1XSSC, 2XSSC, 0.03% SDS, and 0.2% SDS solutions for 2 min each. Arrays were scanned with a GenePrix 4000A microarray scanner (Axon Instruments).

Data processing

Since gene expression in EVCTs and VCTs was detected using different microarray platforms, different procedures were followed for data processing. For EVCT gene expression, which was quantified using the GeneChip (U133A 2.0, Affymetrix) application, data processing used the following filter thresholds: i) percentage of missing data was no more than 50%; ii) threshold to identify up- and downregulated genes for statistical comparison was set to a fold-change of 1.5; and iii) maximum false discovery rate (FDR) was set to 5% [157]. For VCT gene expression,

which was measured using the SHDZ GeneChip/Stanford University (GPL21609) application, data processing used the following filter thresholds: i) background-corrected data were log2-transformed and subjected to the Loess normalization method [11]; ii) differentially expressed genes (DEGs) were determined via the significance analysis of microarrays (SAM) method [220]; and iii) the maximum false discovery rate (FDR) was set to 1%, without a fold-change threshold imposed [221].

GO and pathway enrichment analyses

ClueGo is a Cytoscape plug-in application for the functional classification of genes [222]. Our analysis used Cytoscape version 3.7.1 (The Cytoscape Consortium, New York, NY) and ClueGo version 2.5.4 (released 28 Feb 2019), with the simultaneous update of gene ontology (GO) terms. Using ClueGO, we recovered the GO terms associated with the dataset of all DEGs as well as of up- or downregulated DEGs only; this same application was also used for KEGG & Reactome pathway analysis. GO terms were compared between EVCTs and VCTs using the R package clusterProfiler (version 3.9, synced to latest GO terms and pathways) [181]. For term comparison in clusterProfiler, 10 category terms for each group were selected for inclusion in charts. Instead, ClueGo analyses were based on approximately 30 terms per group in order to generate more detailed visualizations. P-values lower than 0.05 identified significant enrichment.

Protein-Protein Interaction (PPI) network

The STRING database (http://string-db.org) was used to analyze the interactions of DEG-encoded proteins and construct a PPI network. For this, the significant confidence score was set to greater than 0.4. Cytoscape was used to visualize and organize the PPI network. Proteins interacting with PPARγ or RXRα were indicated by

different colors, and shapes were used to represent different groups. Binding site interactions between transcription factors and target genes were predicted by the Cytoscape plug-in iRegulon (based on the TRANSFAC database; version 1.3). Putative regulatory regions were defined as 10 kb around transcription starting sites. The FDR was set to 0.1% to verify the interaction. Cytoscape was used to modify the resulting chart, with red indicating upregulated genes and blue indicating downregulated genes.

3.1.4 Results

1. Gene expression profiling of RT-EVCTs and RT-VCTs

Microarrays were used to characterize gene expression in EVCTs and VCTs with or without rosiglitazone treatment. Our microarray data have been deposited in the Gene Expression Omnibus public repository (https://www.ncbi.nlm.nih.gov/geo/); EVCT microarray data under accession number GSE28426, VCT microarray data under accession number GSE137434). Gene expression profiles of the rosiglitazone-treated (TRT) samples of EVCTs and VCTs were normalized (Figure 1a). Four of the five independent RT-EVCT samples yielded consistent results, with one sample appearing slightly different; instead, all five independent RT-VCT samples yielded similar results. Next, DEGs were detected based on thresholds for both fold-change in expression levels and FDR. In RT-EVCTs, a total of 139 genes were identified as DEGs ($p < 0.05$), of which 114 genes were upregulated (red) and 25 genes were downregulated (blue). In RT-VCTs, a total of 197 genes were identified as DEGs ($p < 0.05$), of which 181 genes were upregulated (red) and 16 genes were downregulated (blue) (Figure 2b).

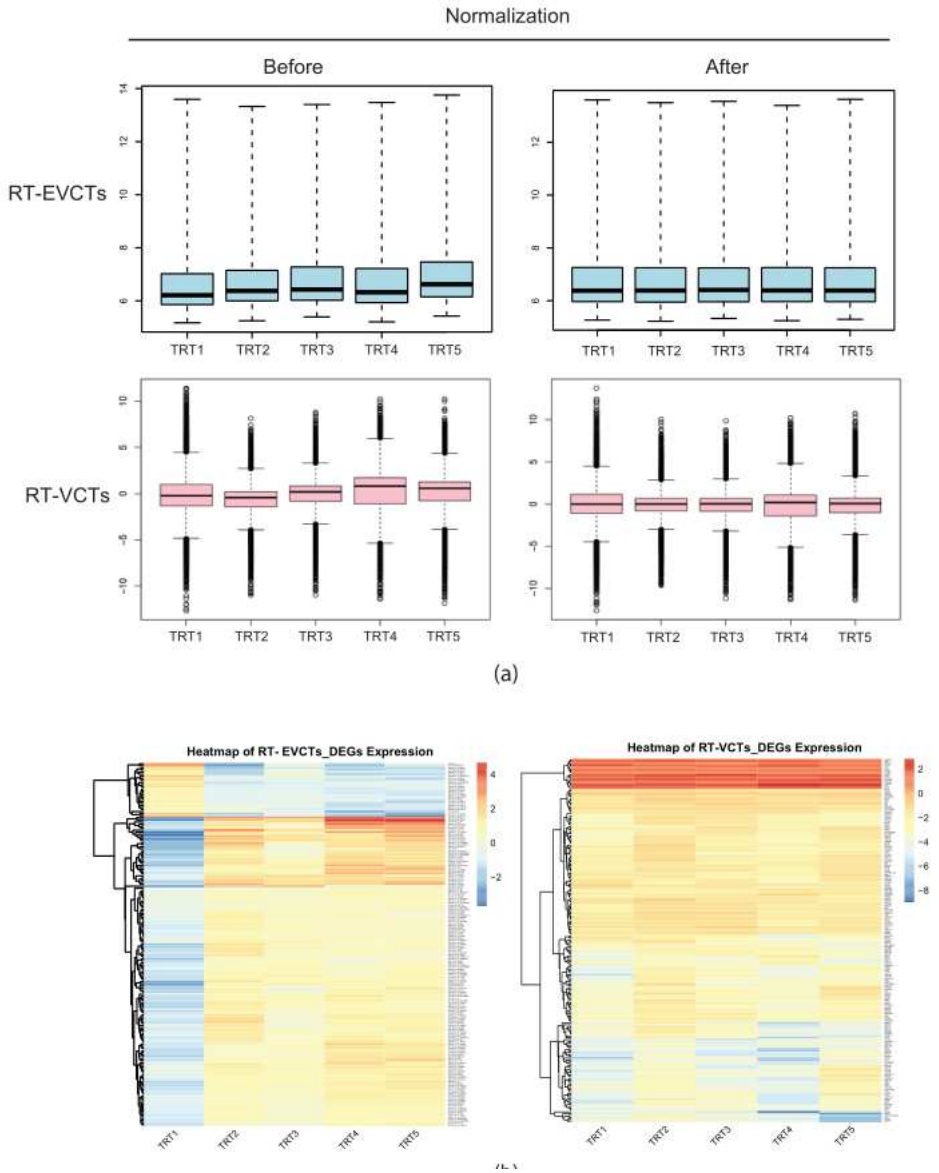


Figure 22 Microarray data normalization and DEG heatmap of RT-EVCTs and RT-VCTs

Fig. 2. Microarray data normalization and DEG heatmap of RT-EVCTs and RT-VCTs. (a) RT-EVCT gene expression microarray was performed with the Affymetrix GeneChip while the RT-VCT microarray used the SHDZ/Stanford University chip. DEGs were detected based on the thresholds of 1.5-fold change and 5% FDR for the RT-EVCT microarray matrix; a threshold of 1% FDR was applied for the RT-VCT microarray matrix. The Loess normalization method was used to normalize both datasets. Box plots represent microarray data before and after normalization, with blue indicating data from RT-EVCTs and pink data from RT-VCTs. (b) Heatmaps of five

independent samples of RT-EVCTs and RT-VCTs. Upregulated DEGs are represented in red; downregulated DEGs in blue. DEGs: differentially expressed genes; RT-EVCTs: rosiglitazone-treated extravillous cytotrophoblasts; RT-VCTs: rosiglitazone-treated villous cytotrophoblasts; FDR: false discovery rate; TRT: treated.

2. Gene ontology and pathway terms of all DEGs from RT-EVCTs and RT-VCTs

The entire set of DEGs from RT-EVCTs and RT-VCTs were separated by cell-type of origin and submitted independently to ClueGO with the default parameters. GO and pathway enrichment were set up for analysis. DEGs were classified three ways: by GO biological process, GO molecular function, and GO cellular component. Enriched pathways were identified through a search of the KEGG and Reactome databases. The results are visualized in Figure 3.

Among the DEGs identified in RT-EVCTs, the main GO biological processes represented were “negative regulation of epithelial cell apoptotic process”, “long-chain fatty-acyl-CoA biosynthetic process”, and “phosphatidylcholine biosynthetic process”. For the same group of DEGs, the GO molecular functions were mainly classified as “alpha-tubulin binding”, “wide pore channel activity”, “positive regulation of cold-induced thermogenesis”, “glutathione transferase activity”, “long-chain fatty acid binding”, “regulation of cell adhesion mediated by integrin”, and “positive regulation of non-motile cilium assembly”. Finally, the GO cellular component that was most associated with these DEGs was “desmosome”. In the pathway enrichment analysis of RT-EVCTs, DEGs were mainly associated with the terms “HIF-1 signaling pathway”, “p53 signaling pathway”, “glutathione metabolism”, “NRAGE signals death through

JNK", "PPAR signaling pathway", "plasma lipoprotein assembly", and "remodeling and clearance".

In the analysis of GO terms associated with the RT-VCT dataset, DEGs were mainly involved in the following biological processes: "regulation of receptor biosynthetic process", "negative regulation of nucleotide metabolic process", "cyclic nucleotide biosynthetic process", and "negative regulation of B cell apoptotic process". The molecular functions of this same group of DEGs were mainly linked to "negative regulation of DNA replication", "regulation of protein deacetylation", "ubiquitin-like protein conjugating enzyme activity", "Hsp90 protein binding", "negative regulation of intracellular protein transport", and "positive regulation of phosphoprotein". With respect to GO cellular components, DEGs were mainly associated with the terms "NuRD complex", "cellular metabolic compound salvage", and "immunological synapse". Finally, the pathway enrichment analysis of RT-VCTs revealed that DEGs were mainly involved in "tight junction", "regulation of HIF by oxygen", "unfolded protein response", "HIF-1 signaling pathway", "nuclear receptor transcription pathway", and "plasma lipoprotein remodeling".

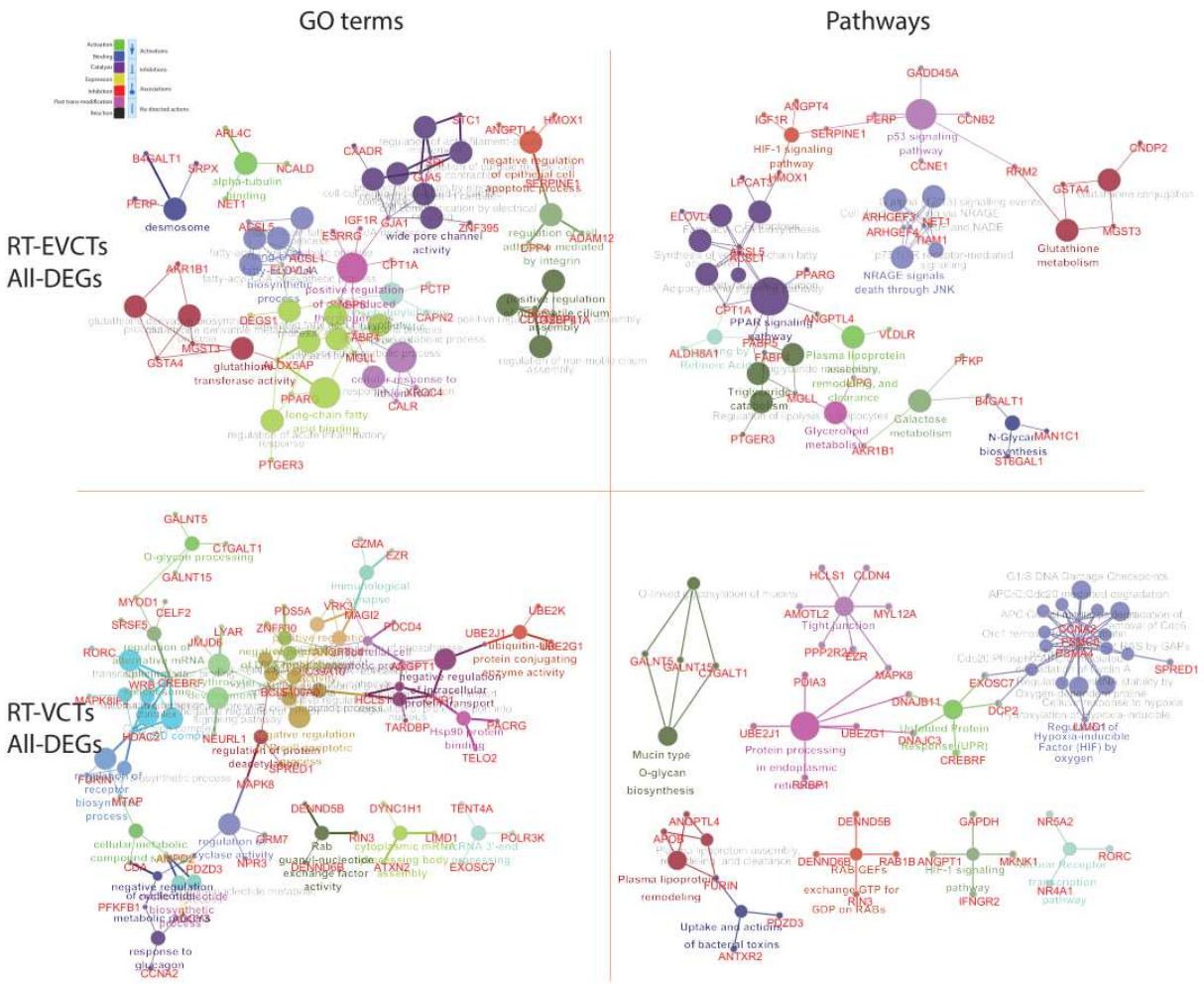


Figure 23 GO and pathway terms associated with all DEGs in RT-EVCTs and RT-VCTs

Fig. 3. GO and pathway terms associated with all DEGs in RT-EVCTs and RT-VCTs. All DEGs were submitted separately according to their cell-type of origin to ClueGO with the default parameters. GO and pathway enrichment were set up for analysis. DEGs were classified three ways: by GO biological process, GO molecular function, and GO cellular component. The KEGG&Reactome database was consulted to determine pathway enrichment. An exhaustive list of all terms (including those not shown above) can be found in supplementary materials (Tables S1-S8). DEGs: differentially expressed genes; RT-EVCTs: rosiglitazone-treated extravillous cytotrophoblasts; RT-VCTs: rosiglitazone-treated villous cytotrophoblasts; GO: gene ontology; KEGG: Kyoto encyclopedia of genes and genomes.

3. GO and pathway terms associated with upregulated DEGs in RT-EVCTs and RT-VCTs

DEGs that were upregulated in RT-EVCTs and RT-VCTs were submitted separately to ClueGO following the same procedure as described above. The results are visualized in Figure 4. In RT-EVCTs, upregulated DEGs were mainly associated with the GO biological processes “fatty acid derivative biosynthetic process” and “negative regulation of epithelial cell apoptotic process”, and the GO molecular functions “positive regulation of insulin secretion”, “temperature homeostasis”, “wide pore channel activity”, “nuclear receptor activity”, and “regulation of plasma lipoprotein particles levels”. The main GO cellular components implicated in the activity of these DEGs were “desmosome” and “intrinsic component of mitochondrial membrane”. Finally, the pathway enrichment analysis indicated that upregulated DEGs in RT-EVCTs were mainly involved in the “p53 signaling pathway”, “HIF-1 signaling pathway”, “peptide hormone metabolism”, “PPAR signaling pathway”, “p57 NTR receptor-mediated signaling”, and “signaling by retinoic acid”.

Instead, from the DEGs that were upregulated in RT-VCTs, no significant GO biological process was identified. In the classification of GO molecular functions, these DEGs were mainly linked with “positive regulation of cell cycle” and “mitotic DNA damage checkpoint”, and the most significant GO cellular component was “transcription factor complex”. In the pathway enrichment analysis, upregulated DEGs in RT-VCTs were mainly associated with the terms “mTOR signaling pathway”, “cell cycle checkpoints”, “DNA repair”, “developmental biology”, “metabolism”, and “vesicle-mediated transport”.

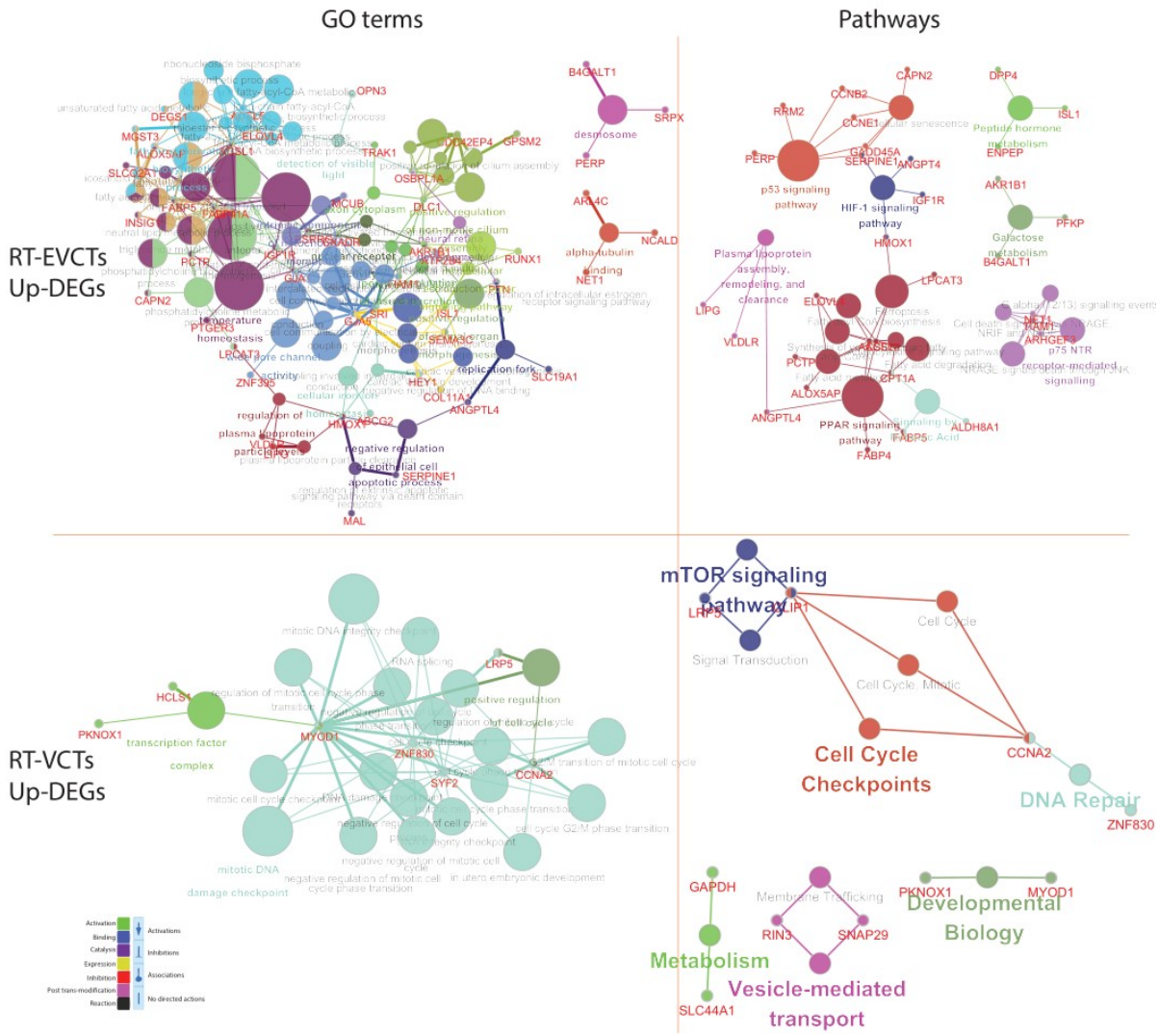


Figure 24 GO and pathway terms associated with DEGs that were upregulated in RT-EVCTs and RT-VCTs

Fig. 4. GO and pathway terms associated with DEGs that were upregulated in RT-EVCTs and RT-VCTs. These DEGs were submitted separately to ClueGO by their cell-type of origin with the default parameters. GO and pathway enrichment were set up for analysis. Upregulated DEGs were classified three ways: by their GO biological process, GO molecular function, and GO cellular component. The KEGG&Reactome databases were consulted to determine pathway enrichment. All additional enrichment terms (not shown above) can be found in supplementary materials (Tables S1-S8). DEGs: differentially expressed genes; RT-EVCTs: rosiglitazone-treated extravillous

cytotrophoblasts; RT-VCTs: rosiglitazone-treated villous cytotrophoblasts; GO: gene ontology; KEGG: Kyoto encyclopedia of genes and genomes.

4. GO and pathway terms associated with downregulated DEGs in RT-EVCTs and RT-VCTs

DEGs that were downregulated in RT-EVCTs and RT-VCTs with respect to controls were submitted to ClueGO using the same procedure as described above. Results are visualized in Figure 5. In RT-EVCTs, downregulated DEGs were mainly associated with the GO biological process "positive regulation of small molecular metabolic process"; the GO molecular functions "membrane fusion", "regulation of epithelial cell migration", "response to estriol", and "protein kinase binding"; and the GO cellular component "phosphorylase kinase complex". From the analysis of pathway enrichment based on the KEGG&Reactome database, downregulated DEGs in RT-EVCTs appeared to be mainly associated with pathways linked with "glycogen breakdown", "influenza infection", "protein processing in endoplasmic reticulum", and "regulation of actin cytoskeleton".

Instead, DEGs that were downregulated in RT-VCTs were mainly involved in the GO biological processes "cyclic nucleotide biosynthetic process", "negative regulation of nucleotide metabolic process", "ncRNA 3'-end processing", and "O-glycan processing"; the GO molecular functions "nuclear receptor activity", "histone deacetylation", "regulation of TOR signaling", "Hsp90 protein binding", "ion channel regulator activity", "nuclear envelope organization", and "peptidyl-threonine modification"; and the GO cellular component "organellar ribosome". In the pathway enrichment analysis of RT-VCTs, downregulated DEGs were mainly associated with the "HIF-1 signaling pathway", "transfer of ubiquitin from E1 to E3", "cell-cell

communication", "transcription regulation of RUNX3", and "formation of NR-MED1 coactivator complex".

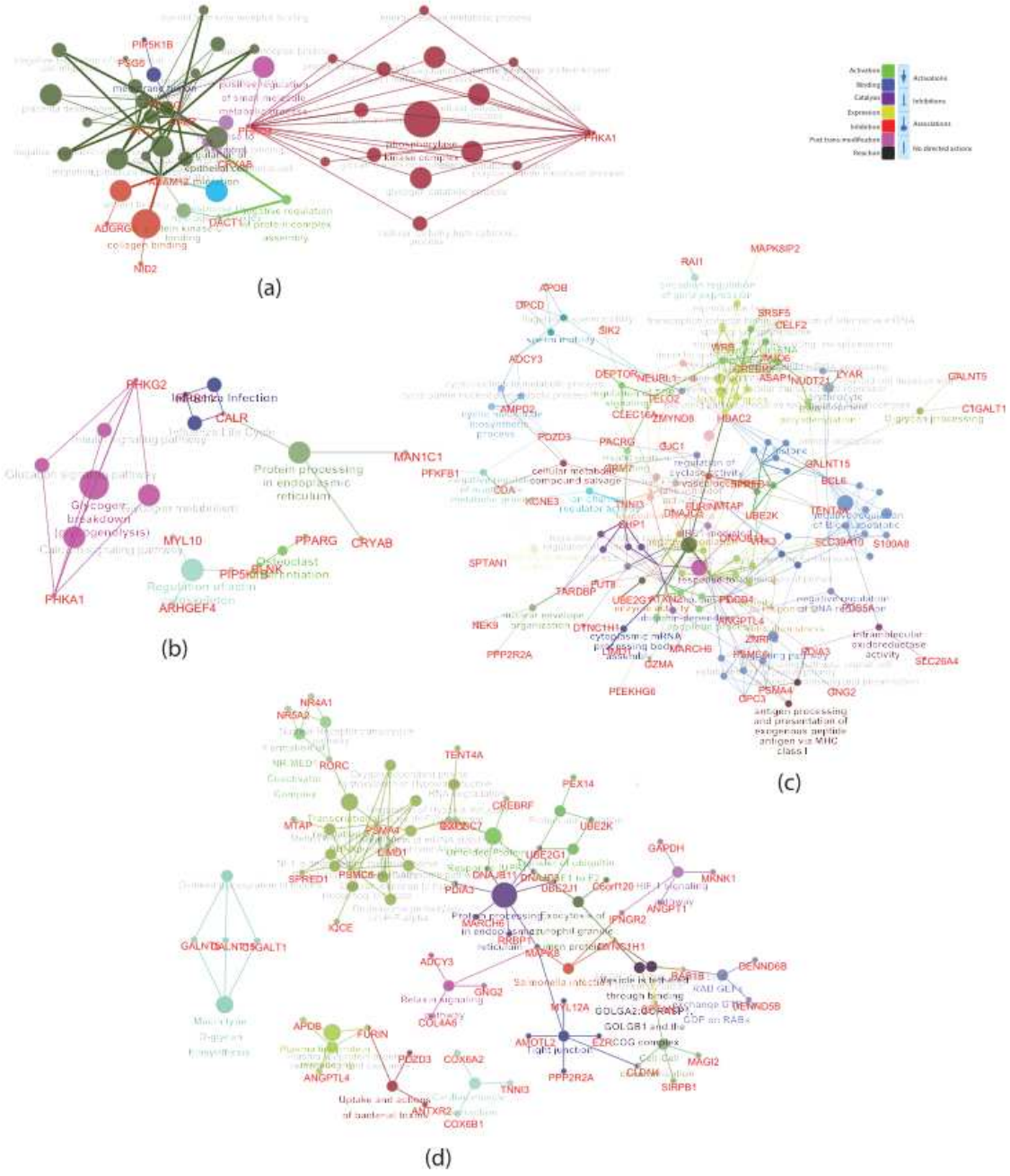


Figure 25 GO and pathway terms associated with DEGs that were downregulated in RT-EVCTs and RT-VCTs, respectively

Fig. 5. GO and pathway terms associated with DEGs that were downregulated in RT-EVCTs and RT-VCTs, respectively, compared to controls. Downregulated DEGs of RT-EVCTs (a, b) and RT-VCTs (c, d) were submitted to ClueGO separately, with the default

parameters. GO and pathway enrichment were set up for analysis. DEGs were classified three ways: by GO biological process, GO molecular function, and GO cellular component (a, c). Pathway enrichment was determined via comparison with the KEGG database (b, d). (a) The GO terms most-associated with DEGs that were downregulated in RT-EVCTs. (b) The pathways that were most enriched among the downregulated DEGs in RT-EVCTs. (c) The GO terms most-associated with DEGs that were downregulated in RT-VCTs. (d) The pathways that were most enriched among the downregulated DEGs in RT-VCTs. An exhaustive list of all associated terms (including those not pictured above) can be found in supplementary materials (Tables S1-S8). DEGs: differentially expressed genes; RT-EVCTs: rosiglitazone-treated extravillous cytotrophoblasts; RT-VCTs: rosiglitazone-treated villous cytotrophoblasts; GO: gene ontology; KEGG: Kyoto encyclopedia of genes and genomes.

5. Comparison of GO terms associated with tissue-specific or tissue-generalist DEGs

Next, we wanted to determine the extent to which the cellular processes affected by rosiglitazone treatment were specific to either EVCTs or VCTs, and which instead were present in both tissue types. To do this, we characterized the up- and downregulated DEGs of RT-EVCTs and RT-VCTs separately using clusterProfiler, using information from the GO and KEGG databases, as well as the Disease Ontology (DO) and Disease Gene Network (DisGeNET) databases. Terms appearing in at least three columns were thought important in both, while terms appearing only in the RT-EVCT or RT-VCT dataset were labeled tissue-specific; significance was determined by p-values less than 0.05.

In both RT-EVCTs and RT-VCTs, the GO biological processes “regulation of endothelial cell migration”, “non-canonical Wnt signaling pathway”, “receptor

metabolic process", "negative regulation of protein phosphorylation", and "metabolism process" appeared to play important roles. Instead, processes specific to RT-EVCTs included "glycogen catabolic process", "cellular carbohydrate catabolic process", "embryo implantation", "fatty acid derivative biosynthetic process", and "long-chain fatty-acyl-CoA biosynthetic process", while those specific to RT-VCTs were "cytoplasmic mRNA processing body assembly", "ribonucleoprotein complex biogenesis", "positive regulation of phosphoprotein phosphatase activity", and "negative regulation of nucleotide metabolic process" (Figure 6a).

The GO molecular functions "nuclear hormone receptor binding", "long-chain fatty acid binding", "fatty acid binding", "nuclear activity", and "transcription factor activity" seemed to be important in both RT-EVCTs and RT-VCTs. Functions specific to RT-EVCTs included "steroid hormone receptor binding", "eicosanoid receptor activity", "phosphatidylinositol phosphate kinase activity", and "fatty acid ligase activity", while those specific to RT-VCTs were "Wnt-activated receptor activity", "cyclin-dependent protein kinase activity", "transferase activity", and "ubiquitin-specific protease activity" (Figure 6b).

Both tissue types shared the significant GO cellular components "smooth endoplasmic reticulum", "ruffle", "transcription factor complex", "apical plasma membrane", "lumen", and "cell-cell junction". Instead, the component terms "beta-catenin destruction complex", "M band", "integral component of lumenal side of endoplasmic reticulum membrane", and "A band" were found only in RT-EVCTs, while "spliceosomal complex", "Wnt signalosome", "pronucleus", "microtubule end", and "autophagosome membrane" appeared to be specific to RT-VCTs (Figure 6c).

Through a search of the KEGG database, the following pathways appeared to be important in both tissue types: "protein processing in endoplasmic reticulum", "glucagon signaling pathway", "Epstein-Barr virus infection", "PPAR signaling

pathway", "HIF-1 signaling pathway", "progesterone-mediated oocyte maturation", and "mTOR signaling pathway". Pathway terms specific to RT-EVCTs included "primary immunodeficiency", and "fatty acid metabolism", while those specific to RT-VCTs were linked with "bacterial invasion of epithelial cells", and "parathyroid hormone synthesis, secretion, and action" (Figure 6d).

From the Disease Ontology database, the terms "pre-eclampsia", "HELLP syndrome", "spinocerebellar ataxia", "familial hyperlipidemia", "lipid metabolism disorder", and "musculoskeletal system cancer" were important in both EVCTs and VCTs. Terms specific to RT-EVCTs included "breast benign neoplasm", "thoracic benign neoplasm", "lipomatous cancer", "amyloidosis", and "vein disease", while those specific to RT-VCTs were "alveolar rhabdomyosarcoma", "osteopetrosis", "giant cell tumor", and "germ cell and embryonal cancer" (Figure 6e).

From a search of the DisGeNET database, the terms "pre-eclampsia", "hypertrophic cardiomyopathy", "immunologic deficiency syndromes", "diabetes mellitus", "vascular inflammations", "hematopoietic neoplasms", "non-alcoholic fatty liver disease", "vascular disease", "ischemic cardiomyopathy", and "triploidy syndrome" were significant for both tissue types. Instead, "chronic neutrophilic leukemia", "glycogen storage disease", and "myeloid, chronic, atypical, and BCR-ABL negative leukemia" were specific to RT-EVCTs, and "alport syndrome" and "aggressive non-Hodgkin lymphoma" were specific to RT-VCTs (Figure 6f).

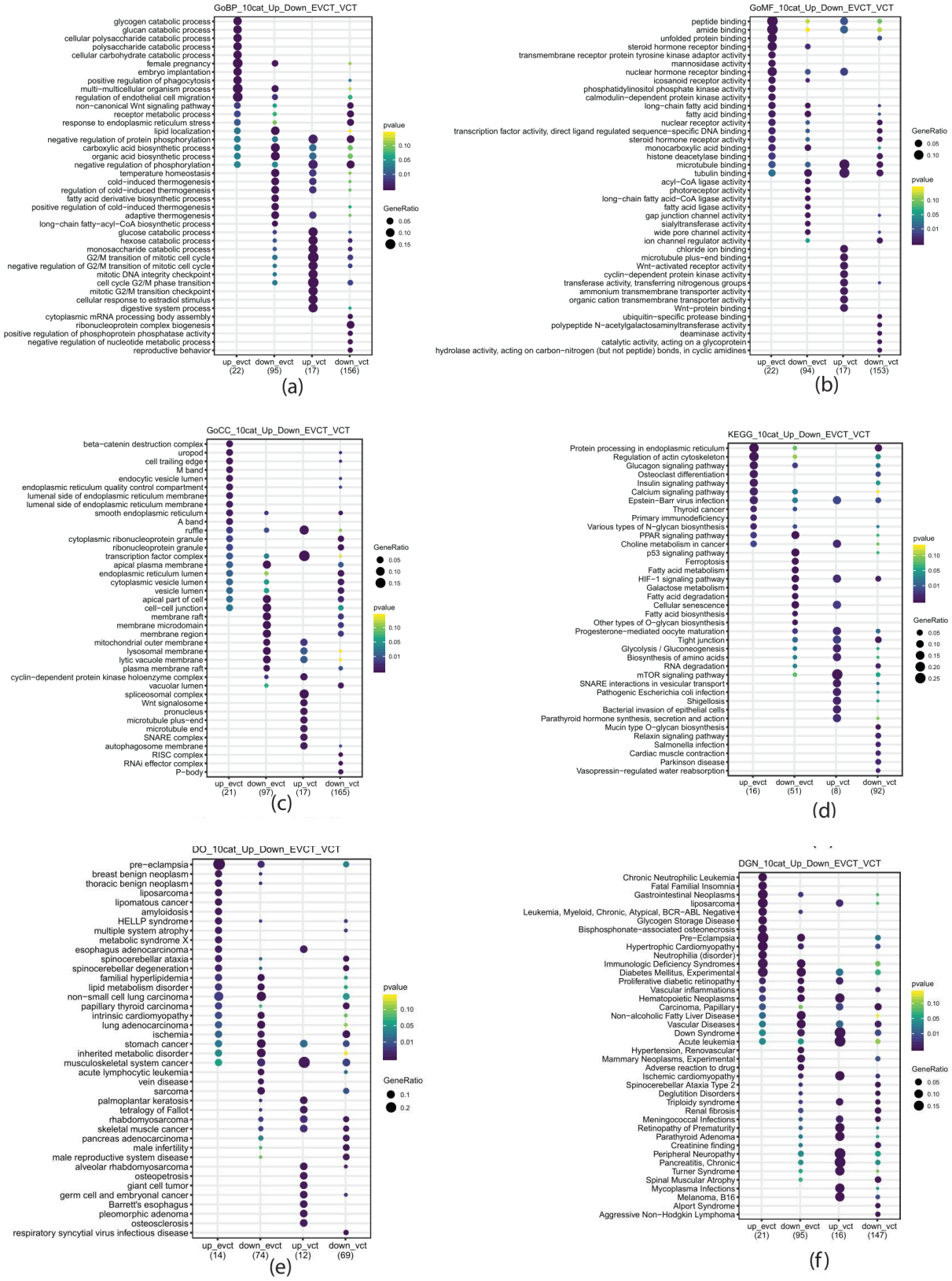


Figure 26 Comparison of enriched GO terms between RT-EVCTs and RT-VCTs

Fig. 6. Comparison of enriched GO terms between RT-EVCTs and RT-VCTs. Up- and downregulated DEGs of RT-EVCTs and RT-VCTs were submitted separately to

analysis in clusterProfiler, for a total of four groups. GO and pathway enrichment were set up for analysis. DEGs were classified by their associated (a) GO biological process, (b) GO molecular function, and (c) GO cellular component. DEGs were further compared with the (d) KEGG database to characterize pathway enrichment, (e) the Disease Ontology (DO) gene set, and (f) the Disease Gene Network (DisGeNET) database. For the purpose of visualization, the top ten categories of enriched terms were included for each gene set. A p-value less than 0.05 determined significance. DEGs: differentially expressed genes; RT-EVCTs: rosiglitazone-treated extravillous cytotrophoblasts; RT-VCTs: rosiglitazone-treated cytotrophoblasts; GO: gene ontology; KEGG: Kyoto encyclopedia of genes and genomes.

6. PPARγ interactions with DEGs of RT-EVCTs and RT-VCTs

Since the gene expression changes we observed here were caused by the activation of PPARγ by rosiglitazone, we next attempted to predict (i) the protein-protein interactions (PPI) of PPARγ with DEG-encoded proteins, and (ii) the transcription factor-target gene (TF-TG) interactions of PPARγ with DEG promoters. In RT-EVCTs (Figure 7, upper left panel), the following proteins appeared to interact directly with the PPARγ complex: MGLL, FABP5, HMOX1, SERPINE1, ABCG2, PHC1, VLDLR, INSIG1, DPP4, ANGPTL4, FAPB4, ACSL1, and CPT1A. Instead, ACSL5, PFKP, AKR1B1, LOX, GXTA4, SOWAHC, GJA1, SLC19A1, RUNX1, PERP, ENPEP, SLFN12, CDC42EP, and LIPG participated in secondary interactions. In RT-VCTs (Figure 7, upper right panel), the PPARγ complex interacted directly with MYOD1, MAPK8, HDAC2, GAPDH, APOB, ANGPTL4, and PDCD4, and secondarily with PDIA3, MAPK8IP2, NR4A1, GNG2, and CCR1. Our analysis of TF-TG interactions in RT-EVCTs (Figure 7, lower left panel) predicted that the target genes of the PPARγ

complex were the upregulated DEGs DLC1, SEMA3C, ARL6IP5, PCTP, ISL1, ZNF395, SR1, DPP4, ALOX5AP, ANGPL4, CDC42EP4, GKN1, ATXN1, CAPN2, LPCAT3, SERPINE1, NET1, LPCAT3, CPT1A, RAB30, GADD45A, MMP19, FHL1, MMD, CCNE1, and ESRRG, as well as the downregulated DEGs ADAM12, GSTA4, PSG5 and DACT1. The same analysis of RT-VCTs (Figure 7, lower right panel) predicted that the target genes of the PPARγ complex were the upregulated DEGs CLIP1, GAPDH, and LPP, as well as the downregulated DEGs CELF2, ZNF512B, SLC39A10, WDR7, FURIN, RRBP1, ATXN1, MRPL4, INNPP4B, ZMYND8, BCL6, ASAP1, UBE2K, RORC, RGL2, ADCY3, FUT8, ANKRD11, SPTAN1, and BAZ2B.

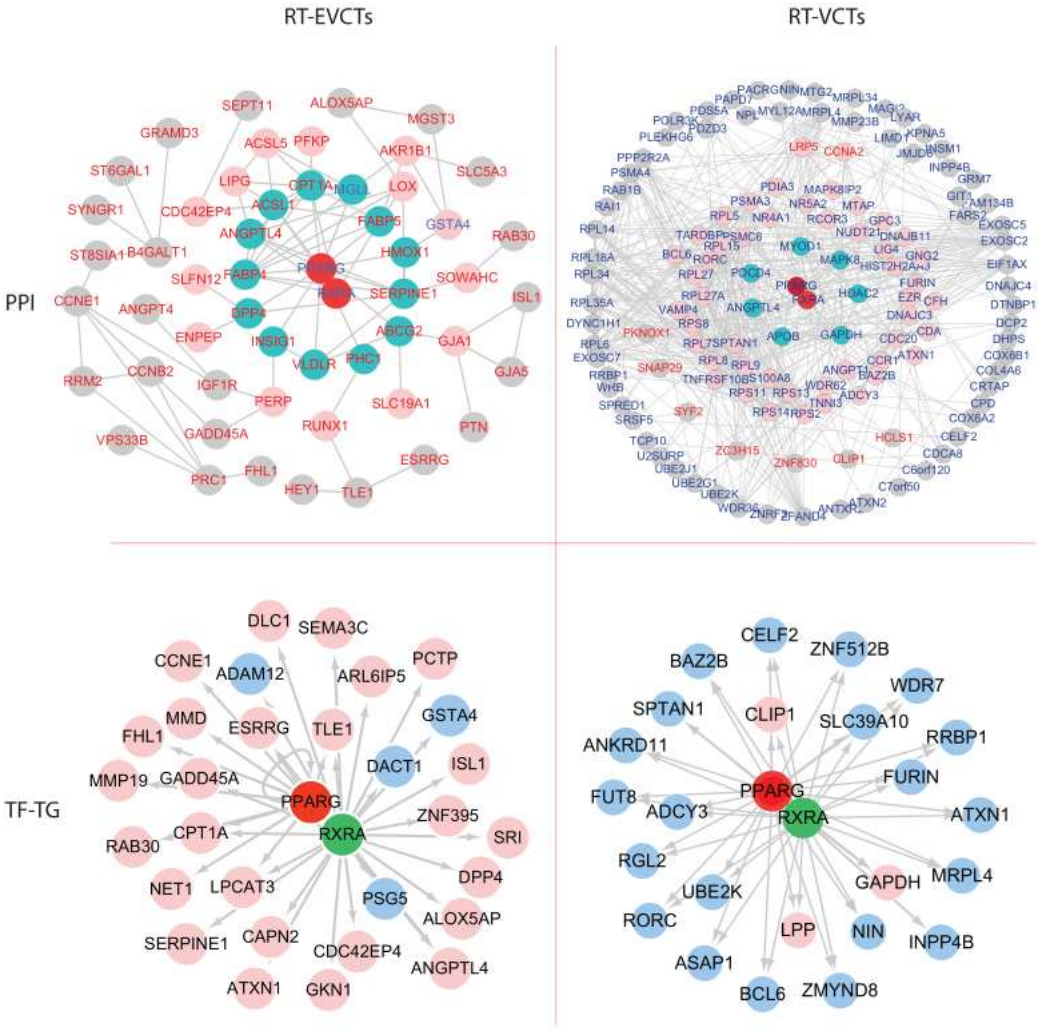


Figure 27 Interactions of the PPARγ & RXRα complex with DEGs of RT-EVCTs and RT-VCTs

Fig. 7. Interactions of the PPARγ & RXRα complex with DEGs of RT-EVCTs and RT-VCTs. Predictions were made of protein-protein interactions between PPARγ and DEG-encoded proteins, as well as of the transcription factor-target gene (TF-TG) interactions of PPARγ with DEG promoters. The PPARγ and RXRα (heterodimeric nuclear receptor partner of PPARγ) complex, together with the DEGs recovered in this study, were submitted to the STRING online tool. Visualizations were modified in Cytoscope to depict hierarchical interactions and gene expression. For hierarchical protein-protein interactions, red text represents upregulated genes, blue text represents downregulated genes, blue circles represent direct interactions with the PPARγ complex, red circles represent second-level interactions, and grey circles represent plus-level interactions. TF-TG interactions of the PPARγ complex with DEG promoters were predicted by iRegulon based on the TRANSFAC database. Red circles represent upregulated DEGs and blue circles represent downregulated DEGs. PPARγ: peroxisome proliferator-activated receptor-γ; RXRα: retinoid x receptor-α; RT-EVCTs: rosiglitazone-treated extravillous cytotrophoblasts; RT-VCTs: rosiglitazone-treated villous cytotrophoblasts; DEGs: differentially expressed genes; PPI: protein-protein-interaction; TF-TG: transcription factor-target gene.

7. Expression of genes targeted by PPARγ in RT-EVCTs and RT-VCTs

We next filtered our datasets to examine only the DEGs targeted directly by the PPARγ complex, based on the TF-TG predictions described above. The filtered RT-EVCT database contained 26 upregulated and 4 downregulated DEGs (Figure 8a), while the filtered RT-VCT database contained 3 upregulated and 21 downregulated DEGs (Figure 8b). Only one target gene, ATXN1, was present in both datasets; it was upregulated in RT-EVCTs and downregulated in RT-VCTs (Figure 8).

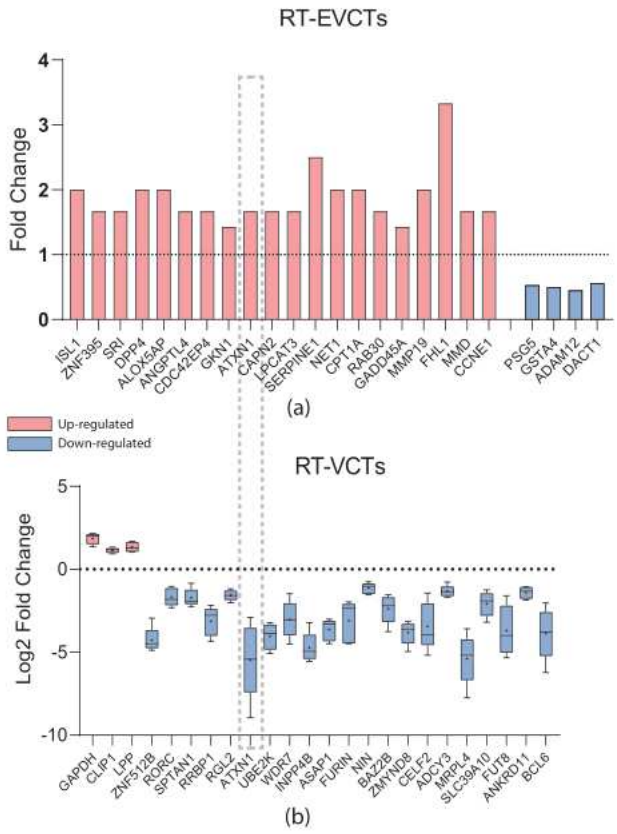


Figure 28 Expression of genes targeted by PPARγ & RXRα in RT-EVCTs and RT-VCTs

Fig. 8. Expression of genes targeted by PPARγ & RXRα in RT-EVCTs and RT-VCTs. Gene symbols were retrieved from the normalized gene expression matrix, together with the log2 fold-change values in each sample. With these values, boxplots were graphed for (a) RT-EVCTs and (b) RT-VCTs, with upregulation represented in red and downregulation in blue. The grey dashed box indicates the only gene found in both tissue types. PPARγ: peroxisome proliferator-activated receptor-γ; RXRα: retinoid x receptor-α; RT-EVCTs: rosiglitazone-treated extravillous cytotrophoblasts; RT-VCTs: rosiglitazone-treated cytotrophoblasts.

3.1.5 Discussion

The human placenta is a critical bridge between mother and fetus, facilitating nutrient exchange and various endocrine and immunological processes. As the cells

that form the main part of placenta, trophoblasts undergo extensive cell differentiation, including invasion, migration, and fusion. Abnormalities in these physiological processes can lead to a series of gestational diseases such as preeclampsia or intrauterine growth restriction. Specifically, both of these disorders appear to be associated with irregularities in the invasion of EVCTs into the maternal uterus, a biological process that is tightly controlled both spatially and temporally [223, 224]. However, the underlying mechanism linking EVCT invasion to gestational dysfunction has yet to be fully investigated. Our team has previously shown the critical influence of activated PPARγ on trophoblasts via treatment of the natural ligands of PPARγ or its specific agonist rosiglitazone [154, 157, 195, 225]. Rosiglitazone is the first synthetic chemical compound to be developed that demonstrates high selectivity for PPARγ (*K*d approximately 40 nM); concentrations of up to 100 μM of this compound have been reported to activate only PPARγ (including the PPARα/β/δ complex [226]). Moreover, our previous research revealed that a concentration of only 1 μM rosiglitazone led to significant alterations in trophloblast differentiation, with more than 50% inhibition of EVCT invasion [154]. In this study we treated EVCTs and VCTs with 1 μM rosiglitazone in order to more fully understand the effects of PPARγ on gene expression in these tissues.

Our microarray results for EVCTs were published previously with the aim of identifying significant DEGs for further study [157]. However, this work provided little information about the relative enrichment of pathways and processes among these DEGs, and did not include any comparisons with RT-VCTs. To more broadly determine the key genes, biological processes, and pathways affected by activated PPARγ in trophoblasts, in this study we also analyzed gene expression changes in VCTs using microarray profiling, and, through various approaches, identified the enriched processes that were linked with these DEGs in RT-EVCTs and RT-VCTs. We were thus

able to compare the similarities and differences between EVCTs and VCTs affected by activated PPARγ. In total, there were 139 DEGs in RT-EVCTs and 197 DEGs in RT-VCTs, and these were associated with enrichment in more than 200 GO and pathway terms (Tables S1-S8). Of these terms, the most significant and relevant are depicted in the figures. The majority of the terms recovered in our analysis were consistent with reports from the existing literature. For example, the terms “long-chain fatty-acyl-CoA biosynthetic process”, “regulation of plasma lipoprotein particle levels”, “plasma lipoprotein remodeling”, and “PPAR signaling pathway” are all associated with “fatty acid transport” which in the placenta is known to demonstrate sex-specific differences due to the PPARγ-dependent response of genes involved in lipogenesis [227]. Signaling molecules and dynamic regulation of the cytoskelton are required in trophoblast invasion [228-230], which are related to such terms as “regulation of epithelial cell migration”, “regulation of actin cytoskeleton”, and “tight junction”. Among the specific pathways highlighted, the HIF-1 signaling pathway is known to participate in PPARγ-mediated placental angiogenesis [213]; the P53 signaling pathway mediates trophoblast apoptosis via ligand-specific activation of PPARγ [145]; the JNK signaling pathway plays an essential role in blood-placental barrier formation [231, 232], as well as in EVCT migration and endothelial-like tube formation [233]; and the mTOR signaling pathway regulates adipogenic proteins in the placenta, with mTOR acting as a decidual nutrient sensor in histotrophic nutrition, which is crucial to embryo viability as well as early placental and fetal development [234]. Furthermore, our results were also consistent with the post-transcriptional modifications involved in placentation, with the terms “positive regulation of phosphoprotein”, “regulation of protein deacetylation”, “histone deacetylation”, and “ubiquitin-like protein conjugating enzyme activity” all known from previous reports. Indeed, different subtypes of trophoblast vary in phosphorylation status depending on the stage of placental

development and differentiation. For example, EVCTs require Smad2/3 phosphorylation for differentiation while the absence of pSmad2C is necessary for VCTs [235]. Downregulation of histone deacetylase-9 can repress trophoblast migration and invasion [236], and likewise, inhibition of histone acetylation in human endometrial stromal cells limits trophoblast invasion [237]. Ubiquitination of amino acid transporters expressed specifically in the plasma membrane of the trophoblast can decrease amino acid uptake, leading to abnormal development of the placenta and restricted fetal growth [238, 239]. In addition, PPARγ can be phosphorylated through activation of the downstream ERKs 1/2 or p38/c-JNK pathways [100, 240]. Rosiglitazone blocks the acetylation of lysine residues of PPARγ at positions K268ac and K293ac [99]. Atypical poly-ubiquitination of PPARγ reduces proteasomal degradation and guarantees the stabilization of PPARγ [241, 242].

A major aim of this study was to compare patterns of enrichment between RT-EVCTs and RT-VCTs. During EVCT invasion, non-invasive EVCTs undergo an epithelial-mesenchymal transition to acquire the invasive phenotype [243]. Invasive EVCTs then migrate away from the placenta up to the first third of the endometrium and colonize the maternal spiral arteries. We found a comparison of these two types of trophoblasts to be particulary compelling given the number of studies that have focused on their differences and similarities. For example, the transformation of non-invasive EVCTs into invasive EVCTs involves expression differences in adhesion molecules, which manifest themselves when EVCTs escape from the anchoring column and invade into the endometrium (decidua, spiral arteries, and myometrium) [244]. Other studies have examined differences between EVCTs and VCTs with respect to hCG secretion for the normal maintenance of pregnancy [245] and placental cytokine secretion [246]. These biological processes are apparent in the terms recovered here that were associated with “regulation of endothelial cell migration”, “embryo

implantation", "steroid hormone receptor binding", "secretion and action", and "pre-eclampsia". The main point is that these biological processes have all been reported to be regulated by PPARγ. For example, the activation of PPARγ has been found to prevent the TGF-β–induced epithelial-mesenchymal transition via inhibition of transcription of the E-cadherin and N-cadherin promoters [247]. Furthermore, PPARγ was reported to modulate basal levels of the hCGα and hCGβ subunits, resulting in differences in expression between EVCTs and VCTs [248]. Additional evidence has been obtained from studies with rosiglitazone; for example, treatment with 1 μM of the PPARγ agonist was found to decrease and increase, respectively, the number of transcripts of TGFβ2 and IL1β [225]. Such regulatory changes might be represented here by the terms "regulation of endothelial cell migration", "Wnt signaling pathway", "negative regulation of protein phosphorylation", "transcription factor activity", "PPAR signaling pathway", and "HIF-1 signaling pathway".

In general, our datasets revealed an abundance of biological processes or pathways affected by PPARγ, many of which are consistent with previous reports. This concordance should increase confidence in our results and indicate avenues for further study.

Finally, we attempted to predict the protein-protein interactions between the DEGs recovered here and PPARγ, in order to facilitate study of the mechanisms behind the molecular interaction. Our results provided evidence for direct protein-protein and protein-promoter interaction with the PPARγ complex. Among the proteins that appear to interact directly with PPARγ, several have been experimentally verified, including ANGPTL4 [249-251], ABCG2 [252], APOB [253], CCNE1 [254], CPT1B [255], FABP4 [256-259], HMOX1 [260], and SERPINE1 [261, 262]. Many of the TF-TG interactions, which were predicted using the position weight matrix algorithm, have not been previously reported and await further verification. Our interaction matrix

(Figure 7) also revealed more extensive upstream-to-downstream signaling pathways, such as the PPARγ-MAPK-MMP signaling pathway. Commonly, phosphorylated PPARγ stimulates the MAPK-activated pathway, leading to the activation of extracellular signal-regulated kinases (ERKs) that then induce the upregulation of matrix metalloproteinase (MMP) [263-265]. Here, only a single DEG, ATXN1, was found in both types of rosiglitazone-treated trophoblast; this gene was upregulated in RT-EVCTs and downregulated in RT-VCTs. It has been reported that the ATXN1 protein family can regulate remodeling of the extracellular matrix [266], which indicates a potential involvement in trophoblast differentiation. In addition to the direct target genes predicted here, the genes in secondary relationships should be paid equal attention in terms of potential regulation by other target genes. For example, our previous research has shown the key role of LOX1, through secondary interactions, in cytotrophoblast invasion [157].

Conclusions

To our knowledge, our results reveal for the first time the widespread effects of PPARγ activation in EVCTs and VCTs, highlighting extensive changes in gene expression and the biological processes and pathways affected. This study provides a broad perspective of PPARγ-influenced biological processes in trophoblasts, and facilitates further study, particularly into potential drug-targeted genes or pathways in the human placenta.

Data Availability

Our microarray data have been deposited in the Gene Expression Omnibus public repository (https://www.ncbi.nlm.nih.gov/geo/); EVCT microarray data under

accession number GSE28426, VCT microarray data under accession number GSE137434).

Conflicts of Interest

The authors declare that there are no conflicts of interest that could affect the impartiality of the reported research.

Funding Statement

This work was carried out with the funding support of Inserm, University of Paris Descartes, and RSI Professions Libérales Provinces, 44 boulevard de la Bastille, 75578 Paris Cedex 12.

Acknowledgments

The authors wish to thank the consenting patients and the clinical staff midwives of Cochin Port Royal for providing placental tissues, as well as Dr Danièle Evain-Brion, former director of the Unit, for her support.

Supplementary Materials

All differentially expressed genes are submitted to ClueGo for the functional classification. An exhaustive list of all terms (including those not shown above) can be found in supplementary materials (Tables S1-S8).

Part III

4 Part Three

Lead-in

This study is paralleled to the previous study because the scientific question was intrigued by the data mining study. In the previous research, we studied the activated PPARγ in two cell types in the human placenta because we discovered the PPAR signaling pathway and the core regulator, the PPARγ. And we validated the existence of PPARγ by detecting the lipid droplet in the human placenta. When we looked into the details of the lipid droplet in the first trimester, we found that the gestational age was negatively related to the amount of the lipid from gestational week 8-14. This period is just overlapped to the time when the environmental oxygen levels in the human placenta increase dramatically. Oxygen level change affects a lot of the cells in the human placenta with respect to differentiation, migration, invasion, etc. The evolution of these processes is complex and intricately regulated by O2 tension. Considering the overlapped period, we propose if there exists a connection between the oxygen level and the lipid metabolism in the first trimester. Namely, since oxygen-related genes or pathways can be regulated by the hypoxia-inducible factor (HIF) and PPARγ can be regulated by its cofactors, if there exist the common elements among the HIF targets and PPARγ cofactors, regulation of PPARγ through its cofactors would be probably achieved by HIF which is sensitive to the oxygen change. In this part, we are going to sequence the samples collected from the early and late first trimester of the human placenta and try to classify the HIF targets from the filtered significant genes before comparing them with the PPARγ cofactors.

4.1 Age and Sex-Related Changes in Human First-Trimester Placenta Transcriptome in Response to Increased Environmental Oxygen Levels

Fulin Liu [1], Christelle Simasotchi [1,2], Françoise Vibert [1], Wencan Zhu [3], Sophie Gil [1,2], Séverine A. Degrelle [1,4,*] Thierry Fournier [1,*]

1 Université de Paris, INSERM, «Pathophysiology & Pharmacotoxicology of the Human Placenta, Pre & Postnatal Microbiota», 3PHM, Paris, F-75006, France

2 Fondation PremUp, Paris, F-75006, France

3 UMR Applied Mathematics & Informatics, AgroParisTech-Université Paris-Saclay, 75005, Paris, France

4 Inovarion, Paris, F-75005, France

* These authors contributed equally to this work.

Correspondence should be addressed to Thierry Fournier; thierry.fournier@inserm.fr

4.1.1 Abstract

Physiological oxygen tension rises dramatically in the placenta between 8 and 14 weeks of gestation. Abnormalities in this period can lead to gestational diseases, whose underlying mechanisms remain unclear. We explored the changes at mRNA level by comparing the transcriptomes of human placentas at 8-10 gestational weeks and 12-14 gestational weeks. A total of 20 samples were collected and divided equally into four groups based on sex and age. Cytotrophoblasts were isolated and sequenced using RNAseq. Key genes were identified using two different methods: DESeq2 and weighted gene co-expression network analysis (WGCNA). We also constructed a local database

of known targets of hypoxia-inducible factor (HIF) subunits alpha and beta to investigate expression patterns likely linked with changes in oxygen. Patterns of gene enrichment in and among the four groups were analyzed based on annotations of gene ontology (GO) and KEGG pathways. We characterized the similarities and differences between the enrichment patterns revealed by the two methods and the two conditions (age & sex), as well as those associated with HIF targets. Our results provide a broad perspective of the processes that are active in cytotrophoblasts during the rise in physiological oxygen, which should benefit efforts to discover possible drug-targeted genes or pathways in the human placenta.

Keywords: Placenta; Trophoblast; Oxygen; RNAseq; WGCNA; DESeq2

4.1.2 Introduction

The human placenta plays a pivotal role in development by regulating the exchange of nutrients, gas, and waste between the mother and the fetus. Normal development of the placenta can be divided into three trimesters, of which the first lays the foundation for all subsequent processes. Indeed, at the very beginning of the first trimester, it is the yolk sac that first establishes the supply of oxygen and nutrients to the embryo, and this role is gradually taken over by the growing placenta. From day 13 post-coitum, the placental villus begins to form and branch, and is subsequently infiltrated by the allantoic blood vessels [267]. This process involves interactions between fetal and maternal tissues and specifically involves the trophectoderm stem cells, which then evolve into several different trophoblast cell lineages that constitute the main part of the placenta [70, 268]. The evolution of this process is complex and intricately regulated by O_2 tension, especially in the period from 8 to 12 gestational

weeks (GW) [269]. Defects in this process can lead to diseases of pregnancy such as spontaneous abortion, preterm birth, and preeclampsia [270, 271].

The oxygen tension in the intervillous space rises dramatically from 2–3% around 8-10 GW to more than 6% after 12 weeks [269]. Oxygen, and the oxidative stress that accompanies it, plays an important role in the positive or negative development and growth of chorionic villi. Histopathological research has shown that villous cytotrophoblast thrives under the intrauterine hypoxia of the first trimester [85]. Instead, the hyperoxic state in the placenta inhibits capillary branching, the formation of sinusoids, and the differentiation of villous cytotrophoblast [272]. The increase in oxygen tension at 8-12 GW is driven by remodeling of the maternal spiral arteries by invasive extravillous trophoblast. The supply of blood flow allows a rapid increase in O_2 tension in the intervillous space [70, 268], thus satisfying the high oxygen demands of the placenta, which represents at least 30% of the total amount of the utero-placental unit [273, 274].

Oxygen is consumed in the mitochondria, and recent research has uncovered sexually dimorphic aspects of this process, particularly in cases of placental mitochondrial dysfunction [275-277]. For example, male guinea pig placenta is more susceptible to respiratory complex chain disruption under gestational hypoxia [278]. It is well recognized that the placenta acts like an orchestrator, adapting morphologically and hormonally in response to variations in the environment [279]. Hypoxia (2–3% O_2), as an ever-present challenge throughout the duration of pregnancy, functions as a two-edged sword for the development of the placenta. While suitable levels of hypoxia/low O_2 tension induce the remodeling of maternal spiral arteries, extreme hypoxia disables the remodeling and can result in adverse conditions such as preeclampsia [270, 271] and insufficient hypoxia leads to over-consumption of the nutrients delivered into the intervillous space for the nourishment of the fetus

[176]. Taken together, these findings perhaps indicate the existence of different strategies for male and female placentas in coping with variable oxygen conditions.

There are three conventional oxygen-sensitive pathways that regulate gene expression in the placenta under low O_2 tension: hypoxia inducible factors (HIF), unfolded protein response (UPR), and mammalian target of rapamycin (mTOR). Their activation regulates gene expression, metabolic homeostasis, and cell survival [280]. Of these, the most studied signaling pathway is HIF, which is a heterodimeric transcription factor composed of two subunits, HIF-α and HIF-β (alias: ARNT, aryl hydrocarbon receptor nuclear translocator) [281]. HIF-α and HIF-β bind to specific promoter sequence elements to activate target gene transcription and thus enable many different cellular processes to respond to hypoxia during early placentation [282, 283]. Interestingly, the response to hypoxia has been shown to differ between male and female placentas, particularly with respect to energy metabolism and angiogenesis. The female placenta was described to activate more protective mechanisms to increase the availability of nutrients for fetal metabolic development [275], whereas in preeclamptic pregnancies male placentas showed stronger reductions in pro-angiogenic markers than female placentas did [284]. However, such patterns of sexual dimorphism are still poorly documented, and at a fundamental level, the steep increase in oxygen tension in the human placenta remains only incompletely understood.

Here, we applied next-generation sequencing technology to sequence human villous cytotrophoblasts freshly isolated from placentas at either 8-10 GW or 12-14 GW with the goal of assessing changes in gene expression and the influence of placental sex differences in this key physiological period.

4.1.3 Materials and Methods

Sample collection and ethics statement

A total of 20 human placentas were collected from normal gravidas in the first trimester. To characterize the processes that accompany the increase in physiological oxygen tension from 8 to 14 GW, we investigated the early and late stages of this period, 8-10 GW and 12-14 GW, respectively. Equal numbers of female and male placentas were collected for each stage. We used the letter "E" to represent "Early", "L" for "Late", "F" for "Female", and "M" for "Male". The final sample pool contained 5 "EF samples,"5"EM" samples, 5 "LF" samples, and 5 "LM" samples. All placental tissues were obtained from patients who voluntarily and legally chose to terminate pregnancy during the first trimester at the Cochin Port-Royal and Montsouris maternity units (Paris, France). These biological samples were obtained following informed written consent from patients and approval from our local ethics committee (CPP 2015-mai-13909).

Cytotrophoblast isolation and culture

The first step was to isolate villous cytotrophoblasts from each sample. To do this, we purified the villous tissues by mincing them into pieces with forceps, removing the membrane on the surface, and rinsing them with Ca^{2+}-, Mg^{2+}-free Hanks' balanced salt solution. The pieces were then digested in Hanks' balanced salt solution with 4.2 mM $MgSO_4$, 0.25% (wt/vol) trypsin powder (Difco), 5 IU of DNase I per ml, 25 mM HEPES, 100 IU/mL penicillin, and 100 μg/mL streptomycin (Biochemical Industry). Digestion was monitored by invert microscopy; the initial digested solution, which mainly consisted of red blood cells, was discarded and the subsequent digested solution, which

mainly consisted of villous cytotrophoblasts, was retained for stratification. The digested solution was slowly transferred to a discontinuous Percoll gradient (5–70% in steps of 5%) in order to stratify the mixed cells and debris. The layer containing the cytotrophoblasts was retained and washed using Dulbecco's modified Eagle's medium (F12/DMEM) that contained 100 IU/mL penicillin, 100 mg/mL streptomycin, and 2 mM glutamine. Cells were counted using a TC20™ Automated Cell Counter (Biorad). For RNASeq experiments, 1.5×10^6 cells were transferred into a 1.5 ml Eppendorf tube. After centrifugation (a 10 s pulse at 14.000 rpm), the supernatants were discarded and the cell pellets were snap-frozen in liquid nitrogen and stored at -80°C until total RNA extraction, PCR, and RNAseq analyses were conducted.

Fetal-sex determination by PCR

Fetal-sex determination was performed via PCR on stored cytotrophoblast cells, as described previously [285]. Genomic DNA (gDNA) was extracted using the 25mM NaOH/9.2 mM EDTA buffer. The sex of the placenta was genetically determined by PCR according to the sex-linked chromosome genes *ZFX* (GenBank Acc.No. NG_021253, NM_003410) and *ZFY* (GenBank Acc.No. NG_008113, NM_003411). The primers were ChrX-Y_F 5'-ATTTGTTCTAAGTCGCCATATTCTCT-3', ChrX_R 5'-GAACACACTACT-GAGCAAAATGTATA-3', and ChrY_R 5'-CATCTTTACAAGCTTGTAGACACACT-3'. Reagents for PCR reactions included gDNA (100-300ng) 10 µl 5×green GoTaq reaction buffer, 2 µl 10 mM dNTPs, 2.5 µl 10 µM ChrX-Y_F, 2.5 µl 10 µM ChrX_R, 2.5 µl 10 µM ChrY_R, 0.2 µl [5 u/µl] GoTaq DNA polymerase (Promega), and water to reach a total volume of 50 µl. Amplification was conducted in a Perkin Elmer Applied Biosystems GeneAmp PCR Thermal Cycler System 2700, with the following cycling parameters: initial denaturation at 94°C for

10min, 35 cycles of denaturation at 94°C for 45s, annealing at 50°C for 45s, and synthesis at 72°C for 30s, and an extension of 5 min at the end of the final cycle. Amplification products (10µl) were directly analyzed on 2% agarose gel and evaluated under UV light. Primers ChrX-Y_F/ ChrY_R were present only in male samples and primers ChrX-Y_F/ChrX_R were present in all tested samples. For validation, we also analyzed the expression of the sex-linked genes *XIST* and *DDX3Y* in the RNAseq dataset, following the procedures described in [286].

RNA-seq experiment and data processing

RNAseq analyses were performed as described in [287]. The RNeasy Micro Kit (Qiagen) was used to extract total RNA from villous cytotrophoblasts isolated from placentas. DNAse was used to degrade genomic DNA, following the RNeasy Micro Kit protocol. RNA sequencing was performed by the Genom'IC lab facility of the Institut Cochin (Paris, France). RNA concentrations were quantified using a Nanodrop device (Thermo Fisher Scientific, USA) and the quality of the RNA was measured on an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA). A total of 800 ng RNA sample was used to construct each RNAseq library using the TruSeq Stranded mRNA kit (Illumina). RNAseq libraries were quantified by RT-qPCR using the KAPA Library Quantification Kit for Illumina Libraries (KapaBiosystems, Wilmington, MA) and corresponding profiles were evaluated using the DNA High Sensitivity LabChip kit on an Agilent Bioanalyzer. RNAseq libraries were sequenced on an Illumina Nextseq 500 instrument using 75 base-length reads and V2 chemistry, in paired-end mode. AOZAN software (ENS, Paris) was applied to demultiplex and characterize the raw data (based on FastQC module / version 0.11.5), and the obtained fastq sequence files were aligned using the STAR algorithm (version 2.5.2b). Raw reads were counted using

Featurecount (version Rsubread 1.24.1) and processed as follows: i) rows with reads equal to zero in more than 10 samples were excluded; ii) all read counts were increased by 1 and log2-transformed; and iii) data were normalized using the DESeq2 package [288].

Clustering analysis and construction of co-expression modules of human placental RNAseq data

To detect outliers, sample clustering was performed based on Euclidean distance, principal component analysis (PCA), and t-distributed stochastic neighbor embedding (t-SNE). To identify differentially abundant genes, DESeq2 and WGCNA were applied separately and the results were then merged for a combined analysis. For DESeq2, gene expression was compared between the different groups and the criteria for defining key genes were a fold-change of 2 and a P-value less than 0.05. For WGCNA, normalized RNAseq reads from all samples were submitted directly to the WGCNA package (version: 1.68) [180] in R (version: 4.0.1) to evaluate correlations in gene expression. We initially assessed the optimal soft thresholding power value by using a range of power values from 1 to 20; the optimal value is the one for which the measurement of scale independence surpasses the threshold of 0.9. This optimal value was then used to reduce the background noise of the correlations in the adjacency matrix. The correlations of eigengenes from the adjacency matrix based on the default unsigned network were used to construct co-expression modules. The default minimum module size of 30 was used to increase the reliability of the results.

Analysis of co-expression modules in human placental RNAseq data

Each module—a set of topologically correlated genes—was represented by a different color. A clustering dendrogram was created based on the correlations between the genes of different modules and a module-trait heatmap was constructed based on the correlations between the module eigengenes and traits of interest (sex and age, *i.e.*, the difference between 8-10 GW and 12-14 GW). For each module, we created scatterplots of the modules' eigengenes in which the x-coordinate represented module membership (the correlation coefficient between a gene's expression profile and the module eigengene) and the y-coordinate represented gene significance (GS; the correlation coefficient between the genes' expression and the traits of interest). Only correlation coefficients higher than 0.6, with a P-value less than 0.05, were included for analyses of module membership.

Enrichment and differential abundance of key genes

Venn diagrams were created to depict the intersections between different datasets using the online tool Draw Venn diagram (http://bioinformatics.psb.ugent.be/webtools/Venn/). Enrichment in gene ontology (GO) terms and pathways was analyzed using the R package clusterProfiler [181] (version 3.9, synced to latest GO terms and pathways). The top 15 category terms for each group were assembled for comparison; those with a P-value lower than 0.05 were identified as being significantly enriched. WebGestalt was used to visualize the terms in volcano format and perform the network topology analysis [289]. Finally, to investigate the involvement of HIF (including both the alpha and beta subunits) in the biological processes under consideration, we retrieved all published or predicted targets of HIF-1α and HIF-1β from the literature and public databases using the R

package *tftargets* (https://github.com/slowkow/tftargets). These were then assembled into a local database of HIF targets.

4.1.4 Results

1. Identification of outliers in samples

Our RNAseq dataset has been deposited in the Gene Expression Omnibus public repository (https://www.ncbi.nlm.nih.gov/geo/) under the accession number GSE163023 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE163023). The quality report generated by FastQC indicated that the raw data were appropriate for further use: the majority of plots were well above phred score 30 (green region), the average quality was in single peak, a well overlapped single hump was shown in the center, and the data remained consistent before and after normalization (Figure S1-1). First, we retrieved the expression of *XIST* and *DDX3Y* in each sample and used this information to validate the sex designations obtained with PCR. The results of the two methods were in agreement for the vast majority of samples, with the exception of samples "LM5" and "EF3" (Figure 1a). Next, to evaluate the quality of the samples and detect outliers, we clustered the samples for classification. The optimal number of sample clusters was suggested to be two according to random k-means clustering (not shown), while the use of distance clustering identified sample "LM5" as an outlier (Figure 1b). "LM5" was also classified as an outlier according to separate PCA analyses of age and sex, while no difference was indicated by the t-SNE analysis (Figure 1c). "EF3" was identified as an outlier in a subsequent analysis of distance clustering in

WGCNA (Figure 1d). Therefore, “LM5” and “EF3” were excluded from the dataset, and the remaining samples were subjected to further gene expression analyses.

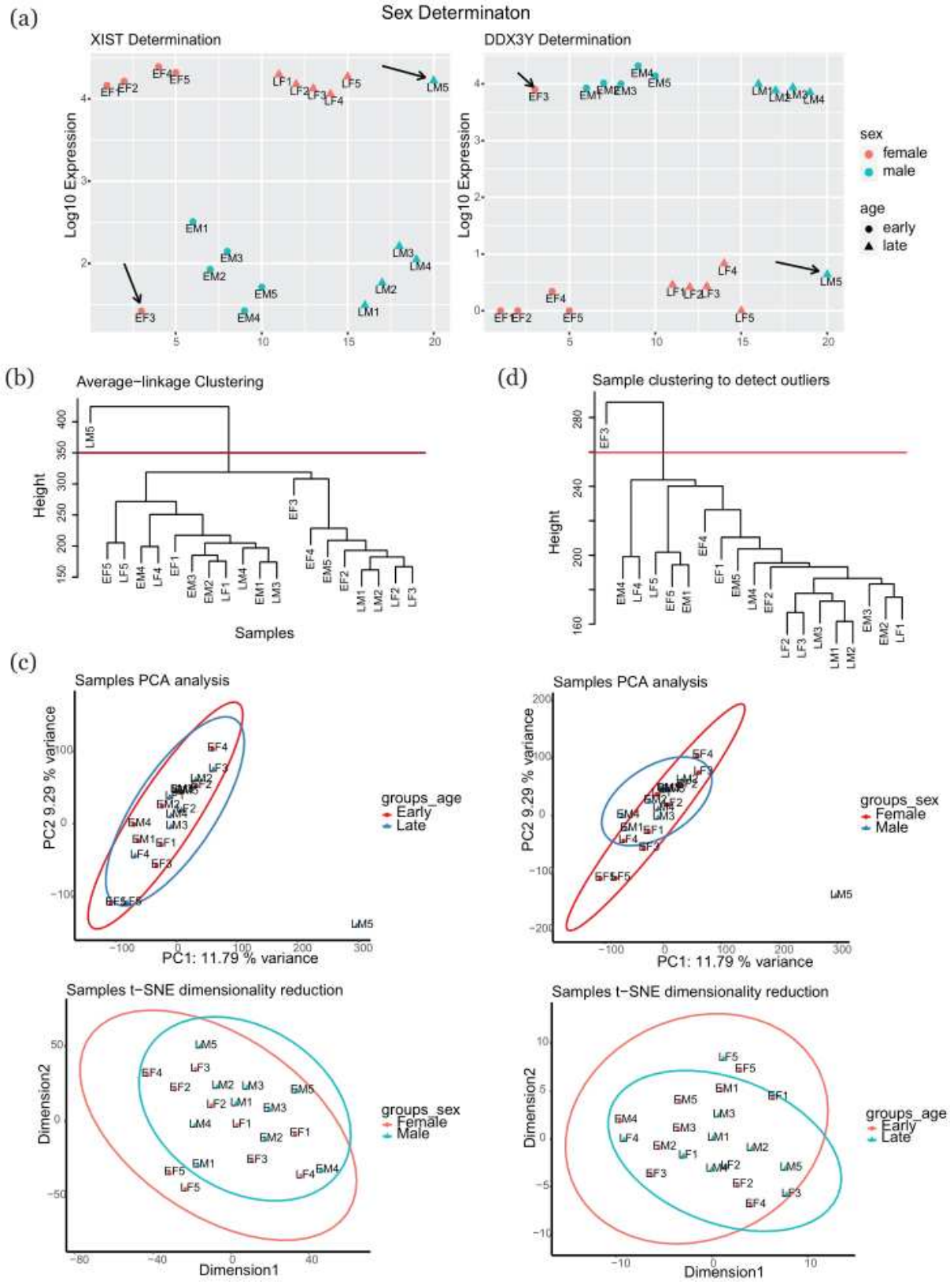


Figure 29 Sex determination and sample clustering

Fig. 1. Sex determination and sample clustering. (a) Expression of the sex-linked genes *XIST* and *DDX3Y* in samples. (b) Optimal number of clusters for samples and sample clustering according to K-means clustering algorithm. (c) PCA analysis and t-SNE dimensionality reduction of samples. (d) Sample re-clustering using WGCNA method without LM5. PCA: principal components analysis; t-SNE: t-distributed stochastic neighbor embedding; EF: early female; EM: early male; LF: late female; LM: late male; WGCNA: weighted gene co-expression network analysis.

2. Identification of key genes and term enrichment using DESeq2

The DESeq2 method, based on the DESeq2 R package, was applied to detect key genes that were differentially expressed between different groups. This analysis identified 15 key genes with expression differences in the comparison of *Female* versus *Male*, 457 key genes for *Early* versus *Late*, 45 key genes for *LF* versus *LM*, 41 key genes for *EF* versus *EM*, 157 key genes for *EF* versus *LF*, and 801 key genes for *EM* versus *LM*. The expression patterns were then used to construct heatmaps (Figure S2-1a-f). The Venn diagram in Figure 2a depicts the intersection of key genes among comparisons. These key genes were then submitted to analyses of enrichment and term comparison, specifically with respect to GO biological processes and KEGG pathways (Figure 2b-c). In addition, we also analyzed enrichment in GO cellular components and GO molecular functions, and these comparisons are detailed in the supplementary materials (Figure S2-2).

For GO biological processes, the early-stage placentas demonstrated enrichment in aspects of organic development such as “camera-type eye development”, “lymphocyte differentiation”, and “hindbrain development”, while later-stage tissues were mainly enriched in processes related to biological regulation, such as “histone

demethylation", "protein demethylation", and "protein dealkylation". When we instead compared the tissues based on sex, we found that the male and female placentas shared many of the same terms for cellular processes and biological regulation, such as "response to transforming growth factor beta", "extracellular matrix organization", and "response to forskolin". We did, however, detect additional enrichment in males in certain processes related to organic development, such as "eye development", "lymphocyte differentiation", and "hindbrain development" (Figure 2b).

For KEGG pathways, when we only examined the effect of age, we observed that early-stage samples were characterized by enrichment in regulatory pathways such as the "notch signaling pathway", "hippo signaling pathway", and "calcium signaling pathways" while in late-stage samples, enrichment was noted in pathways associated with "staphylococcus aureus infection", "estrogen signaling pathway", and "RNA polymerase". Instead, when we compared the male and female samples, we found that the sexes were very similar in their biological regulation; both sexes demonstrated enrichment in terms such as "regulation of lipolysis in adipocytes", "endocrine resistance", "cGMP-PKG signaling pathway", "apelin signaling pathway", and "PI3K-Akt signaling pathway" (Figure 2c).

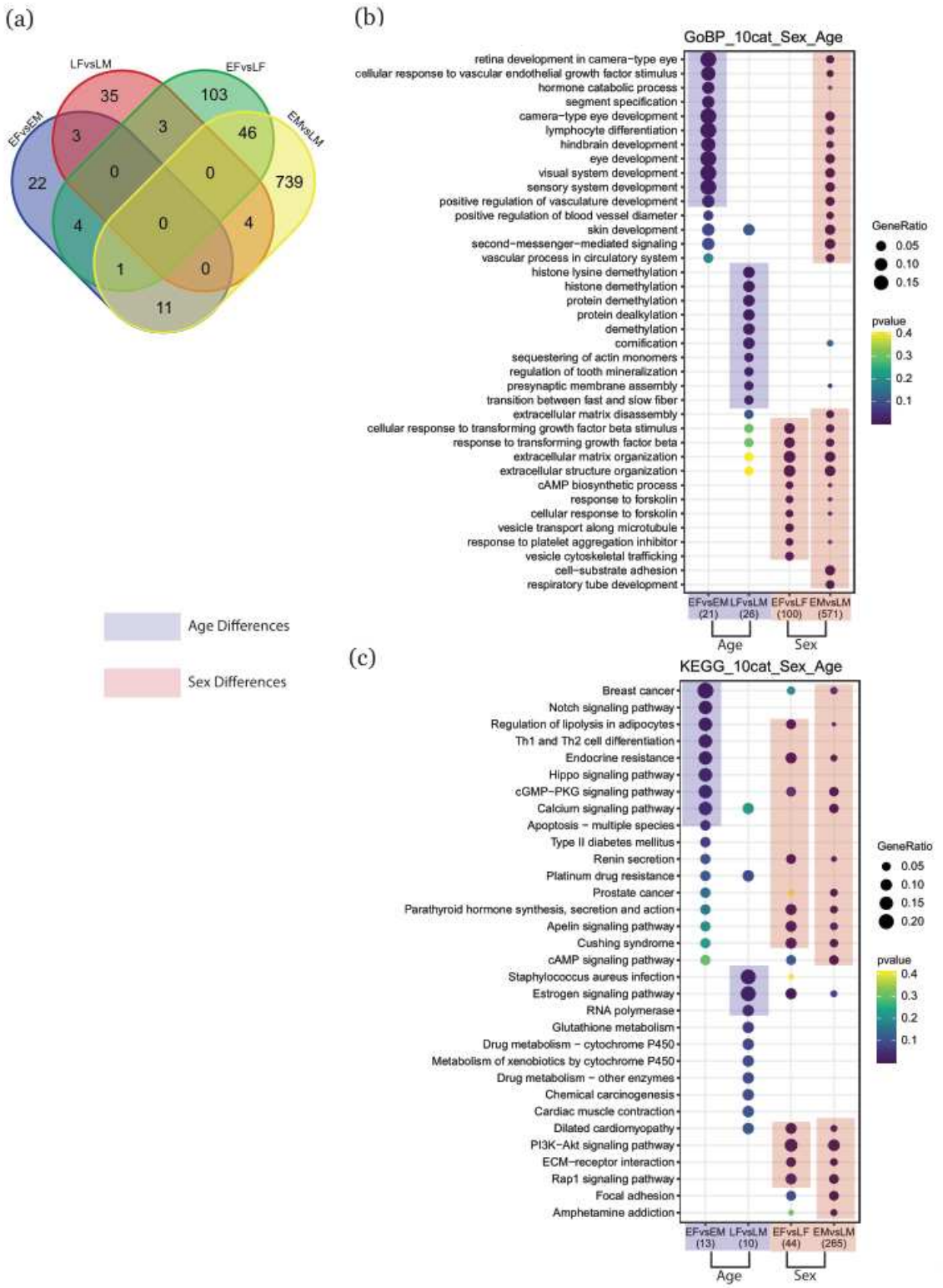


Figure 30 Identification of differentially expressed genes (DEGs) based on the DESeq2 method and assessment of term enrichment

Fig. 2. Identification of differentially expressed genes (DEGs) based on the DESeq2 method and assessment of term enrichment. (a) Venn diagram showing DEGs from

different comparisons. (b&c) Enrichment and comparison of GO terms and KEGG pathways, respectively, based on DEGs. The size of a point indicates the magnitude of enrichment; plum shading represents an association with age and coral shading represents an association with sex. “E” stands for “Early”, “L” for “Late”, “F” for “Female”, and “M” for “Male”. GO: gene ontology; KEGG: Kyoto encyclopedia of genes and genomes.

3. Selection of key genes based on WGCNA method

We also used the WGCNA method to detect key genes that demonstrated differences in expression between groups. The optimal soft thresholding power value was set as 16, which was the value at which measurements of scale independence exceeded the required threshold of 0.9 (Figure 3a). The original clustered modules, which each represented a set of eigengenes, and the merged modules are presented in a cluster dendrogram (Figure 3b). The correlation coefficients between the merged modules and the traits under consideration (age, sex) are shown in the form of a heatmap, in which red denotes a positive relationship and blue denotes a negative one (Figure 3c). Following our selection criteria (correlation coefficient higher than 0.6 and P value less than 0.05), the “blue” and “dark turquoise” modules were retained for their association with age (Figure 3d&3e), while the “steel blue” module was linked with sex (Figure 3f). Within the two modules associated with age, there were 2015 key genes, while the sex-related module contained 233. Figure S3 contains a heatmap depicting the expression of genes in the modules for each of the traits (Figure S3-1a&b).

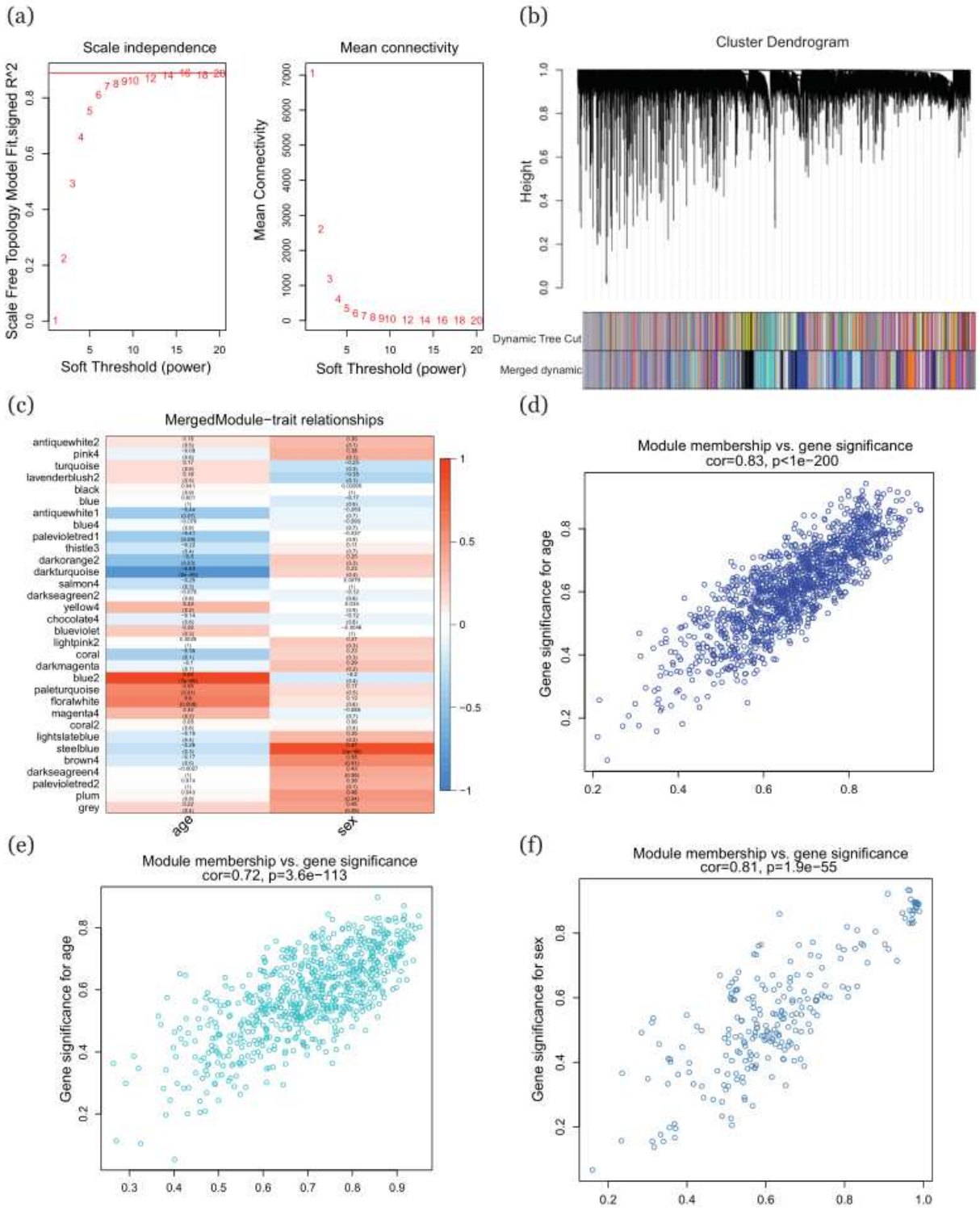


Figure 31 Selection of key genes using the WGCNA method

Fig. 3. Selection of key genes using the WGCNA method. (a) Soft-thresholding power selection from a range of power values from 1 to 20. (b) Cluster dendrogram of genes obtained through dissimilarity clustering based on consensus topological overlap. The

first row below the dendrogram represents the unmerged colored modules while the second depicts the merged colored modules. (c) Heatmap of correlation coefficients of the association between each module and each trait, with rows representing the modules and columns representing the traits. (d&e) Scatterplots for merged modules (blue, dark turquoise) associated with age. (f) Scatterplot for merged module (steel blue) associated with sex. The criteria for selecting merged modules were a correlation coefficient higher than 0.6 and a P value less than 0.05. WGCNA: weighted gene co-expression network analysis.

4. Comparison of enriched terms between DESeq2 and WGCNA

To compare the two methods, we extracted the key genes from each (DESeq2_Sex: 15, DESeq2_Age: 457; WGCNA_Sex: 233, WGCNA_Age: 2015) and compared the patterns of enrichment in GO biological processes and KEGG pathways. The Venn diagram in Figure 4a shows the intersection among groups from different methods and conditions. Using the key sex-related (Figure 4b&4c) or age-related (Figure 4d&4e) genes highlighted by the two methods, we compared the resulting patterns of enrichment in GO biological processes and KEGG pathways revealed by each approach. We performed a similar analysis of enrichment with respect to GO cellular components and GO molecular functions, which is included in the supplementary materials (Figure S3-2).

For GO biological processes, both methods revealed a relationship between sex and enrichment in “dosage compensation by inactivation of X chromosome” and “purine nucleotide catabolic process”. Instead, the WGCNA_Sex dataset highlighted enrichment in regulatory processes such as “histone lysine demethylation”, “demethylation”, and “histone demethylation”, and the DESeq2_Sex dataset was

mainly characterized by metabolic processes such as “nucleotide catabolic process”, “nucleotide phosphate catabolic process”, and “ribonucleotide catabolic process” (Figure 3b). With respect to age, both methods revealed enrichment in “hemidesmosome assembly”, “anion homeostasis”, and “reactive oxygen species metabolic process”, while the WGCNA_Age dataset was mainly characterized by cellular processes such as “protein localization to endoplasmic reticulum”, “protein targeting to membrane”, and “RP-dependent cotranslational protein targeting to membrane” and the DESeq2_Age dataset highlighted cellular processes such as “response to glucagon”, “anion transmembrane transport”, and “extracellular matrix organization” (Figure 3d).

For KEGG pathways, there was no overlap between the WGCNA and DESeq2 datasets for sex-related genes. The analysis of enrichment in the WGCNA_Sex dataset revealed the importance of metabolic processes such as “glycine, serine, and threonine metabolism”, and “cysteine and methionine metabolism” while the DESeq2_Sex dataset was characterized by biological processes such as “staphylococcus aureus infection”, and “estrogen signaling pathway” (Figure 3c). With respect to age-related genes, both methods highlighted enrichment in “Focal adhesion”, “PI3K-Akt signaling pathway”, and “Tight junction” while the WGCNA_Age dataset was mainly characterized by metabolic processes such as “glycan degradation”, “glycosaminoglycan biosynthesis”, and “lysosome” and the DESeq2_Age dataset featured regulatory pathways such as “oxytocin signaling pathway”, “morphine addiction”, and “inflammatory mediator regulation of TRP channels” (Figure 3e).

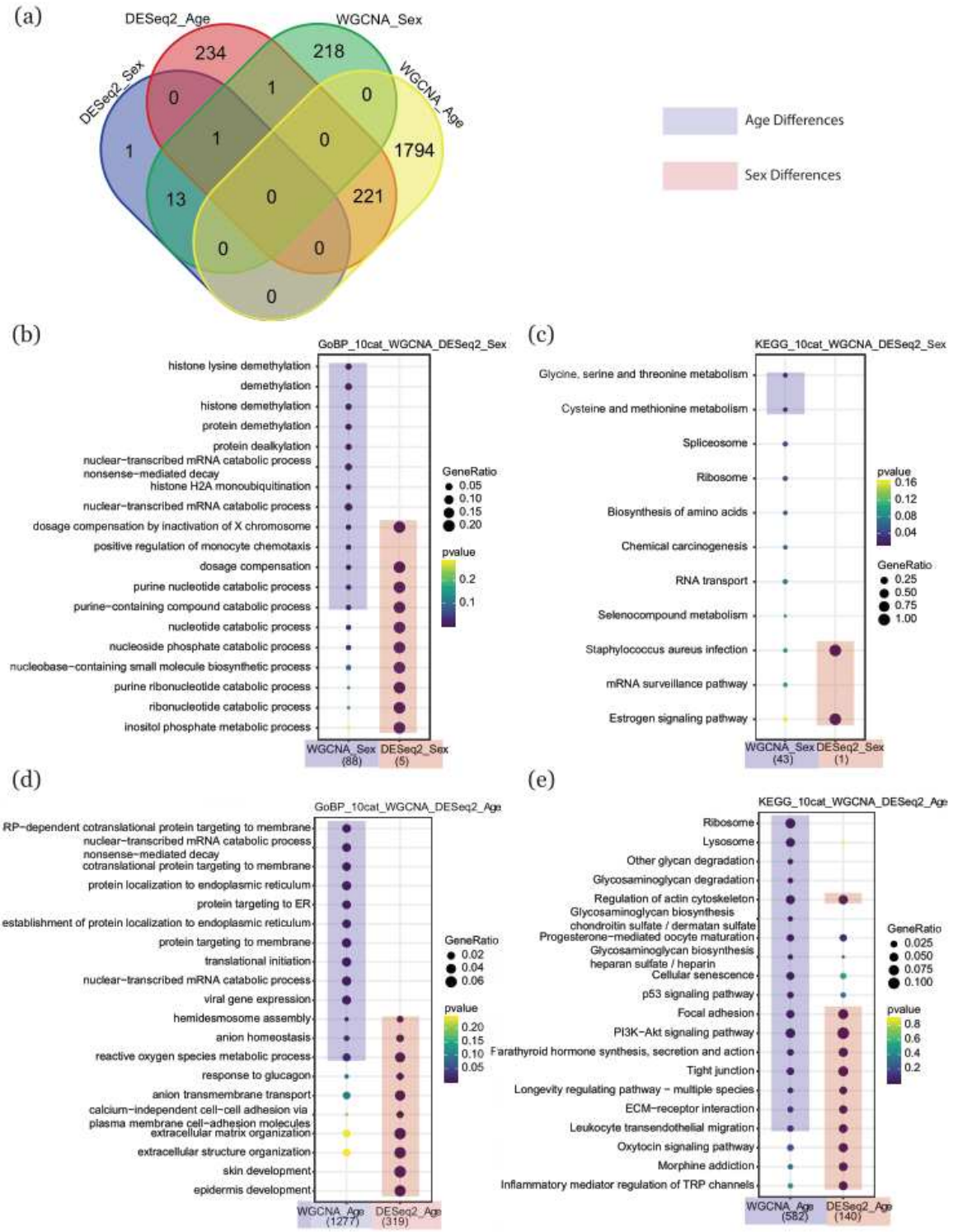


Figure 32 Comparison of enriched GO terms and KEGG pathways in separate sets of genes selected by DESeq2 and WGCNA

Fig. 4. Comparison of enriched GO terms and KEGG pathways in separate sets of genes selected by DESeq2 and WGCNA. (a) Venn diagram showing overlap between

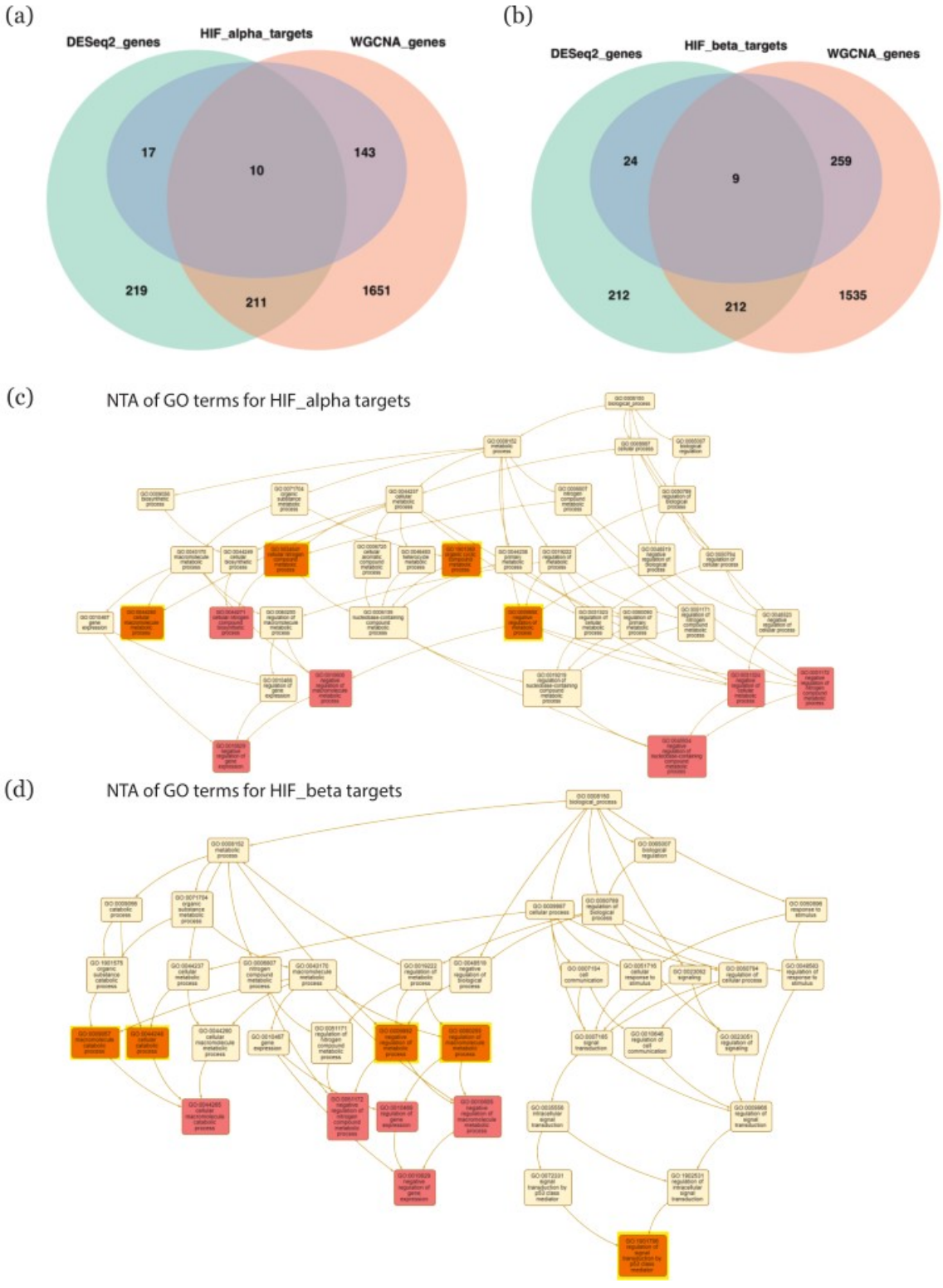
the two methods in the key genes associated with age- and sex-related differences. (b&c) Sex-associated enrichment in GO terms and KEGG pathways identified by the two methods. (d&e) Age-associated enrichment in GO terms and KEGG pathways identified by the two methods. The size of each point represents the magnitude of enrichment, plum shading represents an association with age and coral shading represents an association with sex. GO: gene ontology; KEGG: Kyoto encyclopedia of genes and genomes; WGCNA: weighted gene co-expression network analysis.

5. Enrichment in HIF targets

Because all samples were collected during a period in which cytotrophoblasts are adapting to an increase in O_2 levels, we specifically investigated the involvement of HIF in the biological processes in these cells. HIF transcriptional activity depends on its alpha and beta subunits; we therefore assembled a database of all published or predicted targets of HIF-1α or HIF-1β from the literature and public databases (details in Supplementary Table S1). For HIF-1α, we retrieved a total of 2213 targets, of which 170 corresponded to key genes identified by our analyses (Figure 5a). For HIF-1β, we retrieved a total of 3134 targets, of which 294 corresponded to key genes (Figure 5b). Of these 464 key genes, 461 were age-associated and 3 were sex-associated. Using the groups of HIF-1α or HIF-1β targets, we next performed network topology analyses of the GO biological processes associated with each group. For targets of HIF-1α, the resulting hierarchy of GO biological processes highlighted, in order, the terms “cellular macromolecule metabolic process”, “cellular nitrogen compound metabolic process”, “organic cyclic compound metabolic process”, and “negative regulation of metabolic process” (Figure 5c). For targets of HIF-1β, instead, the dominant hierarchy of terms featured “regulation of signal transduction of p53 class mediator”, “regulation of

macromolecule metabolic process", "negative regulation of metabolic process", "cellular catabolic process" and "macromolecule catabolic process".

Using clusterProfiler, we also analyzed enrichment in GO biological processes and KEGG pathways within each group of targets. Figures 5e&5g depict enrichment patterns for the alpha subunit, while Figures 5f&5h provide the same information for the beta subunit. With respect to GO biological processes, targets of HIF-1α demonstrated enrichment in terms that were mainly oriented around metabolic processes, such as "sterol metabolic process", "cholesterol metabolic process", and "alcohol metabolic process" (Figure 5e). Instead, targets of HIF-1β were mostly associated with terms linked with biological regulation such as "autophagy", "regulation of signal transduction by p53 class mediator", and "regulation of protein catabolic process" (Figure 5f). With respect to KEGG pathways, targets of the alpha subunit were associated with enrichment in regulatory pathways such as "human T-cell leukemia virus 1 infection", "lysosome", and "p53 signaling pathway" (Figure 5g), while targets of the beta subunit were linked with pathways such as "lysosome", "mTOR signaling pathway", and "autophagy" (Figure 5h).


(a)
DESeq2_genes
HIF_alpha_targets
WGCNA_genes
17
10
143
219
211
1651
(b)
DESeq2_genes
HIF_beta_targets
WGCNA_genes
24
9
259
212
212
1535
(c)
NTA of GO terms for HIF_alpha targets
(d)
NTA of GO terms for HIF_beta targets

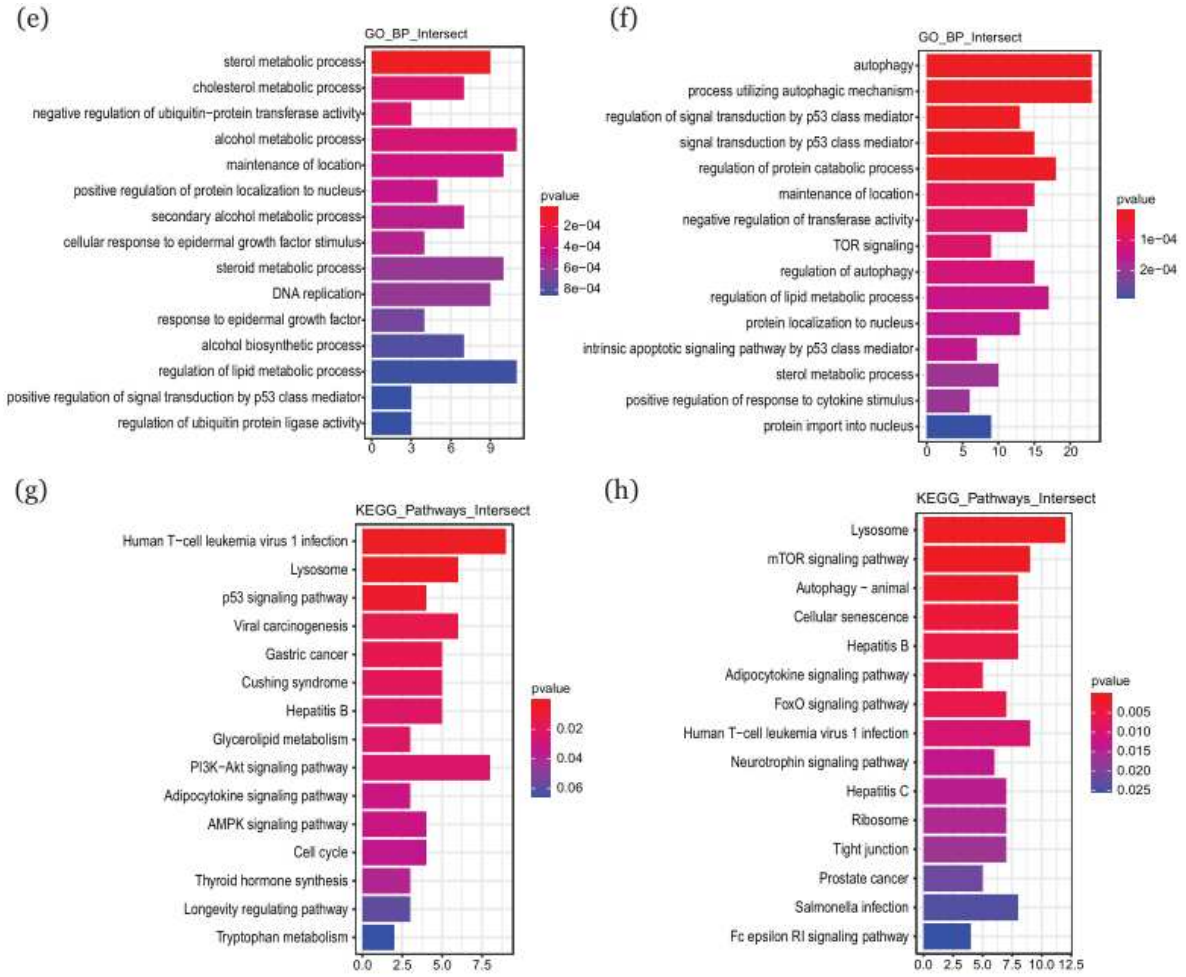


Figure 33 Enrichment patterns in HIF targets

Fig. 5. Enrichment patterns in HIF targets. (a&b) Venn diagram showing intersections between key genes selected by DESeq2 or WGCNA and targets of HIF-1α or HIF-1β, respectively. (c&d) Network topology analysis for enriched GO biological processes associated with targets of HIF-1α or HIF-1β, respectively. (e&g) Enriched GO terms and KEGG pathways, respectively, associated with targets of HIF-1α (f&h) Enriched GO terms and KEGG pathways, respectively, associated with targets of HIF-1β. Significance was defined as P value less than 0.05. HIF: hypoxia-inducible factor; GO: gene ontology; KEGG: Kyoto encyclopedia of genes and genomes; WGCNA: weighted gene co-expression network analysis.

6. Selection of GO terms and associated genes most-affected by HIF

To identify the pathways and genes that were most-affected by HIF, we restricted our set of enriched terms by applying more stringent selection criteria: an FDR threshold less than 0.05 (-log10 of FDR more than 0.69) and log2 enrichment ratio higher than 2. Of all the GO terms and KEGG pathways that were linked with HIF targets, the only ones that met these criteria were two GO terms associated with HIF-1β: "regulation of signal transduction by p53 class mediator" and "TOR signaling" (Figure 6a, in the top-right quadrant). Additional information regarding further evaluation of GO terms and KEGG pathways can be found in the supplementary materials (Figure S4). We then visualized the associations between these pathways and their target genes in a heatmap, with the y-axis representing the pathways and the x-axis representing the genes involved (in red, Figure 6b). Expression data were extracted for the genes associated with these terms, and the changes with respect to age were plotted in violin graphs (Figures 6c&6d).

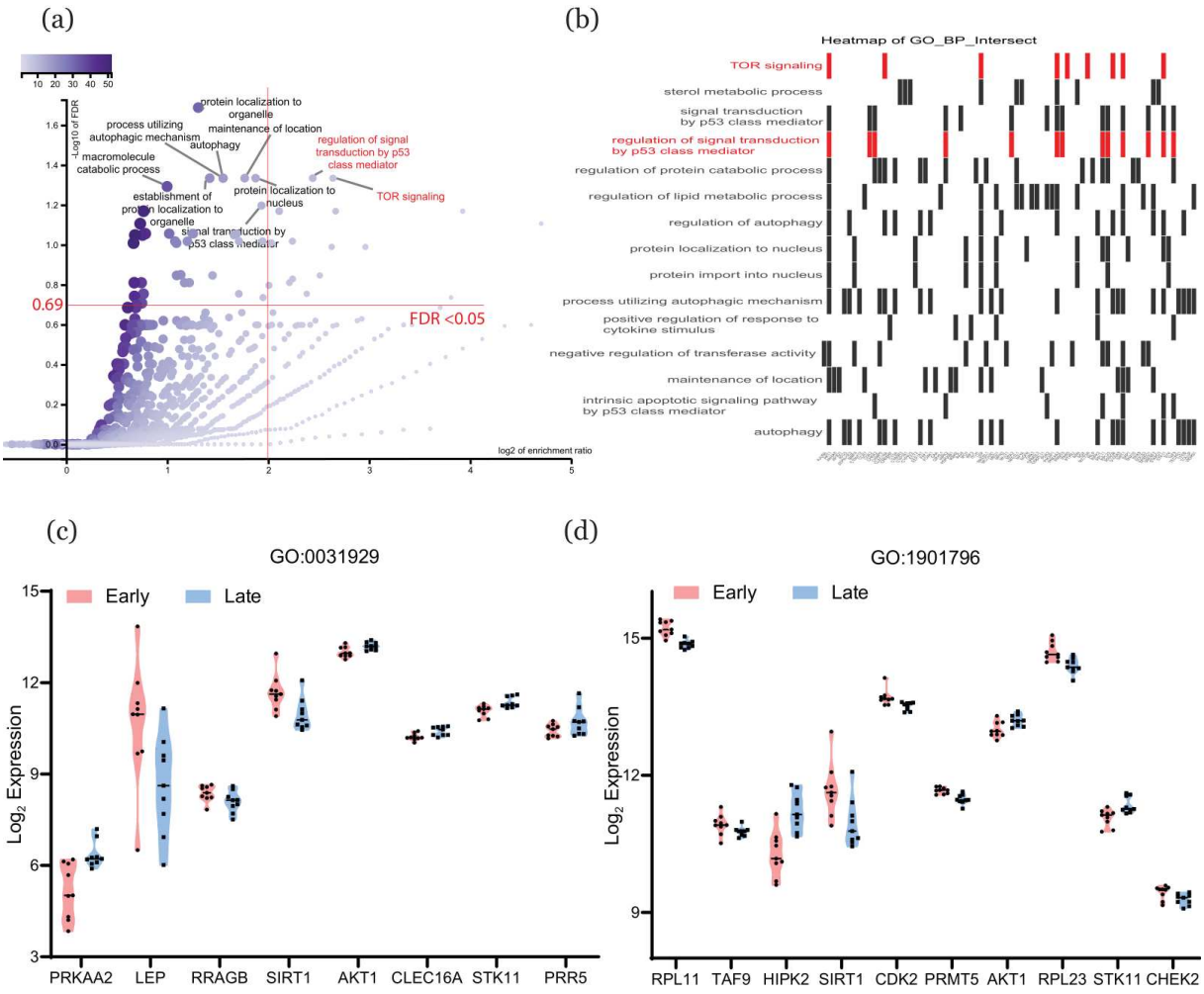


Figure 34 Selection of GO terms and associated genes that were most affected by HIF

Fig. 6. Selection of GO terms and associated genes that were most affected by HIF. (a) Scatterplot of GO terms, with the y-axis representing -log10 FDR and the x-axis representing log2 enrichment ratio. (b) Heatmap of enriched GO terms among HIF targets, with GO terms on the y-axis and the genes associated with each term on the x-axis. (c&d) Expression patterns in early and late placenta samples of genes linked to "regulation of signal transduction by p53 class mediator" (GO:0031929) and "TOR signaling" (GO:1901796), respectively. GO: gene ontology; HIF: hypoxia-inducible factors; FDR: false discovery rate.

4.1.5 Discussion

The human placenta functions as a biological barrier between the mother and the fetus that facilitates the exchange of gases, nutrients, and wastes. Throughout pregnancy, perfusion of the chorionic villi by blood from the uterine spiral arteries is essential for placental development and therefore for the exchanges between maternal and fetal blood. Physiological obturation of the uterine spiral arteries by extravillous trophoblasts plugs in the early first trimester is critical for the growth of chorionic villi and development of feto-placental vasculatures. Moreover, the low physiological O_2 tension within the intervillous space before 10 GW limits the oxidative stress of the chorionic villi, which at this point do not express enzymes that protect from ROS. From 10 to 12 GW, the trophoblastic plugs disintegrate, allowing oxygenated maternal blood into the intervillous space and dramatically raising the oxygen tension. Proper blood flow is thus shaped by oxygen conditions in early placentation, especially in the period from 8-12 GW when oxygen levels increase dramatically [269]. Oxygen levels also regulate the invasion of extravillous cytotrophoblasts into the maternal uterus [223, 224]. Abnormalities in these early physiological processes can lead to pregnancy diseases such as spontaneous abortion, preterm birth, intrauterine growth restriction and preeclampsia [270, 271]. In our study, we focused on the development of normal villous cytotrophoblasts from 8-14 GW, using next generation sequencing technology to investigate changes in gene expression between the early (8-10 GW) and late (12-14 GW) stage of this period.

Although several sequencing datasets have been published for cytotrophoblasts in this period, they have all been focused on different hypotheses. Our data provide what is to our knowledge the first overview of the changes in gene expression that accompany the dramatic increase in oxygen levels from 8-14 GW. A search of publicly available GEO DataSets (https://www.ncbi.nlm.nih.gov/gds/) with the key words "human AND placenta" returned a total of 320 accessions that contained microarray

or RNAseq data from *Homo sapiens*. Of these, however, only a few had performed tissue-specific sequencing or share even superficial similarities with our study [177, 290-293]. For example, Sitras et al. [292] compared microarray-based transcriptomes of first trimester and term human placentas, but their analysis focused only on gestational age and did not investigate the effects of sex. Soncin et al. [177] performed genome-wide expression profiling of human placentas from 4 to 16 GW and at 39 GW with the goal of performing a comparative analysis of mouse and human placentas across gestation, but they did not specifically investigate oxygen-related mechanisms or sexual dimorphism. Of the studies that have examined the effects of hypoxia in human placentas, the samples that were sequenced—term trophoblasts [294], first trimester trophoblasts [295], and extravillous trophoblast cultured at different concentrations of oxygen [296]—all differed from the present study.

To investigate sexual dimorphism in the gene expression of human placentas, Braun et al. [290] surveyed the human chorionic villus transcriptome from 11 to 16 GW for sex-linked signatures, with the goal of characterizing genes that are differentially expressed within the first window of increasing testis-derived androgen production in the male fetus. That study was similar to two others [291, 293] that also focused on sex-based differences in the human placental transcriptome in the late first trimester, with minor differences in the cells or tissues examined. Here, we considered not only differences based on sex but also those based on age in our evaluation of the effect of oxygen tension. In sum, although previous studies have examined the same period of development as the present study, the purposes of their investigations differed substantially. Furthermore, no previous study has taken into consideration the limitations of the use of only one method in exploring these kinds of data. With respect to this kind of analysis, the Limma, edgeR or DESeq2 methods represent improvements over Student's test, but ignore the data connections in the matrix that

arise as a result of topology [297]. To address these limitations, our analysis followed the example of several previous studies [298-301] and combined the methods of DESeq2 and WGCNA.

With regard to gestational age, the results of our analysis of term enrichment were largely consistent with reports from the literature. For example, our results indicated the involvement of the "PI3K-Akt signaling pathway", which has been implicated in the decidualization of trophoblasts in early pregnancy [191]; the "hippo signaling pathway", which has been reported to control the self-renewal of cytotrophoblasts and protect against early pregnancy loss in humans [302]; the "cAMP signaling pathway" and "rap1 signaling pathway", which regulate placental cell fusion [193, 194]; and the "notch signaling pathway", which plays a critical role in the motility and differentiation of cytotrophoblasts [303]. Our analyses also highlighted the terms "estrogen signaling pathway" and "protein targeting to ER", which may reflect reports that increased estrogen levels have a major role in regulating placental secretion of macrophage migration inhibitory factor, a proinflammatory cytokine involved in pregnancy [304]. Finally, we compared our data to those of Soncin et al. [177] by extracting the corresponding samples (8-10 GW and 12-14 GW) from GSE100051 and performing the DESeq2 analysis. This highlighted a total of 185 key genes (Figure S5b), representing enrichment in 145 terms of GO biological processes, compared to 387 in our study (Figure S5c&d). This discrepancy could be an indicator of a higher degree of resolution and accuracy in our study than in this earlier work.

From the DESeq2 analysis that considered only differences related to sex, and not age, we obtained only 15 key genes. What was interesting, though, was that this number apparently increased after we divided the groups based on sex and compared the early and late stages separately (41 genes in early and 44 genes in late). That is, the difference between sexes was partly obscured in the age-mixed set of samples, which indicated

that, in the first trimester of pregnancy, gestational age exerted a stronger influence on the development of the placenta than sexual dimorphism did. This was similar to a report that age appeared to be more dominant than sex in affecting early fetal lung developments from 54-127 days post-conception [305]. When we examined these 85 (41+44) key genes, we found that many of them related directly to the sex chromosome (either X-linked or Y-linked). Within this set of genes, we detected enrichment in GO terms that were mainly associated with catabolic processes, as well as the KEGG pathway "estrogen signaling pathway", which was similar to previously published results [291]. When this approach was combined with WGCNA, the scope of enriched activities was extended to post-transcriptional modifications, such as "ubiquitination", "demethylation", and "dealkylation", which was also consistent with previous research [306-309].

In the placenta, oxygen-sensitive pathways are regulated by the actions of HIF on downstream genes. We thus specifically examined the key genes highlighted by our analyses to identify potential HIF targets. Of the 248 key genes linked with sex-based differences, only 3 were HIF targets: FA complementation group C (FANCC), asparaginyl-tRNA synthetase 2 (NARS2), and RAB38; the latter two genes are known to be active in mitochondria, which could explain their potential correlations with oxygen metabolism. Overall, though, the small number of sexually dimorphic HIF targets could suggest that there is little difference in hypoxia-related biological processes between early male and female placentation. We thus excluded the sex-related genes from our enrichment analyses and focused only on HIF targets that demonstrated age-related expression changes. For the targets of HIF-1α, enrichment analyses highlighted GO terms that were predominantly associated with metabolic processes, and to a lesser extent with biological regulation, while targets of HIF-1β demonstrated the opposite pattern. This could indicate the existence of

complementary roles for HIF-1α and HIF-1β under hypoxic conditions [310]. With respect to enriched KEGG pathways (e.g., "p53 signaling pathway", "PI3K-Akt pathway", "mTOR signaling pathway", and "autophagy), our results were largely in accordance with previous studies. Inhibition of the mTOR signaling pathway has been linked to hypoxia-induced cellular energy conservation, *i.e.*, a decrease in ATP consumption when oxygen is limited [280]. Autophagy plays a critical role in maintaining homeostasis by balancing HIF1α-mediated cellular energy consumption [311], and downregulation of the p53 signaling pathway was reported to drive autophagy in the syncytiotrophoblast [312]. Inhibition of the PI3K-Akt pathway predisposed first-trimester trophoblasts to oxygen-induced cell death [313]. From this set of terms, we wanted to select the most critical for further exploration; to do this, we restricted the FDR threshold to less than 0.05 (-log10 of FDR more than 0.69) and set the log2 enrichment ratio as more than 2. These stringent criteria filtered out all terms except the GO biological processes "regulation of signal transduction by p53 class mediator" and "TOR signaling". If we relaxed the threshold for log2 enrichment ratio to 1, then many other terms also met the requirement, including "mTOR signaling pathway", "autophagy", and "adipocytokine signaling pathway", which, as mentioned above, have all been shown to be involved in gene regulation under hypoxia. This could suggest that the other pathways highlighted by this analysis, such as "AMPK signaling pathway", "adipocytokine signaling pathway", and "FoxO signaling pathway" have potential for further study.

Nevertheless, this study does present some shortcomings that should be addressed in the future. Firstly, we focused on patterns of sexual differentiation only in villous cytotrophoblasts from first-trimester placentas, while previous research has revealed such patterns in different cell types, such as trophoblast epithelium and villous vessel endothelium from term placenta [314]. Secondly, we encountered the

same problem as Soncin et al. [177], namely, that the collected samples did not cluster into distinct early and late groups (Figures 1c&S5a). This might be due to variations within groups or similarities between groups, which would have weakened the accuracy of the filtering process for key genes. Lastly, because we aimed only to provide an overview of the changes in a specific period of pregnancy, we did not conduct any manipulative experiments to verify the role of selected key genes, although we did compare the individual expression patterns of certain critical genes (see Figure 6).

Conclusions

In conclusion, our results provide a broad perspective of the biological processes that are active in trophoblasts during the surge in physiological oxygen availability, specifically with regard to differences over time and between the sexes, which should open new avenues for future research and contribute to the discovery of possible drug-targeted genes or pathways in the human placenta.

Supplementary Materials: The FastQC quality report for the dataset is in Figure S1, including information on phred scores, the average quality, the peaks and the normalization. Heatmaps of gene expression for the key genes identified by DESeq2 for the comparisons of different groups are shown in Figure S2-1a-f. The enriched terms of GO cellular components and GO molecular functions for the DESeq2 comparison are shown in Figure S2-2. From the WGCNA analysis, the expression of key genes in the modules is shown in separate heatmaps based on age and sex in Figure S3-1. The enriched terms of GO cellular components and GO molecular functions for the combined comparison are shown in Figure S3-2. Details of the evaluation of GO and KEGG pathway enrichment in the HIF targets are shown in Figure S4. A summary of dataset GSE100051 is shown in Figure S5, including the clustering of the collected samples, the expression of selected key genes, and a comparison of the terms. A

detailed list of sources for the targets of HIF-1α and HIF-1β retrieved from the literature and public databases can be found in Supplementary Table S1.

Author Contributions: Conceptualization, Thierry Fournier and Séverine A. Degrelle; methodology, Fulin Liu, Françoise Vibert, and Christelle Simasotchi; software, Fulin Liu; validation, Wencan Zhu and Sophie Gil; formal analysis, Fulin Liu; investigation, Sophie Gil; data curation, Séverine A. Degrelle; writing—original draft preparation, Fulin LIU; writing—review and editing, Séverine A. Degrelle and Thierry Fournier; visualization, Fulin Liu; supervision, Thierry Fournier; project administration, Thierry Fournier. All authors have read and agreed to the published version of the manuscript.

Funding: This work was carried out with the funding support of the China Scholarship Council (CSC), Chegongzhuang Avenue, Beijing 100044, P.R China, Inserm and the Université de Paris; and the Campus France.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Written informed consent has been obtained from the patient(s) to publish this paper.

Data Availability Statement: Our RNAseq dataset has been deposited in the Gene Expression Omnibus public repository (https://www.ncbi.nlm.nih.gov/geo/) under the accession number GSE163023 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE163023).

Acknowledgments: The authors wish to thank the consenting patients and clinical staff of the orthogeny service of the Cochin Port-Royal, and Montsouris Hospitals for providing placental tissues, the Genom'IC facility of the Institut Cochin (INSERM U1016, UMR CNRS8104, Université Paris Descartes) for their technical and scientific

expertise in RNASeq, and Lindsay Higgins for English editing (http://www.englishservicesforscientists.com).

Conflicts of Interest: The authors declare no conflict of interest.

Part IV

5 Part Four

Lead-in

In the first part, we have confirmed the importance of the PPAR signaling pathway as well as the existence of PPARγ. In the second part, we also discovered the involved genes and pathways that are important to placental development and that have been regulated by PPARγ. Even though we failed to find the common elements between the HIF targets and the PPARγ cofactors, it doesn't affect the crucial role that PPARγ has functioned in the processes of placentation. Since we have known the significance of PPARγ in placentation, we were wondering if we could find a way to apply the PPARγ in real clinical practice. One way that occurred to our mind is to apply the single nucleotide polymorphisms (SNPs) of PPARγ in disease prediction. SNPs, known for their contribution to the susceptibility of diseases, can be used to predict diseases by combining with other clinical historical information from the previous research. Therefore, if PPARγ is that critical to embryo development, its SNPs should also be possible to be of some predictability for pregnant disease diagnosis. In this part, we are going to collect variants of PPARγ and the clinical information of the pregnant from the normal and preeclamptic. By selecting the important characteristics using different statistical methods and applying machine learning algorithms, we are going to build various models to predict the disease and choose the optimum for the following prediction. Meanwhile, we are going to build a practicable flowchart for clinical use. Just imagine, here comes a new arrival to the hospital, we just need to detect the variants of PPARγ and ask for some information, and then we can predict the occurrence of the disease. This would help a lot for the diagnosis and future treatment.

5.1 C1431T variant of PPARγ is associated with preeclampsia in pregnant women

Fulin Liu [1], Christine Rouault [2], Karine Clément [2,3], Wencan Zhu [4], Séverine A. Degrelle [1,5], Marie-Aline Charles [6,7], Barbara Heude [6*], Thierry Fournier [1*]

[1] *Université de Paris, INSERM, «Pathophysiology & Pharmacotoxicology of the Human Placenta, Pre & Postnatal Microbiota», 3PHM, Paris, F-75006, France*

[2] *Sorbonne Université, INSERM, "Nutrition et Obesités: Approches Systémiques Research Unit", Paris F-75013, France*

[3] *Assistance Publique hôpitaux de Paris, Pitié-Salpêtrière hospital, Nutrition department, Paris F-75013, France*

[4] *Université Paris-Saclay, AgroParisTech, INRAE, UMR MIA-Paris, 75005, Paris, France*

[5] *Inovarion, Paris, F-75005, France*

[6] *Université de Paris, Centre for Research in Epidemiology and Statistics (CRESS), INSERM, INRAE, F-75004 Paris, France*

[7] *Unité mixte Inserm-Ined-EFS ELFE, Ined, Paris, France.*

Correspondence:

* Contribute equally.

5.1.1 Abstract

Introduction: Peroxisome proliferator-activated receptor γ (PPARγ) is essential for placental development, whose polymorphisms increase susceptibility to some pregnancy-related diseases. For example, the placental dysfunction associated with preeclampsia has been linked to disturbance in PPARγ in which genetic polymorphism could play a role. Our aim was to investigate the genetic risk factors for this condition

and build a pragmatic model for preeclampsia prediction.

Methods: Data were collected from a total of 1648 women from the EDEN mother-child cohort study. Numerous clinical characteristics were recorded, along with genotype data for three PPARγ polymorphisms: Pro12Ala, C1431T, and C681G. Univariate analysis was performed to compare the 35 preeclamptic patients to the 1613 control women. Highly correlated characteristics of interest were identified by using three methods of feature selection methods and manual curation; eight different machine learning algorithms were then applied to create predictive models. Model performance was evaluated based on metrics of accuracy and the area under the receiver operating characteristic curve (AUC).

Results: The C1431T polymorphism of PPARγ was the only factor that was significantly associated with preeclampsia ($p < 0.05$) in univariate analyses, with an odds ratio ranging from 4.90 to 8.75. The process of feature selection and manual curation also suggested the inclusion of maternal C1431T and C681G variants as factors, as well as clinical characteristics associated with pregnancy or delivery times, body mass index, education, and cigarette use. Of the machine-learning algorithms tested, the boost tree-based model performed the best, with accuracy and AUC values in the training set as 0.971±0.002 and 0.991±0.001, respectively, and in the testing set as 0.951 and 0.701, respectively. A flowchart of the final tree was constructed to depict the procedure for preeclampsia prediction.

Conclusion: Our results show for the first time that the C1431T variant of PPARγ can play a role in determining susceptibility to preeclampsia. The decision tree created here—based on multiple predictive factors, including the C1431T and C681G variant of PPARγ, pregnancy or delivery times, body mass index, education, and cigarette use—could have applications in the clinical prediction of preeclampsia in the very early

stages of pregnancy.

Keywords: PPARγ; SNPs; machine learning; models; preeclampsia

5.1.2 Introduction

Preeclampsia, which is characterized by high blood pressure and concurrent proteinuria, is a complication of pregnancy that usually manifests after 20-25 weeks of pregnancy [315]. This disease is highly associated with morbidity and mortality for both the mother and the fetus because of its serious risks to fetal maturity and the maternal cardiovascular system [316]. Preeclampsia occurs in 5% to 7% of all pregnant women, leading to over 70 000 maternal deaths and 500 000 fetal deaths worldwide every year [317]. Between 2010 and 2016, an estimated 5.2% of pregnancies in France were affected by gestational hypertension, with 2% of pregnancies developing preeclampsia/eclampsia [318]. Many attempts have been made to accurately diagnose preeclampsia in the early stages; typically, these are based on strategies, such as analyses of metabolomic pathways and combined metabolomic-proteomic data [319-321]. Currently, the most promising method for diagnosis of preeclampsia is the detection of combined biomarkers such as soluble FMS-like tyrosine kinase 1 (sFLT1), soluble Endoglin (sEng), and placental growth factor (PlGF) This has achieved 89% predictive accuracy in pregnancies < 32 gestational weeks, 75% in pregnancies < 37 weeks, and 47% in pregnancies ≥ 37 weeks [322, 323]. Despite the advances that have been made, then, there is still clearly room for improvement.

One novel diagnostic tool may be the use of genetic analysis. The protein product of the FLT1 gene, at locus rs4769613, has been identified as a pathogenetic factor for

susceptibility to preeclampsia in pregnancy [324], as has locus rs9478812, located in an intronic region of the protein PLEKHGI [325]. Another relevant candidate could be peroxisome proliferator-activated receptor γ (PPARγ), of which multiple variants have been implicated in the development of numerous disorders. For example, C1431T has been associated with susceptibility to obesity in the European population [326-328]; C161T (rs3856806) has been linked to the risk of essential hypertension and premature acute myocardial infarction [329-331]; C681G was associated with accelerated growth in young schoolchildren and increased adult height [332]; and the PPARγ rs3856806 C-T substitution polymorphism was found to increase the risk of colorectal cancer [333]. Moreover, previous work by our group has demonstrated an association between two variants of PPARγ-- Pro12Ala and C1431T-- and gestational diabetes [334]. However, the association between PPARγ polymorphisms and the risk of preeclampsia has been merely investigated.

PPARγ is a member of the nuclear hormone receptor subfamily that functions as a transcription factor by binding to target genes, many of which are involved in metabolic processes such as adipogenesis and lipogenesis, insulin sensitivity [335], and immunological processes such as inflammation and differentiation [336]. Complete knock-out of the PPARγ gene can lead to embryonic lethality [167], while deficiency results in insufficient trophoblast differentiation and abnormal vasculogenesis in mice [147]. It also suggested that PPARγ played a role in elevating blood pressure, proteinuria, endothelial dysfunction, platelet aggregation *i.e.*, key features of preeclampsia in rats [337]. Given that single nucleotide polymorphisms (SNPs) in the PPARγ gene have been implicated in a wide range of diseases, we hypothesized that such variants may also play a role in preeclampsia.

In the current study, we aimed to investigate the association of the Pro12Ala, C1431T, and C681G polymorphisms of the PPARγ gene with the risk of preeclampsia. Together with clinical characteristics from our nation-wide EDEN cohort study [338], these data were used to build a model for preeclampsia prediction. The conventional way to build a model is to apply generalized linear models, which are easy and fast to implement. However, erroneous specification of model parameters or assumptions can lead to biases in the results. Here, instead, we applied new machine-learning methods that are able to fully consider complex relationships between the predictors and the outcome with the fine arguments tuning. In this way, potential bias can be, to some extent, diminished. A summary of the study procedure is shown in Figure 1.

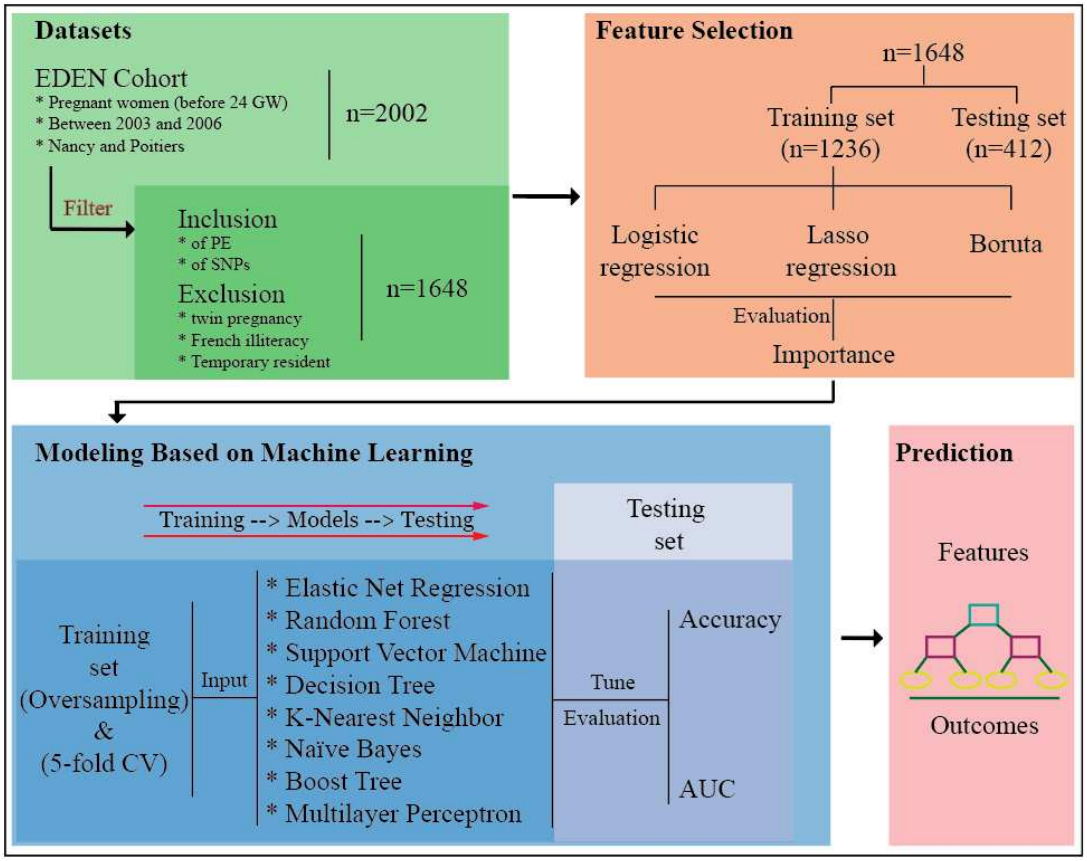


Figure 35 Schematic diagram of the study

Figure 1. Schematic diagram of the study. Pregnant women from the EDEN mother-child cohort study who satisfied the inclusion and exclusion criteria were

recruited. The dataset was then randomly divided into two parts: a training set and a testing set, according to a 3:1 ratio. Three methods were used to evaluate the importance of highly correlated features within the training set: logistic regression, lasso regression, and the Boruta algorithm. Eight machine-learning models were then built, tuned, and trained on an oversampled training set with 5-fold cross-validation (CV); this was followed by validation on the testing set. The performance of the models was evaluated using metrics of accuracy and the area under the receiver operating characteristic curve (AUC). The final model was used to build a decisive procedure for predicting preeclampsia. GW: gestational week; PE: preeclampsia; SNPs: single nucleotide polymorphisms.

5.1.3 Methods

Study population

The EDEN study (Study of pre-and post-natal determinants of children's growth and development) is an ongoing mother-child cohort study that was set up in two locations in France, Nancy and Poitiers (France). A total of 2002 pregnant women were enrolled, and successfully delivered babies will be followed until their fifth birthday. More details about the EDEN study are available in [338]. The study received approval from the ethics committee (CCPPRB, N°02-70, 12 December 2002) of Kremlin Bicêtre and from CNIL (Commission Nationale Informatique et Liberté), the French data privacy institution. Written informed consent was obtained twice from parents, once at enrollment and again after the child's birth. The study was approved by the ethics research committee (Comité Consultatif de Protection des Personnes dans la Recherche Biomédicale) of Bicêtre Hospital and by the Data Protection Authority (Commission Nationale de l'Informatique et des Libertés). All research was performed

in accordance with the relevant guidelines and regulations. Of the 2002 pregnancies, 1648 met the inclusion/exclusion criteria (Figure 1) to be included in the present work.

Socio-demographic and clinical features

At 24-28 gestational weeks, each mother was clinically examined and asked to complete a self-administered questionnaire. Physical characteristics such as maternal weight and height were measured during the examination, while data on personal history such as weight before pregnancy, educational level, and smoking habits were collected during an initial interview. Additional clinical characteristics such as gestational age at delivery and the number of previous pregnancies were extracted from clinical records. Body mass index (BMI) was calculated according to the formula: BMI = kg/m^2, where kg is the weight in kilograms and m^2 is the height in meters squared. Data are presented as mean ± S.D. for continuous features and as percentages (N) for discrete factors.

Genotyping

Maternal blood samples were collected during pregnancy by up to two technicians and stored in -80°C freezers with alarm control. DNA was extracted from leukocytes using the QIAamp DNA Blood Mini Kit (QIAGEN) according to standard procedures. Genotyping of the Pro12Aa polymorphism was conducted using one of two techniques. For the first 729 women enrolled in the study, a LightCycler apparatus (Roche Diagnostics, Meylan, France) and hybridization probes were used, with Pro12Ala primers and probes designed and synthesized by TIB MOLBIOL (Berlin, Germany). The PCR mixture (10 µl total volume) contained 20 ng of DNA, 1X Fast Start DNA master hybridization probes, 0.5 µM of primers, 0.15 µM of probes, and 3 mM $MgCl_2$. Melting curve analysis was applied to monitor SNP genotyping. For the second group

of women in the cohort (1024 mothers), TaqMan (Applied Biosystems, Foster City, CA) was used. Following similar reagent preparation, the results of TaqMan assays were read on a 7900HT Fast Real-Time PCR System (Applied Biosystems, Foster City, CA), and alleles were called using the SDS software (Applied Biosystems, Foster City, CA). For the C1431T and C681G polymorphisms, TaqMan procedure was applied under the same standard operating protocol. DNA samples were amplified by PCR on a 96-well plate with the following cycling parameters: denaturation at 95°C for 10 min, and, 40 cycles at 92°C for 15 sec, 60°C for 1min. The genotyping call rate of the three SNPs was above 98% in each case, including for with the duplicate controls. Details on the genotyping primers, probes, and PCR conditions are available from the corresponding author.

Basic statistical analyses

Maternal clinical characteristics were described separately in women with and without preeclampsia. Student's t-test was used to compare continuous features between groups, while a chi-square test was used for discrete features. Fisher's exact test was applied when any of the cell values of a contingency table were below five. Multivariate logistic regression was used to calculate the odds ratio for preeclampsia, followed by log-transformation. A p-value less than 0.05 was considered statistically significant. Since each SNP can represent either a major allele (M) or a minor allele (m), the genotype can be a major allele homozygote (MM), a heterozygote (Mm), or a minor allele homozygote (mm). Thus, we performed the comparison of allele frequencies among groups using one of three models: a dominant model (MM versus Mm + mm), a recessive model (MM + Mm versus mm), and a co-dominant model (MM + mm versus Mm).

Feature selection

Following the random division of the original dataset into the training and validation sets, important features were selected in the training set using three algorithms: logistic regression, lasso regression, and the Boruta algorithm. Soft thresholds were determined based on the principle that the importance of the features should be more than the mean importance of all features. Therefore, the soft threshold was set as 0.5 in lasso regression, 1 in logistic regression, and 2 with the Boruta method. The features that were highlighted by the three methods were then curated manually based on the preliminary screening of clinical characteristics using univariable logistic regression analysis, odds ratios, and clinical knowledge.

Modeling and evaluation

Using the features selected by the previous step, we chose eight widely used machine learning algorithms (elastic net regression, support vector machine, random forest, boost tree, decision tree, k-nearest neighbor, naïve Bayes, and multilayer perceptron) to build and evaluate models, along with argument tuning using the 1000-candidate maximum entropy design, an optimal design of argument combination based on Shannon's definition of entropy as the amount of information. Before modeling, the data in the training set were balanced to obtain a ratio of positive to negative cases of 3:5. The balanced training set was subsequently resampled with five-fold cross-validation, accompanied by two sets of repeats. To evaluate the performance of the models, the receiver operating characteristic (ROC) curves and the area under the receiver operating characteristic curve (AUC) values were used. Specifically, the closer the AUC value is to 1, the better the performance. The quality of each model was also evaluated using metrics of accuracy, sensitivity, specificity, and the adjusted F1-score

which were calculated based on the confusion matrix in Table 1 and equations 1-5. The testing set was retained for final validation.

Table 1 Confusion matrix

Table 3 Confusion matrix

		True Condition	
		Condition positive	Condition negative
Predicted condition	Predicted condition positive	True positive (TP)	False positive (FP)
	Predicted condition negative	False negative (FN)	True negative (TN)

$$\text{Accuracy} = (TP + TN) / (TP + FN + TN + FP) \quad (1)$$

$$\text{Sensitivity (Recall)} = (TP) / (TP + FN) \quad (2)$$

$$\text{Specificity} = (TN) / (TN + FP) \quad (3)$$

$$\text{Precision} = TP/(TP+FP) \quad (4)$$

$$F1 = 2 \times \text{Precision} \times \text{Recall} / (\text{Precision} + \text{Recall}) \quad (5)$$

Applied R packages

Statistical analyses, feature selection, and modeling were conducted in R software (version 4.0.4) with basic packages in Rstudio (PBC, Boston, MA, http://www.rstudio.com/), an integrated development environment for R. The distribution of missing data was visualized using the R package mice (version 3.11) [339], and imputation of missing data was carried out using the R package missForest (version 1.4) [340]. To account for the imbalance of positive and negative cases, the R package imbalance (version 1.0.2) was used to oversample the smaller population

[341]. The Boruta algorithm, was executed using the Boruta package (version 7.0.0) [342]. Principal component analysis of features and individuals was carried out using FactoMineR (version 2.3) and visualized by factoextra (version 1.0.7) [343]. Machine-learning model building was performed with the tidymodels series of packages (https://www.tidymodels.org/) written by the Rstudio team, including tidymodels (version 0.1.2), vip package (version 0.3.2), discrim (version 0.1.1), modelr (version 0.1.8), yardstick (version 0.0.7), workflows (version 0.2.1), tune (version 0.1.2), rsample (version 0.0.8), recipes (version 0.1.15), and parsnip (version 0.1.4). For the reproducible codes, an R script with detailed comments is provided in the Supplement Materials (Text S1).

5.1.4 Results

1 Overview of maternal clinical characteristics

Table 2 presents a summary of maternal clinical characteristics of the control and preeclampsia groups. The only factor that was significantly different between the two groups was maternal expression of the C1431T variant of PPARγ (Table 2). Instead, logistic regression and analysis of the log odds ratio found that expression of this variant in the mother played a significant role in the development of preeclampsia (p-value less than 0.05; Figure 2). The value of the odds ratio value for maternal C1431T ranged from 4.90 to 8.75 (Supplementary Table S2). Similarly, a comparison of the three genotype models confirmed that the maternal C1431T variant was the only factor that made a significant difference (Table 3). Lastly, we impute missing values, which represented 9.14% of the full data set (supplementary Figure S1). A summary of the data before and after imputation, showing no differences between the imputed and non-imputation summary tables, was showed in Table S1.

Table 2. Maternal and fetal clinical characteristics

Table 4 Maternal and fetal clinical characteristics

Clinical characteristics	Factor name	Controls N=1613	Preeclampsia N=35	p
Maternal age at delivery (year)	ageinc	29.6 ± 4.8	28.3 ± 6.1	0.232
Maternal height (cm)	c24_Height	163.0±6.23	165.0±6.35	0.108
Maternal education (level)	c24_edu	6.55±2.47	6.05±2.86	0.358
BMI (kg/m²)	bmi	23.2 ± 4.6	25.2 ± 5.8	0.067
Primiparous	primidelv	44 (703)	60 (21)	0.079
Number of pregnancies	ob_Nbpreg	1.35±1.49	1.37±1.77	0.956
Number of deliveries	ob_Nbdelv	0.834±0.971	0.657±0.906	0.261
Obesity	obese	9(137)	18(6)	0.109
Cigarette use (no.)	nbcig	1.48±3.44	1.09±3.37	0.524
Maternal polymorphism (% (N))				
P12A	carrier12a_M	19.7 (333)	26.3 (10)	0.493
C1431T	carrier1431_M	21.3 (349)	42.9 (15)	0.004
C681G	carrier681_M	39.3 (663)	39.5 (15)	1

Table 3. Comparison for allele genotypes

Table 5 Comparison for allele genotypes

Clinical characteristics	Controls N=1613			Preeclampsia N=35			p1[a]	p2[b]	p3[c]
	c/c	c/t	t/t	c/c	c/t	t/t			
genot1431_M	78.9(1272)	20.2(326)	0.9(15)	57.1(20)	40(14)	2.9(1)	0.004	0.29	0.008
genot681_M	60.3(967)	35.2(565)	4.5(72)	60(21)	34.3(12)	5.7(2)	1	0.67	1
genotp12a_M	80.4(1297)	18.7(301)	0.9(15)	74.2(26)	22.9(8)	2.9(1)	0.49	0.29	0.68

a: p1 for the dominant model

b: p2 for the recessive model

c: p3 for the co-dominant model

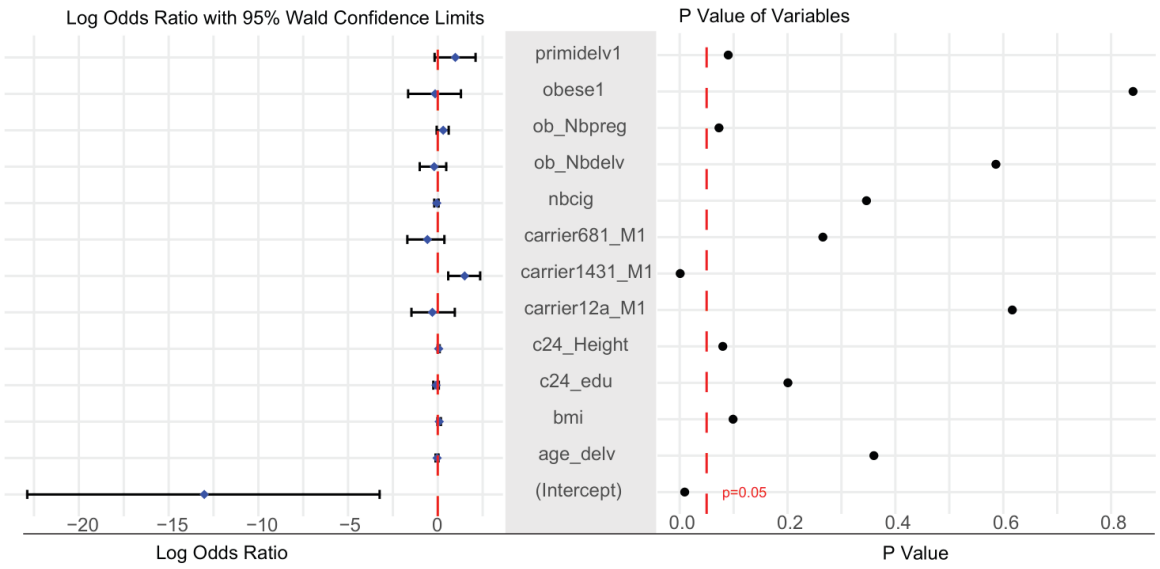


Figure 36 Clinical characteristics evaluation

Figure 2. Evaluation of clinical characteristics. The log odds ratio of all characteristics, with corresponding p-values. Bars indicate mean and 95% confidence interval.

2 Feature selection using three methods: Boruta algorithm, lasso regression, and logistic regression

The Boruta algorithm highlighted six features as important: education, maternal and C681G and C1341T variants, and obesity. When the inclusion criteria were relaxed slightly, delivery age, BMI, and maternal P12A variant were also included (Figure 3A). The features singled out by Lasso regression were maternal C1341T and C681G, and primary delivery (Figure 3B). The features identified as important by logistic regression were maternal carrying of C1341T, number of pregnancies, primary delivery, number of cigarettes, education, and BMI (Figure 3C). Therefore, the final features that included in the model-building process were maternal C681G and C1341T variants, obesity, BMI, number of pregnancies, primary delivery, number of cigarettes, and education.

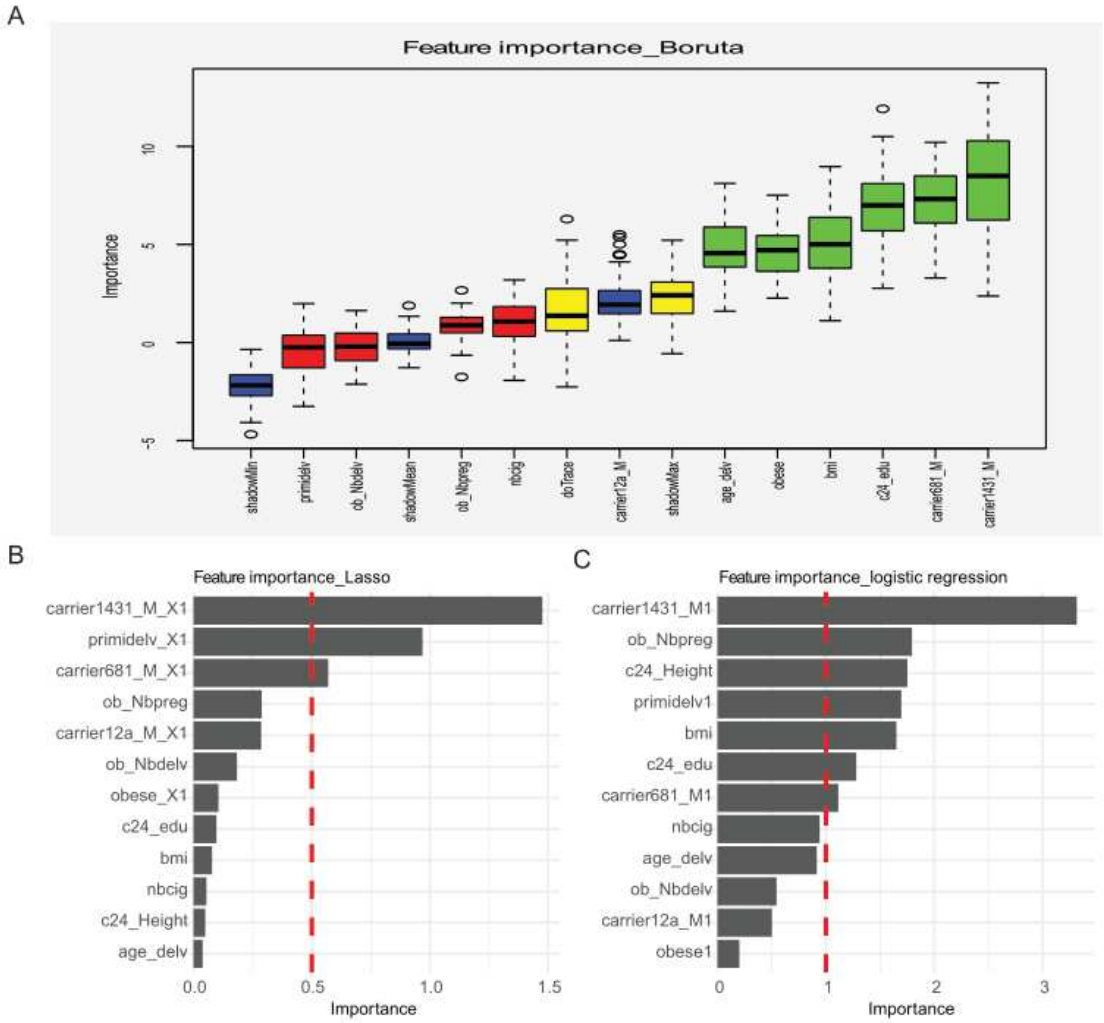


Figure 37 Feature selection

Figure 3. Feature selection. (A) Feature selection with the Boruta method. Blue boxplots indicate different weighted thresholds for selecting features. Red boxplots represent features that were found to be unimportant, while green boxplots are the opposite. Yellow boxplots show features that may be important depending on the criteria used. (B) Selection based on lasso regression with the soft threshold for importance as more than 0.5. (C) Selection based on logistic regression with the soft threshold for importance as more than 1.

3 Modeling based on machine learning

With the selected features, we divided the data set into two parts, the training set, and the testing set. The training set was then balanced with respect to the incidence of preeclampsia, as described in the Methods. Supplementary Table S3 details the clinical characteristics of the original and split datasets; there were no differences between the training and testing set with respect to the representation of categorical factor and the mean and standard deviation of the numeric factors. There was a wide degree of overlap between the distribution of positive and negative cases in the training set before and after balancing (supplementary Figure S2D&E). The optimal combination of features was selected following a thorough process of tuning based on the optimal AUC of models (Supplementary Figure S6).

The results of the eight final machine-learning models with respect to model accuracy and AUC are shown in Table 4 for both the training and testing sets. The optimal model was the boost tree, whose values for accuracy and AUC in the training set were 0.971 and 0.991, respectively, and 0.951 and 0.701, respectively, in the testing set. The diagnostic performance of each of the machine-learning models (AUCs) is depicted in Figure 4.

Table 4. Prediction of the 8 Models by ML Analysis

Table 6 Prediction of the 8 Models by ML Analysis

	Train		Test	
	Accuracy	AUC	Accuracy	AUC
Elastic Net Regression	0.661±0.005	0.695±0.006	0.857	0.784
Random Forest	0.913±0.006	0.969±0.003	0.896	0.723
Support Vector Machine	0.772±0.003	0.847±0.004	0.862	0.545

Decision Tree	0.849±0.007	0.919±0.006	0.874	0.579
K-Nearest Neighbor	0.826±0.006	0.917±0.006	0.801	0.725
Naïve Bayes	0.693±0.005	0.787±0.007	0.930	0.619
Boost Tree	**0.971±0.002**	**0.991±0.001**	**0.951**	**0.701**
Multilayer Perceptron	0.899±0.007	0.919±0.006	0.811	0.670

AUC, area under the receiver operating characteristic curve

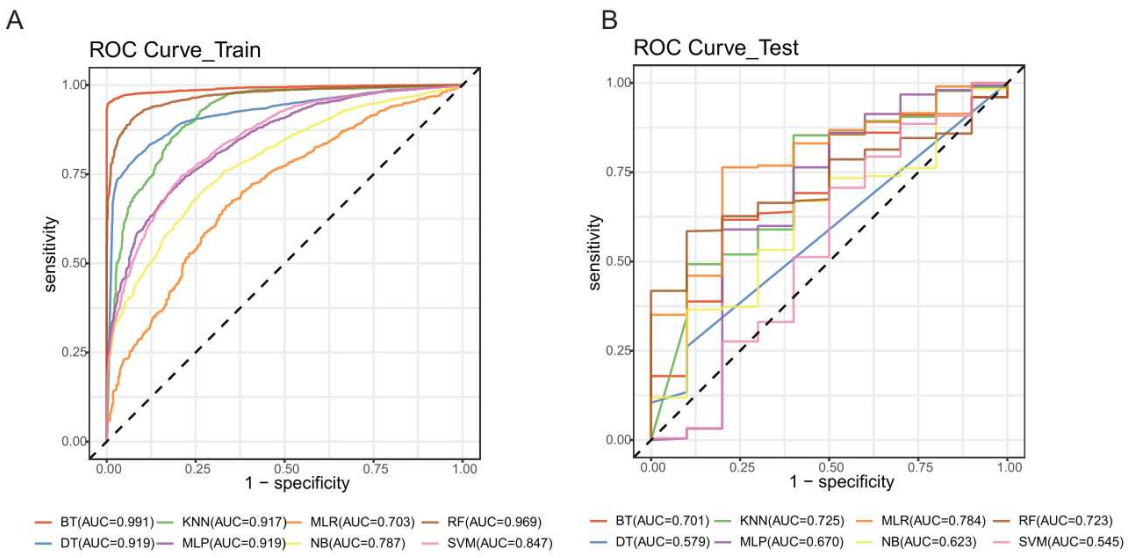


Figure 38 ROC curve of different algorithms

Figure 4. ROC curve of different algorithms. (A) ROC curves with the training set. (B) ROC curves with testing set. The values were shown in the legends. AUC: area under the receiver operating characteristic curve; BT: boost tree; DT: decision tree; ENR: elastic net regression; KNN: k-nearest neighbor; MLP: multilayer perceptron; NB: naïve Bayes; RF: random forest; SVM: support vector machine; ROC: receiver operating characteristic.

4 Prediction procedures of boost tree

We then constructed the boost tree, along with a heatmap of scaled feature values,using the balanced training set (Figure 5). The clinical characteristics that were determined

to be important for prediction included the maternal PPARγ genotypes, primary delivery, number of pregnancies, obesity, BMI, and education. As expected from the univariate tests, the maternal C1431T variant was the first key branching node of the tree. Primary delivery and maternal C681G functioned as the second nodes, followed by the number of pregnancies and BMI, while education played a less important role in the final decision. For the decision branching, for genotypes, "1" represents no mutation in the allele, "2" means a single mutation, and "3" a double mutation; for primary delivery, "1" means "no" while "2" means "yes"; and for obesity, "1" means "no" while "2" means "yes". The corresponding threshold values are shown on the branches. Integrated values of the clinical characteristics of individuals are presented in the heatmap, corresponding to both positive and negative end outcomes. The simplicity of

this procedure is intended to facilitate its possible use in clinical practice for the prediction of preeclampsia.

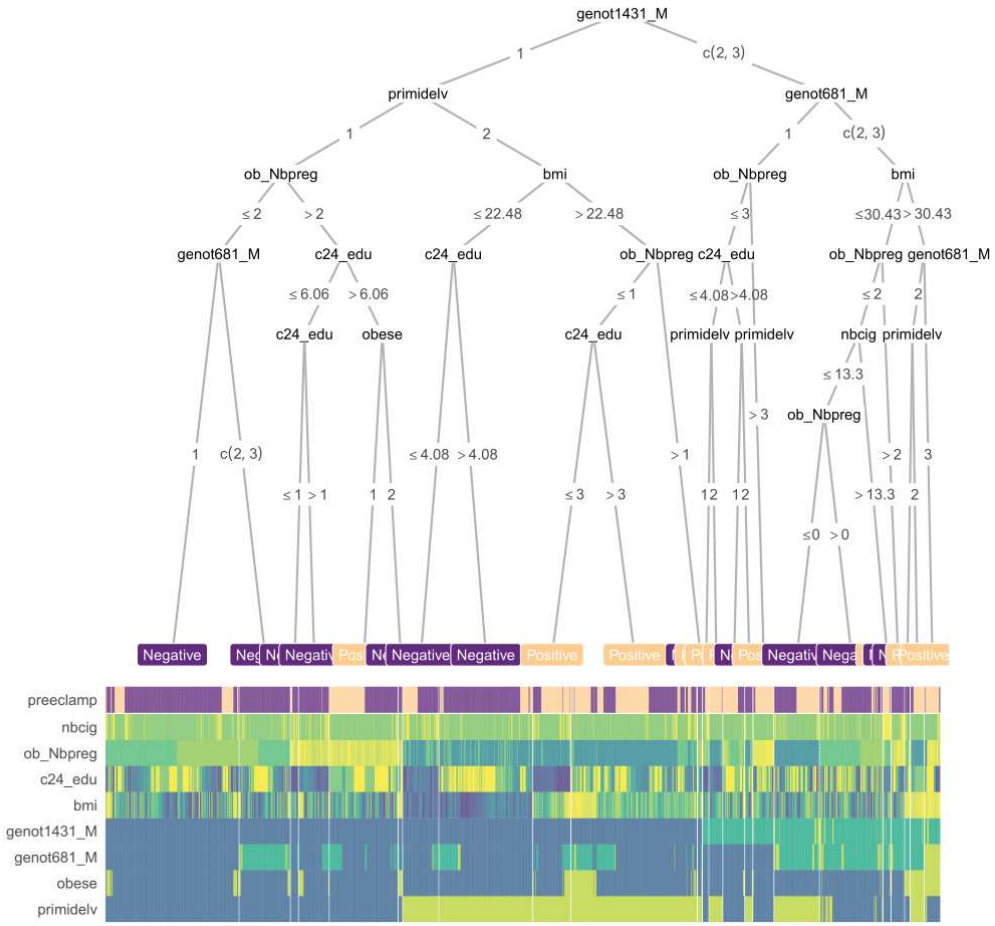


Figure 39 Boost tree-heatmap for predicting preeclampsia

Figure 5. Boost tree-heatmap for predicting preeclampsia. The leaves of boost tree contain the contributive features while branches contain the indicating values. The first row of the heatmap presents the outcomes while the rest presents the predictors. The colors present the scaled value of a sample on each feature.

5.1.5 Discussion

Preeclampsia is a common complication in pregnancy that contributes substantially to morbidity and mortality for both mothers and fetuses. Despite its severity, early diagnosis is often challenging, and there are notable deficiencies in existing counseling

and monitoring systems. The lack of adequate screening methods means that all women who are suspected to have preeclampsia must undergo a series of intensive tests, usually accompanied by time-consuming hospitalizations [344]. A newly proposed method for the diagnosis of preeclampsia relies on combined detection of biomarkers such as sFLT1, sEng, and PlGF, and performs well around 34 weeks gestation [345-347]. However, this method is not universally applicable, as preeclampsia occurring before 32 weeks accounted for only 17% of cases in France [318]. We thus aimed to develop a tool that could be used for earlier diagnosis of preeclampsia. In the present study, we first revealed a significant association between the C1431T polymorphism of PPARγ and preeclampsia, which is different with the previous conclusion that no association between PPARγ polymorphism and the occurrence of preeclampsia [348]. More importantly, we also built a decision tree that represents a possible diagnostic procedure for pragmatic preeclampsia prediction, which could benefit early diagnosis regardless of gestational age.

Multiple reports have links variants of PPARγ with diseases such as coronary heart disease, cancer, metabolic syndrome, and, especially obesity and diabetes [334, 349-352]. The PPARγ gene is located in human chromosome 3p25 and comprises 9 exons. One of the most common structural polymorphisms is a proline (Pro) to alanine (Ala) substitution, Pro12Ala (rs1805192), which results from the mutation of cytosine to guanosine [329]. Compared with the normal Pro allele, the Ala-substituted allele leads to a reduction in activity of PPARγ [353], which can be a high-risk factor for the occurrence of obesity and type 2 diabetes [354, 355]. In addition to the common Pro12Ala variant, the C1431T variant, located in exon 6 of PPARγ, has been associated with the susceptibility to leptin concentrations [326] and body mass index[327, 328], and the C681G variant was associated with accelerated growth in young schoolchildren

and increased adult height [332]. In our study, the presence of C1431T variant of PPARγ in the mother was shown to play a significant role in distinguishing between preeclamptic and normal pregnancies (chi-square test). Instead, no significant association was detected between Pro12Ala/C681G and preeclampsia in either the mother or the fetus. However, even though the chi-square test found no evidence of a link between C681G and preeclampsia, the inclusion of this variant in the final predictive model was found to improve both the accuracy and AUC values of the model.

Machine-learning algorithms are widely used to obtain better prediction accuracy compared to conventional generalized linear models in decision-making scenarios. They offer alternative strategies for the diagnosis of diseases based on clinical characteristics [356-359]. Currently, there are eight machine-learning algorithms in wide use for modeling and building diagnostic procedures based on appropriate medical history [360, 361]. Several studies have compared different machine-learning methods for disease prediction under various clinical conditions, and the results have been mixed [362-364]; this suggests that the optimal algorithm may vary depending on context. In our study, we applied and compared the eight machine-learning algorithms, and in this process, we addressed two common challenges in modeling: insufficiency and overfitting of the models. For the former, we oversampled the positive cases to prevent inaccuracy due to the imbalance between positive and negative cases [365, 366]; both the balanced and unbalanced versions of the training and testing sets were evaluated. For the latter, we used different approaches for preprocessing our dataset and applied 5-fold CV to the training set, followed by validation on the testing set.

First, the original dataset was split into a training set and a testing set without balancing; in this case, the boost tree was the optimal model, with values of accuracy and AUC as 0.99 and 0.92 in the training set, and 0.98 and 0.77 in the testing set, respectively (Table S4 & Figure S3). We then repeated this procedure, but first balanced the original dataset before splitting it into the training set and testing set. The boost tree remained the optimal method with accuracy and AUC values of 0.957 and 0.990 for the training set and 0.975 and 0.996 in the testing set, respectively (Table S5 & Figure S4). We suspected that overfitting may have influenced this model owing to the internal relationship between the training set and the testing set that resulted from the data simulation. Lastly, we balanced the training set only by oversampling the positive cases, and kept the testing set as it was after the split of the original dataset. Those results are shown in Table 4 & Figure 4 that we presented above. Additionally, in the final model, we performed failure mode and effects analysis and calculated the F-score to verify the suitability of accuracy as a metric. We obtained high values for both training and testing sets (Table S6), which were generally in line with the accuracy values. However, the AUC value of the testing set in the final boost-tree model was not high enough to be considered as a convinced example, which we hypothesized due to a shortage of positive cases in the testing set. Further studies with larger datasets are needed to resolve this question.

In our study, the boost tree consistently yielded the highest accuracy and AUC value for both the training and testing sets regardless of the methods used for preprocessing. For this reason, we used this approach to build a clinical flowchart for the evaluation of preeclampsia. This model outperformed both the screening methods currently recommended by the National Institute for Health and Care Excellence (a combination of maternal factors, uterine-artery pulsatility index, mean arterial pressure, and PlGF;

41% accuracy) and ACOG guidelines (94% accuracy) [367], and methods based on biomarkers such as soluble fms-like tyrosine kinase 1 (sFlt1) and PIGF (77% accuracy) [368]. In addition, our model can be used to evaluate women before they become pregnant, as all of the predictors can be examined pre-pregnancy; this thus facilitates earlier diagnosis than existing alternatives [319, 361, 367, 369, 370]. However, despite the high degree of accuracy achieved here, the clear procedure for prediction, and potential for earlier diagnosis, our model also has some deficiencies to improve. First, further studies in other regions or nations are needed since patterns of polymorphism can vary among populations, which can lead to inconsistent conclusions [371]. Second, certain clinical data were missing for some of our study subjects, and future studies on larger samples may be able to avoid this problem. Lastly, a larger number of positive cases should be included to balance the representation of preeclamptic and healthy pregnancies. Even though an appropriate algorithm was used to account for the imbalance here, it is possible that the difference between simulation and real cases may subtly influence the model performance.

In conclusion, by comparing preeclamptic and healthy patients, our study reveals a significant role for a variant of PPARγ, C1431T. By combining data on this variant with information on several clinical characteristics, including the C681G variant of PPARγ, pregnancy or delivery times, body mass index, education, and cigarette use, we built an efficient boost-tree model that is able to predict preeclampsia in very early pregnancy. This model could be invaluable in screening high-risk pregnancies in clinical practice and could serve as a decision-making reference for clinicians.

Declaration of competing interest

The authors declare that they have no competing interests.

Acknowledgments

The authors thank the EDEN mother-child cohort study group, whose members are I. Annesi-Maesano, J.Y. Bernard, M.A. Charles, P. Dargent-Molina, B. de Lauzon-Guillain, P. Ducimetière, M. de Agostini, B. Foliguet, A. Forhan, X. Fritel, A. Germa, V. Goua, R. Hankard, B. Heude, M. Kaminski, B. Larroquey, N. Lelong, J. Lepeule, G. Magnin, L. Marchand, C. Nabet, F Pierre, R. Slama, M.J. Saurel-Cubizolles, M. Schweitzer, and O. Thiebaugeorges.

Supplementary Materials

The supplementary materials provide more details about the analysis process and describe additional analyses that were conducted in order to support the results shown in the text. There is information on the procedures that were used to clean up the dataset (Table S1, Figure S1), the evaluation of clinical characteristics (Figure S2), a summary of the odds ratios of clinical characteristics (Table S2), more information on modeling based on machine learning (Table S3-6, Figure S3&S4, Figure S5), and appendix materials regarding parameter tuning (Figure S6). The R script with detailed comments is provided in Text S1.

Funding Statement

The EDEN study was supported by the Foundation for Medical Research (FRM), the National Agency for Research (ANR), the National Institute for Research in Public Health (IRESP: TGIR cohorte santé 2008 program), the French Ministry of Health (DGS), the French Ministry of Research, the INSERM Bone and Joint Diseases (PRO-A), and Human Nutrition National Research Programs, Paris-Sud University, Nestlé, the French National Institute for Population Health Surveillance (InVS), the French

National Institute for Health Education (INPES), the European Union FP7 Programme (FP7/2007–2013, HELIX, ESCAPE, ENRIECO, Medall projects), the Diabetes National Research Program (through a collaboration with the French Association of Diabetic Patients (AFD)), the French Agency for Environmental Health Safety (now ANSES), Mutuelle Générale de l'Education Nationale (a complementary health insurance; MGEN), the French National Agency for Food Security, and the French-Speaking Association for the Study of Diabetes and Metabolism (ALFEDIAM). This work was also carried out with the funding support of the China Scholarship Council (CSC; Chegongzhuang Avenue, Beijing 100044, P.R China), Inserm, the Université de Paris, and Campus France.

6 Conclusions and Perspectives

My Ph.D. project is studying the role of PPARγ in human trophoblast differentiation and placenta development. Indeed, PPARG is a nuclear receptor that is essential for placentation in mice and humans. In the first part, we mined the datasets downloaded from GEO database and confirmed the importance of the PPAR signaling pathway in placentation. In the second part, with the aid of the microarray technique, we sequenced the isolated placental cells, trophoblasts, that were treated with PPARγ agonist, rosiglitazone, in order to detect the differentially expressed genes or biological pathways that were affected by PPARγ. The change in lipid metabolism was then verified by lipid droplet detection throughout gestation age in placental tissues. We also additionally studied the potential relationship between hypoxia and PPARγ using the RNAseq technique since the placenta suffers from a dramatic increase of oxygen in the first trimester. Lastly, with respect to the importance of PPARγ, we considered the clinical practice of PPARγ in pregnant disease prediction. A series of single nucleotide polymorphisms of PPARγ were thus examined, along with the clinical characteristics, which were subsequently applied to the predictive model building via machine learning algorithms. To sum up, our study integrated the application of clinical data, sequencing data, computer science, and biological validation.

The rapid development of nucleic acid sequencing technologies over the past four decades has improved the capacity to detect genomics in individuals. DNA sequencing gradually evolved from low throughput DNA fragment sequencing to high throughput next-generation (NGS) and third-generation sequencing techniques including rapid ways for genome-wide characterization and profiling of mRNAs, small RNAs,

transcription factor regions, the structure of chromatin, and DNA methylation patterns [169]. Recently, single-cell sequencing technologies have also been rapidly developed for observing the multilayered status of single cells. Single-cell sequencing has the power to elucidate genomic, epigenomic, and transcriptomic heterogeneity in cellular populations, and the changes at these levels, which is more accurate than general sequencing analysis with the measurement of only the average transcript expression in a cell population [372]. No matter how advanced the techniques are, their purpose is to detect the expression of genes, which is inevitable to perform the enrichment of the genes to figure out the critical biological functions. The further research was mostly focused on the study of specific mechanisms in order to discover valuable drug targets or seeking links between the DEGs and diseases for diagnosis and prognosis in clinical practice [172].

These functions were generally provided via matching to the Gene Ontology (GO) knowledgebase, which is the world's canonical and largest source of information on the functions of genes [170]. The GO knowledgebase was widely regarded as a reference to guide further research as soon as the differentially expressed genes (DEGs) were enriched. GO terms represent the comprehensive aspect of the function of the genes and gene products. A series of complicated biological processes would be involved in the disease occurrence, owing to the outcome happening inside or between cells [171]. The processes were summed up in the GO knowledgebase which contains 44,085 terms, 7,931,218 annotations, and 1,564,454 gene products to 4,743 different biological organisms as of February 2021. Among the terms, it contains 28748 biological process terms, 11153 molecular function terms, and 4184 cellular component terms. The abundance and diversity of GO terms indicate the possibility of disease characterization, which means a more general way, at the biological process level,

should not be underestimated in clinical practice. Specifically, instead of linking a single gene or genes with diseases, it is potentially reasonable to use GO terms to predict diseases.

As the case of most studies, our work mainly focused on the role of a single gene, PPARγ, in placentation. With the help of the GO knowledgebase, we obtained a series of terms, with which we were guided to study the corresponding biological processes related to PPARγ and the normal or abnormal development of the human placenta. Interestingly, in this process, we found that the enriched GO terms varied in different types of cells and cells with different treatments. It means that it is one-sided to use GO knowledgebase in such a way with a single gene or genes being applied to link with diseases. We are therefore wondering if the terms could be used to characterize the status of cells which resulted from a series of combined biological processes. This work will be carried on in the future.

Bibliographie

[1] J.M. Foidart, J. Hustin, M. Dubois, J.P. Schaaps, The human placenta becomes haemochorial at the 13th week of pregnancy, Int J Dev Biol 36(3) (1992) 451-3.
[2] E. Jauniaux, L. Poston, G.J. Burton, Placental-related diseases of pregnancy: Involvement of oxidative stress and implications in human evolution, Hum Reprod Update 12(6) (2006) 747-55.
[3] K. Benirschke, P. Kaufmann, Early Development of the Human Placenta, in: K. Benirschke, P. Kaufmann (Eds.), Pathology of the Human Placenta, Springer New York, New York, NY, 1995, pp. 49-56.
[4] L. Selwood, M.H. Johnson, Trophoblast and hypoblast in the monotreme, marsupial and eutherian mammal: evolution and origins, Bioessays 28(2) (2006) 128-45.
[5] W.P. Luckett, Origin and differentiation of the yolk sac and extraembryonic mesoderm in presomite human and rhesus monkey embryos, Am J Anat 152(1) (1978) 59-97.
[6] S. Schlafke, A.C. Enders, Cellular basis of interaction between trophoblast and uterus at implantation, Biol Reprod 12(1) (1975) 41-65.
[7] K. Benirschke, P. Kaufmann, Early Development of the Human Placenta, in: K. Benirschke, P. Kaufmann (Eds.), Pathology of the Human Placenta, Springer New York, New York, NY, 2000, pp. 42-49.
[8] A.T. Hertig, J. Rock, E.C. Adams, A description of 34 human ova within the first 17 days of development, Am J Anat 98(3) (1956) 435-93.
[9] W.J. Hamilton, J.D. Boyd, Development of the human placenta in the first three months of gestation, J Anat 94 (1960) 297-328.
[10] G.J. Burton, E. Jauniaux, A.L. Watson, Maternal arterial connections to the placental intervillous space during the first trimester of human pregnancy: the Boyd collection revisited, Am J Obstet Gynecol 181(3) (1999) 718-24.
[11] G.J. Burton, A.L. Watson, J. Hempstock, J.N. Skepper, E. Jauniaux, Uterine glands provide histiotrophic nutrition for the human fetus during the first trimester of pregnancy, J Clin Endocrinol Metab 87(6) (2002) 2954-9.
[12] T. Cotechini, M. Komisarenko, A. Sperou, S. Macdonald-Goodfellow, M.A. Adams, C.H. Graham, Inflammation in rat pregnancy inhibits spiral artery remodeling leading to fetal growth restriction and features of preeclampsia, J Exp Med 211(1) (2014) 165-79.
[13] C. Robin, K. Bollerot, S. Mendes, E. Haak, M. Crisan, F. Cerisoli, I. Lauw, P. Kaimakis, R. Jorna, M. Vermeulen, M. Kayser, R. van der Linden, P. Imanirad, M. Verstegen, H. Nawaz-Yousaf, N. Papazian, E. Steegers, T. Cupedo, E. Dzierzak, Human placenta is a potent hematopoietic niche containing hematopoietic stem and progenitor cells throughout development, Cell Stem Cell 5(4) (2009) 385-95.
[14] R. Demir, P. Kaufmann, M. Castellucci, T. Erbengi, A. Kotowski, Fetal vasculogenesis and angiogenesis in human placental villi, Acta Anat (Basel) 136(3)

(1989) 190-203.
[15] E.W. Dempsey, The development of capillaries in the villi of early human placentas, Am J Anat 134(2) (1972) 221-37.
[16] E. Jauniaux, D. Jurkovic, S. Campbell, Current topic: in vivo investigation of the placental circulations by Doppler echography, Placenta 16(4) (1995) 323-31.
[17] E. Jauniaux, A. Watson, G. Burton, Evaluation of respiratory gases and acid-base gradients in human fetal fluids and uteroplacental tissue between 7 and 16 weeks' gestation, Am J Obstet Gynecol 184(5) (2001) 998-1003.
[18] F. Rodesch, P. Simon, C. Donner, E. Jauniaux, Oxygen measurements in endometrial and trophoblastic tissues during early pregnancy, Obstet Gynecol 80(2) (1992) 283-5.
[19] G.J. Burton, Oxygen, the Janus gas; its effects on human placental development and function, J Anat 215(1) (2009) 27-35.
[20] G.J. Burton, J. Hempstock, E. Jauniaux, Oxygen, early embryonic metabolism and free radical-mediated embryopathies, Reprod Biomed Online 6(1) (2003) 84-96.
[21] R. Pijnenborg, J.M. Bland, W.B. Robertson, G. Dixon, I. Brosens, The pattern of interstitial trophoblastic invasion of the myometrium in early human pregnancy, Placenta 2(4) (1981) 303-16.
[22] R. Pijnenborg, G. Dixon, W.B. Robertson, I. Brosens, Trophoblastic invasion of human decidua from 8 to 18 weeks of pregnancy, Placenta 1(1) (1980) 3-19.
[23] E. Jauniaux, B. Gulbis, Fluid compartments of the embryonic environment, Hum Reprod Update 6(3) (2000) 268-78.
[24] E. Jauniaux, J. Hempstock, N. Greenwold, G.J. Burton, Trophoblastic oxidative stress in relation to temporal and regional differences in maternal placental blood flow in normal and abnormal early pregnancies, Am J Pathol 162(1) (2003) 115-25.
[25] M. Castellucci, M. Scheper, I. Scheffen, A. Celona, P. Kaufmann, The development of the human placental villous tree, Anat Embryol (Berl) 181(2) (1990) 117-28.
[26] M.R. Jackson, T.M. Mayhew, P.A. Boyd, Quantitative description of the elaboration and maturation of villi from 10 weeks of gestation to term, Placenta 13(4) (1992) 357-70.
[27] E. Jauniaux, G.J. Burton, G.J. Moscoso, J. Hustin, Development of the early human placenta: a morphometric study, Placenta 12(3) (1991) 269-76.
[28] E. Jauniaux, T. Cindrova-Davies, J. Johns, C. Dunster, J. Hempstock, F.J. Kelly, G.J. Burton, Distribution and transfer pathways of antioxidant molecules inside the first trimester human gestational sac, J Clin Endocrinol Metab 89(3) (2004) 1452-8.
[29] R. Leiser, M. Luckhardt, P. Kaufmann, E. Winterhager, U. Bruns, The fetal vascularisation of term human placental villi. I. Peripheral stem villi, Anat Embryol (Berl) 173(1) (1985) 71-80.
[30] A.L. Karimu, G.J. Burton, The distribution of microvilli over the villous surface of the normal human term placenta is homogenous, Reprod Fertil Dev 7(5) (1995) 1269-73.
[31] F. Teasdale, G. Jean-Jacques, Morphometry of the microvillous membrane of the human placenta in maternal diabetes mellitus, Placenta 7(1) (1986) 81-8.
[32] M. Berryman, R. Gary, A. Bretscher, Ezrin oligomers are major cytoskeletal

components of placental microvilli: a proposal for their involvement in cortical morphogenesis, J Cell Biol 131(5) (1995) 1231-42.
[33] O.A. Vanderpuye, H.C. Edwards, A.G. Booth, Proteins of the human placental microvillar cytoskeleton. alpha-Actinin, Biochem J 233(2) (1986) 351-6.
[34] C.J. Jones, H. Fox, An ultrastructural and ultrahistochemical study of the human placenta in maternal pre-eclampsia, Placenta 1(1) (1980) 61-76.
[35] N.S. Rote, B.R. Wei, C. Xu, L. Luo, Caspase 8 and human villous cytotrophoblast differentiation, Placenta 31(2) (2010) 89-96.
[36] J.D. Aplin, C.J. Jones, L.K. Harris, Adhesion molecules in human trophoblast - a review. I. Villous trophoblast, Placenta 30(4) (2009) 293-8.
[37] C.J. Jones, H. Fox, Ultrastructure of the normal human placenta, Electron Microsc Rev 4(1) (1991) 129-78.
[38] J.K. Cleal, R.M. Lewis, The mechanisms and regulation of placental amino acid transport to the human foetus, J Neuroendocrinol 20(4) (2008) 419-26.
[39] M. Mori, G. Ishikawa, S.-S. Luo, T. Mishima, T. Goto, J.M. Robinson, S. Matsubara, T. Takeshita, H. Kataoka, T. Takizawa, The Cytotrophoblast Layer of Human Chorionic Villi Becomes Thinner but Maintains Its Structural Integrity During Gestation1, Biology of Reproduction 76(1) (2007) 164-172.
[40] A.R. Midgley, G.B. Pierce, G.A. Deneau, J.R.G. Gosling, Morphogenesis of Syncytiotrophoblast in vivo: An Autoradiographic Demonstration, Science 141(3578) (1963) 349.
[41] M. Castellucci, A. Celona, H. Bartels, B. Steininger, V. Benedetto, P. Kaufmann, Mitosis of the Hofbauer cell: possible implications for a fetal macrophage, Placenta 8(1) (1987) 65-76.
[42] G.J. Burton, The fine structure of the human placental villus as revealed by scanning electron microscopy, Scanning Microsc 1(4) (1987) 1811-28.
[43] M. Castellucci, D. Zaccheo, G. Pescetto, A three-dimensional study of the normal human placental villous core. I. The Hofbauer cells, Cell Tissue Res 210(2) (1980) 235-47.
[44] M.R. Jackson, A.J. Walsh, R.J. Morrow, J.B. Mullen, S.J. Lye, J.W. Ritchie, Reduced placental villous tree elaboration in small-for-gestational-age pregnancies: relationship with umbilical artery Doppler waveforms, Am J Obstet Gynecol 172(2 Pt 1) (1995) 518-25.
[45] E.A. Linton, B. Rodriguez-Linares, F. Rashid-Doubell, D.J. Ferguson, C.W. Redman, Caveolae and caveolin-1 in human term villous trophoblast, Placenta 24(7) (2003) 745-57.
[46] T.W. Lyden, C.L. Anderson, J.M. Robinson, The endothelium but not the syncytiotrophoblast of human placenta expresses caveolae, Placenta 23(8-9) (2002) 640-52.
[47] L. Myatt, R.B. Rosenfield, A.L. Eis, D.E. Brockman, I. Greer, F. Lyall, Nitrotyrosine residues in placenta. Evidence of peroxynitrite formation and action, Hypertension 28(3) (1996) 488-93.
[48] D. Heinrich, J. Metz, E. Raviola, W.G. Forssmann, Ultrastructure of perfusion-fixed fetal capillaries in the human placenta, Cell and Tissue Research 172(2) (1976)

157-169.
[49] T.H. Hung, J.N. Skepper, G.J. Burton, In vitro ischemia-reperfusion injury in term human placenta as a model for oxidative stress in pathological pregnancies, Am J Pathol 159(3) (2001) 1031-43.
[50] G.J. Burton, C.J. Jones, Syncytial knots, sprouts, apoptosis, and trophoblast deportation from the human placenta, Taiwan J Obstet Gynecol 48(1) (2009) 28-37.
[51] E. Asan, F.F. Kaymaz, A.N. Cakar, A. Dagdeviren, M.S. Beksac, Vasculogenesis in early human placental villi: an ultrastructural study, Ann Anat 181(6) (1999) 549-54.
[52] K.M. Downs, S. Gifford, M. Blahnik, R.L. Gardner, Vascularization in the murine allantois occurs by vasculogenesis without accompanying erythropoiesis, Development 125(22) (1998) 4507-20.
[53] E.A. te Velde, N. Exalto, P. Hesseling, H.C. van der Linden, First trimester development of human chorionic villous vascularization studied with CD34 immunohistochemistry, Hum Reprod 12(7) (1997) 1577-81.
[54] R.H.F. van Oppenraaij, A.H.J. Koning, B.A. Lisman, K. Boer, M.J.B. van den Hoff, P.J. van der Spek, E.A.P. Steegers, N. Exalto, Vasculogenesis and angiogenesis in the first trimester human placenta: an innovative 3D study using an immersive Virtual Reality system, Placenta 30(3) (2009) 220-222.
[55] T.M. Mayhew, E. Wadrop, R.A. Simpson, Proliferative versus hypertrophic growth in tissue subcompartments of human placental villi during gestation, J Anat 184 (Pt 3) (1994) 535-43.
[56] E.G. Zhang, G.J. Burton, S.K. Smith, D.S. Charnock-Jones, Placental vessel adaptation during gestation and to high altitude: changes in diameter and perivascular cell coverage, Placenta 23(10) (2002) 751-62.
[57] G. Kohnen, S. Kertschanska, R. Demir, P. Kaufmann, Placental villous stroma as a model system for myofibroblast differentiation, Histochem Cell Biol 105(6) (1996) 415-29.
[58] R. Demir, G. Kosanke, G. Kohnen, S. Kertschanska, P. Kaufmann, Classification of human placental stem villi: review of structural and functional aspects, Microsc Res Tech 38(1-2) (1997) 29-41.
[59] M.G. Tuuli, M.S. Longtine, D.M. Nelson, Review: Oxygen and trophoblast biology--a source of controversy, Placenta 32 Suppl 2 (2011) S109-18.
[60] J.C. Kingdom, P. Kaufmann, Oxygen and placental villous development: origins of fetal hypoxia, Placenta 18(8) (1997) 613-21; discussion 623-6.
[61] M. Kadyrov, G. Kosanke, J. Kingdom, P. Kaufmann, Increased fetoplacental angiogenesis during first trimester in anaemic women, Lancet 352(9142) (1998) 1747-9.
[62] N.A. Beischer, R. Sivasamboo, S. Vohra, S. Silpisornkosal, S. Reid, Placental hypertrophy in severe pregnancy anaemia, J Obstet Gynaecol Br Commonw 77(5) (1970) 398-409.
[63] B. Piotrowicz, T.K. Niebroj, G. Sieron, The morphology and histochemistry of the full term placenta in anaemic patients, Folia Histochem Cytochem (Krakow) 7(4) (1969) 435-44.
[64] L.G. Moore, M. Shriver, L. Bemis, B. Hickler, M. Wilson, T. Brutsaert, E. Parra, E.

Vargas, Maternal adaptation to high-altitude pregnancy: an experiment of nature--a review, Placenta 25 Suppl A (2004) S60-71.
[65] L.M. Postovit, M.A. Adams, G.E. Lash, J.P. Heaton, C.H. Graham, Oxygen-mediated regulation of tumor cell invasiveness. Involvement of a nitric oxide signaling pathway, J Biol Chem 277(38) (2002) 35730-7.
[66] J. Espinoza, N.J. Sebire, F. McAuliffe, E. Krampl, K.H. Nicolaides, Placental villus morphology in relation to maternal hypoxia at high altitude, Placenta 22(6) (2001) 606-8.
[67] G.J. Burton, O.S. Reshetnikova, A.P. Milovanov, O.V. Teleshova, Stereological evaluation of vascular adaptations in human placental villi to differing forms of hypoxic stress, Placenta 17(1) (1996) 49-55.
[68] T.G. Smith, P.A. Robbins, P.J. Ratcliffe, The human side of hypoxia-inducible factor, Br J Haematol 141(3) (2008) 325-34.
[69] G.L. Semenza, Hypoxia-inducible factor 1: master regulator of O2 homeostasis, Curr Opin Genet Dev 8(5) (1998) 588-94.
[70] J. Patel, K. Landers, R.H. Mortimer, K. Richard, Regulation of hypoxia inducible factors (HIF) in hypoxia and normoxia during placental development, Placenta 31(11) (2010) 951-7.
[71] L.A. Palmer, B. Gaston, R.A. Johns, Normoxic stabilization of hypoxia-inducible factor-1 expression and activity: redox-dependent effect of nitrogen oxides, Mol Pharmacol 58(6) (2000) 1197-203.
[72] K.G. Pringle, K.L. Kind, A.N. Sferruzzi-Perri, J.G. Thompson, C.T. Roberts, Beyond oxygen: complex regulation and activity of hypoxia inducible factors in pregnancy, Hum Reprod Update 16(4) (2010) 415-31.
[73] T. Lofstedt, E. Fredlund, L. Holmquist-Mengelbier, A. Pietras, M. Ovenberger, L. Poellinger, S. Pahlman, Hypoxia inducible factor-2alpha in cancer, Cell Cycle 6(8) (2007) 919-26.
[74] J.L. James, P.R. Stone, L.W. Chamley, The regulation of trophoblast differentiation by oxygen in the first trimester of pregnancy, Hum Reprod Update 12(2) (2006) 137-44.
[75] C.P. Chen, J.D. Aplin, Placental extracellular matrix: gene expression, deposition by placental fibroblasts and the effect of oxygen, Placenta 24(4) (2003) 316-25.
[76] W. Droge, Free radicals in the physiological control of cell function, Physiol Rev 82(1) (2002) 47-95.
[77] T. Cindrova-Davies, Gabor Than Award Lecture 2008: pre-eclampsia - from placental oxidative stress to maternal endothelial dysfunction, Placenta 30 Suppl A (2009) S55-65.
[78] T. Cindrova-Davies, O. Spasic-Boskovic, E. Jauniaux, D.S. Charnock-Jones, G.J. Burton, Nuclear factor-kappa B, p38, and stress-activated protein kinase mitogen-activated protein kinase signaling pathways regulate proinflammatory cytokines and apoptosis in human placental explants in response to oxidative stress: effects of antioxidant vitamins, Am J Pathol 170(5) (2007) 1511-20.
[79] P.W. Conrad, L. Conforti, S. Kobayashi, D. Beitner-Johnson, R.T. Rust, Y. Yuan, H.W. Kim, R.H. Kim, K. Seta, D.E. Millhorn, The molecular basis of O2-sensing and

hypoxia tolerance in pheochromocytoma cells, Comp Biochem Physiol B Biochem Mol Biol 128(2) (2001) 187-204.
[80] S.B. Cullinan, J.A. Diehl, Coordination of ER and oxidative stress signaling: the PERK/Nrf2 signaling pathway, Int J Biochem Cell Biol 38(3) (2006) 317-32.
[81] H.P. Harding, Y. Zhang, H. Zeng, I. Novoa, P.D. Lu, M. Calfon, N. Sadri, C. Yun, B. Popko, R. Paules, D.F. Stojdl, J.C. Bell, T. Hettmann, J.M. Leiden, D. Ron, An integrated stress response regulates amino acid metabolism and resistance to oxidative stress, Mol Cell 11(3) (2003) 619-33.
[82] B.G. Wouters, M. Koritzinsky, Hypoxia signalling through mTOR and the unfolded protein response in cancer, Nat Rev Cancer 8(11) (2008) 851-64.
[83] D. Ron, P. Walter, Signal integration in the endoplasmic reticulum unfolded protein response, Nat Rev Mol Cell Biol 8(7) (2007) 519-29.
[84] M. Schroder, R.J. Kaufman, The mammalian unfolded protein response, Annu Rev Biochem 74 (2005) 739-89.
[85] K.Z. Ali, Stereological study of the effect of altitude on the trophoblast cell populations of human term placental villi, Placenta 18(5-6) (1997) 447-50.
[86] O. Genbacev, Y. Zhou, J.W. Ludlow, S.J. Fisher, Regulation of human placental development by oxygen tension, Science 277(5332) (1997) 1669-72.
[87] Y. Kudo, C.A. Boyd, I.L. Sargent, C.W. Redman, Hypoxia alters expression and function of syncytin and its receptor during trophoblast cell fusion of human placental BeWo cells: implications for impaired trophoblast syncytialisation in pre-eclampsia, Biochim Biophys Acta 1638(1) (2003) 63-71.
[88] H. Alvarez, W.L. Benedetti, R.L. Morel, M. Scavarelli, Trophoblast development gradient and its relationship to placental hemodynamics, Am J Obstet Gynecol 106(3) (1970) 416-20.
[89] G.J. Burton, A.W. Woods, E. Jauniaux, J.C. Kingdom, Rheological and physiological consequences of conversion of the maternal spiral arteries for uteroplacental blood flow during human pregnancy, Placenta 30(6) (2009) 473-82.
[90] G.E. Lash, Molecular Cross-Talk at the Feto-Maternal Interface, Cold Spring Harb Perspect Med 5(12) (2015).
[91] J.L. James, P.R. Stone, L.W. Chamley, The effects of oxygen concentration and gestational age on extravillous trophoblast outgrowth in a human first trimester villous explant model, Hum Reprod 21(10) (2006) 2699-705.
[92] K.D. Cowden Dahl, B.H. Fryer, F.A. Mack, V. Compernolle, E. Maltepe, D.M. Adelman, P. Carmeliet, M.C. Simon, Hypoxia-inducible factors 1alpha and 2alpha regulate trophoblast differentiation, Mol Cell Biol 25(23) (2005) 10479-91.
[93] I. Caniggia, J.L. Winter, Adriana and Luisa Castellucci Award lecture 2001. Hypoxia inducible factor-1: oxygen regulation of trophoblast differentiation in normal and pre-eclamptic pregnancies--a review, Placenta 23 Suppl A (2002) S47-57.
[94] I. Caniggia, H. Mostachfi, J. Winter, M. Gassmann, S.J. Lye, M. Kuliszewski, M. Post, Hypoxia-inducible factor-1 mediates the biological effects of oxygen on human trophoblast differentiation through TGFbeta(3), J Clin Invest 105(5) (2000) 577-87.
[95] S. Natanson-Yaron, E.Y. Anteby, C. Greenfield, D. Goldman-Wohl, Y. Hamani, D. Hochner-Celnikier, S. Yagel, FGF 10 and Sprouty 2 modulate trophoblast invasion and

branching morphogenesis, Mol Hum Reprod 13(7) (2007) 511-9.
[96] C.H. Graham, L.M. Postovit, H. Park, M.T. Canning, T.E. Fitzpatrick, Adriana and Luisa Castellucci award lecture 1999: role of oxygen in the regulation of trophoblast gene expression and invasion, Placenta 21(5-6) (2000) 443-50.
[97] K.L. Howe, P. Achuthan, J. Allen, J. Allen, J. Alvarez-Jarreta, M.R. Amode, I.M. Armean, A.G. Azov, R. Bennett, J. Bhai, K. Billis, S. Boddu, M. Charkhchi, C. Cummins, L. Da Rin Fioretto, C. Davidson, K. Dodiya, B. El Houdaigui, R. Fatima, A. Gall, C. Garcia Giron, T. Grego, C. Guijarro-Clarke, L. Haggerty, A. Hemrom, T. Hourlier, O.G. Izuogu, T. Juettemann, V. Kaikala, M. Kay, I. Lavidas, T. Le, D. Lemos, J. Gonzalez Martinez, J.C. Marugan, T. Maurel, A.C. McMahon, S. Mohanan, B. Moore, M. Muffato, D.N. Oheh, D. Paraschas, A. Parker, A. Parton, I. Prosovetskaia, M.P. Sakthivel, A.I.A. Salam, B.M. Schmitt, H. Schuilenburg, D. Sheppard, E. Steed, M. Szpak, M. Szuba, K. Taylor, A. Thormann, G. Threadgold, B. Walts, A. Winterbottom, M. Chakiachvili, A. Chaubal, N. De Silva, B. Flint, A. Frankish, S.E. Hunt, I.I. GR, N. Langridge, J.E. Loveland, F.J. Martin, J.M. Mudge, J. Morales, E. Perry, M. Ruffier, J. Tate, D. Thybert, S.J. Trevanion, F. Cunningham, A.D. Yates, D.R. Zerbino, P. Flicek, Ensembl 2021, Nucleic Acids Res 49(D1) (2021) D884-D891.
[98] T. Fournier, V. Tsatsaris, K. Handschuh, D. Evain-Brion, PPARs and the placenta, Placenta 28(2-3) (2007) 65-76.
[99] L. Qiang, L. Wang, N. Kon, W. Zhao, S. Lee, Y. Zhang, M. Rosenbaum, Y. Zhao, W. Gu, S.R. Farmer, D. Accili, Brown remodeling of white adipose tissue by SirT1-dependent deacetylation of Ppargamma, Cell 150(3) (2012) 620-32.
[100] E. Hu, J.B. Kim, P. Sarraf, B.M. Spiegelman, Inhibition of adipogenesis through MAP kinase-mediated phosphorylation of PPARgamma, Science 274(5295) (1996) 2100-3.
[101] G. Pascual, A.L. Fong, S. Ogawa, A. Gamliel, A.C. Li, V. Perissi, D.W. Rose, T.M. Willson, M.G. Rosenfeld, C.K. Glass, A SUMOylation-dependent pathway mediates transrepression of inflammatory response genes by PPAR-gamma, Nature 437(7059) (2005) 759-63.
[102] M. Anbalagan, B. Huderson, L. Murphy, B.G. Rowan, Post-translational modifications of nuclear receptors and human disease, Nucl Recept Signal 10 (2012) e001.
[103] S.R. Fuhs, J. Meisenhelder, A. Aslanian, L. Ma, A. Zagorska, M. Stankova, A. Binnie, F. Al-Obeidi, J. Mauger, G. Lemke, J.R. Yates, 3rd, T. Hunter, Monoclonal 1- and 3-Phosphohistidine Antibodies: New Tools to Study Histidine Phosphorylation, Cell 162(1) (2015) 198-210.
[104] E. Meyers, F.L. Weisenborn, F.E. Pansy, D.S. Slusarchyk, M.H. Von Saltza, M.L. Rathnum, W.L. Parker, Janiemycin, a new peptide antibiotic, J Antibiot (Tokyo) 23(10) (1970) 502-7.
[105] R. Geiss-Friedlander, F. Melchior, Concepts in sumoylation: a decade on, Nature Reviews Molecular Cell Biology 8(12) (2007) 947-956.
[106] D. Mukhopadhyay, H. Riezman, Proteasome-independent functions of ubiquitin in endocytosis and signaling, Science 315(5809) (2007) 201-5.
[107] J.D. Schnell, L. Hicke, Non-traditional functions of ubiquitin and ubiquitin-

binding proteins, J Biol Chem 278(38) (2003) 35857-60.
[108] M.H. Glickman, A. Ciechanover, The ubiquitin-proteasome proteolytic pathway: destruction for the sake of construction, Physiol Rev 82(2) (2002) 373-428.
[109] N. Weder, H. Zhang, K. Jensen, B.Z. Yang, A. Simen, A. Jackowski, D. Lipschitz, H. Douglas-Palumberi, M. Ge, F. Perepletchikova, K. O'Loughlin, J.J. Hudziak, J. Gelernter, J. Kaufman, Child abuse, depression, and methylation in genes involved with stress, neural plasticity, and brain circuitry, J Am Acad Child Adolesc Psychiatry 53(4) (2014) 417-24 e5.
[110] H. Sun, I.M. Berquin, R.T. Owens, J.T. O'Flaherty, I.J. Edwards, Peroxisome proliferator-activated receptor gamma-mediated up-regulation of syndecan-1 by n-3 fatty acids promotes apoptosis of human breast cancer cells, Cancer Res 68(8) (2008) 2912-9.
[111] H. Sun, I.M. Berquin, I.J. Edwards, Omega-3 polyunsaturated fatty acids regulate syndecan-1 expression in human breast cancer cells, Cancer Res 65(10) (2005) 4442-7.
[112] I.J. Edwards, I.M. Berquin, H. Sun, T. O'Flaherty J, L.W. Daniel, M.J. Thomas, L.L. Rudel, R.L. Wykle, Y.Q. Chen, Differential effects of delivery of omega-3 fatty acids to human cancer cells by low-density lipoproteins versus albumin, Clin Cancer Res 10(24) (2004) 8275-83.
[113] Z.H. Yang, H. Miyahara, Y. Iwasaki, J. Takeo, M. Katayama, Dietary supplementation with long-chain monounsaturated fatty acids attenuates obesity-related metabolic dysfunction and increases expression of PPAR gamma in adipose tissue in type 2 diabetic KK-Ay mice, Nutr Metab (Lond) 10(1) (2013) 16.
[114] M. Heim, J. Johnson, F. Boess, I. Bendik, P. Weber, W. Hunziker, B. Fluhmann, Phytanic acid, a natural peroxisome proliferator-activated receptor (PPAR) agonist, regulates glucose metabolism in rat primary hepatocytes, FASEB J 16(7) (2002) 718-20.
[115] M.A. Deeg, M.H. Tan, Pioglitazone versus Rosiglitazone: Effects on Lipids, Lipoproteins, and Apolipoproteins in Head-to-Head Randomized Clinical Studies, PPAR Res 2008 (2008) 520465.
[116] M.H. Tan, Current treatment of insulin resistance in type 2 diabetes mellitus, Int J Clin Pract Suppl (113) (2000) 54-62.
[117] J. Dormandy, M. Bhattacharya, A.R. van Troostenburg de Bruyn, P.R. investigators, Safety and tolerability of pioglitazone in high-risk patients with type 2 diabetes: an overview of data from PROactive, Drug Saf 32(3) (2009) 187-202.
[118] S.E. Nissen, K. Wolski, Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes, N Engl J Med 356(24) (2007) 2457-71.
[119] A. Krishnaswami, S. Ravi-Kumar, J.M. Lewis, Thiazolidinediones: a 2010 perspective, Perm J 14(3) (2010) 64-72.
[120] A.L. Peters Harmel, D.M. Kendall, J.B. Buse, P.J. Boyle, A. Marchetti, H. Lau, Impact of adjunctive thiazolidinedione therapy on blood lipid levels and glycemic control in patients with type 2 diabetes, Curr Med Res Opin 20(2) (2004) 215-23.
[121] M. Diamant, R.J. Heine, Thiazolidinediones in type 2 diabetes mellitus: current clinical evidence, Drugs 63(13) (2003) 1373-405.

[122] K.G. Tolman, The safety of thiazolidinediones, Expert Opin Drug Saf 10(3) (2011) 419-28.
[123] A.E. Caballero, R. Saouaf, S.C. Lim, O. Hamdy, K. Abou-Elenin, C. O'Connor, F.W. Logerfo, E.S. Horton, A. Veves, The effects of troglitazone, an insulin-sensitizing agent, on the endothelial function in early and late type 2 diabetes: a placebo-controlled randomized clinical trial, Metabolism 52(2) (2003) 173-80.
[124] A. Rogue, C. Lambert, R. Josse, S. Antherieu, C. Spire, N. Claude, A. Guillouzo, Comparative gene expression profiles induced by PPARgamma and PPARalpha/gamma agonists in human hepatocytes, PLoS One 6(4) (2011) e18816.
[125] A. Rogue, C. Spire, M. Brun, N. Claude, A. Guillouzo, Gene Expression Changes Induced by PPAR Gamma Agonists in Animal and Human Liver, PPAR Res 2010 (2010) 325183.
[126] L.A. Marlow, L.A. Reynolds, A.S. Cleland, S.J. Cooper, M.L. Gumz, S. Kurakata, K. Fujiwara, Y. Zhang, T. Sebo, C. Grant, B. McIver, J.T. Wadsworth, D.C. Radisky, R.C. Smallridge, J.A. Copland, Reactivation of suppressed RhoB is a critical step for the inhibition of anaplastic thyroid cancer growth, Cancer Res 69(4) (2009) 1536-44.
[127] A. Rubenstrunk, R. Hanf, D.W. Hum, J.C. Fruchart, B. Staels, Safety issues and prospects for future generations of PPAR modulators, Biochim Biophys Acta 1771(8) (2007) 1065-81.
[128] M.C. Carmona, K. Louche, B. Lefebvre, A. Pilon, N. Hennuyer, V. Audinot-Bouchez, C. Fievet, G. Torpier, P. Formstecher, P. Renard, P. Lefebvre, C. Dacquet, B. Staels, L. Casteilla, L. Penicaud, R. Consortium of the French Ministry of, Technology, S 26948: a new specific peroxisome proliferator activated receptor gamma modulator with potent antidiabetes and antiatherogenic effects, Diabetes 56(11) (2007) 2797-808.
[129] F.L. Dunn, L.S. Higgins, J. Fredrickson, A.M. DePaoli, I.N.T.s. group, Selective modulation of PPARgamma activity can lower plasma glucose without typical thiazolidinedione side-effects in patients with Type 2 diabetes, J Diabetes Complications 25(3) (2011) 151-8.
[130] B.R. Henke, Peroxisome proliferator-activated receptor alpha/gamma dual agonists for the treatment of type 2 diabetes, J Med Chem 47(17) (2004) 4118-27.
[131] C. Grommes, G.E. Landreth, M.T. Heneka, Antineoplastic effects of peroxisome proliferator-activated receptor gamma agonists, Lancet Oncol 5(7) (2004) 419-29.
[132] G.G. Long, V.L. Reynolds, A. Lopez-Martinez, T.E. Ryan, S.L. White, S.R. Eldridge, Urothelial carcinogenesis in the urinary bladder of rats treated with naveglitazar, a gamma-dominant PPAR alpha/gamma agonist: lack of evidence for urolithiasis as an inciting event, Toxicol Pathol 36(2) (2008) 218-31.
[133] S.E. Nissen, K. Wolski, E.J. Topol, Effect of muraglitazar on death and major adverse cardiovascular events in patients with type 2 diabetes mellitus, JAMA 294(20) (2005) 2581-6.
[134] G. Brusotti, R. Montanari, D. Capelli, G. Cattaneo, A. Laghezza, P. Tortorella, F. Loiodice, F. Peiretti, B. Bonardo, A. Paiardini, E. Calleri, G. Pochetti, Betulinic acid is a PPARgamma antagonist that improves glucose uptake, promotes osteogenesis and inhibits adipogenesis, Sci Rep 7(1) (2017) 5777.
[135] J.M. Seargent, E.A. Yates, J.H. Gill, GW9662, a potent antagonist of

PPARgamma, inhibits growth of breast tumour cells and promotes the anticancer effects of the PPARgamma agonist rosiglitazone, independently of PPARgamma activation, Br J Pharmacol 143(8) (2004) 933-7.
[136] A. Asteian, A.L. Blayo, Y. He, M. Koenig, Y. Shin, D.S. Kuruvilla, C.A. Corzo, M.D. Cameron, L. Lin, C. Ruiz, S. Khan, N. Kumar, S. Busby, D.P. Marciano, R.D. Garcia-Ordonez, P.R. Griffin, T.M. Kamenecka, Design, Synthesis, and Biological Evaluation of Indole Biphenylcarboxylic Acids as PPARgamma Antagonists, ACS Med Chem Lett 6(9) (2015) 998-1003.
[137] D.P. Marciano, D.S. Kuruvilla, S.V. Boregowda, A. Asteian, T.S. Hughes, R. Garcia-Ordonez, C.A. Corzo, T.M. Khan, S.J. Novick, H. Park, D.J. Kojetin, D.G. Phinney, J.B. Bruning, T.M. Kamenecka, P.R. Griffin, Pharmacological repression of PPARgamma promotes osteogenesis, Nat Commun 6 (2015) 7443.
[138] I. Cesari, M. Hoerle, C. Simoes-Pires, P. Grisoli, E.F. Queiroz, C. Dacarro, L. Marcourt, P.F. Moundipa, P.A. Carrupt, M. Cuendet, G. Caccialanza, J.L. Wolfender, G. Brusotti, Anti-inflammatory, antimicrobial and antioxidant activities of Diospyros bipindensis (Gurke) extracts and its main constituents, J Ethnopharmacol 146(1) (2013) 264-70.
[139] I. Takada, M. Makishima, Peroxisome proliferator-activated receptor agonists and antagonists: a patent review (2014-present), Expert Opin Ther Pat 30(1) (2020) 1-13.
[140] Y. Zhang, Y. Huo, W. He, S. Liu, H. Li, L. Li, Visfatin is regulated by interleukin6 and affected by the PPARgamma pathway in BeWo cells, Mol Med Rep 19(1) (2019) 400-406.
[141] A. Blitek, M. Szymanska, Peroxisome proliferator-activated receptor beta/delta and gamma agonists differentially affect prostaglandin E2 and cytokine synthesis and nutrient transporter expression in porcine trophoblast cells during implantation, Theriogenology 152 (2020) 36-46.
[142] Y. Zhang, L. Hu, Y. Cui, Z. Qi, X. Huang, L. Cai, T. Zhang, Y. Yin, Z. Lu, J. Xiang, Roles of PPARgamma/NF-kappaB signaling pathway in the pathogenesis of intrahepatic cholestasis of pregnancy, PLoS One 9(1) (2014) e87343.
[143] A. Tarrade, K. Schoonjans, J. Guibourdenche, J.M. Bidart, M. Vidaud, J. Auwerx, C. Rochette-Egly, D. Evain-Brion, PPAR gamma/RXR alpha heterodimers are involved in human CG beta synthesis and human trophoblast differentiation, Endocrinology 142(10) (2001) 4504-14.
[144] M. Ruebner, M. Langbein, P.L. Strissel, C. Henke, D. Schmidt, T.W. Goecke, F. Faschingbauer, R.L. Schild, M.W. Beckmann, R. Strick, Regulation of the human endogenous retroviral Syncytin-1 and cell-cell fusion by the nuclear hormone receptors PPARgamma/RXRalpha in placentogenesis, J Cell Biochem 113(7) (2012) 2383-96.
[145] W.T. Schaiff, M.G. Carlson, S.D. Smith, R. Levy, D.M. Nelson, Y. Sadovsky, Peroxisome proliferator-activated receptor-gamma modulates differentiation of human trophoblast in a ligand-specific manner, J Clin Endocrinol Metab 85(10) (2000) 3874-81.
[146] W.T. Schaiff, I. Bildirici, M. Cheong, P.L. Chern, D.M. Nelson, Y. Sadovsky, Peroxisome proliferator-activated receptor-gamma and retinoid X receptor signaling

regulate fatty acid uptake by primary human placental trophoblasts, J Clin Endocrinol Metab 90(7) (2005) 4267-75.
[147] Y. Barak, M.C. Nelson, E.S. Ong, Y.Z. Jones, P. Ruiz-Lozano, K.R. Chien, A. Koder, R.M. Evans, PPAR gamma is required for placental, cardiac, and adipose tissue development, Mol Cell 4(4) (1999) 585-95.
[148] K. Nadra, L. Quignodon, C. Sardella, E. Joye, A. Mucciolo, R. Chrast, B. Desvergne, PPARgamma in placental angiogenesis, Endocrinology 151(10) (2010) 4969-81.
[149] L. Liu, X. Zhuang, M. Jiang, F. Guan, Q. Fu, J. Lin, ANGPTL4 mediates the protective role of PPARgamma activators in the pathogenesis of preeclampsia, Cell Death Dis 8(9) (2017) e3054.
[150] R. Mayama, T. Izawa, K. Sakai, N. Suciu, M. Iwashita, Improvement of insulin sensitivity promotes extravillous trophoblast cell migration stimulated by insulin-like growth factor-I, Endocr J 60(3) (2013) 359-68.
[151] T. Workalemahu, D.A. Enquobahrie, B. Gelaye, T.A. Thornton, F. Tekola-Ayele, S.E. Sanchez, P.J. Garcia, H.G. Palomino, A. Hajat, R. Romero, C.V. Ananth, M.A. Williams, Abruptio placentae risk and genetic variations in mitochondrial biogenesis and oxidative phosphorylation: replication of a candidate gene association study, Am J Obstet Gynecol 219(6) (2018) 617 e1-617 e17.
[152] L. Pavan, A. Tarrade, A. Hermouet, C. Delouis, M. Titeux, M. Vidaud, P. Therond, D. Evain-Brion, T. Fournier, Human invasive trophoblasts transformed with simian virus 40 provide a new tool to study the role of PPARgamma in cell invasion process, Carcinogenesis 24(8) (2003) 1325-36.
[153] T. Fournier, L. Pavan, A. Tarrade, K. Schoonjans, J. Auwerx, C. Rochette-Egly, D. Evain-Brion, The role of PPAR-gamma/RXR-alpha heterodimers in the regulation of human trophoblast invasion, Ann N Y Acad Sci 973 (2002) 26-30.
[154] A. Tarrade, K. Schoonjans, L. Pavan, J. Auwerx, C. Rochette-Egly, D. Evain-Brion, T. Fournier, PPARgamma/RXRalpha heterodimers control human trophoblast invasion, J Clin Endocrinol Metab 86(10) (2001) 5017-24.
[155] B. Rauwel, B. Mariame, H. Martin, R. Nielsen, S. Allart, B. Pipy, S. Mandrup, M.D. Devignes, D. Evain-Brion, T. Fournier, C. Davrinche, Activation of peroxisome proliferator-activated receptor gamma by human cytomegalovirus for de novo replication impairs migration and invasiveness of cytotrophoblasts from early placentas, J Virol 84(6) (2010) 2946-54.
[156] K. Handschuh, J. Guibourdenche, M. Guesnon, I. Laurendeau, D. Evain-Brion, T. Fournier, Modulation of PAPP-A expression by PPARgamma in human first trimester trophoblast, Placenta 27 Suppl A (2006) S127-34.
[157] N. Segond, S.A. Degrelle, S. Berndt, E. Clouqueur, C. Rouault, B. Saubamea, P. Dessen, K.S. Fong, K. Csiszar, J. Badet, D. Evain-Brion, T. Fournier, Transcriptome analysis of PPARgamma target genes reveals the involvement of lysyl oxidase in human placental cytotrophoblast invasion, PLoS One 8(11) (2013) e79413.
[158] S.J. Li, T. Shang, S.Y. Li, Q.L. Li, [Effects of peroxisome proliferator-activated receptor gamma and its ligands on cytotrophoblast invasion in first trimester of pregnancy and mechanism thereof], Zhonghua Yi Xue Za Zhi 87(3) (2007) 174-8.

[159] F. Gao, W. Hu, Y. Li, H. Shen, J. Hu, Mono-2-ethylhexyl phthalate inhibits human extravillous trophoblast invasion via the PPARgamma pathway, Toxicol Appl Pharmacol 327 (2017) 23-29.
[160] I. Bildirici, C.R. Roh, W.T. Schaiff, B.M. Lewkowski, D.M. Nelson, Y. Sadovsky, The lipid droplet-associated protein adipophilin is expressed in human trophoblasts and is regulated by peroxisomal proliferator-activated receptor-gamma/retinoid X receptor, J Clin Endocrinol Metab 88(12) (2003) 6056-62.
[161] T.J. Mariani, V. Budhraja, B.H. Mecham, C.C. Gu, M.A. Watson, Y. Sadovsky, A variable fold change threshold determines significance for expression microarrays, FASEB J 17(2) (2003) 321-3.
[162] P. Diaz, J. Harris, F.J. Rosario, T.L. Powell, T. Jansson, Increased placental fatty acid transporter 6 and binding protein 3 expression and fetal liver lipid accumulation in a mouse model of obesity in pregnancy, Am J Physiol Regul Integr Comp Physiol 309(12) (2015) R1569-77.
[163] R.L. Schild, W.T. Schaiff, M.G. Carlson, E.J. Cronbach, D.M. Nelson, Y. Sadovsky, The activity of PPAR gamma in primary human trophoblasts is enhanced by oxidized lipids, J Clin Endocrinol Metab 87(3) (2002) 1105-10.
[164] Z.Y. Wang, [Chemoprevention in the high incidence area of lung cancer], Zhonghua Zhong Liu Za Zhi 11(3) (1989) 207-10.
[165] C.R. Cawyer, D. Horvat, D. Leonard, S.R. Allen, R.O. Jones, D.C. Zawieja, T.J. Kuehl, M.N. Uddin, Hyperglycemia impairs cytotrophoblast function via stress signaling, Am J Obstet Gynecol 211(5) (2014) 541 e1-8.
[166] N. Kubota, Y. Terauchi, H. Miki, H. Tamemoto, T. Yamauchi, K. Komeda, S. Satoh, R. Nakano, C. Ishii, T. Sugiyama, K. Eto, Y. Tsubamoto, A. Okuno, K. Murakami, H. Sekihara, G. Hasegawa, M. Naito, Y. Toyoshima, S. Tanaka, K. Shiota, T. Kitamura, T. Fujita, O. Ezaki, S. Aizawa, T. Kadowaki, et al., PPAR gamma mediates high-fat diet-induced adipocyte hypertrophy and insulin resistance, Mol Cell 4(4) (1999) 597-609.
[167] S.Z. Duan, C.Y. Ivashchenko, S.E. Whitesall, L.G. D'Alecy, D.C. Duquaine, F.C. Brosius, 3rd, F.J. Gonzalez, C. Vinson, M.A. Pierre, D.S. Milstone, R.M. Mortensen, Hypotension, lipodystrophy, and insulin resistance in generalized PPARgamma-deficient mice rescued from embryonic lethality, J Clin Invest 117(3) (2007) 812-22.
[168] I.L. Aye, T.L. Powell, T. Jansson, Review: Adiponectin--the missing link between maternal adiposity, placental transport and fetal growth?, Placenta 34 Suppl (2013) S40-5.
[169] J.A. Reuter, D.V. Spacek, M.P. Snyder, High-throughput sequencing technologies, Mol Cell 58(4) (2015) 586-97.
[170] C. Gene Ontology, The Gene Ontology project in 2008, Nucleic Acids Res 36(Database issue) (2008) D440-4.
[171] L. du Plessis, N. Skunca, C. Dessimoz, The what, where, how and why of gene ontology--a primer for bioinformaticians, Brief Bioinform 12(6) (2011) 723-35.
[172] H. Zafar, Y. Wang, L. Nakhleh, N. Navin, K. Chen, Monovar: single-nucleotide variant detection in single cells, Nat Methods 13(6) (2016) 505-7.
[173] M.A. Costa, Scrutinising the regulators of syncytialization and their expression in pregnancy-related conditions, Mol Cell Endocrinol 420 (2016) 180-93.

[174] R.J. Reiter, S.A. Rosales-Corral, L.C. Manchester, D.X. Tan, Peripheral reproductive organ health and melatonin: ready for prime time, Int J Mol Sci 14(4) (2013) 7231-72.
[175] G.J. Burton, E. Jauniaux, What is the placenta?, Am J Obstet Gynecol 213(4 Suppl) (2015) S6 e1, S6-8.
[176] G.J. Burton, A.L. Fowden, The placenta: a multifaceted, transient organ, Philos Trans R Soc Lond B Biol Sci 370(1663) (2015) 20140066.
[177] F. Soncin, M. Khater, C. To, D. Pizzo, O. Farah, A. Wakeland, K. Arul Nambi Rajan, K.K. Nelson, C.W. Chang, M. Moretto-Zita, D.R. Natale, L.C. Laurent, M.M. Parast, Comparative analysis of mouse and human placentae across gestation reveals species-specific regulators of placental development, Development 145(2) (2018).
[178] V.D. Winn, R. Haimov-Kochman, A.C. Paquet, Y.J. Yang, M.S. Madhusudhan, M. Gormley, K.T. Feng, D.A. Bernlohr, S. McDonagh, L. Pereira, A. Sali, S.J. Fisher, Gene expression profiling of the human maternal-fetal interface reveals dramatic changes between midgestation and term, Endocrinology 148(3) (2007) 1059-79.
[179] H.L. Seal, Studies in the history of probability and statistics. XV. The historical velopment of the Gauss linear model, Biometrika 54(1) (1967) 1-24.
[180] P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis, BMC Bioinformatics 9 (2008) 559.
[181] G. Yu, L.G. Wang, Y. Han, Q.Y. He, clusterProfiler: an R package for comparing biological themes among gene clusters, OMICS 16(5) (2012) 284-7.
[182] S.A. Founds, Y.P. Conley, J.F. Lyons-Weiler, A. Jeyabalan, W.A. Hogge, K.P. Conrad, Altered global gene expression in first trimester placentas of women destined to develop preeclampsia, Placenta 30(1) (2009) 15-24.
[183] X. Jiang, H.Y. Bar, J. Yan, S. Jones, P.M. Brannon, A.A. West, C.A. Perry, A. Ganti, E. Pressman, S. Devapatla, F. Vermeylen, M.T. Wells, M.A. Caudill, A higher maternal choline intake among third-trimester pregnant women lowers placental and circulating concentrations of the antiangiogenic factor fms-like tyrosine kinase-1 (sFLT1), FASEB J 27(3) (2013) 1245-53.
[184] F. Liu, C. Rouault, M. Guesnon, W. Zhu, K. Clement, S.A. Degrelle, T. Fournier, Comparative Study of PPARgamma Targets in Human Extravillous and Villous Cytotrophoblasts, PPAR Res 2020 (2020) 9210748.
[185] A.M. Mikheev, T. Nabekura, A. Kaddoumi, T.K. Bammler, R. Govindarajan, M.F. Hebert, J.D. Unadkat, Profiling gene expression in human placentae of different gestational ages: an OPRU Network and UW SCOR Study, Reprod Sci 15(9) (2008) 866-77.
[186] Z. Bar-Joseph, Analyzing time series gene expression data, Bioinformatics 20(16) (2004) 2493-503.
[187] J.C.H. Tsang, J.S.L. Vong, L. Ji, L.C.Y. Poon, P. Jiang, K.O. Lui, Y.B. Ni, K.F. To, Y.K.Y. Cheng, R.W.K. Chiu, Y.M.D. Lo, Integrative single-cell and cell-free plasma RNA transcriptomics elucidates placental cellular dynamics, Proc Natl Acad Sci U S A 114(37) (2017) E7786-E7795.
[188] B. Zhang, S. Horvath, A general framework for weighted gene co-expression network analysis, Stat Appl Genet Mol Biol 4 (2005) Article17.

[189] S. Horvath, J. Dong, Geometric interpretation of gene coexpression network analysis, PLoS Comput Biol 4(8) (2008) e1000117.
[190] P. Langfelder, S. Horvath, Eigengene networks for studying the relationships between co-expression modules, BMC Syst Biol 1 (2007) 54.
[191] X. Zhang, L.J. Fu, X.Q. Liu, Z.Y. Hu, Y. Jiang, R.F. Gao, Q. Feng, X. Lan, Y.Q. Geng, X.M. Chen, J.L. He, Y.X. Wang, Y.B. Ding, nm23 regulates decidualization through the PI3K-Akt-mTOR signaling pathways in mice and humans, Hum Reprod 31(10) (2016) 2339-51.
[192] W. Yang, L.D. Klaman, B. Chen, T. Araki, H. Harada, S.M. Thomas, E.L. George, B.G. Neel, An Shp2/SFK/Ras/Erk signaling pathway controls trophoblast stem cell survival, Dev Cell 10(3) (2006) 317-27.
[193] C.W. Chang, M.L. Cheong, G.D. Chang, M.S. Tsai, H. Chen, Involvement of Epac1/Rap1/CaMKI/HDAC5 signaling cascade in the regulation of placental cell fusion, Mol Hum Reprod 19(11) (2013) 745-55.
[194] P. Gerbaud, K. Tasken, G. Pidoux, Spatiotemporal regulation of cAMP signaling controls the human trophoblast fusion, Front Pharmacol 6 (2015) 202.
[195] T. Fournier, P. Therond, K. Handschuh, V. Tsatsaris, D. Evain-Brion, PPARgamma and early human placental development, Curr Med Chem 15(28) (2008) 3011-24.
[196] F. Gilardi, C. Winkler, L. Quignodon, J.G. Diserens, B. Toffoli, M. Schiffrin, C. Sardella, F. Preitner, B. Desvergne, Systemic PPARgamma deletion in mice provokes lipoatrophy, organomegaly, severe type 2 diabetes and metabolic inflexibility, Metabolism 95 (2019) 8-20.
[197] L. Zhao, X. Zheng, J. Liu, R. Zheng, R. Yang, Y. Wang, L. Sun, PPAR signaling pathway in the first trimester placenta from in vitro fertilization and embryo transfer, Biomed Pharmacother 118 (2019) 109251.
[198] S.S. Chassen, V. Ferchaud-Roucher, M.B. Gupta, T. Jansson, T.L. Powell, Alterations in placental long chain polyunsaturated fatty acid metabolism in human intrauterine growth restriction, Clin Sci (Lond) 132(5) (2018) 595-607.
[199] M. Kolodziej, S. Strauss, A. Lazaridis, V. Bucan, J.W. Kuhbier, P.M. Vogt, S. Konneker, Influence of glucose and insulin in human adipogenic differentiation models with adipose-derived stem cells, Adipocyte 8(1) (2019) 254-264.
[200] C.M. Scifres, B. Chen, D.M. Nelson, Y. Sadovsky, Fatty acid binding protein 4 regulates intracellular lipid accumulation in human trophoblasts, J Clin Endocrinol Metab 96(7) (2011) E1083-91.
[201] Fulin Liu, Christine Rouault, Mickael Guesnon, Wencan Zhu, Karine Clément, Séverine A. Degrelle, Thierry Fournier, Comparative Study of PPARγ Targets in Human Extravillous and Villous Cytotrophoblasts, PPAR Research In press (2020).
[202] T. Garin-Shkolnik, A. Rudich, G.S. Hotamisligil, M. Rubinstein, FABP4 attenuates PPARgamma and adipogenesis and is inversely correlated with PPARgamma in adipose tissues, Diabetes 63(3) (2014) 900-11.
[203] G.L. Victor Han, Jennifer Ballard, Aaron Booy, Nicholas Carruthers, Delfina Siroen Early onset preeclampsia is characterized by altered placental lipid metabolism and a premature increase in circulating FABP4, Nature Procedings (2010) 1-23.

[204] J. Buchrieser, S.A. Degrelle, T. Couderc, Q. Nevers, O. Disson, C. Manet, D.A. Donahue, F. Porrot, K.H. Hillion, E. Perthame, M.V. Arroyo, S. Souquere, K. Ruigrok, A. Dupressoir, T. Heidmann, X. Montagutelli, T. Fournier, M. Lecuit, O. Schwartz, IFITM proteins inhibit placental syncytiotrophoblast formation and promote fetal demise, Science 365(6449) (2019) 176-180.
[205] C.A. Labarrere, H.L. DiCarlo, E. Bammerlin, J.W. Hardin, Y.M. Kim, P. Chaemsaithong, D.M. Haas, G.S. Kassab, R. Romero, Failure of physiologic transformation of spiral arteries, endothelial and trophoblast cell activation, and acute atherosis in the basal plate of the placenta, Am J Obstet Gynecol 216(3) (2017) 287 e1-287 e16.
[206] L. Ji, J. Brkic, M. Liu, G. Fu, C. Peng, Y.L. Wang, Placental trophoblast cell differentiation: physiological regulation and pathological relevance to preeclampsia, Mol Aspects Med 34(5) (2013) 981-1023.
[207] V. Chandra, P. Huang, Y. Hamuro, S. Raghuram, Y. Wang, T.P. Burris, F. Rastinejad, Structure of the intact PPAR-gamma-RXR- nuclear receptor complex on DNA, Nature 456(7220) (2008) 350-6.
[208] Q. Zhao, D. Yang, L. Gao, M. Zhao, X. He, M. Zhu, C. Tian, G. Liu, L. Li, C. Hu, Downregulation of peroxisome proliferator-activated receptor gamma in the placenta correlates to hyperglycemia in offspring at young adulthood after exposure to gestational diabetes mellitus, J Diabetes Investig 10(2) (2019) 499-512.
[209] J.R. Chen, O.P. Lazarenko, M.L. Blackburn, S. Rose, R.E. Frye, T.M. Badger, A. Andres, K. Shankar, Maternal Obesity Programs Senescence Signaling and Glucose Metabolism in Osteo-Progenitors From Rat and Human, Endocrinology 157(11) (2016) 4172-4183.
[210] L. Kadam, B. Kilburn, D. Baczyk, H.R. Kohan-Ghadr, J. Kingdom, S. Drewlo, Rosiglitazone blocks first trimester in-vitro placental injury caused by NF-kappaB-mediated inflammation, Sci Rep 9(1) (2019) 2018.
[211] L. Kadam, N. Gomez-Lopez, T.N. Mial, H.R. Kohan-Ghadr, S. Drewlo, Rosiglitazone Regulates TLR4 and Rescues HO-1 and NRF2 Expression in Myometrial and Decidual Macrophages in Inflammation-Induced Preterm Birth, Reprod Sci 24(12) (2017) 1590-1599.
[212] H.R. Kohan-Ghadr, B.A. Kilburn, L. Kadam, E. Johnson, B.L. Kolb, J. Rodriguez-Kovacs, M. Hertz, D.R. Armant, S. Drewlo, Rosiglitazone augments antioxidant response in the human trophoblast and prevents apoptosisdagger, Biol Reprod 100(2) (2019) 479-494.
[213] J. Zhang, X. Peng, A. Yuan, Y. Xie, Q. Yang, L. Xue, Peroxisome proliferatoractivated receptor gamma mediates porcine placental angiogenesis through hypoxia inducible factor, vascular endothelial growth factor and angiopoietinmediated signaling, Mol Med Rep 16(3) (2017) 2636-2644.
[214] Y. Lin, K.M. Bircsak, L. Gorczyca, X. Wen, L.M. Aleksunes, Regulation of the placental BCRP transporter by PPAR gamma, J Biochem Mol Toxicol 31(5) (2017).
[215] T. Fournier, J. Guibourdenche, K. Handschuh, V. Tsatsaris, B. Rauwel, C. Davrinche, D. Evain-Brion, PPARgamma and human trophoblast differentiation, J Reprod Immunol 90(1) (2011) 41-9.

[216] T. Fournier, K. Handschuh, V. Tsatsaris, J. Guibourdenche, D. Evain-Brion, Role of nuclear receptors and their ligands in human trophoblast invasion, J Reprod Immunol 77(2) (2008) 161-70.
[217] H.J. Kliman, J.E. Nestler, E. Sermasi, J.M. Sanger, J.F. Strauss, 3rd, Purification, characterization, and in vitro differentiation of cytotrophoblasts from human term placentae, Endocrinology 118(4) (1986) 1567-82.
[218] E. Alsat, V. Mirlesse, C. Fondacci, M. Dodeur, D. Evain-Brion, Parathyroid hormone increases epidermal growth factor receptors in cultured human trophoblastic cells from early and term placenta, J Clin Endocrinol Metab 73(2) (1991) 288-95.
[219] C. Henegar, J. Tordjman, V. Achard, D. Lacasa, I. Cremer, M. Guerre-Millo, C. Poitou, A. Basdevant, V. Stich, N. Viguerie, D. Langin, P. Bedossa, J.D. Zucker, K. Clement, Adipose tissue transcriptomic signature highlights the pathological relevance of extracellular matrix in human obesity, Genome Biol 9(1) (2008) R14.
[220] V.G. Tusher, R. Tibshirani, G. Chu, Significance analysis of microarrays applied to the ionizing radiation response, Proc Natl Acad Sci U S A 98(9) (2001) 5116-21.
[221] C. Rouault, K. Clement, M. Guesnon, C. Henegar, M.A. Charles, B. Heude, D. Evain-Brion, S.A. Degrelle, T. Fournier, Transcriptomic signatures of villous cytotrophoblast and syncytiotrophoblast in term human placenta, Placenta 44 (2016) 83-90.
[222] G. Bindea, B. Mlecnik, H. Hackl, P. Charoentong, M. Tosolini, A. Kirilovsky, W.H. Fridman, F. Pages, Z. Trajanoski, J. Galon, ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks, Bioinformatics 25(8) (2009) 1091-3.
[223] B. Huppertz, Placental origins of preeclampsia: challenging the current hypothesis, Hypertension 51(4) (2008) 970-5.
[224] J.M. Roberts, D.W. Cooper, Pathogenesis and genetics of pre-eclampsia, Lancet 357(9249) (2001) 53-6.
[225] T. Fournier, K. Handschuh, V. Tsatsaris, D. Evain-Brion, Involvement of PPARgamma in human trophoblast invasion, Placenta 28 Suppl A (2007) S76-81.
[226] J.M. Lehmann, L.B. Moore, T.A. Smith-Oliver, W.O. Wilkison, T.M. Willson, S.A. Kliewer, An antidiabetic thiazolidinedione is a high affinity ligand for peroxisome proliferator-activated receptor gamma (PPAR gamma), J Biol Chem 270(22) (1995) 12953-6.
[227] M. Gimpfl, J. Rozman, M. Dahlhoff, R. Kubeck, A. Blutke, B. Rathkolb, M. Klingenspor, M. Hrabe de Angelis, S. Oner-Sieben, A. Seibt, A.A. Roscher, E. Wolf, R. Ensenauer, Modification of the fatty acid composition of an obesogenic diet improves the maternal and placental metabolic environment in obese pregnant mice, Biochim Biophys Acta Mol Basis Dis 1863(6) (2017) 1605-1614.
[228] A.J. Ridley, M.A. Schwartz, K. Burridge, R.A. Firtel, M.H. Ginsberg, G. Borisy, J.T. Parsons, A.R. Horwitz, Cell migration: integrating signals from front to back, Science 302(5651) (2003) 1704-9.
[229] C. Chakraborty, L.M. Gleeson, T. McKinnon, P.K. Lala, Regulation of human trophoblast migration and invasiveness, Can J Physiol Pharmacol 80(2) (2002) 116-24.

[230] P. Bischof, A. Meisser, A. Campana, Paracrine and autocrine regulators of trophoblast invasion--a review, Placenta 21 Suppl A (2000) S55-60.
[231] V. Nadeau, J. Charron, Essential role of the ERK/MAPK pathway in blood-placental barrier formation, Development 141(14) (2014) 2825-37.
[232] J. Charron, V. Bissonauth, V. Nadeau, Implication of MEK1 and MEK2 in the establishment of the blood-placenta barrier during placentogenesis in mouse, Reprod Biomed Online 25(1) (2012) 58-67.
[233] N. Lala, G.V. Girish, A. Cloutier-Bosworth, P.K. Lala, Mechanisms in decorin regulation of vascular endothelial growth factor-induced human trophoblast migration and acquisition of endothelial phenotype, Biol Reprod 87(3) (2012) 59.
[234] S.L. Roberti, R. Higa, V. White, T.L. Powell, T. Jansson, A. Jawerbaum, Critical role of mTOR, PPARgamma and PPARdelta signaling in regulating early pregnancy decidual function, embryo viability and feto-placental growth, Mol Hum Reprod 24(6) (2018) 327-340.
[235] S. Haider, V. Kunihs, C. Fiala, J. Pollheimer, M. Knofler, Expression pattern and phosphorylation status of Smad2/3 in different subtypes of human first trimester trophoblast, Placenta 57 (2017) 17-25.
[236] D. Xie, J. Zhu, Q. Liu, J. Li, M. Song, K. Wang, Q. Zhou, Y. Jia, T. Li, Dysregulation of HDAC9 Represses Trophoblast Cell Migration and Invasion Through TIMP3 Activation in Preeclampsia, Am J Hypertens 32(5) (2019) 515-523.
[237] C. Estella, I. Herrer, S.P. Atkinson, A. Quinonero, S. Martinez, A. Pellicer, C. Simon, Inhibition of histone deacetylase activity in human endometrial stromal cells promotes extracellular matrix remodelling and limits embryo invasion, PLoS One 7(1) (2012) e30508.
[238] F.J. Rosario, K.G. Dimasuay, Y. Kanai, T.L. Powell, T. Jansson, Regulation of amino acid transporter trafficking by mTORC1 in primary human trophoblast cells is mediated by the ubiquitin ligase Nedd4-2, Clin Sci (Lond) 130(7) (2016) 499-512.
[239] Y.Y. Chen, F.J. Rosario, M.A. Shehab, T.L. Powell, M.B. Gupta, T. Jansson, Increased ubiquitination and reduced plasma membrane trafficking of placental amino acid transporter SNAT-2 in human IUGR, Clin Sci (Lond) 129(12) (2015) 1131-41.
[240] H.S. Camp, S.R. Tafuri, Regulation of peroxisome proliferator-activated receptor gamma activity by mitogen-activated protein kinase, J Biol Chem 272(16) (1997) 10811-6.
[241] J.J. Li, R. Wang, R. Lama, X. Wang, Z.E. Floyd, E.A. Park, F.F. Liao, Ubiquitin Ligase NEDD4 Regulates PPARgamma Stability and Adipocyte Differentiation in 3T3-L1 Cells, Sci Rep 6 (2016) 38550.
[242] M. Watanabe, H. Takahashi, Y. Saeki, T. Ozaki, S. Itoh, M. Suzuki, W. Mizushima, K. Tanaka, S. Hatakeyama, The E3 ubiquitin ligase TRIM23 regulates adipocyte differentiation via stabilization of the adipogenic activator PPARgamma, Elife 4 (2015) e05615.
[243] S. DaSilva-Arnold, J.L. James, A. Al-Khan, S. Zamudio, N.P. Illsley, Differentiation of first trimester cytotrophoblast to extravillous trophoblast involves an epithelial-mesenchymal transition, Placenta 36(12) (2015) 1412-8.
[244] L.K. Harris, C.J. Jones, J.D. Aplin, Adhesion molecules in human trophoblast -

a review. II. extravillous trophoblast, Placenta 30(4) (2009) 299-304.
[245] K. Handschuh, J. Guibourdenche, V. Tsatsaris, M. Guesnon, I. Laurendeau, D. Evain-Brion, T. Fournier, Human chorionic gonadotropin expression in human trophoblasts from early placenta: comparative study between villous and extravillous trophoblastic cells, Placenta 28(2-3) (2007) 175-84.
[246] K. Naruse, B.A. Innes, J.N. Bulmer, S.C. Robson, R.F. Searle, G.E. Lash, Secretion of cytokines by villous cytotrophoblast and extravillous trophoblast in the first trimester of human pregnancy, J Reprod Immunol 86(2) (2010) 148-50.
[247] A.K. Reka, H. Kurapati, V.R. Narala, G. Bommer, J. Chen, T.J. Standiford, V.G. Keshamouni, Peroxisome proliferator-activated receptor-gamma activation inhibits tumor metastasis by antagonizing Smad3-mediated epithelial-mesenchymal transition, Mol Cancer Ther 9(12) (2010) 3221-32.
[248] K. Handschuh, J. Guibourdenche, M. Cocquebert, V. Tsatsaris, M. Vidaud, D. Evain-Brion, T. Fournier, Expression and regulation by PPARgamma of hCG alpha- and beta-subunits: comparison between villous and invasive extravillous trophoblastic cells, Placenta 30(12) (2009) 1016-22.
[249] S. Yamagishi, T. Matsui, K. Nakamura, M. Takeuchi, H. Inoue, Telmisartan inhibits advanced glycation end products (AGEs)-elicited endothelial cell injury by suppressing AGE receptor (RAGE) expression via peroxisome proliferator-activated receptor-gammaactivation, Protein Pept Lett 15(8) (2008) 850-3.
[250] J.G. Chen, X. Li, H.Y. Huang, H.L. Liu, D.G. Liu, T.J. Song, C.G. Ma, D. Ma, H.Y. Song, Q.Q. Tang, Identification of a peroxisome proliferator responsive element (PPRE)-like cis-element in mouse plasminogen activator inhibitor-1 gene promoter, Biochem Biophys Res Commun 347(3) (2006) 821-6.
[251] S. Mandard, F. Zandbergen, N.S. Tan, P. Escher, D. Patsouris, W. Koenig, R. Kleemann, A. Bakker, F. Veenman, W. Wahli, M. Muller, S. Kersten, The direct peroxisome proliferator-activated receptor target fasting-induced adipose factor (FIAF/PGAR/ANGPTL4) is present in blood plasma as a truncated protein that is increased by fenofibrate treatment, J Biol Chem 279(33) (2004) 34411-20.
[252] I. Szatmari, G. Vamosi, P. Brazda, B.L. Balint, S. Benko, L. Szeles, V. Jeney, C. Ozvegy-Laczka, A. Szanto, E. Barta, J. Balla, B. Sarkadi, L. Nagy, Peroxisome proliferator-activated receptor gamma-regulated ABCG2 expression confers cytoprotection to human dendritic cells, J Biol Chem 281(33) (2006) 23812-23.
[253] N. Xu, B. Ahren, J. Jiang, P. Nilsson-Ehle, Down-regulation of apolipoprotein M expression is mediated by phosphatidylinositol 3-kinase in HepG2 cells, Biochim Biophys Acta 1761(2) (2006) 256-60.
[254] W.K. Leung, A.H. Bai, V.Y. Chan, J. Yu, M.W. Chan, K.F. To, J.R. Wu, K.K. Chan, Y.G. Fu, F.K. Chan, J.J. Sung, Effect of peroxisome proliferator activated receptor gamma ligands on growth and gene expression profiles of gastric cancer cells, Gut 53(3) (2004) 331-8.
[255] A. Baldan, J. Relat, P.F. Marrero, D. Haro, Functional interaction between peroxisome proliferator-activated receptors-alpha and Mef-2C on human carnitine palmitoyltransferase 1beta (CPT1beta) gene activation, Nucleic Acids Res 32(16) (2004) 4742-9.

[256] U.H. Park, S.K. Yoon, T. Park, E.J. Kim, S.J. Um, Additional sex comb-like (ASXL) proteins 1 and 2 play opposite roles in adipogenesis via reciprocal regulation of peroxisome proliferator-activated receptor {gamma}, J Biol Chem 286(2) (2011) 1354-63.
[257] L. Sarov-Blat, R.S. Kiss, B. Haidar, N. Kavaslar, M. Jaye, M. Bertiaux, K. Steplewski, M.R. Hurle, D. Sprecher, R. McPherson, Y.L. Marcel, Predominance of a proinflammatory phenotype in monocyte-derived macrophages from subjects with low plasma HDL-cholesterol, Arterioscler Thromb Vasc Biol 27(5) (2007) 1115-22.
[258] Y. Rival, A. Stennevin, L. Puech, A. Rouquette, C. Cathala, F. Lestienne, E. Dupont-Passelaigue, J.F. Patoiseau, T. Wurch, D. Junquero, Human adipocyte fatty acid-binding protein (aP2) gene promoter-driven reporter assay discriminates nonlipogenic peroxisome proliferator-activated receptor gamma ligands, J Pharmacol Exp Ther 311(2) (2004) 467-75.
[259] D.B. Savage, C.P. Sewter, E.S. Klenk, D.G. Segal, A. Vidal-Puig, R.V. Considine, S. O'Rahilly, Resistin / Fizz3 expression in relation to obesity and peroxisome proliferator-activated receptor-gamma action in humans, Diabetes 50(10) (2001) 2199-202.
[260] G. Kronke, A. Kadl, E. Ikonomu, S. Bluml, A. Furnkranz, I.J. Sarembock, V.N. Bochkov, M. Exner, B.R. Binder, N. Leitinger, Expression of heme oxygenase-1 in human vascular cells is regulated by peroxisome proliferator-activated receptors, Arterioscler Thromb Vasc Biol 27(6) (2007) 1276-82.
[261] J.C. Yoon, T.W. Chickering, E.D. Rosen, B. Dussault, Y. Qin, A. Soukas, J.M. Friedman, W.E. Holmes, B.M. Spiegelman, Peroxisome proliferator-activated receptor gamma target gene encoding a novel angiopoietin-related protein associated with adipose differentiation, Mol Cell Biol 20(14) (2000) 5343-9.
[262] N. Marx, T. Bourcier, G.K. Sukhova, P. Libby, J. Plutzky, PPARgamma activation in human endothelial cells increases plasminogen activator inhibitor type-1 expression: PPARgamma as a potential mediator in vascular disease, Arterioscler Thromb Vasc Biol 19(3) (1999) 546-51.
[263] X. Lan, L.J. Fu, J. Zhang, X.Q. Liu, H.J. Zhang, X. Zhang, M.F. Ma, X.M. Chen, J.L. He, L.B. Li, Y.X. Wang, Y.B. Ding, Bisphenol A exposure promotes HTR-8/SVneo cell migration and impairs mouse placentation involving upregulation of integrin-beta1 and MMP-9 and stimulation of MAPK and PI3K signaling pathways, Oncotarget 8(31) (2017) 51507-51521.
[264] L. Liu, Y. Wang, C. Shen, J. He, X. Liu, Y. Ding, R. Gao, X. Chen, Benzo(a)pyrene inhibits migration and invasion of extravillous trophoblast HTR-8/SVneo cells via activation of the ERK and JNK pathway, J Appl Toxicol 36(7) (2016) 946-55.
[265] M. Cohen, A. Meisser, L. Haenggeli, P. Bischof, Involvement of MAPK pathway in TNF-alpha-induced MMP-9 expression in human trophoblastic cells, Mol Hum Reprod 12(4) (2006) 225-32.
[266] Y. Lee, J.D. Fryer, H. Kang, J. Crespo-Barreto, A.B. Bowman, Y. Gao, J.J. Kahle, J.S. Hong, F. Kheradmand, H.T. Orr, M.J. Finegold, H.Y. Zoghbi, ATXN1 protein family and CIC regulate extracellular matrix remodeling and lung alveolarization, Dev Cell 21(4) (2011) 746-57.

[267] F. Soncin, D. Natale, M.M. Parast, Signaling pathways in mouse and human trophoblast differentiation: a comparative review, Cell Mol Life Sci 72(7) (2015) 1291-302.
[268] J.L. James, A.M. Carter, L.W. Chamley, Human placentation from nidation to 5 weeks of gestation. Part I: What do we know about formative placental development following implantation?, Placenta 33(5) (2012) 327-34.
[269] E. Jauniaux, A.L. Watson, J. Hempstock, Y.P. Bao, J.N. Skepper, G.J. Burton, Onset of maternal arterial blood flow and placental oxidative stress. A possible factor in human early pregnancy failure, Am J Pathol 157(6) (2000) 2111-22.
[270] B. Sibai, G. Dekker, M. Kupferminc, Pre-eclampsia, Lancet 365(9461) (2005) 785-99.
[271] D. Goldman-Wohl, S. Yagel, Regulation of trophoblast invasion: from normal implantation to pre-eclampsia, Mol Cell Endocrinol 187(1-2) (2002) 233-8.
[272] T. Todros, A. Sciarrone, E. Piccoli, C. Guiot, P. Kaufmann, J. Kingdom, Umbilical Doppler waveforms and placental villous angiogenesis in pregnancies complicated by fetal growth restriction, Obstet Gynecol 93(4) (1999) 499-503.
[273] H. Schneider, Placental oxygen consumption. Part II: in vitro studies--a review, Placenta 21 Suppl A (2000) S38-44.
[274] A.M. Carter, Placental oxygen consumption. Part I: in vivo studies--a review, Placenta 21 Suppl A (2000) S31-7.
[275] C.N. Miller, J.A. Dye, A.R. Henriquez, E.J. Stewart, K.S. Lavrich, G.K. Carswell, H. Ren, D.L. Freeborn, S.J. Snow, M.C. Schladweiler, J.H. Richards, P.R.S. Kodavanti, A. Fisher, B.N. Chorley, U.P. Kodavanti, Ozone-induced fetal growth restriction in rats is associated with sexually dimorphic placental and fetal metabolic adaptation, Mol Metab 42 (2020) 101094.
[276] A.N. Sferruzzi-Perri, J.S. Higgins, O.R. Vaughan, A.J. Murray, A.L. Fowden, Placental mitochondria adapt developmentally and in response to hypoxia to support fetal growth, Proc Natl Acad Sci U S A 116(5) (2019) 1621-1626.
[277] O. Holland, M. Dekker Nitert, L.A. Gallo, M. Vejzovic, J.J. Fisher, A.V. Perkins, Review: Placental mitochondrial function and structure in gestational disorders, Placenta 54 (2017) 2-9.
[278] H. Song, B.P. Telugu, L.P. Thompson, Sexual dimorphism of mitochondrial function in the hypoxic guinea pig placenta, Biol Reprod 100(1) (2019) 208-216.
[279] A.N. Sferruzzi-Perri, E.J. Camm, The Programming Power of the Placenta, Front Physiol 7 (2016) 33.
[280] C.W. Chang, A.K. Wakeland, M.M. Parast, Trophoblast lineage specification, differentiation and their regulation by oxygen tension, J Endocrinol 236(1) (2018) R43-R56.
[281] M. Mandl, R. Depping, Hypoxia-inducible aryl hydrocarbon receptor nuclear translocator (ARNT) (HIF-1beta): is it a rare exception?, Mol Med 20 (2014) 215-20.
[282] J.W. Lee, S.H. Bae, J.W. Jeong, S.H. Kim, K.W. Kim, Hypoxia-inducible factor (HIF-1)alpha: its protein stability and biological functions, Exp Mol Med 36(1) (2004) 1-12.
[283] D.M. Adelman, M. Gertsenstein, A. Nagy, M.C. Simon, E. Maltepe, Placental cell

fates are regulated in vivo by HIF-mediated hypoxia responses, Genes Dev 14(24) (2000) 3191-203.
[284] S. Muralimanoharan, A. Maloyan, L. Myatt, Evidence of sexual dimorphism in the placental function with severe preeclampsia, Placenta 34(12) (2013) 1183-9.
[285] S.A. Degrelle, T. Fournier, Fetal-sex determination of human placental tissues, Placenta 61 (2018) 103-105.
[286] D. Hoch, B. Novakovic, S. Cvitic, R. Saffery, G. Desoye, A. Majali-Martinez, Sex matters: XIST and DDX3Y gene expression as a tool to determine fetal sex in human first trimester placenta, Placenta 97 (2020) 68-70.
[287] L. Rachdi, A. Maugein, S. Pechberty, M. Armanet, J. Hamroune, P. Ravassard, S. Marullo, O. Albagli, R. Scharfmann, Regulated expression and function of the GABAB receptor in human pancreatic beta cell line and islets, Sci Rep 10(1) (2020) 13469.
[288] M.I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, Genome Biol 15(12) (2014) 550.
[289] Y. Liao, J. Wang, E.J. Jaehnig, Z. Shi, B. Zhang, WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs, Nucleic Acids Res 47(W1) (2019) W199-W205.
[290] A.E. Braun, K.L. Muench, B.G. Robinson, A. Wang, T.D. Palmer, V.D. Winn, Examining Sex Differences in the Human Placental Transcriptome During the First Fetal Androgen Peak, Reprod Sci (2020).
[291] T.L. Gonzalez, T. Sun, A.F. Koeppel, B. Lee, E.T. Wang, C.R. Farber, S.S. Rich, L.W. Sundheimer, R.A. Buttle, Y.I. Chen, J.I. Rotter, S.D. Turner, J. Williams, 3rd, M.O. Goodarzi, M.D. Pisarska, Sex differences in the late first trimester human placenta transcriptome, Biol Sex Differ 9(1) (2018) 4.
[292] V. Sitras, C. Fenton, R. Paulssen, A. Vartun, G. Acharya, Differences in gene expression between first and third trimester human placenta: a microarray study, PLoS One 7(3) (2012) e33294.
[293] T. Sun, T.L. Gonzalez, N. Deng, R. DiPentino, E.L. Clark, B. Lee, J. Tang, Y. Wang, B.R. Stripp, C. Yao, H.R. Tseng, S.A. Karumanchi, A.F. Koeppel, S.D. Turner, C.R. Farber, S.S. Rich, E.T. Wang, J. Williams, M.D. Pisarska, Sexually Dimorphic Crosstalk at the Maternal-Fetal Interface, J Clin Endocrinol Metab 105(12) (2020).
[294] O. Brew, M.H.F. Sullivan, Oxygen and tissue culture affect placental gene expression, Placenta 55 (2017) 13-20.
[295] C.E. Cross, M.F. Tolba, C.M. Rondelli, M. Xu, S.Z. Abdel-Rahman, Oxidative Stress Alters miRNA and Gene Expression Profiles in Villous First Trimester Trophoblasts, Biomed Res Int 2015 (2015) 257090.
[296] J. Treissman, V. Yuan, J. Baltayeva, H.T. Le, B. Castellana, W.P. Robinson, A.G. Beristain, Low oxygen enhances trophoblast column growth by potentiating differentiation of the extravillous lineage and promoting LOX activity, Development 147(2) (2020).
[297] F. Liu, W. Zhu, H. Shoaito, A. Chissey, S.A. Degrelle, T. Fournier, Mining of combined human placental gene expression data across pregnancy, applied to PPAR signaling pathway, Placenta 99 (2020) 157-165.
[298] L. Qi, B. Liu, X. Chen, Q. Liu, W. Li, B. Lv, X. Xu, L. Wang, Q. Zeng, J. Xue, Z.

Xue, Single-Cell Transcriptomic Analysis Reveals Mitochondrial Dynamics in Oocytes of Patients With Polycystic Ovary Syndrome, Front Genet 11 (2020) 396.
[299] Z. Wang, W. Tang, J. Yuan, B. Qiang, W. Han, X. Peng, Integrated Analysis of RNA-Binding Proteins in Glioma, Cancers (Basel) 12(4) (2020).
[300] M.J. Koh, D.H. Shin, S.J. Lee, C.S. Hwang, H.J. Lee, A. Kim, W.Y. Park, J.H. Lee, K.U. Choi, J.Y. Kim, C.H. Lee, M.Y. Sol, Gastric-type gene expression and phenotype in non-terminal respiratory unit type adenocarcinoma of the lung with invasive mucinous adenocarcinoma morphology, Histopathology 76(6) (2020) 898-905.
[301] S.R. Broderick, S. Wijeratne, A.J. Wijeratn, L.J. Chapin, T. Meulia, M.L. Jones, RNA-sequencing reveals early, dynamic transcriptome changes in the corollas of pollinated petunias, BMC Plant Biol 14 (2014) 307.
[302] B. Saha, A. Ganguly, P. Home, B. Bhattacharya, S. Ray, A. Ghosh, M.A.K. Rumi, C. Marsh, V.A. French, S. Gunewardena, S. Paul, TEAD4 ensures postimplantation development by promoting trophoblast self-renewal: An implication in early human pregnancy loss, Proc Natl Acad Sci U S A 117(30) (2020) 17864-17875.
[303] S. Haider, G. Meinhardt, P. Velicky, G.R. Otti, G. Whitley, C. Fiala, J. Pollheimer, M. Knofler, Notch signaling plays a critical role in motility and differentiation of human first-trimester cytotrophoblasts, Endocrinology 155(1) (2014) 263-74.
[304] F. Ietta, N. Bechi, R. Romagnoli, J. Bhattacharjee, M. Realacci, M. Di Vito, C. Ferretti, L. Paulesu, 17{beta}-Estradiol modulates the macrophage migration inhibitory factor secretory pathway by regulating ABCA1 expression in human first-trimester placenta, Am J Physiol Endocrinol Metab 298(3) (2010) E411-8.
[305] A.T. Kho, D. Chhabra, S. Sharma, W. Qiu, V.J. Carey, R. Gaedigk, C.A. Vyhlidal, J.S. Leeder, K.G. Tantisira, S.T. Weiss, Age, Sexual Dimorphism, and Disease Associations in the Developing Human Fetal Lung Transcriptome, Am J Respir Cell Mol Biol 54(6) (2016) 814-21.
[306] D. Baczyk, M.C. Audette, E. Coyaud, B. Raught, J.C. Kingdom, Spatiotemporal distribution of small ubiquitin-like modifiers during human placental development and in response to oxidative and inflammatory stress, J Physiol 596(9) (2018) 1587-1600.
[307] C. Bebington, F.J. Doherty, S.D. Fleming, Ubiquitin cross-reactive protein gene expression is increased in decidualized endometrial stromal cells at the initiation of pregnancy, Mol Hum Reprod 5(10) (1999) 966-72.
[308] L. Anton, A.G. Brown, M.S. Bartolomei, M.A. Elovitz, Differential methylation of genes associated with cell adhesion in preeclamptic placentas, PLoS One 9(6) (2014) e100148.
[309] N. Mohammad, A. Yaqinuddin, F. Kakal, L. Sheikh, R. Qureshi, M. Somani, Frequent hypomethylation of PTGS2 gene promoter in human term placenta, Ital J Anat Embryol 118(2) (2013) 211-6.
[310] Q. Ke, M. Costa, Hypoxia-inducible factor-1 (HIF-1), Mol Pharmacol 70(5) (2006) 1469-80.
[311] M. Yamanaka-Tatematsu, A. Nakashima, N. Fujita, T. Shima, T. Yoshimori, S. Saito, Autophagy induced by HIF1alpha overexpression supports trophoblast invasion by supplying cellular energy, PLoS One 8(10) (2013) e76605.

[312] M. Gauster, S. Maninger, M. Siwetz, A. Deutsch, A. El-Heliebi, D. Kolb-Lenz, U. Hiden, G. Desoye, F. Herse, A. Prokesch, Downregulation of p53 drives autophagy during human trophoblast differentiation, Cell Mol Life Sci 75(10) (2018) 1839-1855.
[313] J. Perkins, J. St John, A. Ahmed, Modulation of trophoblast cell death by oxygen and EGF, Mol Med 8(12) (2002) 847-56.
[314] S. Cvitic, M.S. Longtine, H. Hackl, K. Wagner, M.D. Nelson, G. Desoye, U. Hiden, The human placental sexome differs between trophoblast epithelium and villous vessel endothelium, PLoS One 8(10) (2013) e79233.
[315] Hypertension in pregnancy. Report of the American College of Obstetricians and Gynecologists' Task Force on Hypertension in Pregnancy, Obstet Gynecol 122(5) (2013) 1122-1131.
[316] E.V. Kuklina, C. Ayala, W.M. Callaghan, Hypertensive disorders and severe obstetric morbidity in the United States, Obstet Gynecol 113(6) (2009) 1299-1306.
[317] M.C. Hogan, K.J. Foreman, M. Naghavi, S.Y. Ahn, M. Wang, S.M. Makela, A.D. Lopez, R. Lozano, C.J. Murray, Maternal mortality for 181 countries, 1980-2008: a systematic analysis of progress towards Millennium Development Goal 5, Lancet 375(9726) (2010) 1609-23.
[318] V. Olié, E. Moutengou, C. Deneux-Tharaux, S. Kretz, A. Vallée, J. Blacher, V. Tsatsaris, G. Plu-Bureau, Prevalence of hypertensive disorders during pregnancy and post-partum in France, Archives of Cardiovascular Diseases Supplements 12(1) (2020) 155-156.
[319] R. Bahado-Singh, L.C. Poon, A. Yilmaz, A. Syngelaki, O. Turkoglu, P. Kumar, J. Kirma, M. Allos, V. Accurti, J. Li, P. Zhao, S.F. Graham, D.R. Cool, K. Nicolaides, Integrated Proteomic and Metabolomic prediction of Term Preeclampsia, Sci Rep 7(1) (2017) 16189.
[320] R.S. Kelly, D.C. Croteau-Chonka, A. Dahlin, H. Mirzakhani, A.C. Wu, E.S. Wan, M.J. McGeachie, W. Qiu, J.E. Sordillo, A. Al-Garawi, K.J. Gray, T.F. McElrath, V.J. Carey, C.B. Clish, A.A. Litonjua, S.T. Weiss, J.A. Lasky-Su, Integration of metabolomic and transcriptomic networks in pregnant women reveals biological pathways and predictive signatures associated with preeclampsia, Metabolomics 13(1) (2017).
[321] S. Agrawal, A.S. Cerdeira, C. Redman, M. Vatish, Meta-Analysis and Systematic Review to Assess the Role of Soluble FMS-Like Tyrosine Kinase-1 and Placenta Growth Factor Ratio in Prediction of Preeclampsia: The SaPPPhirE Study, Hypertension 71(2) (2018) 306-316.
[322] N. O'Gorman, D. Wright, A. Syngelaki, R. Akolekar, A. Wright, L.C. Poon, K.H. Nicolaides, Competing risks model in screening for preeclampsia by maternal factors and biomarkers at 11-13 weeks gestation, Am J Obstet Gynecol 214(1) (2016) 103 e1-103 e12.
[323] N. O'Gorman, D. Wright, L.C. Poon, D.L. Rolnik, A. Syngelaki, A. Wright, R. Akolekar, S. Cicero, D. Janga, J. Jani, F.S. Molina, C. de Paco Matallana, N. Papantoniou, N. Persico, W. Plasencia, M. Singh, K.H. Nicolaides, Accuracy of competing-risks model in screening for pre-eclampsia by maternal factors and biomarkers at 11-13 weeks' gestation, Ultrasound Obstet Gynecol 49(6) (2017) 751-755.
[324] R. McGinnis, V. Steinthorsdottir, N.O. Williams, G. Thorleifsson, S. Shooter, S.

Hjartardottir, S. Bumpstead, L. Stefansdottir, L. Hildyard, J.K. Sigurdsson, J.P. Kemp, G.B. Silva, L.C.V. Thomsen, T. Jaaskelainen, E. Kajantie, S. Chappell, N. Kalsheker, A. Moffett, S. Hiby, W.K. Lee, S. Padmanabhan, N.A.B. Simpson, V.A. Dolby, E. Staines-Urias, S.M. Engel, A. Haugan, L. Trogstad, G. Svyatova, N. Zakhidova, D. Najmutdinova, F. Consortium, G. Consortium, A.F. Dominiczak, H.K. Gjessing, J.P. Casas, F. Dudbridge, J.J. Walker, F.B. Pipkin, U. Thorsteinsdottir, R.T. Geirsson, D.A. Lawlor, A.C. Iversen, P. Magnus, H. Laivuori, K. Stefansson, L. Morgan, Variants in the fetal genome near FLT1 are associated with risk of preeclampsia, Nat Genet 49(8) (2017) 1255-1260.
[325] K.J. Gray, V.P. Kovacheva, H. Mirzakhani, A.C. Bjonnes, B. Almoguera, A.T. DeWan, E.W. Triche, A.F. Saftlas, J. Hoh, D.L. Bodian, E. Klein, K.C. Huddleston, S.A. Ingles, C.J. Lockwood, H. Hakonarson, T.F. McElrath, J.C. Murray, M.L. Wilson, E.R. Norwitz, S.A. Karumanchi, B.T. Bateman, B.J. Keating, R. Saxena, Gene-Centric Analysis of Preeclampsia Identifies Maternal Association at PLEKHG1, Hypertension 72(2) (2018) 408-416.
[326] A. Meirhaeghe, L. Fajas, N. Helbecque, D. Cottel, P. Lebel, J. Dallongeville, S. Deeb, J. Auwerx, P. Amouyel, A genetic polymorphism of the peroxisome proliferator-activated receptor gamma gene influences plasma leptin levels in obese humans, Hum Mol Genet 7(3) (1998) 435-40.
[327] A. Doney, B. Fischer, D. Frew, A. Cumming, D.M. Flavell, M. World, H.E. Montgomery, D. Boyle, A. Morris, C.N. Palmer, Haplotype analysis of the PPARgamma Pro12Ala and C1431T variants reveals opposing associations with body weight, BMC Genet 3 (2002) 21.
[328] R. Valve, K. Sivenius, R. Miettinen, J. Pihlajamaki, A. Rissanen, S.S. Deeb, J. Auwerx, M. Uusitupa, M. Laakso, Two polymorphisms in the peroxisome proliferator-activated receptor-gamma gene are associated with severe overweight among obese women, J Clin Endocrinol Metab 84(10) (1999) 3708-12.
[329] G. Cai, X. Zhang, W. Weng, G. Shi, S. Xue, B. Zhang, Associations between PPARG polymorphisms and the risk of essential hypertension, PLoS One 12(7) (2017) e0181644.
[330] T.H. Chao, Y.H. Li, J.H. Chen, H.L. Wu, G.Y. Shi, P.Y. Liu, W.C. Tsai, H.R. Guo, The 161TT genotype in the exon 6 of the peroxisome-proliferator-activated receptor gamma gene is associated with premature acute myocardial infarction and increased lipid peroxidation in habitual heavy smokers, Clin Sci (Lond) 107(5) (2004) 461-6.
[331] X.L. Wang, J. Oosterhof, N. Duarte, Peroxisome proliferator-activated receptor gamma C161-->T polymorphism and coronary artery disease, Cardiovasc Res 44(3) (1999) 588-94.
[332] J.E. Cecil, B. Fischer, A.S. Doney, M. Hetherington, P. Watt, W. Wrieden, C. Bolton-Smith, C.N. Palmer, The Pro12Ala and C-681G variants of the PPARG locus are associated with opposing growth phenotypes in young schoolchildren, Diabetologia 48(8) (2005) 1496-502.
[333] J. Lin, Y. Chen, W.F. Tang, C. Liu, S. Zhang, Z.Q. Guo, G. Chen, X.W. Zheng, PPARG rs3856806 C>T Polymorphism Increased the Risk of Colorectal Cancer: A Case-Control Study in Eastern Chinese Han Population, Front Oncol 9 (2019) 63.

[334] B. Heude, V. Pelloux, A. Forhan, J.F. Bedel, J.M. Lacorte, K. Clement, M.A. Charles, E.M.-C.C.S. Group, Association of the Pro12Ala and C1431T variants of PPARgamma and their haplotypes with susceptibility to gestational diabetes, J Clin Endocrinol Metab 96(10) (2011) E1656-60.
[335] S. Azhar, Peroxisome proliferator-activated receptors, metabolic syndrome and cardiovascular disease, Future Cardiol 6(5) (2010) 657-91.
[336] L. Peng, H. Yang, Y. Ye, Z. Ma, C. Kuhn, M. Rahmeh, S. Mahner, A. Makrigiannakis, U. Jeschke, V. von Schonfeldt, Role of Peroxisome Proliferator-Activated Receptors (PPARs) in Trophoblast Functions, Int J Mol Sci 22(1) (2021).
[337] F.P. McCarthy, S. Drewlo, F.A. English, J. Kingdom, E.J. Johns, L.C. Kenny, S.K. Walsh, Evidence implicating peroxisome proliferator-activated receptor-gamma in the pathogenesis of preeclampsia, Hypertension 58(5) (2011) 882-7.
[338] B. Heude, A. Forhan, R. Slama, L. Douhaud, S. Bedel, M.J. Saurel-Cubizolles, R. Hankard, O. Thiebaugeorges, M. De Agostini, I. Annesi-Maesano, M. Kaminski, M.A. Charles, E.m.-c.c.s. group, Cohort Profile: The EDEN mother-child cohort on the prenatal and early postnatal determinants of child health and development, Int J Epidemiol 45(2) (2016) 353-63.
[339] S. van Buuren, K. Groothuis-Oudshoorn, mice: Multivariate Imputation by Chained Equations in R, Journal of Statistical Software; Vol 1, Issue 3 (2011) (2011).
[340] D.J. Stekhoven, P. Buhlmann, MissForest--non-parametric missing value imputation for mixed-type data, Bioinformatics 28(1) (2012) 112-8.
[341] I. Cordón, S. García, A. Fernández, F. Herrera, Imbalance: Oversampling algorithms for imbalanced classification in R, Knowledge-Based Systems 161 (2018) 329-341.
[342] M.B. Kursa, W.R. Rudnicki, Feature Selection with the Boruta Package, Journal of Statistical Software 36(i11) (2010).
[343] S. Lê, J. Josse, F. Husson, FactoMineR: An R Package for Multivariate Analysis, Journal of Statistical Software; Vol 1, Issue 1 (2008) (2008).
[344] S. Rana, E. Lemoine, J.P. Granger, S.A. Karumanchi, Preeclampsia: Pathophysiology, Challenges, and Perspectives, Circ Res 124(7) (2019) 1094-1112.
[345] S. Verlohren, L.A. Droge, The diagnostic value of angiogenic and antiangiogenic factors in differential diagnosis of preeclampsia, Am J Obstet Gynecol (2020).
[346] S. Rana, C.E. Powe, S. Salahuddin, S. Verlohren, F.H. Perschel, R.J. Levine, K.H. Lim, J.B. Wenger, R. Thadhani, S.A. Karumanchi, Angiogenic factors and the risk of adverse outcomes in women with suspected preeclampsia, Circulation 125(7) (2012) 911-9.
[347] R. Romero, J.K. Nien, J. Espinoza, D. Todem, W. Fu, H. Chung, J.P. Kusanovic, F. Gotsch, O. Erez, S. Mazaki-Tovi, R. Gomez, S. Edwin, T. Chaiworapongsa, R.J. Levine, S.A. Karumanchi, A longitudinal study of angiogenic (placental growth factor) and anti-angiogenic (soluble endoglin and soluble vascular endothelial growth factor receptor-1) factors in normal pregnancy and patients destined to develop preeclampsia and deliver a small for gestational age neonate, J Matern Fetal Neonatal Med 21(1) (2008) 9-23.
[348] J. Laasanen, S. Heinonen, M. Hiltunen, A. Mannermaa, M. Laakso,

Polymorphism in the peroxisome proliferator-activated receptor-gamma gene in women with preeclampsia, Early Hum Dev 69(1-2) (2002) 77-82.
[349] S.M. Almeida, J.M. Furtado, P. Mascarenhas, M.E. Ferraz, J.C. Ferreira, M.P. Monteiro, M. Vilanova, F.P. Ferraz, Association between LEPR, FTO, MC4R, and PPARG-2 polymorphisms with obesity traits and metabolic phenotypes in school-aged children, Endocrine 60(3) (2018) 466-478.
[350] J. Jiang, Z. Xie, J. Guo, Y. Wang, C. Liu, S. Zhang, W. Tang, Y. Chen, Association of PPARG rs 1801282 C>G polymorphism with risk of colorectal cancer: from a case-control study to a meta-analysis, Oncotarget 8(59) (2017) 100558-100569.
[351] R.M. Rocha, G.B. Barra, E.C. Rosa, E.C. Garcia, A.A. Amato, M.F. Azevedo, Prevalence of the rs1801282 single nucleotide polymorphism of the PPARG gene in patients with metabolic syndrome, Arch Endocrinol Metab 59(4) (2015) 297-302.
[352] J.S. Ho, S. Germer, C.H. Tam, W.Y. So, M. Martin, R.C. Ma, J.C. Chan, M.C. Ng, Association of the PPARG Pro12Ala polymorphism with type 2 diabetes and incident coronary heart disease in a Hong Kong Chinese population, Diabetes Res Clin Pract 97(3) (2012) 483-91.
[353] S.S. Deeb, L. Fajas, M. Nemoto, J. Pihlajamaki, L. Mykkanen, J. Kuusisto, M. Laakso, W. Fujimoto, J. Auwerx, A Pro12Ala substitution in PPARgamma2 associated with decreased receptor activity, lower body mass index and improved insulin sensitivity, Nat Genet 20(3) (1998) 284-7.
[354] D. Altshuler, J.N. Hirschhorn, M. Klannemark, C.M. Lindgren, M.C. Vohl, J. Nemesh, C.R. Lane, S.F. Schaffner, S. Bolk, C. Brewer, T. Tuomi, D. Gaudet, T.J. Hudson, M. Daly, L. Groop, E.S. Lander, The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes, Nat Genet 26(1) (2000) 76-80.
[355] B.A. Beamer, C.J. Yen, R.E. Andersen, D. Muller, D. Elahi, L.J. Cheskin, R. Andres, J. Roth, A.R. Shuldiner, Association of the Pro12Ala variant in the peroxisome proliferator-activated receptor-gamma2 gene with obesity in two Caucasian populations, Diabetes 47(11) (1998) 1806-8.
[356] M.K. Hoffman, N. Ma, A. Roberts, A machine learning algorithm for predicting maternal readmission for hypertensive disorders of pregnancy, Am J Obstet Gynecol MFM 3(1) (2021) 100250.
[357] H. Sufriyana, Y.W. Wu, E.C. Su, Artificial intelligence-assisted prediction of preeclampsia: Development and external validation of a nationwide health insurance dataset of the BPJS Kesehatan in Indonesia, EBioMedicine 54 (2020) 102710.
[358] L.M. Bodnar, A.R. Cartus, S.I. Kirkpatrick, K.P. Himes, E.H. Kennedy, H.N. Simhan, W.A. Grobman, J.Y. Duffy, R.M. Silver, S. Parry, A.I. Naimi, Machine learning as a strategy to account for dietary synergy: an illustration based on dietary intake and adverse pregnancy outcomes, Am J Clin Nutr 111(6) (2020) 1235-1243.
[359] J.H. Jhee, S. Lee, Y. Park, S.E. Lee, Y.A. Kim, S.W. Kang, J.Y. Kwon, J.T. Park, Prediction model development of late-onset preeclampsia using machine learning-based methods, PLoS One 14(8) (2019) e0221202.
[360] ACOG Practice Bulletin No. 202 Summary: Gestational Hypertension and Preeclampsia, Obstet Gynecol 133(1) (2019) 1.

[361] Committee Opinion No. 638: First-Trimester Risk Assessment for Early-Onset Preeclampsia, Obstet Gynecol 126(3) (2015) e25-e27.
[362] A.R. Brasier, S. Victor, H. Ju, W.W. Busse, D. Curran-Everett, E. Bleecker, M. Castro, K.F. Chung, B. Gaston, E. Israel, S.E. Wenzel, S.C. Erzurum, N.N. Jarjour, W.J. Calhoun, Predicting intermediate phenotypes in asthma using bronchoalveolar lavage-derived cytokines, Clin Transl Sci 3(4) (2010) 147-57.
[363] S. Dreiseitl, L. Ohno-Machado, Logistic regression and artificial neural network classification models: a methodology review, J Biomed Inform 35(5-6) (2002) 352-9.
[364] S. Kuhle, B. Maguire, H. Zhang, D. Hamilton, A.C. Allen, K.S. Joseph, V.M. Allen, Comparison of logistic regression with machine learning methods for the prediction of fetal growth abnormalities: a retrospective cohort study, BMC Pregnancy Childbirth 18(1) (2018) 333.
[365] A. Luque, A. Carrasco, A. Martín, A. de las Heras, The impact of class imbalance in classification performance metrics based on the binary confusion matrix, Pattern Recognition 91 (2019) 216-231.
[366] P. Banerjee, F.O. Dehnbostel, R. Preissner, Prediction Is a Balancing Act: Importance of Sampling Methods to Balance Sensitivity and Specificity of Predictive Models Based on Imbalanced Chemical Data Sets, Frontiers in Chemistry 6(362) (2018).
[367] N. O'Gorman, D. Wright, L.C. Poon, D.L. Rolnik, A. Syngelaki, M. de Alvarado, I.F. Carbone, V. Dutemeyer, M. Fiolna, A. Frick, N. Karagiotis, S. Mastrodima, C. de Paco Matallana, G. Papaioannou, A. Pazos, W. Plasencia, K.H. Nicolaides, Multicenter screening for pre-eclampsia by maternal factors and biomarkers at 11-13 weeks' gestation: comparison with NICE guidelines and ACOG recommendations, Ultrasound Obstet Gynecol 49(6) (2017) 756-760.
[368] S. Rana, S. Salahuddin, A. Mueller, A.H. Berg, R.I. Thadhani, S.A. Karumanchi, Angiogenic biomarkers in triage and risk for preeclampsia with severe features, Pregnancy Hypertens 13 (2018) 100-106.
[369] A.O. Odibo, K.R. Goetzinger, L. Odibo, A.G. Cahill, G.A. Macones, D.M. Nelson, D.J. Dietzen, First-trimester prediction of preeclampsia using metabolomic biomarkers: a discovery phase study, Prenat Diagn 31(10) (2011) 990-4.
[370] R.O. Bahado-Singh, A. Syngelaki, R. Akolekar, R. Mandal, T.C. Bjondahl, B. Han, E. Dong, S. Bauer, Z. Alpay-Savasan, S. Graham, O. Turkoglu, D.S. Wishart, K.H. Nicolaides, Validation of metabolomic models for prediction of early-onset preeclampsia, Am J Obstet Gynecol 213(4) (2015) 530 e1-530 e10.
[371] S. Ding, L. Liu, Q.C. Zhuge, Z. Yu, X. Zhang, J. Xie, W. Hong, S. Wang, Y. Yang, B. Chen, The meta-analysis of the association of PPARG P12A, C161T polymorphism and coronary heart disease, Wien Klin Wochenschr 124(19-20) (2012) 671-7.
[372] J. Lee, D.Y. Hyeon, D. Hwang, Single-cell multiomics: technologies and data analysis methods, Exp Mol Med 52(9) (2020) 1428-1442.

Annexes

Part I: Supplementary materials

Table S1_unsigned. The number of genes in the color modules in dataset1-3.

Table 7 Table S1_unsigned. The number of genes in the color modules in dataset1-3.

Module colors	Counts		
	Dataset1	Dataset2	Dataset3
black	290	132	89
blue	2491	290	190
brown	1320	258	112
green	421	150	99
grey	88	1122	400
pink	208	115	56
red	40	135	91
turquoise	4437	302	395
yellow	469	241	102
magenta	103	106	-

Table S1_signed. The number of genes in the color modules in dataset1-3.

Table 8 Table S1_signed. The number of genes in the color modules in dataset1-3.

Module colors	Counts		
	Dataset1	Dataset2	Dataset3
black	195	156	110
blue	2871	294	209
brown	679	286	149
green	429	216	141
grey	102	1282	489
pink	123	127	58
red	232	93	111
turquoise	3954	295	377
yellow	551	230	146
magenta	140	79	-

Table S2 Completed results of biological process and pathway analysis for differentially expressed genes with four sheets inside.

Table 9 Table S2 Completed results of biological process and pathway analysis for differentially expressed genes with four sheets inside.

ID	Description	GeneRati	BgRatio	pvalue	p.adjust	qvalue	geneID	Count
hsa04072	Phospholipase D signaling patl	23/576	148/7915	0.000401	0.089813	0.083293	1956/5332	23
hsa04115	p53 signaling pathway	14/576	72/7915	0.000578	0.089813	0.083293	896/1019/	14
hsa04014	Ras signaling pathway	30/576	232/7915	0.001397	0.120828	0.112055	801/1956/	30
hsa04810	Regulation of actin cytoskeletc	28/576	213/7915	0.001554	0.120828	0.112055	23365/195	28
hsa04925	Aldosterone synthesis and sec	15/576	98/7915	0.004557	0.227915	0.211368	801/3949/	15
hsa00020	Citrate cycle (TCA cycle)	7/576	30/7915	0.004853	0.227915	0.211368	5160/3417	7
hsa04015	Rap1 signaling pathway	26/576	210/7915	0.00513	0.227915	0.211368	801/1956/	26
hsa00562	Inositol phosphate metabolisn	12/576	74/7915	0.006705	0.234797	0.21775	3635/3707	12
hsa05212	Pancreatic cancer	12/576	75/7915	0.007472	0.234797	0.21775	1956/1019	12
hsa01210	2-Oxocarboxylic acid metaboli	5/576	18/7915	0.007751	0.234797	0.21775	3417/48/1	5
hsa03320	PPAR signaling pathway	12/576	76/7915	0.008305	0.234797	0.21775	6319/2171	12
hsa04510	Focal adhesion	24/576	199/7915	0.009626	0.249486	0.231373	3371/896/	24
hsa04070	Phosphatidylinositol signaling	14/576	99/7915	0.012009	0.287282	0.266425	801/3635/	14
hsa04024	cAMP signaling pathway	25/576	216/7915	0.013689	0.304096	0.282018	801/1080/	25
hsa04151	PI3K-Akt signaling pathway	37/576	354/7915	0.015538	0.322147	0.298759	10971/337	37
hsa04110	Cell cycle	16/576	124/7915	0.017462	0.328844	0.304969	10971/896	16
hsa04022	cGMP-PKG signaling pathway	20/576	167/7915	0.018413	0.328844	0.304969	801/292/2	20
hsa05214	Glioma	11/576	75/7915	0.019033	0.328844	0.304969	801/1956/	11
hsa05210	Colorectal cancer	12/576	86/7915	0.021184	0.346751	0.321576	1956/332/	12
hsa04614	Renin-angiotensin system	5/576	23/7915	0.022527	0.350298	0.324866	290/5550/	5
hsa04010	MAPK signaling pathway	31/576	295/7915	0.023745	0.351659	0.326128	3310/1956	31
hsa04924	Renin secretion	10/576	69/7915	0.026662	0.36291	0.336562	801/5332/	10
hsa05219	Bladder cancer	7/576	41/7915	0.026839	0.36291	0.336562	1956/1019	7
hsa05167	Kaposi sarcoma-associated he	21/576	186/7915	0.028848	0.373818	0.346679	801/7538/	21
hsa04611	Platelet activation	15/576	124/7915	0.035058	0.386954	0.358861	23365/128	15
hsa04912	GnRH signaling pathway	12/576	93/7915	0.036507	0.386954	0.358861	801/1956/	12
hsa00071	Fatty acid degradation	7/576	44/7915	0.03796	0.386954	0.358861	3033/217/	7

... GO Terms for dataset2&3 | KEGG Pathways for dataset1 | KEGG Pathways for dataset2 ...

Table S3 Characteristics of placenta used for logistic regression.

Table 10 Table S3 Characteristics of placenta used for logistic regression.

Sample NO.	Sex	Smoking	Gestational Age (weeks)	Evaluation 1	Evaluation 2
1	Male	-	8	+++	++
2	Female	+	8	+++	++
3	Female	+	12	++	+
4	Male	+	8	+++	+++
5	Male	+	9	+++	+++
6	Female	-	8	+++	++++
7	Female	+	12	++	++
8	Male	-	8	++++	++++
9	Male	-	7	++++	++++
10	Male	+	8	+++	++
11	Female	-	12	++	+++
12	Female	-	9	+++	++
13	Male	+	11	+	++
14	Male	+	12	++	+++
15	Female	-	12	++	++
16	Male	-	12	+	+++
17	Male	+	12	++	++
18	Male	-	13	+	+

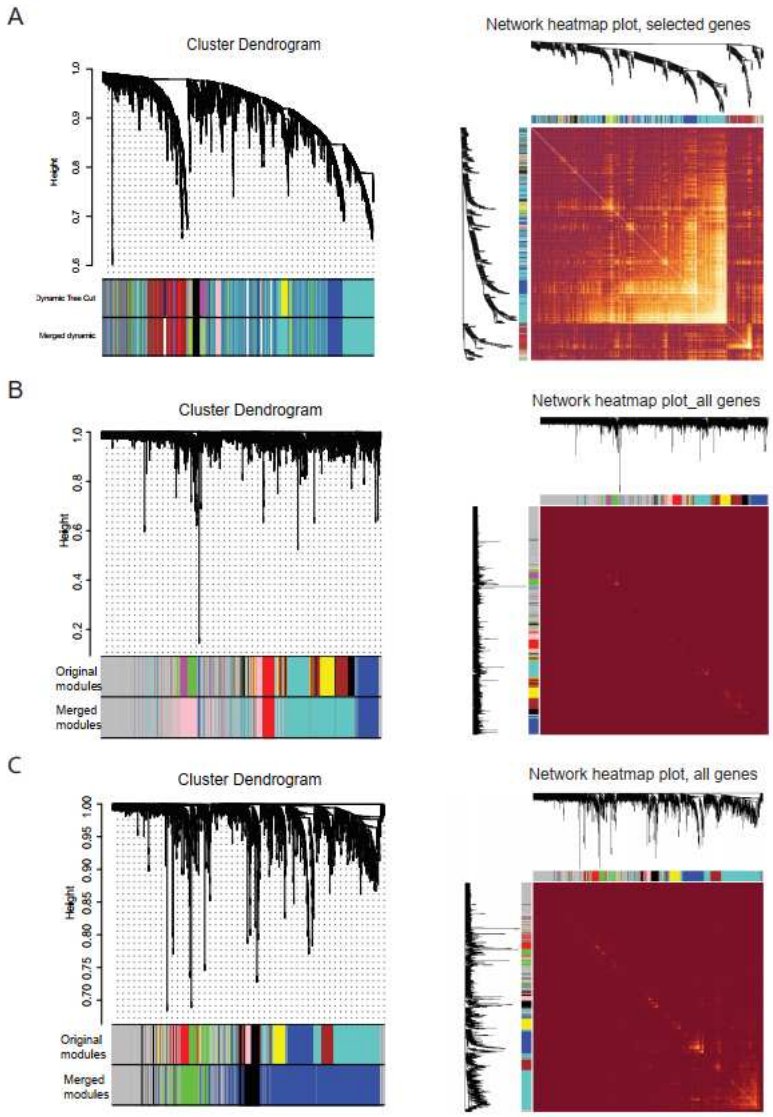


Figure 40 Fig. S1. Construction and visualization of co-expression modules and gene adjacency matrix.

Fig. S1. Construction and visualization of co-expression modules and gene adjacency matrix. Clustering dendrograms of genes were performed according to the dynamic tree cutting method to form original modules which were further used to construct merged modules based on the similarity for dataset1-3, which were shown in the left panels of graph A&B&C, respectively. The heatmaps in the right panels subsequently depict the Topological Overlap Matrix (TOM) of selected genes in dataset1 (A) and all genes in dataset2 (B) and datset3 (C). Light color in heatmap means high overlap value and darker red color represents low overlap value.

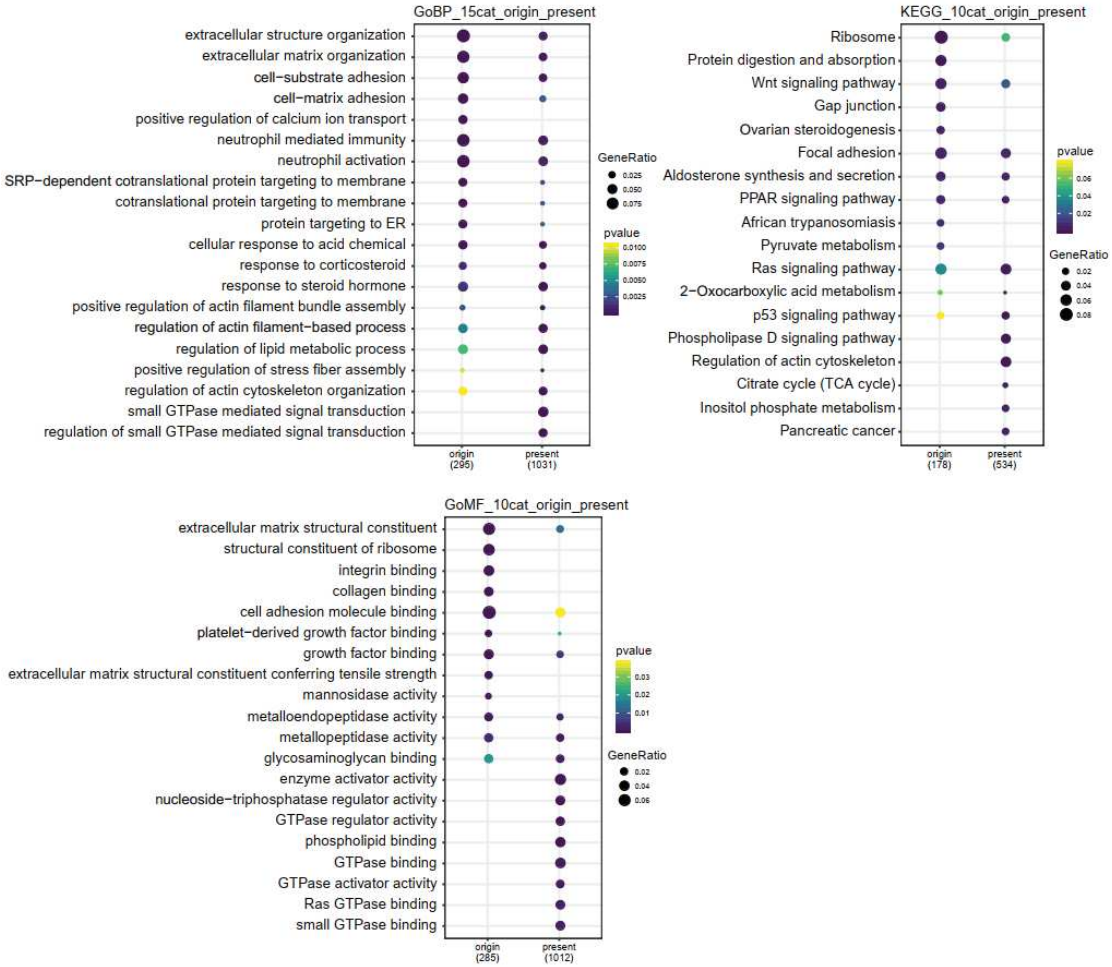


Figure 41 Fig. S2. Comparison of the top terms of biological process, molecular function and pathway between the original research and our present study.

Fig. S2. Comparison of the top terms of biological process, molecular function and pathway between the original research and our present study. GoBP: gene ontology biological process; GoMF: gene ontology molecular function; KEGG: kyoto encyclopedia genes and genome pathways.

Part II: Supplementary materials

Table Ss. An exhaustive list of all terms.

Tableau 1 Table Ss. An exhaustive list of all terms.

Table 11 Table Ss. An exhaustive list of all terms.

Name	Date modified	Type	Size
Table S1_EVCT_DEGs_Go_Biological Process	21-Aug-19 13:53	TSV File	48 KB
Table S1_EVCT_DEGs_Go_Biological Process	20-Feb-20 08:16	Microsoft Excel 工作表	22 KB
Table S2_EVCT_DEGs_Go_Cellular Component	21-Aug-19 13:53	TSV File	17 KB
Table S2_EVCT_DEGs_Go_Cellular Component	20-Feb-20 08:17	Microsoft Excel 工作表	14 KB
Table S3_EVCT_DEGs_Go_Molecular Function	21-Aug-19 13:53	TSV File	5 KB
Table S3_EVCT_DEGs_Go_Molecular Function	20-Feb-20 08:17	Microsoft Excel 工作表	12 KB
Table S4_EVCT_DEGs_Pathways	22-Aug-19 14:55	TSV File	3 KB
Table S4_EVCT_DEGs_Pathways	20-Feb-20 08:18	Microsoft Excel 工作表	12 KB
Table S5_VCT_DEGs_Go_Biological Process	19-Aug-19 14:34	TSV File	85 KB
Table S5_VCT_DEGs_Go_Biological Process	20-Feb-20 08:18	Microsoft Excel 工作表	24 KB
Table S6_VCT_DEGs_Go_Cellular Component	19-Aug-19 14:35	TSV File	32 KB
Table S6_VCT_DEGs_Go_Cellular Component	20-Feb-20 08:18	Microsoft Excel 工作表	15 KB
Table S7_VCT_DEGs_Go_Molecular Function	19-Aug-19 14:35	TSV File	4 KB
Table S7_VCT_DEGs_Go_Molecular Function	20-Feb-20 08:18	Microsoft Excel 工作表	12 KB
Table S8_VCT_DEGs_Pathways	19-Aug-19 14:35	TSV File	17 KB
Table S8_VCT_DEGs_Pathways	20-Feb-20 08:19	Microsoft Excel 工作表	15 KB

Part III: Supplementary materials

Table S1. Datasets for extracting hypoxia-inducible factors (HIF) targets.

Table 12 Table S1. Datasets for extracting hypoxia-inducible factors (HIF) targets.

Dataset	Gene Identifier	Number of TF	Number of TF Targets	Reference
TRED	ENTREZ	133	7066	Jiang et al.
ITFP	HGNC Symbol/Alias	1974	67154	Zheng et al.
ENCODE	ENTREZ	157	20428	Encode
Neph2012	HGNC Symbol/Alias	536	16484	Neph et al.
TRRUST	HGNC Symbol/Alias	748	8215	Han et al.
Marbach2016	HGNC Symbol/Alias	643	1305782	Marbach et al.

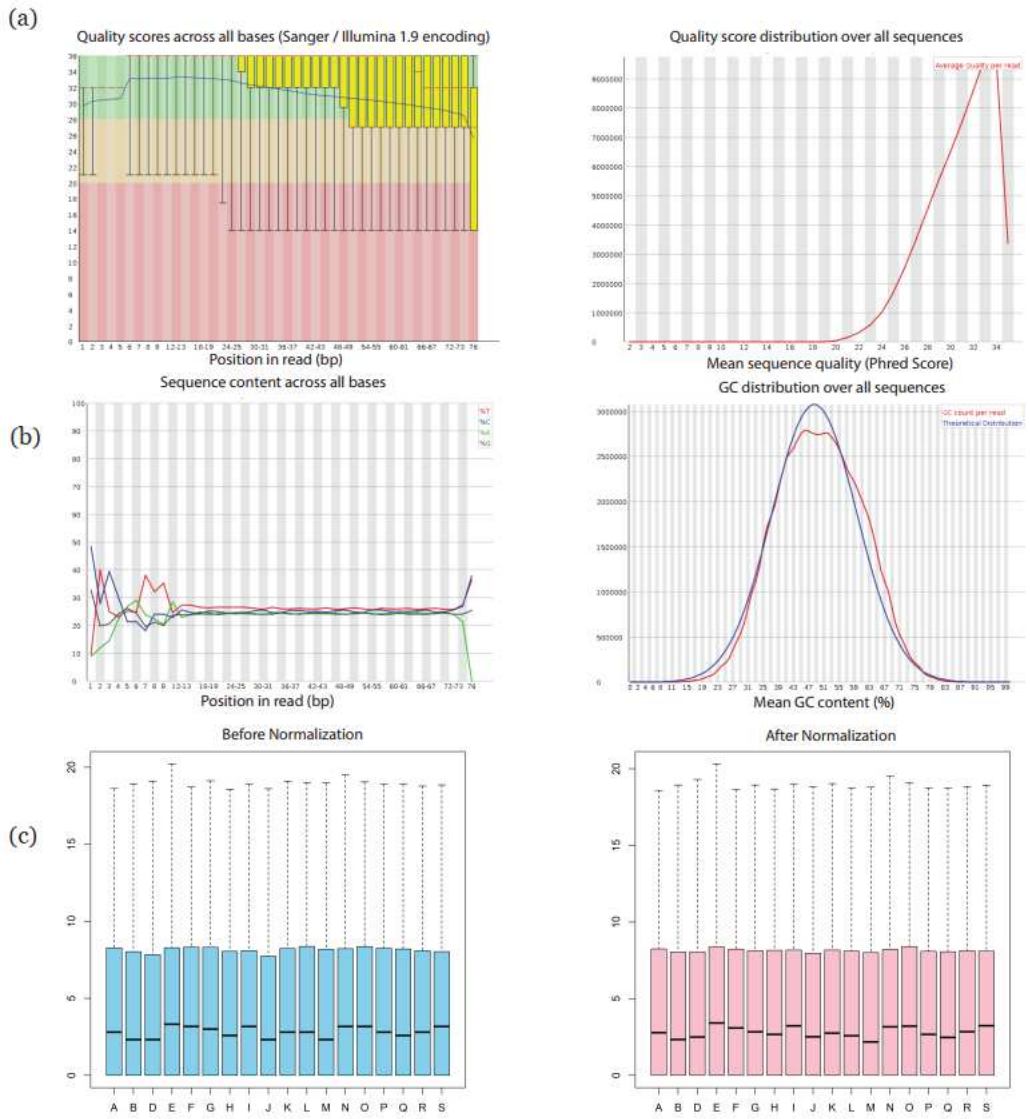


Figure 42 Fig. S1. Quality report of raw data

Fig. S1. Quality report of raw data. (a) Average quality of sequencing for each position within the reads and the distribution according to their average quality. (b) Representation of the average proportion of each base for all reads and distribution of readings according to their composition in GC. (c) Normalization of raw data: before and after.

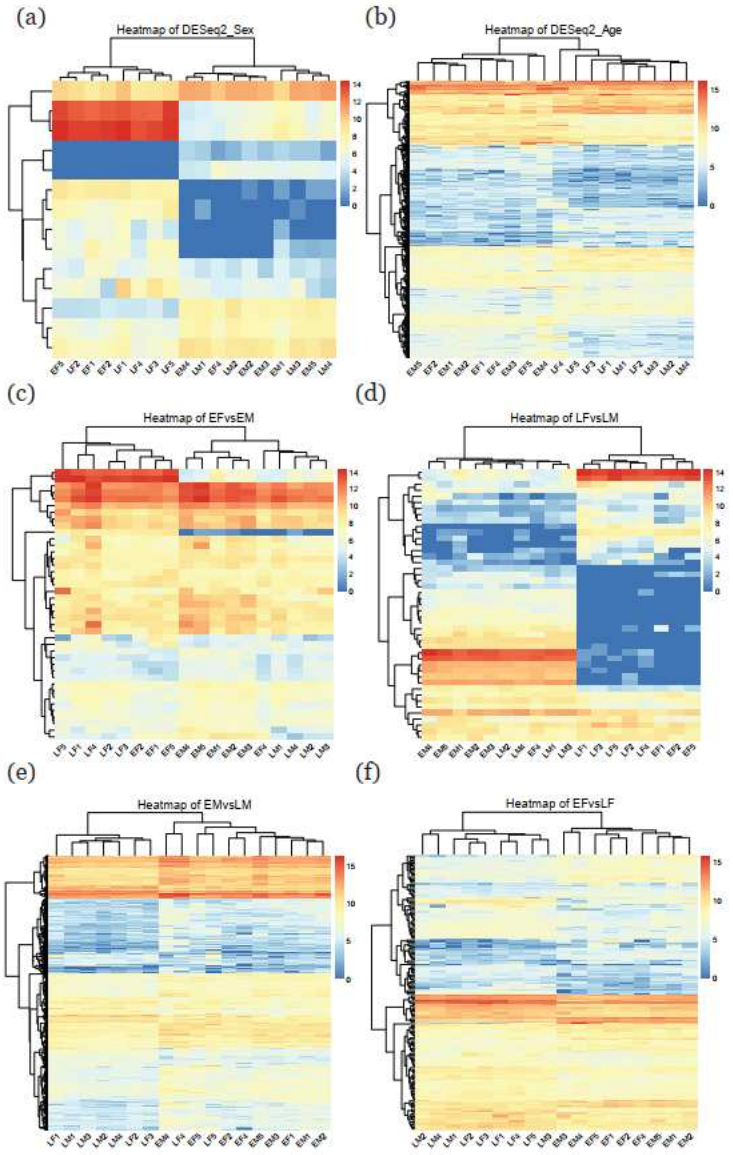


Figure 43 Fig. S2. The heatmaps of gene expression for the key genes identified using DESeq2

Fig. S2. The heatmaps of gene expression for the key genes identified using DESeq2 under the comparison of different groups, including Female vesus Male (a), Early versus Late (b), EF versus EM (c), LF versus LM (d), EM versus LM (e), and EF versus LF (f). Alphabet "E" represents the abbreviation of "Early", "L" for "Late", "F" for "Female" and "M" for "Male" in groups.

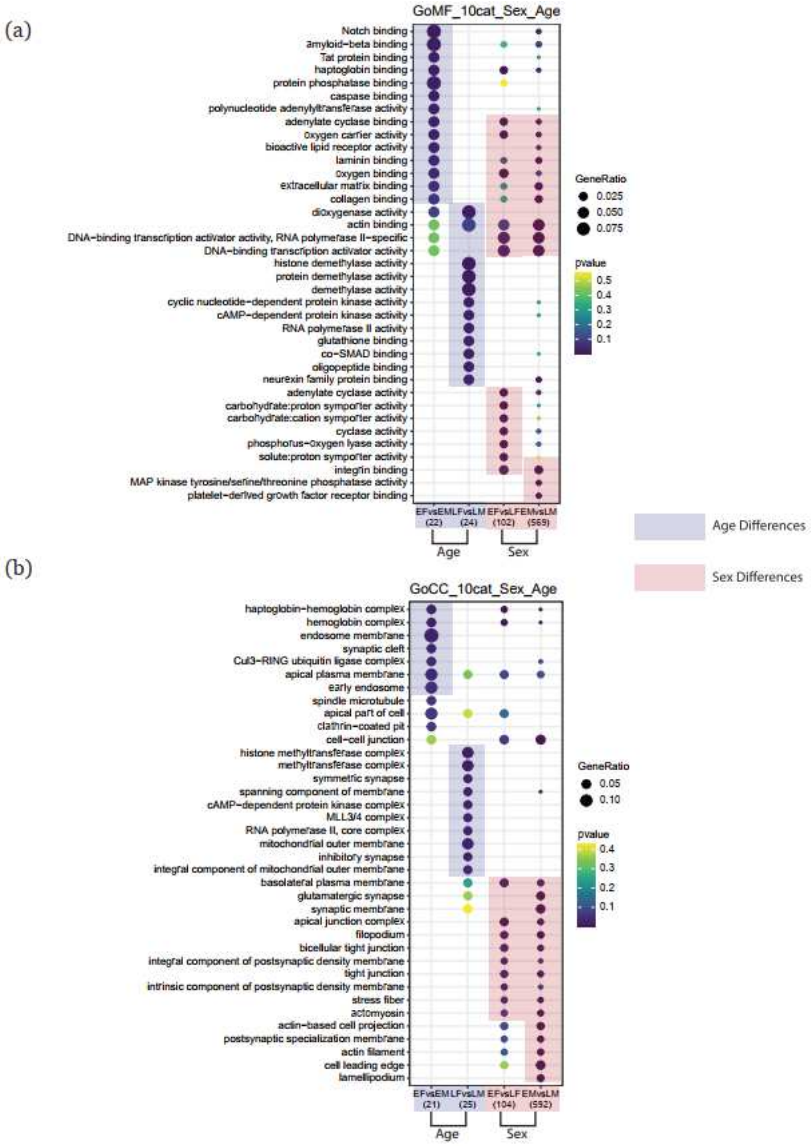


Figure 44 Fig. S3. The enriched terms of GO cellular component (a) and GO molecular function (b) for the DESeq2 method's comparison

Fig. S3. The enriched terms of GO cellular component (a) and GO molecular function (b) for the DESeq2 method's comparison. Weighted points filled with plum represent the age difference and coral represent the sexual difference. Alphabet "E" represents the abbreviation of "Early", "L" for "Late", "F" for "Female" and "M" for "Male" in groups. GO: gene ontology; KEGG: Kyoto encyclopedia of genes and genomes.

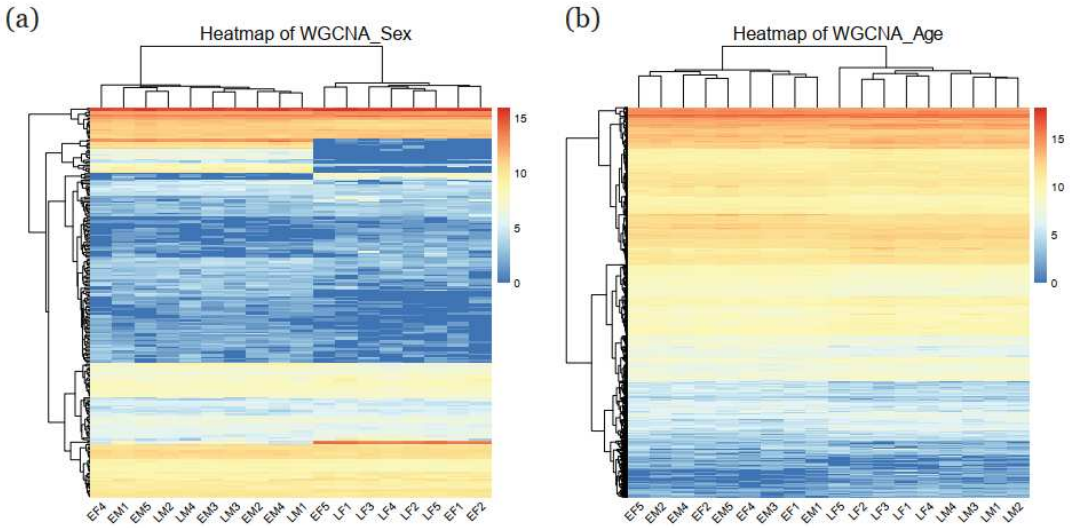


Figure 45 Fig. S4. The expressions of key genes in the modules are shown in heatmap separately

Fig. S4. The expressions of key genes in the modules are shown in heatmap separately using WGCNA based on the traits of age (a) and sexes (b). WGCNA: weighted gene co-expression network analysis.

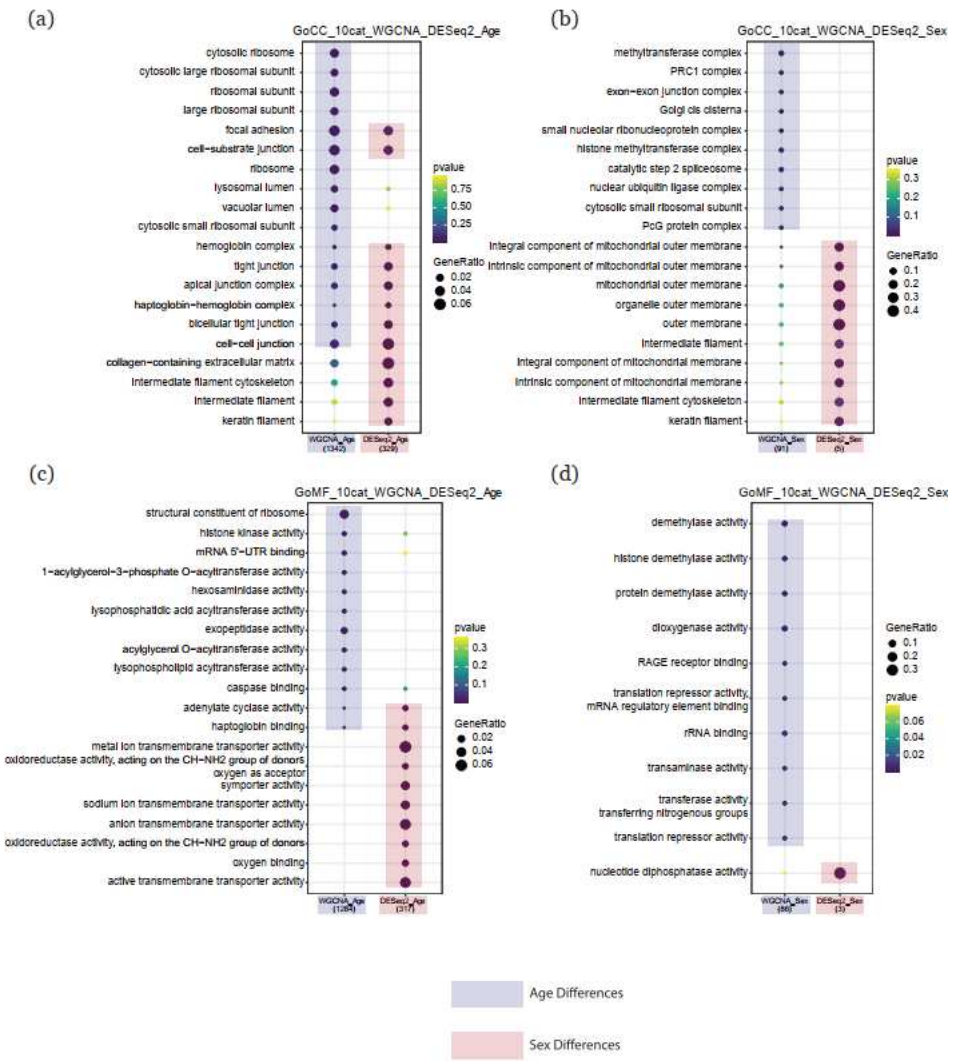


Figure 46 Fig. S5. Comparison of enriched GO cellular component (CC) and GO molecular function (MF)

Fig. S5. Comparison of enriched GO cellular component (CC) and GO molecular function (MF) based on genes selected by DESeq2 and WGCNA methods separately. (a&c) Enrichment and comparison of terms under age difference. (b&d) Enrichment and comparison of terms nder sexual difference. Weighted points filled with plum represent the age difference and coral represent the sexual difference. GO: gene ontology; WGCNA: weighted gene co-expression network analysis.

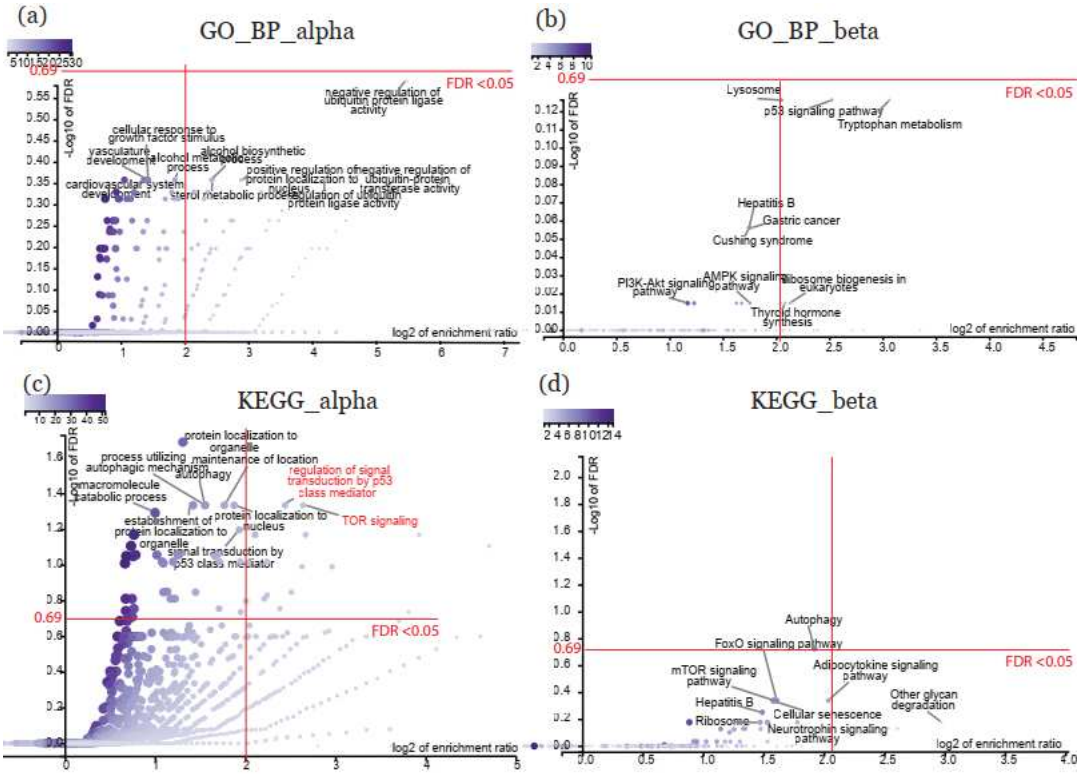


Figure 47 Fig. S6. Details of evaluation for the terms of GO and KEGG pathways for the HIF targets

Fig. S6. Details of evaluation for the terms of GO and KEGG pathways for the HIF targets. (a) Scatterplot of GO_BP_alpha. (b) Scatterplot of GO_BP_beta. (c) Scatterplot of KEGG_alpha. (d) Scatterplot of KEGG_beta. Y coordinate represents -log10 of FDR and x coordinate represents log2 enrichment ratio. GO: gene ontology; BP: biological process; HIF: hypoxia-inducible factor; KEGG: Kyoto encyclopedia of genes and genomes.

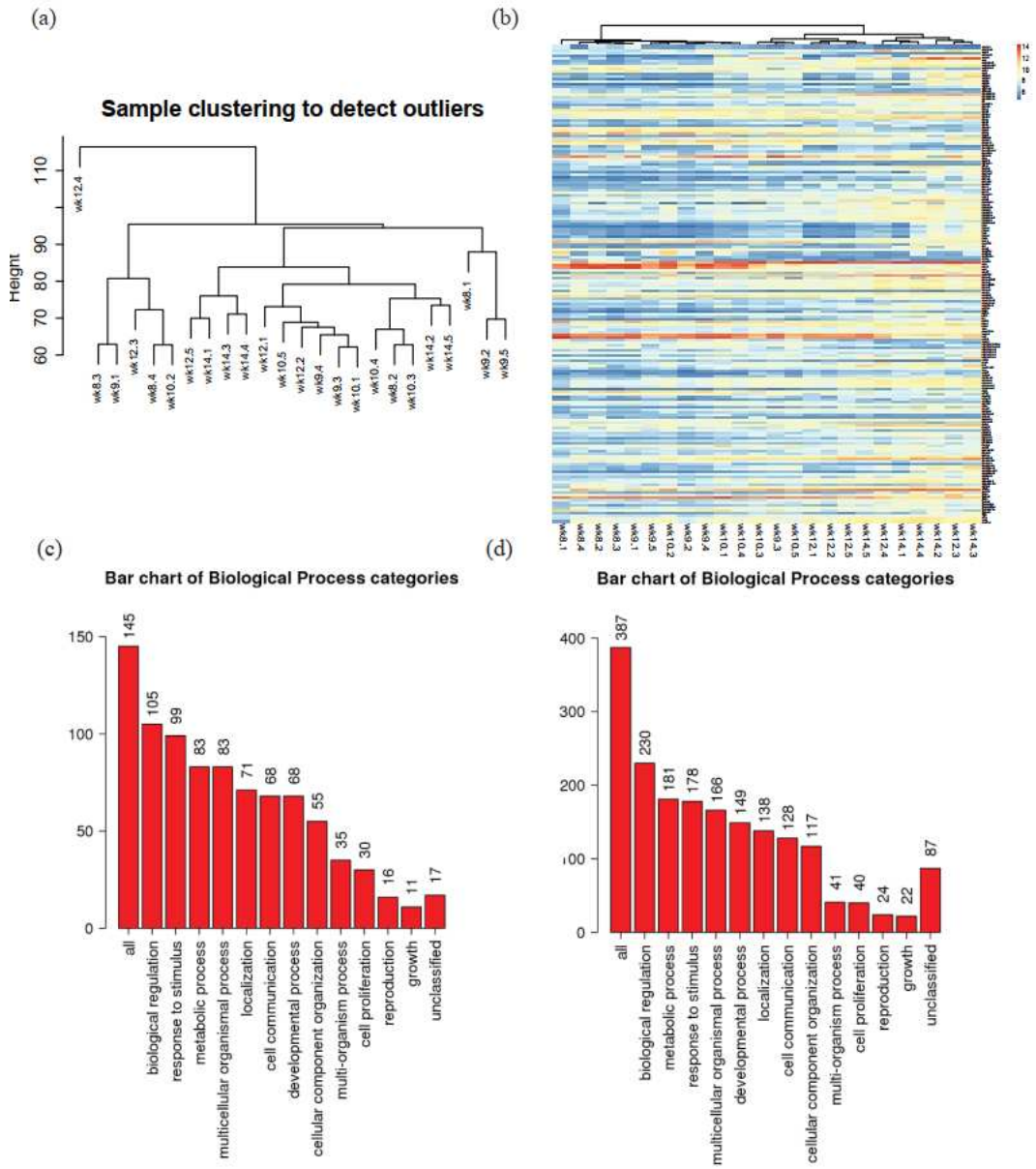


Figure 48 Fig. S7. Summary of dataset GSE100051

Fig. S7. Summary of dataset GSE100051. The clustering of the collected samples from GSE100051 (a). The expressions of the selected key genes. (b) The comparison of the terms from GSE100051 and our results using the DESeq2 analysis (c&d).

Part IV: Supplementary materials

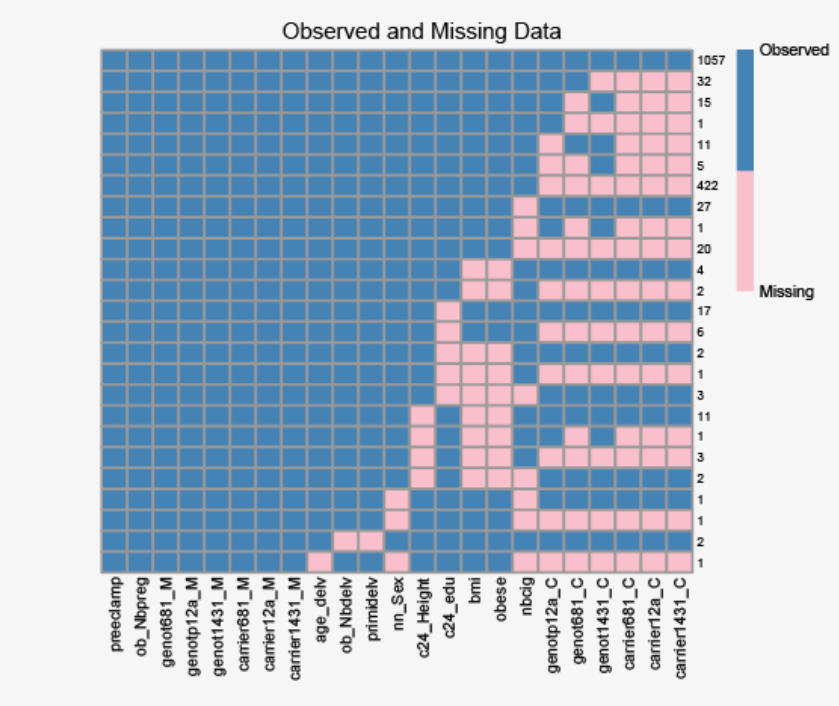


Figure 49 Fig. S1. Overview of missing data in original data

Fig. S1. Overview of missing data in original data. Columns represent clinical characteristics. Rows indicate the missing characteristics and the corresponding number of individuals. Pink grids represent missing values while blue grids represent observed values.

Table S1. Imputation comparison for factor type features before and after. *continued*

Table 13 Table S1. Imputation comparison for factor type features before and after.

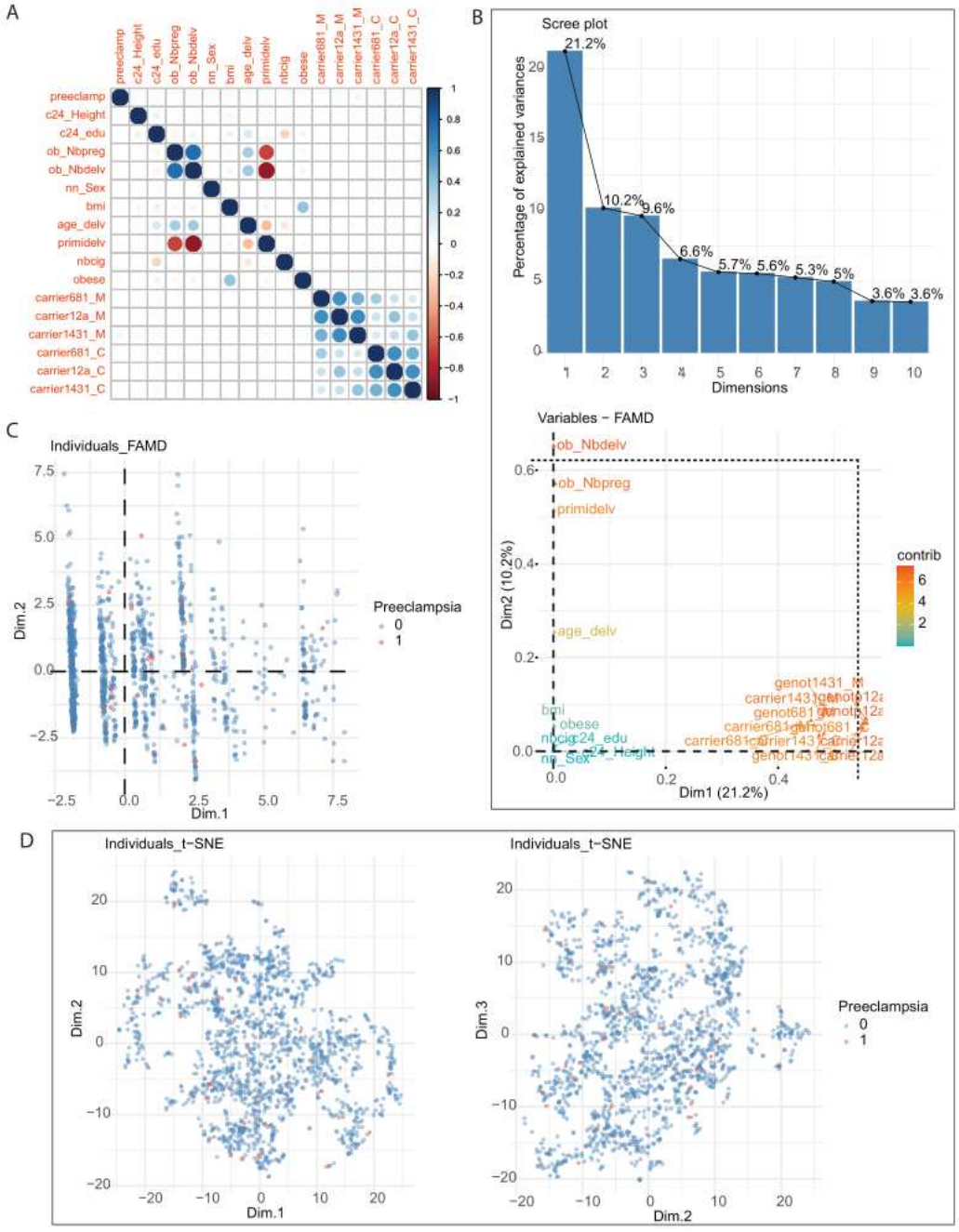
Features	Number of missing values	Imputation	
		Before	After
		Count of factors	Count of factors
preeclamp	0	0: 1613, 1: 35	0: 1613, 1: 35
nn_Sex	3	1: 878, 2: 767	1: 879, 2: 769
primidelv	2	0: 922, 1: 724	0: 924, 1: 724
obese	29	0: 1476, 1: 143	0: 1505, 1: 143

genot681_C	479	11: 698, 12: 419, 22: 52	11: 1175, 12: 420, 22: 53
genot681_M	0	11: 997, 12: 577, 22: 74	11: 997, 12: 577, 22: 74
genotp12a_C	472	11: 946, 12: 219, 22: 11	11: 1418, 12: 219, 22: 11
genotp12a_M	0	11: 1323, 12: 309, 22: 16	11: 1323, 12: 309, 22: 16
genot1431_C	489	11: 880, 12: 264, 22: 15	11: 1369, 12: 264, 22: 15
genot1431_M	0	11: 1292, 12: 340, 22: 16	11: 1292, 12: 340, 22: 16
carrier681_M	0	0: 997, 1: 651	0: 997, 1: 651
carrier12a_M	0	0: 1323, 1: 325	0: 1323, 1: 325
carrier1431_M	0	0: 1292, 1: 356	0: 1292, 1: 356
carrier681_C	522	0: 666, 1: 460	0: 1175, 1: 473
carrier12a_C	522	0: 906, 1: 220	0: 1418, 1: 230
carrier1431_C	522	0: 854, 1: 272	0: 1369, 1: 279

Table S1. Imputation comparison for numeric type features before and after.

Features	Number of missing values	Imputation			
		Before		After	
		mean	sd	mean	sd
c24_Height	17	163.46	6.23	163.46	6.20
c24_edu	29	6.54	2.48	6.53	2.47

ob_Nbpreg	0	1.35	1.49	1.35	1.49
ob_Nbdelv	2	0.83	0.97	0.83	0.97
bmi	29	23.27	4.63	23.26	4.60
age_delv	1	29.55	4.86	29.55	4.86
nbcig	56	1.47	3.44	1.48	3.39



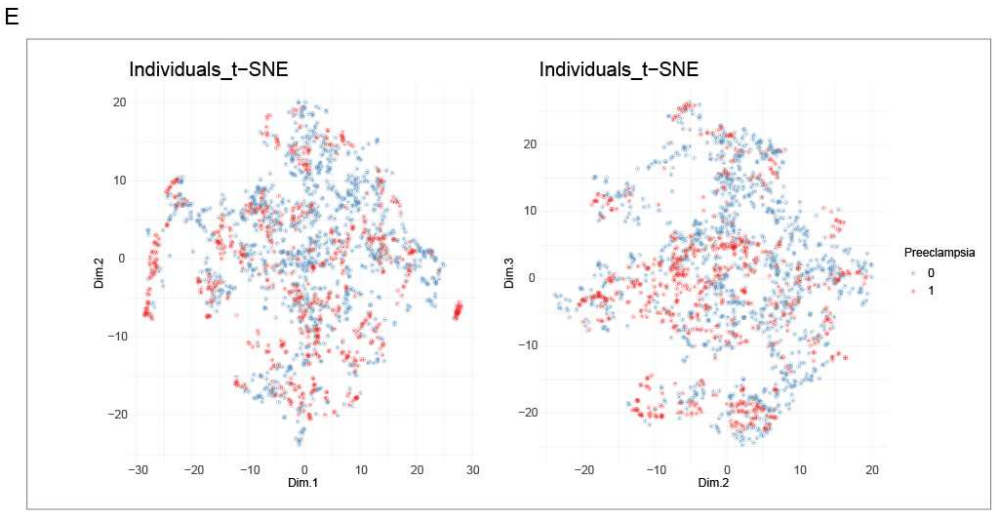


Figure 50 Fig. S2. Clinical characteristics evaluation

Fig. S2. Clinical characteristics evaluation. (A) The matrix of correlation coefficient for all the characteristics. The factorial characteristics were transformed to numerical type, accompanied by adding 0.1 to all the data in order to facilitate the calculation without 0. (B) A principal component analysis with factor analysis of mixed data (FAMD) for all the characteristics with detailed component proportions shown in the right. The first and second dimensions were chosen to draw the graph with colors presenting the contribution. (C) Distribution of individuals by FAMD. (D) Distribution of individuals by t-SNE in the dimensions before balancing. (E) Distribution of individuals by t-SNE in the dimensions after balancing.

Table S2. Clinical characteristics for different data set related to factor features. *continued*

Table 144 Table S2. Clinical characteristics for different data set related to factor features.

Factor features	Number of factors	Total	Train	Test
		Count of factors	Count of factors	Count of factors
nn_Sex	2	1: 879, 2: 769	1: 660, 2: 576	1: 220, 2: 192

primidelv	2	0: 924, 1: 724	0: 713, 1: 523	0: 211, 1: 201
preeclamp	2	0: 1613, 1: 35	0: 1211, 1: 25	0: 402, 1: 10
genot681_C	3	11: 1175, 12: 420, 22: 53	11: 875, 12: 324, 22: 37	11: 300, 12: 97, 22: 15
genot681_M	3	11: 997, 12: 577, 22: 74	11: 753, 12: 420, 22: 63	11: 244, 12: 157, 22: 11
genot1431_C	3	11: 1369, 12: 264, 22: 15	11: 1027, 12: 196, 22: 13	11: 342, 12: 68, 22: 2
genot1431_M	3	11: 1292, 12: 340, 22: 16	11: 973, 12: 250, 22: 13	11: 319, 12: 90, 22: 3
obese	2	0: 1505, 1: 143	0: 1120, 1: 116	0: 384, 1: 28

Table S2. Clinical characteristics for different data set related to numeric features.

Numeric features	Total		Train		Test	
	mean	sd	mean	sd	mean	sd
c24_edu	6.53	2.47	6.50	2.48	6.59	2.43
ob_Nbpreg	1.35	1.49	1.42	1.54	1.15	1.32
bmi	23.26	4.60	23.40	4.67	22.93	4.43
nbcig	1.48	3.39	1.57	3.49	1.24	3.07

Table S3. Prediction of the 8 Models by ML Analysis (without balancing)

Table 155 Table S3. Prediction of the 8 Models by ML Analysis (without balancing)

	Train	Test

	Accuracy	AUC	Accuracy	AUC
Elastic Net Regression	0.98	0.70	0.98	0.76
Random Forest	0.99	0.84	0.98	0.73
Support Vector Machine	0.99	0.89	0.98	0.66
Decision Tree	0.98	0.82	0.97	0.54
K-Nearest Neighbor	0.97	0.66	0.95	0.51
Naïve Bayes	0.99	0.74	0.97	0.63
Boost Tree	**0.99**	**0.92**	**0.98**	**0.77**
Multilayer Perceptron	0.98	0.82	0.95	0.55

AUC, area under the receiver operating characteristic curve

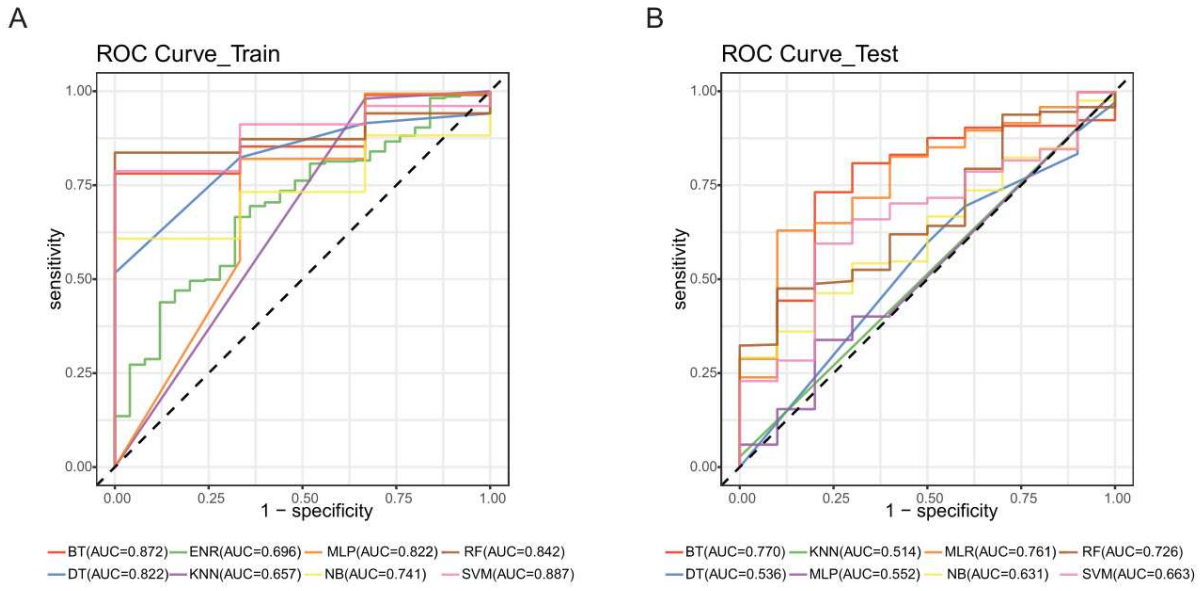


Figure 51 Fig. S3. ROC curve of different algorithms

Fig. S3. ROC curve of different algorithms. (A) ROC curves with training set. (B) ROC curves with testing set. The values were shown in the legends. AUC: area under the receiver operating characteristic curve; BT: boost tree; DT: decision tree; ENR: elastic net regression; KNN: k-nearest neighbor; MLP: multilayer perceptron; NB: naïve bayes; RF: random forest; SVM: support vector machine; ROC: receiver operating characteristic.

Table S4. Prediction of the 8 Models by ML Analysis (with total balancing before split)

Table 166 Table S4. Prediction of the 8 Models by ML Analysis (with total balancing before split)

	Train		Test	
	Accuracy	AUC	Accuracy	AUC
Elastic Net Regression	0.721	0.761	0.705	0.735
Random Forest	0.911	0.966	0.913	0.973
Support Vector Machine	0.783	0.864	0.743	0.807
Decision Tree	0.863	0.926	0.819	0.892
K-Nearest Neighbor	0.830	0.901	0.772	0.876
Naïve Bayes	0.687	0.770	0.702	0.787
Boost Tree	**0.957**	**0.990**	**0.975**	**0.996**
Multilayer Perceptron	0.863	0.926	0.735	0.800

AUC, area under the receiver operating characteristic curve

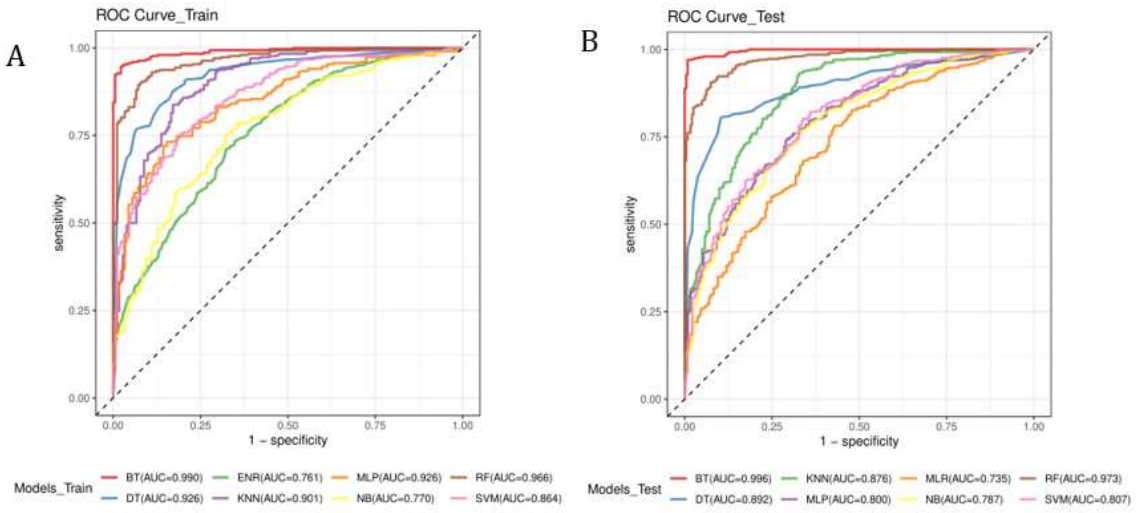


Figure 52 Fig. S4. ROC curve of different algorithms

Fig. S4. ROC curve of different algorithms. (A) ROC curves with training set. (B) ROC curves with testing set. The values were shown in the legends. AUC: area under the receiver operating

characteristic curve; BT: boost tree; DT: decision tree; ENR: elastic net regression; KNN: k-nearest neighbor; MLP: multilayer perceptron; NB: naïve bayes; RF: random forest; SVM: support vector machine; ROC: receiver operating characteristic.

Table S5. Prediction of the 8 Models by ML Analysis

Table 177 Table S5. Prediction of the 8 Models by ML Analysis

	Train		Test	
	Accuracy	AUC	Accuracy	AUC
Elastic Net Regression	0.658	0.707	0.920	0.780
Random Forest	0.948	0.979	0.896	0.677
Support Vector Machine	0.818	0.878	0.808	0.593
Decision Tree	0.899	0.946	0.871	0.545
K-Nearest Neighbor	0.818	0.924	0.796	0.681
Naïve Bayes	0.704	0.846	0.947	0.632
Boost Tree	**0.971**	**0.995**	**0.954**	**0.700**
Multilayer Perceptron	0.899	0.946	0.782	0.521

AUC, area under the receiver operating characteristic curve

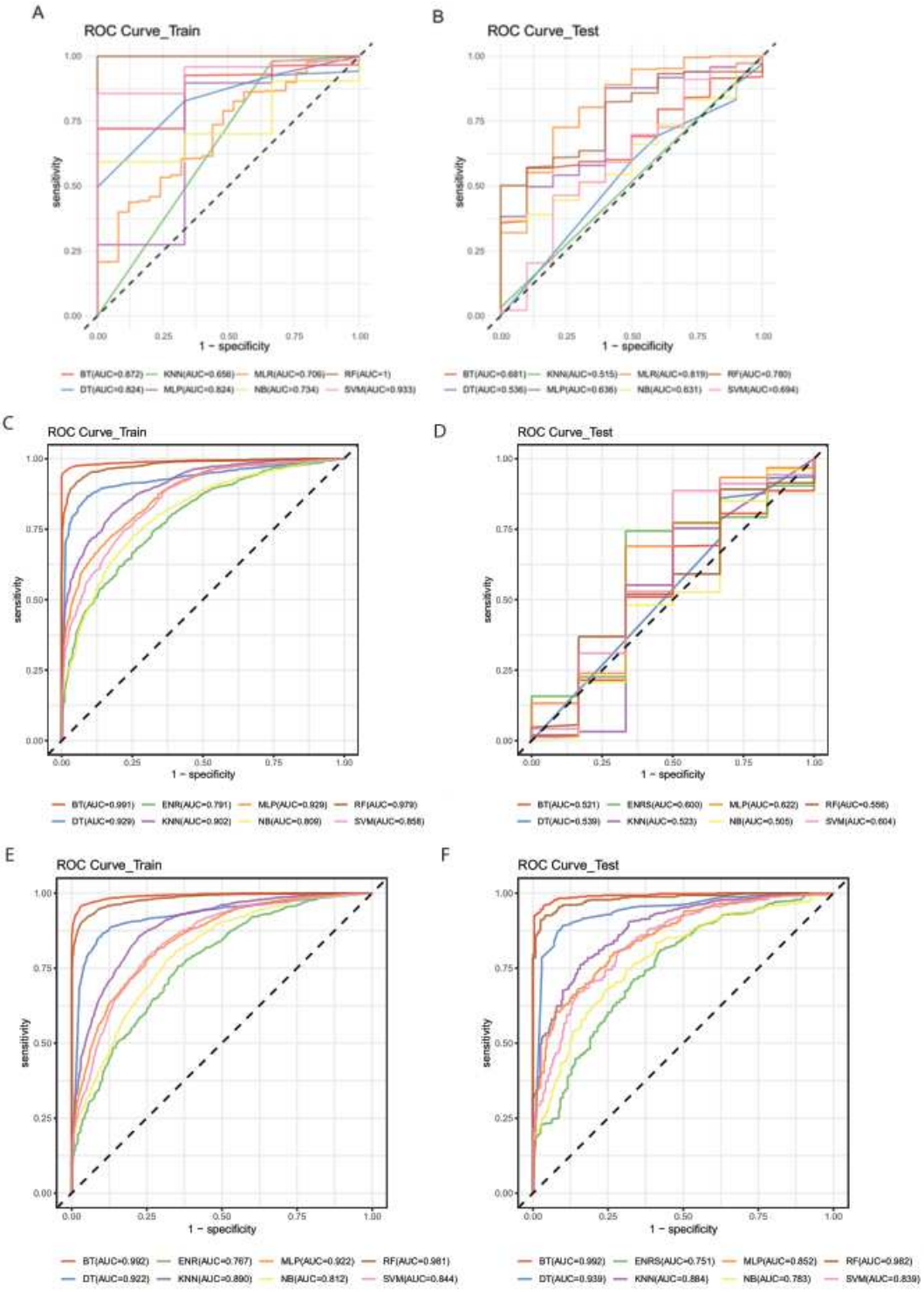


Figure 53 ROC curve of different algorithms before and after balancing in fetal-feature-including models

Figure S5. ROC curve of different algorithms before and after balancing in fetal-feature-including models. (A&B) ROC curves with training set and testing set without balancing. (C&D) ROC curves with training set balanced only. (E&F) ROC curves with training set and testing set both balanced. The values were shown in the legends. AUC: area under the receiver operating characteristic curve; BT: boost tree; DT: decision tree; ENR: elastic net regression; KNN: k-nearest neighbor; MLP: multilayer perceptron; NB: naïve bayes; RF: random forest; SVM: support vector machine; ROC: receiver operating characteristic.

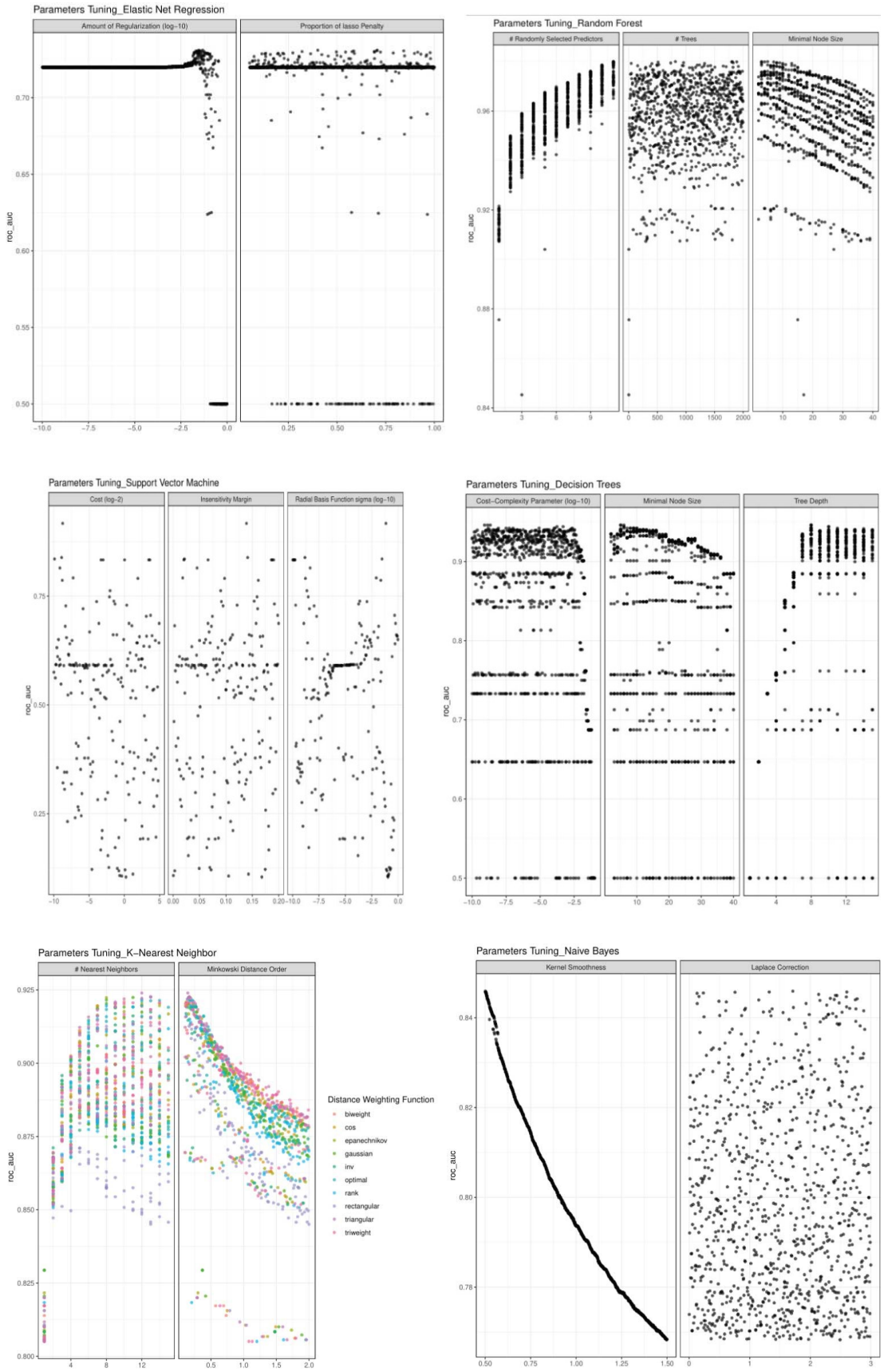


Figure 54 Fig. S6 Arguments tuning of machine learning methods

Fig. S6. We chose 8 widely used machine learning algorithms (elastic net regression, support vector machine, random forest, boost tree, decision tree, k-nearest neighbor, naïve Bayes, and multilayer perceptron) to test models in training set, along with argument tuning using the maximum entropy design with 1000 candidate values. These graphs show the grid research of the optimal combination of arguments from different algorithms. The area under the receiver operating characteristic curve (AUC) values were used to evaluate the performance of models, with the best being filtered.

Text S1 The R script with detailed comments related to machine learning modeling.

```
library(tidymodels)
library(modelr)
library(broom)
library(discrim)
library(factoextra)
library(FactoMineR)
# library(Factoshiny)
# library(mice)
library(missForest)
library(Boruta)
library(vip)
library(imbalance)
options(digits = 4)

ppar <- readxl::read_xlsx(path = "eclaxir.xlsx")
df_names <- read.csv("df_name.csv", header = TRUE)

# 1 Preprocessing ----

pparg <- ppar %>%
  select(c("preeclamp",
           "c24_Height", "c24_edu", "ob_Nbpreg", "ob_Nbdelv",
           "nn_Sex", "bmi","age_delv","primidelv", "nbcig","obese",
           "genot681_C", "genot681_M", "genotp12a_C", "genotp12a_M", "genot1431_C",
"genot1431_M",
           "carrier681_M", "carrier12a_M", "carrier1431_M", "carrier681_C",
"carrier12a_C","carrier1431_C"))

# imputation

set.seed(222)
imp <- missForest(as.data.frame(pparg))
pparg_imp <-
  imp$ximp %>%
  as_tibble() %>%
```

```
  as.data.frame()


# 2 Individual weight----

# PCA cluster for PE patients

famd <- FAMD(pparg_imp[,-1], ncp = 10, graph = FALSE) # without preecalmpsia
fviz_famd_var(
  famd,
  col.var = "contrib",
  gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
  repel = TRUE,
  # Avoid text overlapping
  ggtheme = theme_minimal()
)
fviz_screeplot(famd, addlabels = TRUE)

ggplot(
  data = as.data.frame(famd$ind$coord)[, 1:2] %>%
bind_cols(pparg_imp$preeclamp) %>% rename(preeclamp = ...3),
  aes(x = Dim.1, y = Dim.2, color = preeclamp)
) +
  geom_point(alpha = 0.5) +
  scale_color_manual(values = c("steelblue","red")) +
  theme_minimal() +
  geom_hline(
    yintercept = 0,
    color = "black",
    lty = 2,
    size = 1
  ) +
  geom_vline(
    xintercept = 0,
    color = "black",
    lty = 2,
    size = 1
  ) +
  labs(color = "Preeclampsia") +
  labs(title = "Individuals_FAMD")

# t-SNE

library(Rtsne)
## Curating the database for analysis with both t-SNE and PCA
pparg_imp$preeclamp <- factor(pparg_imp$preeclamp)

## Executing the algorithm on curated data
tsne <-
  Rtsne(
    pparg_imp,
```

```
  dims = 3,
  perplexity = 30,
  verbose = TRUE,
  max_iter = 500
 )

## Plot
tsne$Y %>% as_tibble() %>%
 rename(Dim.1 = V1, Dim.2 = V2, Dim.3 = V3) %>%
 bind_cols(preeclamp = pparg_imp$preeclamp) %>%
 ggplot(aes(Dim.1, Dim.2, color = preeclamp)) +
 geom_point(alpha = 0.5) +
 scale_color_manual(values = c("steelblue", "red")) +
 labs(color = "Preeclampsia") +
 labs(title = "Individuals_t-SNE") +
 theme_minimal()

# Odds ratio
nogene_glm <-
 glm(preeclamp ~ .,
    data = pparg_imp %>% dplyr::select(-starts_with("genot")),
    family = binomial)

# Plot odds ratio
ggplot(nogene_glm %>% tidy(conf.int = TRUE) %>% select(term) %>%
       bind_cols(nogene_glm %>% tidy(conf.int = TRUE) %>% select(-term) %>%
exp())) +
 geom_errorbar(aes(xmin = conf.low, xmax = conf.high, y = term, col = "red"), size =
0.8) +
 geom_point(aes(y = term, x = estimate), shape = 18, size = 3) +
 geom_vline(xintercept = 0, lty = 2, size = 1) +
 #scale_x_log10() +
 labs(title = "Odds Ratio with 95% Wald Confidence Limits", x = "Odds Ratio", y = "")
+
 theme_minimal() +
 theme(legend.position = "none")

# Plot p values
ggplot(nogene_glm %>% tidy(conf.int = TRUE), aes(y = term, x = p.value)) +
 geom_point(size = 2) +
 geom_vline(xintercept = 0.05, color = "red", lty = 2, size = 1) +
 theme_minimal() +
 annotate("text", x = 0.12, y = 1, label = "p=0.05", color = "red") +
 labs(title = "P value of variables", x = "P value", y = "")


# Data set split and balance----
set.seed(123)
splits <- initial_split(pparg_imp, prop = 3 / 4, strata = preeclamp)
pparg_train <- training(splits)
pparg_test <- testing(splits)
```

```r
# Oversampling
imbalanceRatio(pparg_train %>% rename(Class = preeclamp))
# 0.02064

set.seed(1212)
pparg_os <- pparg_train %>%
  rename(Class = preeclamp) %>%
  map_if(is.factor, as.numeric) %>%
  as.data.frame() %>%
  oversample(ratio = 0.6, method = "RACOG", filtering = FALSE) %>%
  mutate_at(vars(-c(c24_Height, c24_edu, ob_Nbdelv, ob_Nbpreg, bmi, age_delv,
nbcig)), as.factor)

pparg_train <- pparg_os %>% rename(preeclamp = Class)


# 3 Feature selection ----

# correlation of variables
library(corrplot)
corrplot::corrplot(cor(pparg_imp[, colnames(pparg_imp) %in% quanti]))
corrplot(cor(
  pparg_imp %>%
    select(-starts_with("genot")) %>%
    # select(-preeclamp) %>%
    map_if(is.factor, as.numeric) %>%
    as_tibble() %>%
    as.matrix() %>%
    + 0.1,
  method = "kendall"
))


# 1) logistic regression, 2) penalized logistic regression, 3) Boruta

# 1) logistic regression

# to confirm mutation's role in predicting, remove genetic variables with 3 levels
# keep carrier variables with 2 levels to exhibit the correlation
lr_fit <-
  glm(preeclamp ~ .,
      data = pparg_train_fs,
      family = binomial)
# variable selection according to AIC principle
library(MASS)
stepAIC(lr_fit)

vip(lr_fit, num_features = 30) +
  geom_hline(yintercept = 1, col = "red", size = 1, lty = 2) +
  labs(title = "Variables importance_logistic regression") +
```

```
  theme_minimal()
# results
# [1] "ob_Nbpreg", "primidelv", "c24_edu", "nn_Sex", "bmi", "carrier1431_M",
"carrier1431_C", "nbcig"

# 2) penalized logistic regression

plr_mod <-
  logistic_reg(mixture = 1) %>%
  set_engine("glmnet")

plr_recipe <-
  recipe(preeclamp ~ ., data = pparg_train_fs) %>%
  step_dummy(all_nominal(),-all_outcomes()) %>%
  step_zv(all_predictors())

plr_wkflow <-
  workflow() %>%
  add_model(plr_mod) %>%
  add_recipe(plr_recipe)

plr_fit <- fit(plr_wkflow, data = pparg_train_fs)

# importance plot
plr_fit %>%
  pull_workflow_fit() %>%
  vip(num_features = 30) +
  geom_hline(yintercept = 0.5, col = "red", size = 1, lty = 2) +
  labs(title = "Variables importance_Lasso") +
  theme_minimal()
# results
# [1] "carrier1431_M", "primidelv", "carrier1431_C", "carrier681_M"

# 3) Boruta

set.seed(111)
boruta_train <- Boruta(preeclamp ~., data = pparg_train_fs, doTrace = 2)
boruta_df <- attStats(boruta_train)

plot(boruta_train, xlab = "", xaxt = "n")
boruta_train$ImpHistory %>%
  as_tibble() %>% filter(across(everything(), is.finite)) %>%
  map(median) %>%
  as_tibble() %>%
  sort() %>%
  colnames() %>%
  axis(side = 1,las=2,labels = ., at = 1:ncol(boruta_train$ImpHistory), cex.axis = 0.7)
  title("Variables importance_Boruta")

getSelectedAttributes(boruta_train)
```

```
# [1] "c24_edu", "obese", (bmi), "carrier681_M", "carrier1431_M", "carrier681_C",
"carrier1431_C"


# 4 Machine learning models----

# to unit the results:
# "ob_Nbpreg","primidelv","c24_edu","nn_Sex","bmi","nbcig","obese"
#
"carrier1431_M"("genot1431_M"),"carrier681_C("genot681_C)","carrier1431_C"("ge
not1431_C"),"carrier681_M"(genot681_M),

pparg_train <- pparg_train %>%

select(preeclamp,genot1431_M,genot681_M,primidelv,ob_Nbpreg,c24_edu,obese,b
mi,nbcig)
#c(genot1431_C,genot681_C,nn_Sex,)
pparg_test <- testing(splits) %>%

select(preeclamp,genot1431_M,genot681_M,primidelv,ob_Nbpreg,c24_edu,obese,b
mi,nbcig) %>%
  map_if(is.factor, as.numeric) %>%
  as.data.frame() %>%
  mutate_at(vars(-c(c24_edu, ob_Nbpreg, bmi, nbcig)), as.factor)

set.seed(234)
# val_set <- validation_split(data = pparg_train,
#                             prop = 3/4,
#                             strata = preeclamp)

val_set <- vfold_cv(data = pparg_train,
                    v = 5,
                    repeats = 2,
                    strata = preeclamp)

# 1) MLR, 2) RF, 3) SVM, 4) DT, 5) KNN, 6) NB, 7) BT, 8a) MLP
# BT, boost tree; DT, decision tree; KNN, k-nearest neighbor; MLR, multivariate
logistic regression;
# MLP, multilayer perceptron; NB, naïve Bayes; RF, random forest;
# SVM, support vector machine.

eval_sets <- metric_set(ppv,npv, accuracy, f_meas, mcc)

# 0 logistic regression ----

lr_mod <-
  logistic_reg() %>% # tuning
  set_engine("glm")


lr_wkflow <-
```

```
  workflow() %>%
  add_model(lr_mod) %>%
  add_formula(preeclamp ~ .)


lr_res <- fit_resamples(lr_wkflow, resamples = val_set, control =
control_resamples(save_pred = TRUE))
collect_predictions(lr_res)
collect_metrics(lr_res)

# result
# accuracy binary      0.677
# roc_auc binary       0.720

lr_fit <- fit(lr_wkflow, data = pparg_train)

# train
lr_pred_train <-
  predict(lr_fit, new_data = pparg_train) %>%
  bind_cols(predict(lr_fit, new_data = pparg_train, type = "prob")) %>%
  bind_cols(pparg_train %>% select(preeclamp))

lr_pred_train %>% accuracy(truth = preeclamp, .pred_class)
lr_pred_train %>% roc_auc(truth = preeclamp, .pred_1)
lr_auc_train <-
  lr_pred_train %>%
  roc_curve(truth = preeclamp, .pred_1) %>%
  mutate(model = "ENR(AUC=0.722)")
# result
# accuracy binary      0.671
# roc_auc binary       0.722

# test
lr_pred_test <-
  predict(lr_fit, new_data = pparg_test) %>%
  bind_cols(predict(lr_fit, new_data = pparg_test, type = "prob")) %>%
  bind_cols(pparg_test %>% select(preeclamp))

lr_pred_test %>% accuracy(truth = preeclamp, .pred_class)
lr_pred_test %>% roc_auc(truth = preeclamp, .pred_1)
lr_auc_test <- lr_pred_test %>%
  roc_curve(truth = preeclamp, .pred_1) %>%
  mutate(model = "MLR(AUC=0.829)")
# result
# accuracy binary      0.976
# roc_auc binary       0.819

# 1) Multivariate Logistic regression----

mlr_mod <-
  logistic_reg(penalty = tune(), mixture = tune()) %>% # tuning
```

```
 set_engine("glmnet")

mlr_recipe <-
 recipe(preeclamp ~., data = pparg_train) %>%
 step_dummy(all_nominal(), -all_outcomes()) %>%
 step_zv(all_predictors())

mlr_wkflow <-
 workflow() %>%
 add_model(mlr_mod) %>%
 add_recipe(mlr_recipe)

# for tuning
set.seed(100)
mlr_tune <-
 mlr_wkflow %>%
 tune_grid(
  resamples = val_set,
  param_info = mlr_mod %>% parameters(),
  grid = 1000,
  metrics = metric_set(roc_auc)
 )

show_best(mlr_tune, n = 5)
select_best(mlr_tune, metric = "roc_auc")

autoplot(mlr_tune) + theme_bw() + ggtitle("Parameters Tuning_Elastic Net
Regression")
ggsave("Parameteris Tuning_Elastic Net Regression.pdf", width = 8, height= 8)


final_mlr_wkflow <-
 mlr_wkflow %>%
 finalize_workflow(mlr_tune %>% select_best(metric = "roc_auc"))


mlr_res <- fit_resamples(final_mlr_wkflow, resamples = val_set, control =
control_resamples(save_pred = TRUE))
collect_predictions(mlr_res)
collect_metrics(mlr_res)

# result
# accuracy binary      0.661
# roc_auc binary       0.795

mlr_fit <- fit(final_mlr_wkflow, data = pparg_train)

# train
mlr_pred_train <-
 predict(mlr_fit, new_data = pparg_train) %>%
 bind_cols(predict(mlr_fit, new_data = pparg_train, type = "prob")) %>%
```

```
  bind_cols(pparg_train %>% select(preeclamp))

mlr_pred_train %>% accuracy(truth = preeclamp, .pred_class)
mlr_pred_train %>% roc_auc(truth = preeclamp, .pred_1)
mlr_auc_train <-
  mlr_pred_train %>%
  roc_curve(truth = preeclamp, .pred_1) %>%
  mutate(model = "MLR(AUC=0.703)")
eval_sets(mlr_pred_train, truth = preeclamp, estimate = .pred_class)
# result
# accuracy binary       0.671
# roc_auc binary        0.703

# test
mlr_pred_test <-
  predict(mlr_fit, new_data = pparg_test) %>%
  bind_cols(predict(mlr_fit, new_data = pparg_test, type = "prob")) %>%
  bind_cols(pparg_test %>% select(preeclamp))

mlr_pred_test %>% accuracy(truth = preeclamp, .pred_class)
mlr_pred_test %>% roc_auc(truth = preeclamp, .pred_1)
mlr_auc_test <- mlr_pred_test %>%
  roc_curve(truth = preeclamp, .pred_1) %>%
  mutate(model = "MLR(AUC=0.784)")
eval_sets(mlr_pred_test, truth = preeclamp, estimate = .pred_class)
# result
# accuracy binary       0.875
# roc_auc binary        0.784

# 2) random forest ----

cores <- parallel::detectCores()
rf_mod <-
  rand_forest(mtry = tune(), trees = tune(), min_n = tune()) %>%
  set_engine("ranger", num.threads = cores, importance = "impurity", keep.inbag =
TRUE) %>%
  set_mode("classification")

rf_recipe <-
  recipe(preeclamp ~ ., data = pparg_train) # equals to "add_formula(preeclamp ~ .,
data = pparg_train)"
# step_dummy(all_nominal()) # for randomForest, it's unnecessary

rf_wkflow <-
  workflow() %>%
  add_model(rf_mod) %>%
  add_recipe(rf_recipe)

# for tuning
set.seed(100)
rf_tune <-
```

```
 rf_wkflow %>%
 tune_grid(
  resamples = val_set,
  param_info = rf_mod %>% parameters(),
  grid = 1000,
  metrics = metric_set(roc_auc),
  control = control_grid(save_pred = TRUE)
 )

show_best(rf_tune, n = 5)
select_best(rf_tune, metric = "roc_auc")

autoplot(rf_tune) + theme_bw() + ggtitle("Parameters Tuning_Random Forest")
ggsave("Parameteris Tuning_Random Forest.pdf", width = 8, height= 8)


final_rf_wkflow <-
 rf_wkflow %>%
 finalize_workflow(rf_tune %>% select_best())

rf_res <-
 fit_resamples(final_rf_wkflow, resamples = val_set, control =
control_resamples(save_pred = TRUE))

collect_metrics(rf_res)
collect_predictions(rf_res)

# result
# accuracy binary      0.913
# roc_auc binary       0.969

# Roc curve for validation set
rf_auc_train <-
 rf_res %>%
 collect_predictions() %>%
 roc_curve(preeclamp, .pred_1) %>%
 mutate(model = "RF(AUC=0.969)")

rf_fit <- fit(final_rf_wkflow, data = pparg_train)

# train
rf_pred_train <-
 predict(rf_fit, new_data = pparg_train) %>%
 bind_cols(predict(rf_fit, new_data = pparg_train, type = "prob")) %>%
 bind_cols(pparg_train %>% select(preeclamp))
eval_sets(rf_pred_train, truth = preeclamp, estimate = .pred_class)

# test
rf_pred <-
 predict(rf_fit, new_data = pparg_test) %>%
 bind_cols(predict(rf_fit, new_data = pparg_test, type = "prob")) %>%
```

```
  bind_cols(pparg_test %>% select(preeclamp)) %>%
  bind_cols(predict(rf_fit, new_data = pparg_test, type = "conf_int"))
eval_sets(rf_pred, truth = preeclamp, estimate = .pred_class)

rf_pred %>% accuracy(truth = preeclamp, .pred_class)
rf_pred %>% roc_auc(truth = preeclamp, .pred_1)
rf_auc_test <-
  rf_pred %>%
  roc_curve(truth = preeclamp, .pred_1) %>%
  mutate(model = "RF(AUC=0.723)")
# result
# accuracy binary       0.896
# roc_auc binary        0.723

# 3) Support vector machine ----

svm_mod <-
  svm_rbf(cost = tune(), rbf_sigma = tune(), margin = tune()) %>%
  set_engine("kernlab") %>%
  set_mode("classification")

svm_recipe <-
  recipe(preeclamp ~ ., data = pparg_train) %>%
  step_dummy(all_nominal(), -all_outcomes()) %>%
  step_zv(all_predictors()) %>%
  step_normalize(-all_nominal())

svm_wkflow <-
  workflow() %>%
  add_model(svm_mod) %>%
  add_recipe(svm_recipe)

set.seed(100)
svm_tune <-
  svm_wkflow %>%
  tune_grid(
    resamples = val_set,
    param_info = svm_mod %>% parameters(),
    grid = 1000,
    metrics = metric_set(roc_auc),
    control = control_grid(save_pred = TRUE)
  )

show_best(svm_tune, n = 5)
select_best(svm_tune, metric = "roc_auc")

autoplot(svm_tune) + theme_bw() + ggtitle("Parameters Tuning_Support Vector
Machine")
ggsave("Parameteris Tuning_Support Vector Machine.pdf", width = 8, height= 8)

final_svm_wkflow <-
```

```
 svm_wkflow %>%
 finalize_workflow(svm_tune %>% select_best())

svm_res <-
 fit_resamples(final_svm_wkflow, val_set, control = control_resamples(save_pred
= TRUE))
collect_metrics(svm_res)
# result for trains
# accuracy binary      0.772
# roc_auc binary       0.847

# Roc curve for validation set
svm_auc_train <-
 svm_res %>%
 collect_predictions() %>%
 roc_curve(preeclamp, .pred_1) %>%
 mutate(model = "SVM(AUC=0.847)")

svm_fit <- fit(final_svm_wkflow, data = pparg_train)

# train
svm_pred_train <-
 predict(svm_fit, new_data = pparg_train) %>%
 bind_cols(predict(svm_fit, new_data = pparg_train, type = "prob")) %>%
 bind_cols(pparg_train %>% select(preeclamp))
eval_sets(svm_pred_train, truth = preeclamp, estimate = .pred_class)

# test
svm_pred <-
 predict(svm_fit, new_data = pparg_test) %>%
 bind_cols(predict(svm_fit, new_data = pparg_test, type = "prob")) %>%
 bind_cols(pparg_test %>% select(preeclamp))
eval_sets(svm_pred, truth = preeclamp, estimate = .pred_class)

svm_pred %>% accuracy(truth = preeclamp, .pred_class)
svm_pred %>% roc_auc(truth = preeclamp, .pred_1)
svm_auc_test <-
 svm_pred %>%
 roc_curve(truth = preeclamp, .pred_1) %>%
 mutate(model = "SVM(AUC=0.545)")
# result for test
# accuracy binary      0.862
# roc_auc binary       0.545


# 4) decision trees ----

dt_mod <-
 decision_tree(tree_depth = tune(), min_n = tune(), cost_complexity = tune()) %>%
 set_engine("rpart") %>%
 set_mode("classification")
```

```r
dt_wkflow <-
  workflow() %>%
  add_model(dt_mod) %>%
  add_formula(preeclamp ~ .)

set.seed(100)
dt_tune <-
  dt_wkflow %>%
  tune_grid(
    resamples = val_set,
    grid = 1000,
    param_info = dt_mod %>% parameters(),
    metrics = metric_set(roc_auc),
    control = control_grid(save_pred = TRUE)
  )

show_best(dt_tune, n = 5)
select_best(dt_tune, metric = "roc_auc")

autoplot(dt_tune) + theme_bw() + ggtitle("Parameters Tuning_Decision Trees")
ggsave("Parameteris Tuning_Decision Trees.pdf", width = 8, height= 8)

final_dt_wkflow <-
  dt_wkflow %>%
  finalize_workflow(dt_tune %>% select_best())

dt_res <-
  fit_resamples(final_dt_wkflow, resamples = val_set, control =
control_resamples(save_pred = TRUE))
collect_metrics(dt_res)
# result for trains
# accuracy binary       0.849
# roc_auc binary        0.919

# Roc curve for validation set
dt_auc_train <-
  dt_res %>%
  collect_predictions() %>%
  roc_curve(preeclamp, .pred_1) %>%
  mutate(model = "DT(AUC=0.919)")

dt_fit <- fit(final_dt_wkflow, data = pparg_train)

# train
dt_pred_train <-
  predict(dt_fit, new_data = pparg_train) %>%
  bind_cols(predict(dt_fit, new_data = pparg_train, type = "prob")) %>%
  bind_cols(pparg_train %>% select(preeclamp))
eval_sets(dt_pred_train, truth = preeclamp, estimate = .pred_class)
```

```r
# test
dt_pred <-
  predict(dt_fit, new_data = pparg_test) %>%
  bind_cols(predict(dt_fit, new_data = pparg_test, type = "prob")) %>%
  bind_cols(pparg_test %>% select(preeclamp))
eval_sets(dt_pred, truth = preeclamp, estimate = .pred_class)

dt_pred %>% accuracy(truth = preeclamp, .pred_class)
dt_pred %>% roc_auc(truth = preeclamp, .pred_2)
dt_auc_test <-
  dt_pred %>%
  roc_curve(truth = preeclamp, .pred_2) %>%
  mutate(model = "DT(AUC=0.579)")
# result
# accuracy binary      0.874
# roc_auc binary       0.579

# 5) K-nearest neighbor ----

knn_mod <-
  nearest_neighbor(neighbors = tune(), weight_func = tune(), dist_power =
tune()) %>%
  set_engine("kknn") %>%
  set_mode("classification")

knn_recipe <-
  recipe(preeclamp ~ ., data = pparg_train) %>%
  step_dummy(all_nominal(), -all_outcomes()) %>%
  step_zv(all_predictors()) %>%
  step_normalize(-all_nominal())

knn_wkflow <-
  workflow() %>%
  add_model(knn_mod) %>%
  add_recipe(knn_recipe)

set.seed(100)
knn_tune <-
  knn_wkflow %>%
  tune_grid(
    resamples = val_set,
    grid = 1000,
    param_info = knn_mod %>% parameters(),
    metrics = metric_set(roc_auc),
    control = control_grid(save_pred = TRUE)
  )

show_best(knn_tune, n = 5)
select_best(knn_tune, metric = "roc_auc")
```

```
autoplot(knn_tune) + theme_bw() + ggtitle("Parameters Tuning_K-Nearest
Neighbor")
ggsave("Parameteris Tuning_K-Nearest Neighbor.pdf", width = 8, height= 8)

final_knn_wkflow <-
  knn_wkflow %>%
  finalize_workflow(knn_tune %>% select_best())

knn_res <-
  fit_resamples(final_knn_wkflow, resamples = val_set, control =
control_resamples(save_pred = TRUE))
collect_metrics(knn_res)
# result for trains
# accuracy binary       0.826
# roc_auc binary        0.917

# Roc curve for validation set
knn_auc_train <-
  knn_res %>%
  collect_predictions() %>%
  roc_curve(preeclamp, .pred_1) %>%
  mutate(model = "KNN(AUC=0.917)")

knn_fit <- fit(final_knn_wkflow, data = pparg_train)

# train
knn_pred_train <-
  predict(knn_fit, new_data = pparg_train) %>%
  bind_cols(predict(knn_fit, new_data = pparg_train, type = "prob")) %>%
  bind_cols(pparg_train %>% select(preeclamp))
eval_sets(knn_pred_train, truth = preeclamp, estimate = .pred_class)

# test
knn_pred <-
  predict(knn_fit, new_data = pparg_test) %>%
  bind_cols(predict(knn_fit, new_data = pparg_test, type = "prob")) %>%
  bind_cols(pparg_test %>% select(preeclamp))
eval_sets(knn_pred, truth = preeclamp, estimate = .pred_class)

knn_pred %>% accuracy(truth = preeclamp, .pred_class)
knn_pred %>% roc_auc(truth = preeclamp, .pred_1)
knn_auc_test <-
  knn_pred %>%
  roc_curve(truth = preeclamp, .pred_1) %>%
  mutate(model = "KNN(AUC=0.725)")
# result
# accuracy binary       0.801
# roc_auc binary        0.725

# 6) naive Bayes ----
```

```
nb_mod <-
  naive_Bayes(smoothness = tune(), Laplace = tune()) %>%
  set_engine("naivebayes") %>%
  set_mode("classification")

nb_recipe <-
  recipe(preeclamp ~ ., data = pparg_train) %>%
  step_zv(all_predictors())

nb_wkflow <-
  workflow() %>%
  add_model(nb_mod) %>%
  add_recipe(nb_recipe)

set.seed(100)
nb_tune <-
  nb_wkflow %>%
  tune_grid(
    resamples = val_set,
    grid = 1000,
    param_info = nb_mod %>% parameters(),
    metrics = metric_set(roc_auc),
    control = control_grid(save_pred = TRUE)
  )

show_best(nb_tune, n = 5)
select_best(nb_tune, metric = "roc_auc")

autoplot(nb_tune) + theme_bw() + ggtitle("Parameters Tuning_Naive Bayes")
ggsave("Parameteris Tuning_Naive Bayes.pdf", width = 8, height= 8)

final_nb_wkflow <-
  nb_wkflow %>%
  finalize_workflow(nb_tune %>% select_best())

nb_res <-
  fit_resamples(final_nb_wkflow, resamples = val_set, control =
control_resamples(save_pred = TRUE))
collect_metrics(nb_res)
# result for trains
# accuracy binary       0.693
# roc_auc binary        0.787

# Roc curve for validation set
nb_auc_train <-
  nb_res %>%
  collect_predictions() %>%
  roc_curve(preeclamp, .pred_1) %>%
  mutate(model = "NB(AUC=0.787)")

nb_fit <- fit(final_nb_wkflow, data = pparg_train)
```

```r
# train
nb_pred_train <-
  predict(nb_fit, new_data = pparg_train) %>%
  bind_cols(predict(nb_fit, new_data = pparg_train, type = "prob")) %>%
  bind_cols(pparg_train %>% select(preeclamp))
eval_sets(nb_pred_train, truth = preeclamp, estimate = .pred_class)

# test
nb_pred <-
  predict(nb_fit, new_data = pparg_test) %>%
  bind_cols(predict(nb_fit, new_data = pparg_test, type = "prob")) %>%
  bind_cols(pparg_test %>% select(preeclamp))
eval_sets(nb_pred, truth = preeclamp, estimate = .pred_class)

nb_pred %>% accuracy(truth = preeclamp, .pred_class)
nb_pred %>% roc_auc(truth = preeclamp, .pred_1)
nb_auc_test <-
  nb_pred %>%
  roc_curve(truth = preeclamp, .pred_1) %>%
  mutate(model = "NB(AUC=0.619)")
# result
# accuracy binary       0.930
# roc_auc binary        0.619

# 7) boost tree ----

bt_mod <-
  boost_tree(tree_depth = tune(),
             trees = tune(),
             learn_rate = tune(),
             mtry = tune(),
             min_n = tune(),
             loss_reduction = tune(),
             sample_size = tune()) %>%
  set_engine("xgboost") %>%
  set_mode("classification")

bt_wkflow <-
  workflow() %>%
  add_model(bt_mod) %>%
  add_formula(preeclamp ~ .)

set.seed(100)
bt_tune <-
  bt_wkflow %>%
  tune_grid(
    resamples = val_set,
    grid = 1000,
    param_info = bt_mod %>% parameters(),
    metrics = metric_set(roc_auc),
```

```
  control = control_grid(save_pred = TRUE)
 )

show_best(bt_tune, n = 5)
select_best(bt_tune, metric = "roc_auc")

autoplot(nb_tune) + theme_bw() + ggtitle("Parameters Tuning_Boost Tree")
ggsave("Parameteris Tuning_Boost Tree.pdf", width = 8, height= 8)

final_bt_wkflow <-
 bt_wkflow %>%
 finalize_workflow(bt_tune %>% select_best())

bt_res <-
 fit_resamples(final_bt_wkflow, resamples = val_set, control =
control_resamples(save_pred = TRUE))
collect_metrics(bt_res)
# result for trains
# accuracy binary      0.967
# roc_auc binary       0.991

# Roc curve for validation set
bt_auc_train <-
 bt_res %>%
 collect_predictions() %>%
 roc_curve(preeclamp, .pred_1) %>%
 mutate(model = "BT(AUC=0.991)")

bt_fit <- fit(final_bt_wkflow, data = pparg_train)

# train
bt_pred_train <-
 predict(bt_fit, new_data = pparg_train) %>%
 bind_cols(predict(bt_fit, new_data = pparg_train, type = "prob")) %>%
 bind_cols(pparg_train %>% select(preeclamp))
eval_sets(bt_pred_train, truth = preeclamp, estimate = .pred_class)

# test
bt_pred <-
 predict(bt_fit, new_data = pparg_test) %>%
 bind_cols(predict(bt_fit, new_data = pparg_test, type = "prob")) %>%
 bind_cols(pparg_test %>% select(preeclamp))
eval_sets(bt_pred, truth = preeclamp, estimate = .pred_class)

bt_pred %>% accuracy(truth = preeclamp, .pred_class)
bt_pred %>% roc_auc(truth = preeclamp, .pred_1)
bt_auc_test <-
 bt_pred %>%
 roc_curve(truth = preeclamp, .pred_1) %>%
 mutate(model = "BT(AUC=0.701)")
# result
```

```
# accuracy binary       0.954
# roc_auc binary        0.700

# plot boost tree
library(treeheatr)
heat_tree(
  pparg_train,
  target_lab = "preeclamp",
  task = "classification",
  label_map = c(`2` = "Positive", `1` = "Negative"),
  feats =
c("genot1431_M","genot681_M","primidelv","ob_Nbpreg","c24_edu","obese","bmi",
"nbcig"),
  lev_fac = 1.2,
  show = "heat-tree",
  heat_rel_height = 0.25,
  panel_space = 0.001
)
ggsave("Boost tree map.pdf", width = 8, height = 8)

# Boost tree diagram

library(DiagrammeR)
xgb.plot.tree(
  feature_names = c(
    "genot1431_M",
    "genot1431_C",
    "genot681_M",
    "genot681_C",
    "primidelv",
    "ob_Nbpreg",
    "c24_edu",
    "nn_Sex",
    "obese",
    "bmi",
    "nbcig"
  ),
  model = xgb.fit.final
)

gr <- xgb.plot.multi.trees(xgb.fit.final,
                  feature_names = c(
                    "genot1431_M",
                    "genot1431_C",
                    "genot681_M",
                    "genot681_C",
                    "primidelv",
                    "ob_Nbpreg",
                    "c24_edu",
                    "nn_Sex",
                    "obese",
```

```
                  "bmi",
                  "nbcig"),
                features_keep = 11,
                render = TRUE
               )
gr

# 8) multilayer perceptron ----

mlp_mod <-
  mlp(hidden_units = tune(), penalty = tune(), epochs = tune()) %>%
  set_engine("nnet") %>%
  set_mode("classification")

mlp_recipe <-
  recipe(preeclamp ~ ., data = pparg_train) %>%
  step_dummy(all_nominal(), -all_outcomes()) %>%
  step_zv(all_predictors()) %>%
  step_normalize(all_predictors())

mlp_wkflow <-
  workflow() %>%
  add_model(mlp_mod) %>%
  add_recipe(mlp_recipe)

set.seed(100)
mlp_tune <-
  mlp_wkflow %>%
  tune_grid(
    resamples = val_set,
    grid = 1000,
    param_info = mlp_mod %>% parameters(),
    metrics = metric_set(roc_auc),
    control = control_grid(save_pred = TRUE)
  )

show_best(mlp_tune, n = 5)
select_best(mlp_tune, metric = "roc_auc")

autoplot(nb_tune) + theme_bw() + ggtitle("Parameters Tuning_Multilayer
Perceptron")
ggsave("Parameteris Tuning_Multilayer Perceptron.pdf", width = 8, height= 8)

final_mlp_wkflow <-
  mlp_wkflow %>%
  finalize_workflow(mlp_tune %>% select_best())
mlp_res <-
  fit_resamples(final_mlp_wkflow, resamples = val_set, control =
control_resamples(save_pred = TRUE))
collect_metrics(dt_res)
# result for trains
```

```r
# accuracy binary       0.849
# roc_auc binary        0.919

# Roc curve for validation set
mlp_auc_train <-
  mlp_res %>%
  collect_predictions() %>%
  roc_curve(preeclamp, .pred_1) %>%
  mutate(model = "MLP(AUC=0.919)")

mlp_fit <- fit(final_mlp_wkflow, data = pparg_train)

# train
mlp_pred_train <-
  predict(mlp_fit, new_data = pparg_train) %>%
  bind_cols(predict(mlp_fit, new_data = pparg_train, type = "prob")) %>%
  bind_cols(pparg_train %>% select(preeclamp))
eval_sets(mlp_pred_train, truth = preeclamp, estimate = .pred_class)

# test
mlp_pred <-
  predict(mlp_fit, new_data = pparg_test) %>%
  bind_cols(predict(mlp_fit, new_data = pparg_test, type = "prob")) %>%
  bind_cols(pparg_test %>% select(preeclamp))
eval_sets(mlp_pred, truth = preeclamp, estimate = .pred_class)

mlp_pred %>% accuracy(truth = preeclamp, .pred_class)
mlp_pred %>% roc_auc(truth = preeclamp, .pred_1)
mlp_auc_test <-
  mlp_pred %>%
  roc_curve(truth = preeclamp, .pred_1) %>%
  mutate(model = "MLP(AUC=0.670)")
# result
# accuracy binary       0.811
# roc_auc binary        0.670

# ROC curve combined plots----

# train
p1 <-
  bind_rows(mlr_auc_train,
            rf_auc_train,
            svm_auc_train,
            dt_auc_train,
            knn_auc_train,
            nb_auc_train,
            bt_auc_train,
            mlp_auc_train) %>%
  ggplot(aes(x = 1 - specificity, y = sensitivity, col = model)) +
  geom_path(lwd = 1, alpha = 0.8) +
  geom_abline(lty = 2) +
```

```
  coord_equal() +
  scale_color_brewer(palette = "Set1") +
  labs(col = "Models_Train") +
  theme_bw() +
  theme(legend.position = "bottom") +
  ggtitle("ROC Curve_Train")

p1

ggsave("c1431t_Roc_curve_Train.pdf", width = 8, height = 6)

# test
p2 <-
  bind_rows(mlr_auc_test,
            rf_auc_test,
            svm_auc_test,
            dt_auc_test,
            knn_auc_test,
            nb_auc_test,
            bt_auc_test,
            mlp_auc_test) %>%
  ggplot(aes(x = 1 - specificity, y = sensitivity, col = model)) +
  geom_path(lwd = 1, alpha = 0.8) +
  geom_abline(lty = 2) +
  coord_equal() +
  scale_color_brewer(palette = "Set1") +
    labs(col = "Models_Test") +
  theme_bw() +
  theme(legend.position = "bottom") +
  ggtitle("ROC Curve_Test")

p2

ggsave("c1431t_Roc_curve_Test.pdf", width = 8, height = 6)
```