

# A genome-wide association study to identify candidate genes for metabolic disorders in offspring by in vitro fertilization

Lihong Liu

Shantou University Medical College

Siyao Ha

Fudan University

Zhiling Li (✉ [stlizhiling@126.com](mailto:stlizhiling@126.com))

Shantou University Medical College

MingQing Li

Fudan University

---

## Research Article

**Keywords:** in vitro fertilization (IVF), offspring health, type 2 diabetes mellitus (T2DM), fatty acid metabolism

**Posted Date:** September 22nd, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-717175/v3>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# **A genome-wide association study to identify candidate genes for metabolic disorders in offspring by in vitro fertilization**

Lihong LIU<sup>1</sup>, Siyao HA<sup>2</sup>, Zhiling Li<sup>1</sup>, and MingQing Li<sup>2</sup>

1. Reproductive Center, The First Affiliated Hospital of Shantou University Medical College, Shantou University, Shantou 515041,China
2. Laboratory for Reproductive Immunology, National Health Commission (NHC) Key Lab of Reproduction Regulation (Shanghai Institute of Planned Parenthood Research), Hospital of Obstetrics and Gynecology, Shanghai Medical School, Fudan University, Shanghai 200082,China

Lihong LIU and Siyao HA contributed equally to this work.

\*Correspondence: Zhiling Li,E-mail: [stlizhiling@126.com](mailto:stlizhiling@126.com)

\*Correspondence: Mingqing Li E-mail: [mqli@fudan.edu.cn](mailto:mqli@fudan.edu.cn)

## **Abstract**

In vitro fertilization (IVF) processes increase offspring's short-term and long-term health risks, but their mechanisms remain unclear. We conducted a bibliometric analysis to determine the landscape of IVF offspring health. Subsequently, a bioinformatics method was utilized to identify the co-genes properties and biological function mechanisms of IVF and type 2 diabetes mellitus (T2DM). Finally, we predicted compounds against key targets and performed multiple validations of the mechanisms underlying IVF offspring health risks. We identified 15 genes associated with T2DM, and their biological functions are primarily associated with lipid metabolism. We also identified the properties of co-genes, modified characteristics, confirmed a conserved motif, identified 3 SNPs sites, and determined the three core genes, APOA1, APOB, and APOE, which were mainly correlated with metabolic and cardiovascular diseases. In addition, we predicted drugs that may improve metabolic abnormalities in IVF offspring. The impact of aberrant lipid metabolism in offspring after IVF therapy warrants additional investigation, particularly in terms of long-term health consequences and possible mechanisms.

**Keywords:** in vitro fertilization (IVF); offspring health; type 2 diabetes mellitus (T2DM); fatty acid metabolism

**Abbreviations:** IVF, In vitro fertilization; T2DM, Type 2 diabetes mellitus; ART, Assisted reproductive technology; DOHAD, Developmental Origin of Adult Health and Disease Hypothesis; FET, Frozen embryo transfer; LBW, low birth weight; LGA, large gestational age; DEGs, differentially expressed genes; GO, gene ontology; CC, cellular component; MF, molecular function; BP, biological process; KEGG, Kyoto Encyclopedia of Genes and Genomes; PPIs, Protein-protein interactions; MW, molecular weight; FGR, fetal growth restriction; MCC, Maximal Clique Centrality; APL, antiphospholipid functional attenuation.

## **Introduction**

Although the success of IVF holds a promise for infertile couples, whether a series of non-physiological operations affect the health of their offspring remains controversial. Assisted reproductive technology (ART), especially IVF, includes a series of non-physiological processes such as superovulation, embryo manipulation, in vitro culture, and embryo transfer before implantation. During this critical period of genetic recombination, gametes and embryos are exposed to altered hormonal environments, temperature changes, pH values, and oxygen tension<sup>[1]</sup>. According to Developmental Origin of Adult Health and Disease Hypothesis (DOHAD)<sup>[2]</sup>, early environmental changes affect adult health and develop phenotypic plasticity, and children born with ART have an increased risk of metabolic diseases. Although most children born with IVF are healthy, the safety of this technique requires additional validation, particularly regarding long-term health risks. Animal and follow-up studies of IVF offspring have an increased risk of tumors<sup>[3]</sup>, early-onset diabetes<sup>[4]</sup> cardiovascular disease<sup>[5]</sup>, dyslipidemia<sup>[6]</sup>, and long-term neurologic morbidity<sup>[7]</sup>.

Numerous studies have indicated that IVF offspring is strongly associated with low birth weight (LBW)<sup>[8,9]</sup>. Fresh embryo transfer after ovarian hyperstimulation is associated with LBW. Although frozen embryo transfer (FET) is expected to reduce the adverse consequences of maternal hormone levels, it has been reported to be linked to large gestational age (LGA)<sup>[10]</sup>. There is a robust and well-documented observational that abnormal birth weight (LBW and LGA) increases the risk of cardiovascular disease and T2DM in advanced life<sup>[11]</sup>. Birth weight is a critical measure for assessing neonatal health. Prenatal fetal growth restriction leads to "catch-up" growth in the early postnatal period, predisposing adults to metabolic abnormalities such as visceral obesity, insulin resistance, and glucose intolerance<sup>[12]</sup>. Since these changes are related to T2DM, this study focused on the risk associated with IVF embryo metabolism changes and T2DM.

Sun's team collected villi from patients with fresh IVF embryo transfer and those with natural conception for transcriptome analysis and found that the two groups had significantly different data in multiple metabolic pathways, including insulin

resistance, fatty acid metabolism, steroid biosynthesis, and other T2DM-related pathways<sup>[13]</sup>. In addition, there is sufficient evidence to prove that excessive oxidative stress during IVF treatment affects the long-term health risk of offspring, and oxidative stress has adverse effects on offspring on fatty acid oxidation and mitochondrial function<sup>[14]</sup>. Abnormal fatty acid oxidation is one of the factors for increased metabolic risk of offspring<sup>[15]</sup>. Since metabolic abnormalities early in the embryo lead to an increased risk of metabolic diseases in adulthood, we explored the association between IVF offspring differences in gene expression and T2DM.

It is necessary to understand the effect of IVF treatment on progeny metabolism and identify related biomarkers and mechanisms to reduce treatment-related risks. In this study, we used data mining methods to compare IVF with villous trantranome data from early natural pregnancy to identify genes associated with generational diseases, identify genes associated with long-term effects of IVF treatment on offspring health, and explore the mechanisms by which IVF affects offspring health.

## **Materials and methods**

### **Data mining**

More attention has been paid to the safety of IVF treatment, especially its effect on offspring health. In this regard, we conducted a bibliometric analysis using VOSviewer tool to explore the current research landscape and identify relevant keywords on this issue. Furthermore, we employed the bioinformatics method to identify IVF differentially expressed genes and subsequently investigated their biofunction.

Bibliometric analysis. VOSviewer is a freely available programme created by Leiden University, displaying large bibliometric maps in an easy-to-interpret manner<sup>[16]</sup>. We constructed and reviewed bibliometric maps of IVF offspring health. All literature data were retrieved from Web of Science Core Collection from 1980 to 2021. The network map developed by VOSviewer eliminated various words that represented different keywords associated with offspring health.

Microarray data processing. IVF-related mRNA expression data were obtained from GEO database (<https://www.ncbi.nlm.nih.gov/geo/>). After a systematic review, the profile GSE122214 was selected since it contained four IVF patients and four normal pregnancy volunteers. In the IVF group, twin-to-singleton selective fetal reduction was performed 30–35 days after embryo transfer, equivalent to 7–8 gestational weeks. Collected villi from double chorionic twin pregnancies during routine abortion are used as a control group. Genes associated with T2DM were obtained from Phenopedia database<sup>[17]</sup> (<https://phgkb.cdc.gov/PHGKB/startPagePhenoPedia.action>). R script was utilized to identify differentially expressed genes (DEGs) between IVF

and natural conception. The adjusted  $p < 0.05$  and  $|\log FC| \geq 2$  were set as cut-off criteria for DEGs screening in the two groups using Bioconductor (<http://www.bioconductor.org/>). Subsequently, T2DM-related genetic data were downloaded from the website<sup>[18]</sup>. All the selected genes were examined using Venn diagram web tool (<https://bioinfogp.cnb.csic.es/tools/venny/index.html>).

Biofunctional analysis. We used DAVID (<https://david.ncifcrf.gov/home.jsp>) (version 6.8)<sup>[19,20]</sup> and Metascape<sup>[21]</sup> (<https://metascape.org>) to perform biological process enrichment analysis with overlapping genes between IVF and T2DM. The DAVID tool was a web-accessible program with categorical data for gene ontology (GO)<sup>[22]</sup>, including cellular component (CC), molecular function (MF), biological process (BP) three parts, and Kyoto Encyclopedia of Genes and Genomes (KEGG)<sup>[23]</sup>. Furthermore, MCODE algorithm was employed to cluster PPI networks to identify the subnetworks, namely potential protein complexes.

### **Properties of selected genes**

We first identified the expression of co-genes in IVF treatment and naturally conceived patients' villi and then analyzed these genes' characteristics and interaction networks in detail.

Physiochemical properties of selected genes. To better understand the co-genes, basic information of common targets was analyzed. The protein and nucleic acid sequences of selected genes were retrieved from Uniprot<sup>[24]</sup> (<https://www.uniprot.org/>) and NCBI databases<sup>[25]</sup> (<https://www.ncbi.nlm.nih.gov/gene/>), respectively. Further-more, the ProtParam tool in ExPASy database<sup>[26]</sup> (<https://web.expasy.org/protparam/>) was utilized to sequence the hydrophilic (GRAVY) isoelectric point (PI), molecular weight (MW), instability index, and total average.

Modification of selected protein. NetNGlyc 1.0 server<sup>[27]</sup> (<http://www.cbs.dtu.dk/services/NetNGlyc/>) was then utilized to predict the potential N-glycosylation and phosphorylation sites predicted by NetPhos 3.1 server<sup>[28]</sup> (<http://www.cbs.dtu.dk/services/NetPhos/>).

Other characteristics. UCSC database<sup>[29]</sup> was deployed to retrieve the promoter sites of selected genes (1500 bp upstream of the transcription start site). The Gapped Local Alignment of Motifs (GLAM2) tool<sup>[30]</sup> was utilized to find common conserved motifs sought at the promoter sites of selected genes. Then, we searched 15 common protein and diabetes-related SNP sites from Uniprot and SNP4 Disease databases (<http://snp4disease.mpi-bn.mpg.de/>), respectively. Venn diagram identified common SNP sites, and Cytoscape presented the network. Finally, target Scan Human<sup>[31]</sup> ([http://www.targetscan.org/vert\\_72/](http://www.targetscan.org/vert_72/)), a software for predicting miRNA binding sites, was applied to identify microRNA (miRNA) conserved sites broadly conserved

among vertebrates. Through this website, we searched for miRNAs that may regulate gene expression.

### **Network establishment and analysis**

Network establishment and hub gene analysis. Protein-protein interactions (PPIs) of the studied genes were assessed using STRING v11 tool<sup>[32]</sup> and visualized by Cytoscape<sup>[33]</sup>. In addition, to further analyze PPI network, cytoHubba<sup>[34]</sup> and MCODE<sup>[35]</sup> applications were used to identify the key nodes, and the correlation between genes was explored through gene expression. We then did reverse validation by looking for genetic diseases. DISEASES is a web asset that integrates evidence on disease-gene associations derived from programmed text mining, physically curated literature, disease information, and genome-wide association researches<sup>[36]</sup>. The key gene-related diseases were obtained by DISEASES, and the gene-disease interaction network was constructed to reversely verify the enriched key targets.

The connection between hub genes and diabetes was performed using Attie Lab Diabetes database (<http://diabetes.wisc.edu>), an open-source database displaying gene expression profiles of different experimental groups (lean and obese BTBR mice at 4 and 10 weeks of age) in different tissues<sup>[37]</sup>. We contrast both lean and obese BTBR mice in different ages (i.e., at 4 and 10 weeks) to verify mRNA levels of hub genes in different tissues (significance was set at  $P < 0.05$ ).

Identification of potential drugs. DrugBank is a bioinformatics and chemoinformatics database that combines detailed drug data with comprehensive drug target information provided by the University of Alberta<sup>[38]</sup>. We used DrugBank to explore molecular substances related to key targets to provide theoretical guidance for improving the health status of offspring.

### **Gene set enrichment analysis (GSEA)**

GSEA is a free software package developed initially to discover changes in the metabolic pathways correlated with human diabetes. Apart from using an arbitrary cut-off in terms of fold-change or significant score, GSEA evaluates all genes in an experiment and determines significance by permuting the class labels to determine gene-gene correlations, resulting in a more realistic null model<sup>[39]</sup>. In addition, GSEA can be employed to identify the pathways correlated with gene expression. Quality sets were considered significantly enriched at predefined  $p$ -qualities and  $FDR < 0.25$ . GSEA programming analyzed and sorted genes according to the algorithm after importing gene annotation files, analyzed the positions of all genes, and accumulated them to obtain enrichment pathways.

## **Results**

### **The current research landscape of IVF offspring health**

Following a WoS search, 585 publications and 9698 citation items were retrieved from 2003 to April 2021. From the trend line, the number of documents increased exponentially and reached a peak in 2020. A relevance score is calculated for all terms based on 24075 keyword terms and 492 records that met the threshold. Finally, 53 keywords related to health problems of offspring after IVF implantation were chosen (Figure 1A), which mainly includes keywords correlated with metabolism such as birth weight, BMI, diabetes, blood pressure, cardiovascular disease, long-term health, obesity, etc.

### **Identification of common genes between IVF and T2DM**

The following transcriptome profile datasets were downloaded from NCBI GEO database: GSE122214. After being analyzed, the villus transcriptome following IVF treatment was significantly different from natural conception. We successfully identified 1,806 DEGs, including 1,064 upregulated and 742 downregulated genes (Figures 1B). In addition, the potential targets associated with T2DM were retrieved from Phenopedia database. Further analysis of these DEGs using Venn diagram revealed 15 common genes as hub genes to identify IVF infants at risk of developing T2DM as adults (Figures 1C), and the common gene heatmap is displayed in Figure 1D.

### **Functional annotation for DEGs via DAVID and Metascape**

To further explore the biological function of target genes, two online databases (David and Metascape) were used to conduct functional analysis of the selected genes. Through DAVID analysis, the results of KEGG pathways indicated that the top canonical pathways associated with target genes included PPAR signaling pathway, HIF-1 signaling pathway, vitamin digestion, and absorption, as well as fat digestion and absorption (Figure 2A). GO analysis revealed that variations in common genes linked to biological processes were mainly enriched in cholesterol metabolic process, positive regulation of PI3K signaling, blood pressure regulation, lipoprotein metabolic process, and so on (Figure 2B).

Target genes linked to cellular components were significantly enriched in plasma membrane, extracellular exosome, very low-density lipoprotein particle, and chylomicron (Figure 2C). Regarding molecular function, these genes were significantly enriched in cholesterol transporter activity, identical protein binding, phospholipid binding, and lipid binding (Figure 2D). Furthermore, the functional enrichment analysis with Metascape revealed that target genes were significantly

enriched in regulating lipid metabolic process, small molecular, metabolic process, oxidoreductase activity, and so on ( $P < 0.05$ , Figures 2E–G).

### **Properties of selected genes**

Gene expression in the sample. A total of 15 common genes were differentially expressed in the villus of IVF and naturally conceived patients. Compared with the villus of naturally conceived patients, eight genes were upregulated, and seven genes were down-regulated in the villus of IVF patients (Figure 3A).

Physicochemical properties of selected genes. The 15 common genes' physicochemical features are listed in Table 1. The table lists the gene name, gene ID, protein length, MW, PI, instability index, and predicted N-glycosylation site. The results illustrated that these selected genes are distributed on chromosomes 1, 2, 3, 6, 7, 11, 12, and 19. The protein sequence length ranged from 167 to 4653 amino acids, with LEP having the shortest sequence at 167 amino acids and APOB having the longest sequence at 4653 amino acids. Besides, the molecular weight (MW) of selected genes ranged from 18.64 to 515.6 kDa. Furthermore, PI ranged from 5.56 to 8.93. Based on protein instability index, most studied proteins (10/15) were unstable. Also, most proteins exhibited negative GRAVY, revealing that they possess hydrophilic properties.

Protein modification information. Posttranscriptional modifications such as phosphorylation and glycosylation are involved in regulating protein stability and protein interactions. Therefore, the potential phosphorylation and glycosylation sites between IVF and T2DM were predicted into the amino acid sequence of selected genes (Table 1, Figure 3B). As predicted, APOB was identified as an N-hyperglycosylated protein that may be mostly glycosylated (Table 1). Also, the results demonstrated that APOB gene exhibits more phosphorylated sites (Figure 3B).

Other characteristics. Three common SNP sites were associated with diabetes mellitus, including two SNP sites of ApoE: rs7412 and rs429358 (Figure 3C). To identify common conserved motifs upstream of the selected genes, GLAM2 tool was utilized to screen promoter sites (1500 bp upstream of the start codon). As a result, there are GC-rich regions identified as conserved motifs in the upstream of selected genes; the conserved sequence may be involved in regulating key targets (Figure 3D). Next, we identified three diabetes-associated SNP sites and the conserved site for miRNA families. Concurrently, multiple miRNAs were identified to be involved in gene expression regulation, and green miRNAs, as central miRNAs, were involved in regulating multiple genes (Figure 3E).

### **Network establishment and analysis**

The protein-protein interaction (PPI) network involves most biological processes,

such as DNA transcription and replication, protein transport, protein degradation, and cell cycle regulation. In this study, STRING v11 was employed to construct a PPI network (Figure 4A), visualized using Cytoscape 3.5.1. First, the Maximal Clique Centrality (MCC) of each node was calculated by CytoHubba, a plugin in Cytoscape. The genes with the top 6 MCC values were considered hub genes. Concurrently, another Cytoscape plugin Mcode analyzes the network and determines the central key nodes, resulting in seven hub genes. Subsequently, an online Metascape enrichment analysis was used to obtain a key target complex: APOA1, APOB, and APOE (Figure 4B). Next, the gene correlation analysis was deployed to further investigate the expression relationship between genes—the stronger the gene correlation, the darker the color (Figure 4C). Finally, the Venn diagram was employed to identify common genes; APOE and APOA1 were identified as two such genes (Figure 4D). When the correlation between genes is considered, it can be concluded that APOA1, APOE, and APOB are strongly correlated and positively associated. As a result, we assumed that APOA1, APOE, and APOB were the key genes.

Gene complex-related diseases. Disease correlation analysis was conducted on the gene complex composed of APOA1, APOB, and APOE to find the diseases related to the complex for reverse verification, and a gene-disease network was constructed (Figure 4E). As the figure displays, diseases closely related to the three genes include metabolic diseases, cardio-cerebrovascular diseases, nervous system diseases, and immune-related diseases, among which metabolic diseases include obesity, diabetes, non-alcoholic fatty liver disease, and lipid metabolism disorders. Cardiovascular and cerebrovascular diseases include hypertension, heart disease, coronary artery disease, cerebrovascular disease, etc. Immune-related diseases mainly include Leukocyte chemotactic factor 2 Amyloidosis, and Apolipoprotein C-III is associated with Amyloidosis and other neurological diseases, including Alzheimer's disease.

Expression levels of hub genes in obesity. We applied Attie Lab Diabetes database to find the correlation between hub genes and diabetes; BTBR mice become severely diabetic with obesity at 10 weeks of age. We discovered that APOA1, APOB, and APOE expressions were expressed differently in different tissues in 4-week and 10-week BTBR obese diabetic mice (Figure 5). In the 4-weeks BTBR obese diabetic mice's adipose, APOA1 was downregulated, and APOE was downregulated in adipose and islet. However, in the 10-weeks BTBR obese diabetic mice's liver, APOA1 was significantly downregulated, while APOB and APOE were significantly increased in islet tissue.

Identification of potential predictive drugs. From DrugBank, we obtained ten drug-mRNA interaction pairs between the three key targets. The basic information and structural formula of drug are presented (Table 2). There were nine drugs regulating lipid metabolism and one antioxidant drug. Rosuvastatin, pitavastatin, and gemfibrozil

were approved for use, whereas gamolenic acid, lovastatin, and mipomersen were approved for use but require additional investigation.

### **GSEA revealed offspring's safety and potential health effects through IVF-ET.**

Functional differences between the two groups were determined from a macroscopic genome-wide perspective, not just DEGs. The most significantly enriched gene sets correlated with natural subjects were maturity-onset diabetes in young, glycine serine and threonine metabolism, fatty acid metabolism, and peroxisomes (Figure 6).

### **Discussion**

Since IVF affects the long-term health and safety of offspring, it has garnered increased attention in reproductive areas, indicating that it gradually became a research hotspot, even into the future. Based on bibliometric analysis results, we found that metabolism-related issues, including birth weight, are the focus of IVF offspring healthy. The question of whether IVF affects newborn birth weight remains uncertain. Some studies have indicated that IVF causes LBW<sup>[40]</sup>, some revealed that the IVF process does not affect fetal birth weight, and others indicated that frozen embryo transplantation leads to LGA fetus<sup>[41]</sup>. A multicenter cohort study identified significant differences in body weight among IVF centers with several confounding factors unrelated to the IVF process<sup>[42]</sup>. However, it is an indisputable fact that adverse intrauterine environments affect children's health, including metabolic diseases<sup>[43]</sup>. As a common cause of fetal growth restriction(FGR), LBW is a major public health problem having a higher risk of developing adipose tissue metabolic abnormalities, insulin resistance, and T2MD in adults compared to normal birth weight<sup>[44]</sup>.

Fatty acid metabolism is closely related to adult metabolic diseases; abnormal fatty acid metabolism leads to lipid deposition, obesity, hepatic insulin resistance, and glucose overproduction<sup>[45]</sup>. Our basic research has also demonstrated that excess palmitic acid (PA) enrichment in the decidua causes glutamine oxidation through TLR4/JNK/NF-kB pathway leading to decidual dysfunction and is associated with several adverse pregnancy outcomes, such as gestational diabetes mellitus, preeclampsia, and preterm delivery, and intrauterine growth restriction<sup>[46]</sup>. To avoid research bias, we need to study the impact of IVF from a global perspective, so we performed GSEA function analysis on all transcriptional genes from the villus of IVF and naturally conceived patients. Consistent with earlier findings, the IVF group demonstrated significant variations in metabolism, particularly fatty acid metabolism. Obviously, abnormal fatty acid metabolism is an important factor in abnormal embryonic development and increased long-term health risk following IVF treatment.

We examined IVF-induced DEGS and T2DM-related genes and discovered 15 co-genes that may contribute to long-term metabolic risk. Then, we investigated their properties and functions to better understand the mechanisms underlying the increased risk of metabolic disease in IVF offspring. Most selected genes were anticipated unsteady proteins, demonstrating that stability of chemicals is low in elaborate cell responses, and most studied proteins have negative GRAVY value, indicating that they are more hydrophilic<sup>[47]</sup>. Furthermore, the functional analysis of 15 differentially expressed genes revealed that they were mainly concentrated in PPAR signaling pathway, HIF-1 signaling pathway, vitamin digestion and absorption, and fat digestion and absorption pathway.

Lipid metabolism plays an important role in pregnancy outcomes and affects fetal growth and development. PPAR signaling pathway mainly regulates lipid metabolism in vivo and anti-inflammatory pathway, participates in maternal and fetal metabolic disorders, intrauterine pro-inflammatory environment and developmental defects, and reduces inflammation and insulin resistance in liver, skeletal muscle, and adipose tissue<sup>[48]</sup>. In the rat model of gestational diabetes mellitus (GDM), PPAR $\gamma$  expression in fetal rats was significantly different by sex, with elevated levels in the liver of male fetuses but not in female fetuses, explaining why there is considerable sexual dimorphism in adult IVF phenotypes<sup>[49]</sup>.

The 15 genes are also associated with lipid metabolism and diabetes. Slc2a1 (GLUT1) is mainly expressed in rodents, and human placenta between maternal-fetal glucose transport and is critical in placenta glucose transport. Excessive and exogenous endoplasmic reticulum stress may lead to abnormal placenta function or partly result in GLUT and change of blood vessel-related gene expression, leading to LBW<sup>[50]</sup>. The analysis of physicochemical characteristics, protein modification effects, and SNPs of these genes revealed SNPs of NOS3 rs1799983, APOE rs7412, and rs429358, which were previously reported to be related to diabetes<sup>[51]</sup>. ApoE polymorphisms are associated with complex metabolic changes, including abnormalities in several novel biomarkers associated with elevated cardiac metabolism and the risk of all-cause mortality<sup>[52]</sup>.

After analyzing these protein interaction networks, we discovered a gene complex composed of APOA1, APOB, and APOE. All three proteins are apolipoproteins, which mainly carry lipid substances and the structure of stable lipoproteins and participate in lipid metabolism and transformation. APOA1 is related to embryo quality, development, and survival. Studies have indicated that APOA1 expression is significantly increased in chorionic villi, decidua, and serum of patients with IVF and early abortion, and Verma's results also proved that APOA1 expression in the villi of IVF patients was increased<sup>[53]</sup>. Furthermore, APOE plays a major role in antiphospholipid functional attenuation (APL), inducing pregnancy complications<sup>[54]</sup>.

Subsequently, the expression of the three gene-related diseases in different tissues of diabetic mice was analyzed to verify our conclusion. Finally, we predicted the potential compounds targeted by the three genes, including rosuvastatin, pitavastatin, gemfibrozil, and simvastatin drugs for clinical use, as well as gamolenic acid, lovastatin, and mipomersen compounds were approved for clinical use and investigational use.

These are clinical studies, and there are significant variances between clinical individuals and no shared mechanisms. However, animal experiments have established that IVF mice do exhibit metabolic and cardiovascular abnormalities<sup>[55]</sup>. To ensure offspring health, it is vital to investigate the influence of IVF-ET on embryo metabolism and related mechanisms. ROS production in assisted reproduction technology not only leads IVF offspring to abnormal lipid metabolism but also plays an important role in the adverse health outcomes during later life. Our previous studies have implied that oxidative stress stimulation during IVF operation can lead to DNA damage in the fertilized egg, affecting offspring health<sup>[56,57]</sup>. Since fatty acids are important energy source for oocyte and embryo development, their abnormal metabolism can affect the normal development of embryo<sup>[58]</sup>. Oxidative stress leads to impaired embryo fatty acid oxidation (FAO) and increased lipid storage and impairs embryo fatty acid oxidation and mitochondrial function<sup>[59]</sup>. In addition, oocyte quality is a critical factor determining embryo quality, such as high lipid levels and ROS increase, resulting in poor oocyte and embryo development quality.

## **Conclusion**

Academics believe that studying embryo metabolism is critical since we now understand that the metabolic mechanism is complex, multi-molecular, multi-target, and networked. Therefore, we studied physicochemical properties, modifications, polymorphisms, and miRNA of 15 targets to determine the mechanisms underlying IVF treatment of increased risk of metabolic diseases in offspring. Indeed, embryo metabolism is a series of complex processes, which require additional research validation so that humans can have a deeper understanding of health problems associated with babies born through assisted reproduction and intervention means.

## **Declaration of Competing Interest**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## **Acknowledgements**

1. This manuscript has been released as a pre-print at Research Square, Lihong Liu, Siyao Ha, Zhiling Li, et al. A genome-wide association study to identify candidate genes for metabolic disorders in offspring by in vitro fertilization, 13 Aug, 2021, PREPRINT (Version 1), URL: <https://www.researchsquare.com/article/rs-717175/v1>. There is no conflict of interests regarding the publication of this article.
2. This study received funding from the following sources: the National Natural Science Foundation of China (No. 81871223, No. 31970798). The funding sources do not play any roles in interpreting the results or inferring the conclusions.

### **Author contributions**

Zhiling Li & MingQing Li: Conceptualization, Writing - Review & Editing. Validation. Lihong Liu & and Siyao HA: Preparation, Methodology, Visualization, Investigation, Software, Data Curation, Formal analysis, Writing- Original draft. Lihong LIU and Siyao HA contributed equally to this work.

### **References**

- 
- [1] Mani S, Ghosh J, Coutifaris C, et al. Epigenetic changes and assisted reproductive technologies[J]. *Epigenetics*, 2020, 15(1-2): 12-25.
  - [2] Wells J. Developmental plasticity as adaptation: adjusting to the external environment under the imprint of maternal capital[J]. *Philos Trans R Soc Lond B Biol Sci*, 2019, 374(1770): 20180122.
  - [3] Bal M H, Harlev A, Sergienko R, et al. Possible association between in vitro fertilization technologies and offspring neoplasm[J]. *Fertil Steril*, 2021, 116(1): 105-13.
  - [4] Cui L, Zhou W, Xi B, et al. Increased risk of metabolic dysfunction in children conceived by assisted reproductive technology[J]. *Diabetologia*, 2020, 63(10): 2150-7.
  - [5] Barbara L, Brown M B, Ethan W, et al. The risk of birth defects with conception by ART[J]. *Hum Reprod*, 2020, 36(1): 116-29.
  - [6] Chronopoulou, E, Harper, et al. IVF culture media: past, present and future[J]. *Hum Reprod Update*, 2015, 21(1): 39-55.
  - [7] Shai Levin, Eyal Sheiner, Tamar Wainstock, et al. Infertility Treatments and Long-Term Neurologic Morbidity of the Offspring[J]. *Am J Perinatol*, 2019, 36(9): 949-54.
  - [8] Kalra S K, Ratcliffe S J, Coutifaris C, et al. Ovarian Stimulation and Low Birth Weight in Newborns Conceived Through In Vitro Fertilization[J]. *Obstet Gynecol*, 2011, 118(4): 863-71.
  - [9] Sunkara S K, Antonisamy B, Redla A C, et al. Female causes of infertility are associated with higher risk of preterm birth and low birth weight: analysis of 117 401 singleton live births following IVF[J]. *Hum Reprod*, 2021, 36(3): 676-82.
  - [10] Kevin, Coetzee, Kemal, et al. Large-for-gestational age is male-gender dependent in artificial frozen embryo transfers cycles: a cohort study of 1295 singleton live births[J]. *Reprod Biomed Online*, 2019, 40(1): 134-41.
  - [11] Moen G H, Brumpton B, Willer C, et al. Mendelian randomization study of maternal influences on birthweight and future cardiometabolic risk in the HUNT cohort[J]. *Nat Commun*, 2020, 11(1): 5404.

- 
- [12] Morrison J L, Duffield J A, Muhlhausler B S, et al. Fetal growth restriction, catch-up growth and the early origins of insulin resistance and visceral obesity[J]. *Pediatr Nephrol*, 2010, 25(4): 669-77.
- [13] Zhao L, Zheng X, J Liu, et al. The placental transcriptome of the first-trimester placenta is affected by in vitro fertilization and embryo transfer[J]. *Reprod Biol Endocrinol*, 2019, 17(1): 50.
- [14] Thomas M M, Haghiac M, Grozav C, et al. Oxidative stress impairs fatty acid oxidation and mitochondrial function in the term placenta[J]. *Reprod Sci*, 2019, 26(7): 972-8.
- [15] Lvarez D, Muoz Y, Ortiz M, et al. Impact of Maternal Obesity on the Metabolism and Bioavailability of Polyunsaturated Fatty Acids during Pregnancy and Breastfeeding[J]. *Nutrients*, 2020, 13(1): 19.
- [16] Eck N, Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping[J]. *Scientometrics*, 2010, 84(2):523-38.
- [17] Yu W, Clyne M, Khoury M J, et al. Phenopedia and Genopedia: disease-centered and gene-centered views of the evolving knowledge of human genetic associations[J]. *Bioinformatics*, 2010,26(1): 145-6.
- [18] W Yu, M Clyne, M J Khoury W, et al. Phenopedia and Genopedia: disease-centered and gene-centered views of the evolving knowledge of human genetic associations[J]. *Bioinformatics*, 2010,26(1): 145-6.
- [19] Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources[J]. *Nat Protoc*. 2009;4(1): 44-57.
- [20] Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists[J]. *Nucleic Acids Res*. 2009;37(1): 1-13.
- [21] Zhou Y, Zhou B, Pache L, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets[J]. *Nat Commun*, 2019, 10(1): 1523.
- [22] Mi H, Dustin E, Anushya M, et al. PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API[J]. *Nucleic Acids Res*, 2020,49(D1): 394-403.
- [23] Minoru K, Miho F, Mao T, et al. KEGG: new perspectives on genomes, pathways, diseases and drugs[J]. *Nucleic Acids Res*, 2017,45(D1): 353-61.
- [24] Consortium U P. UniProt: a hub for protein information[J]. *Nucleic Acids Res*, 2015,43 (Database issue): 204-12.
- [25] Donna M, Jim O, Pruitt K D, et al. Entrez Gene: gene-centered information at NCBI[J]. *Nucleic Acids Res*, 2011, 39(Database issue): D52-7.
- [26] Elisabeth G, Alexandre G, Christine H, et al. ExPASy: the proteomics server for in-depth protein knowledge and analysis[J]. *Nucleic Acids Res*, 2003, 31(13): 3784-8.
- [27] Gupta R, Brunak S. Prediction of glycosylation across the human proteome and the correlation to protein function[J]. *Pac Symp Biocomput*, 2002, 310-22.
- [28] Blom N, Sicheritz-Ponten T, Gupta R, et al. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence[J]. *Proteomics*, 2004,4(6): 1633-49.
- [29] Jairo Navarro Gonzalez, Ann S Zweig, Matthew L Speir, et al. The UCSC Genome Browser database: 2021 update[J]. *Nucleic Acids Res*, 2021,49(D1): 1046-57.
- [30] Bailey T L, Boden M, Buske F A, et al. MEME SUITE: tools for motif discovery and searching[J]. *Nucleic Acids Res*, 2009 (Web Server issue): 202-8.
- [31] Agarwal V, Bell GW, Nam J, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs[J]. *Elife*, 2015, 4: e05005.
- [32] Szklarczyk D, Gable AL, Lyon D, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets[J]. *Nucleic Acids Res*, 2019, 47(D1): 607-13.
- [33] Gang S, Morris JH, Demchak B, et al. Biological Network Exploration with Cytoscape 3[J]. *Curr Protoc Bioinformatics*, 2014, 47:813.1-24.
- [34] Chin C H, Chen S H, Wu H H, et al. cytoHubba: identifying hub objects and sub-networks from complex interactome[J]. *BMC Systems Biology*, 2014, 8(S4): S11.

- 
- [35] Ba Der GD, Hogue C. An Automated Method for Finding Molecular Complexes in Large Protein Interaction Networks[J]. *BMC Bioinformatics*, 2003, 4:2.
- [36] Pletscher-Frankild S, A Pallegà, Tsafo K, et al. DISEASES: Text mining and data integration of disease–gene associations[J]. *Methods*, 2015, 74: 83-9.
- [37] Keller MP, Choi YJ, Ping W, et al. A gene expression network model of type 2 diabetes links cell cycle regulation in islets with diabetes susceptibility[J]. *Genome Res*, 2008, 18(5): 706-16.
- [38] Wishart DS, Feunang YD, An CG, et al. DrugBank 5.0: A major update to the DrugBank database for 2018[J]. *Nucleic Acids Res*, 2018, 46(D1): 1074-82.
- [39] Subramanian A, Tamayo P, Mootha V K, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles[J]. *Proc Natl Acad Sci U S A*, 2005, 102(43): 15545-50.
- [40] Pontesilli M, Hof M H, Ravelli A C J, et al. Effect of parental and ART treatment characteristics on perinatal outcomes[J]. *Hum Reprod*, 2021,36(6): 1640-65.
- [41] Anzola A B, Pauly V, Riviere O, et al. Monitoring IVF children birth weight is still a current issue and still a history of maternal factors[J]. *Reprod Biomed Online*, 2019, 39(6): 990-9.
- [42] Castillo C M, Johnstone E D, Greg H, et al. Associations of IVF singleton birthweight and gestation with clinical treatment and laboratory factors: a multicentre cohort study[J]. *Hum Reprod*, 2020,35(12): 2860-70.
- [43] Lesseur C, Chen J. Adverse Maternal Metabolic Intrauterine Environment and Placental Epigenetics: Implications for Fetal Metabolic Programming[J]. *Curr Environ Health Rep*, 2018, 5(4):531-43.
- [44] Broholm C, Ribel-Madsen R, Hjort L, et al. Epigenome- and Transcriptome-wide Changes in Muscle Stem Cells from Low-Birth-Weight Men[J]. *Endocr Res*, 2019, 45(1): 58-71.
- [45] Li A, Liu Q, Qiang L, et al. Berberine Reduces Pyruvate-driven Hepatic Glucose Production by Limiting Mitochondrial Import of Pyruvate through Mitochondrial Pyruvate Carrier 1[J]. *Ebiomedicine*, 2018, (34):243-55.
- [46] Si-Yao H, Qiu X M, Lai Z Z, et al. Excess palmitate induces decidual stromal cell apoptosis via the TLR4/JNK/NF- $\kappa$ B pathways and possibly through glutamine oxidation[J]. *Mol Hum Reprod*, 2020,26(2):88-100.
- [47] Heidari P, Ahmadizadeh M, Izanlo F, et al. In silico study of the CESA and CSL gene family in *Arabidopsis thaliana* and *Oryza sativa*: Focus on post-translation modifications[J]. *Plant Gene*, 2019, 19: 100189.
- [48] Jawerbaum A, Capobianco E. Effects of PPAR activation in the placenta and the fetus: implications in maternal diabetes[J]. *Placenta*, 2011, 32: S212-7.
- [49] Fornes D, White V, Higa R, et al. Sex-dependent changes in lipid metabolism, PPAR pathways and microRNAs that target PPARs in the fetal liver of rats with gestational diabetes[J]. *Mol Cell Endocrinol*, 2018, 461: 12-21.
- [50] Kawakami T, Yoshimi M, Kadota Y, et al. Prolonged endoplasmic reticulum stress alters placental morphology and causes low birth weight[J]. *Toxicol Appl Pharmacol*, 2014, 275(2):134-44.
- [51] Galal A A, Abd Elmajeed A A, Elbaz R A, et al. Association of Apolipoprotein E gene polymorphism with the risk of T2DM and obesity among Egyptian subjects[J]. *Gene*, 2021, 769: 145223.
- [52] Karjalainen J P, Mononen N, N Hutri-Kähönen, et al. The effect of apolipoprotein E polymorphism on serum metabolome – a population-based 10-year follow-up study[J]. *Sci Rep*, 2019, 9(1): 458.
- [53] Verma P, Nair R R, Singh S, et al. High Level of APOA1 in Blood and Maternal Fetal Interface Is Associated with Early Miscarriage[J]. *Reprod Sci*, 2019, 26(5): 649-56.
- [54] Ulrich, Victoria, Gelber, et al. ApoE Receptor 2 Mediation of Trophoblast Dysfunction and Pregnancy Complications Induced by Antiphospholipid Antibodies in Mice[J]. *Arthritis Rheumatol*, 2016,68(3):730-9.
- [55] Feuer SK, Xiaowei Liu, Annemarie D, et al. Use of a Mouse In Vitro Fertilization Model to Understand the Developmental Origins of Health and Disease

- Hypothesis[J]. *Endocrinology*, 2014,155(5): 1956-69.
- [56] Qian D, Li Z, Zhang Y, et al. Response of Mouse Zygotes Treated with Mild Hydrogen Peroxide as a Model to Reveal Novel Mechanisms of Oxidative Stress-Induced Injury in Early Embryos[J]. *Oxid Med Cell Longev*,2016,1521428.
- [57] Huang Y, Ha S, Li Z, et al. CHK1-CENP B/MAD2 is associated with mild oxidative damage-induced sex chromosome aneuploidy of male mouse embryos during in vitro fertilization[J]. *Free Radic Biol Med*, 2019, 137: 181-93.
- [58] Ishigaki M, Hashimoto K, Sato H, et al. Non-destructive monitoring of mouse embryo development and its qualitative evaluation at the molecular level using Raman spectroscopy[J]. *Sci Rep*, 2017, 7(1): 43942.
- [59] Thomas M M, Haghiac M, Grozav C, et al. Oxidative Stress Impairs Fatty Acid Oxidation and Mitochondrial Function in the Term Placenta[J]. *Reprod Sci*, 2018,26(7): 972-78.

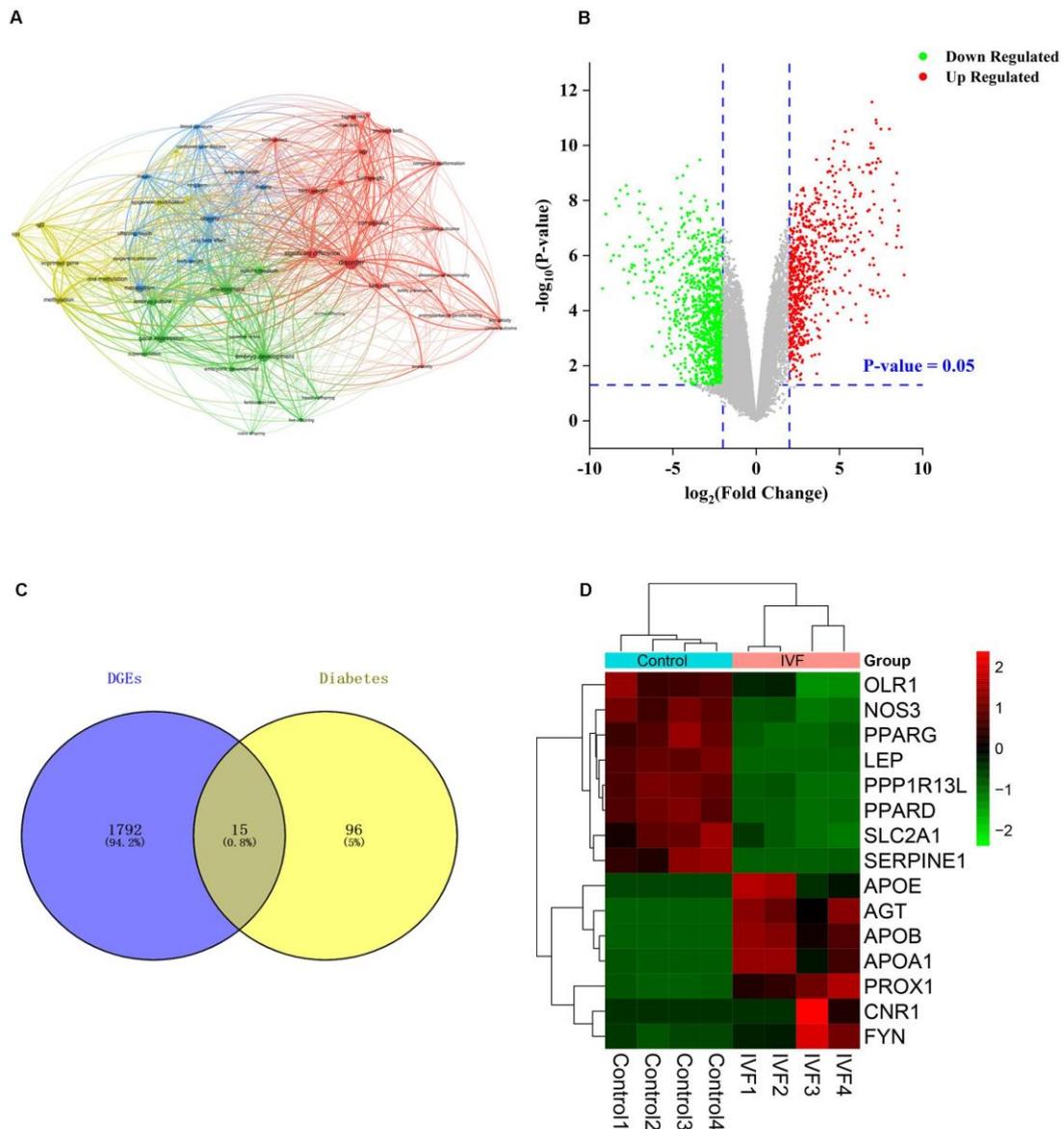
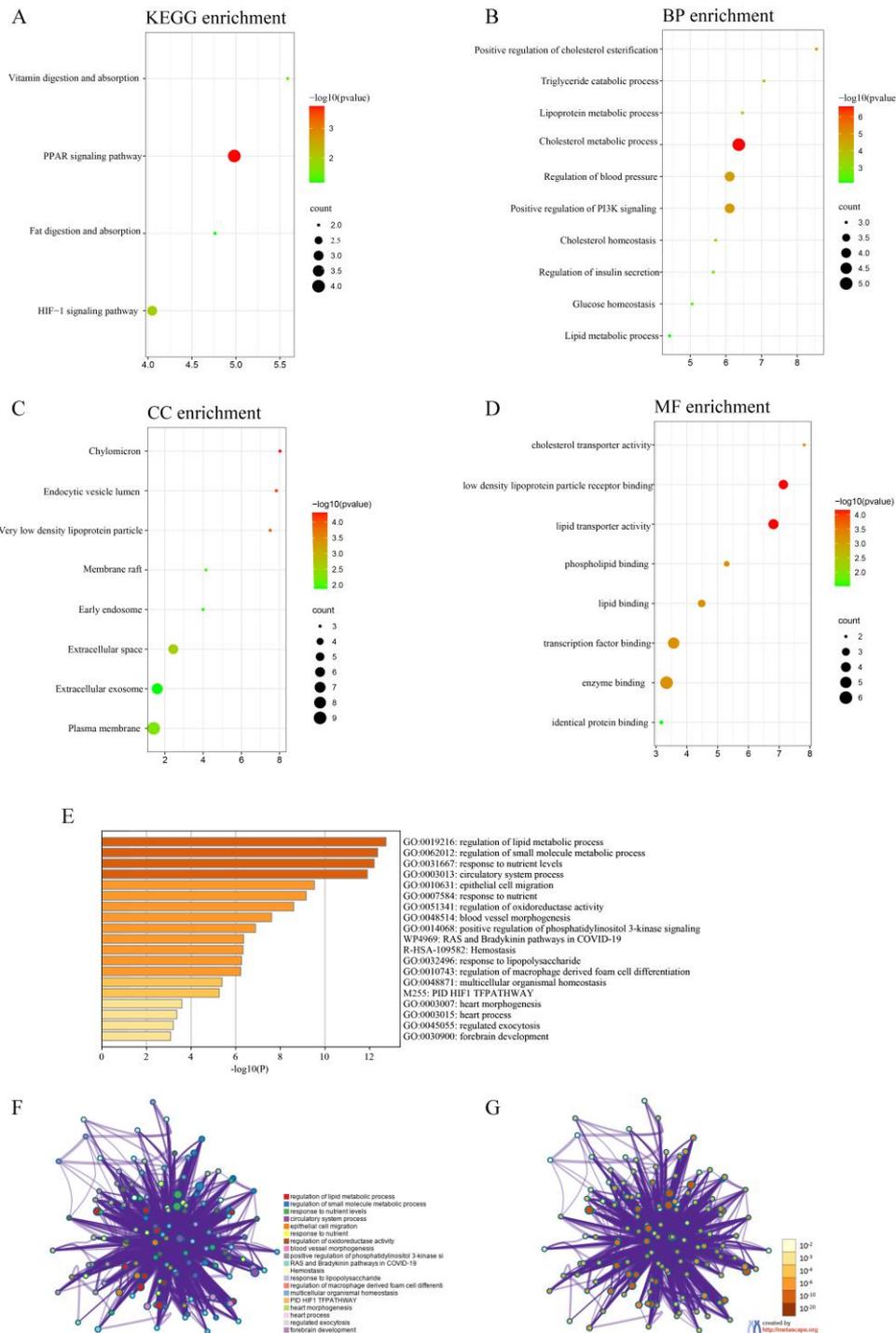
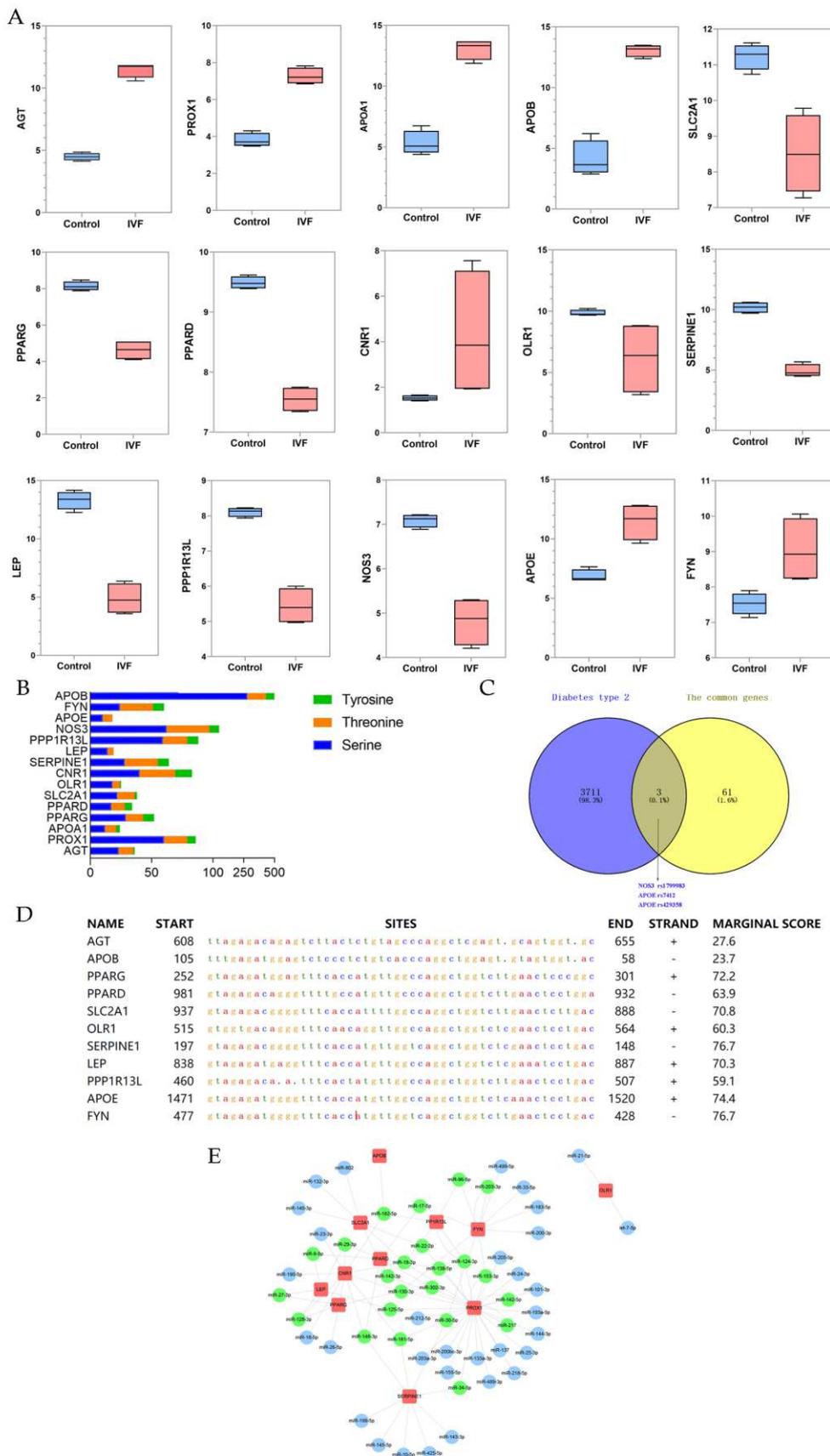


Figure 1. Bibliometric analysis of safety and health of children born through IVF and identification of key genes. (A) Keywords extraction from papers and cited documents, the node size is positively correlated with keyword frequency. (B) IVF and natural conception of different genes' volcano. (C) Differential genes and T2DM-associated genes' Venn map. (D) The key genes' expression heatmap.



---

Figure 2. Target genes' functional enrichment analysis (A) KEGG pathway analysis. (B) Biological processes enrichment. (C) Cellular component enrichment. (D) Molecular function enrichment. (E-G) Metascape biological functions enrichment.



Figures 3. (A) The expression values of 15 key genes in gene microarray. Blue: natural conception

group, Red: IVF treatment group. (B) The phosphorylation site was predicted according to amino acid sequence of selected genes. (C) The prediction of SNP sites of studied genes and T2DM. (D) The position and sequence of a conserved motif in the promoter sequence of selected genes. (E) The prediction of miRNA targets of studied genes.

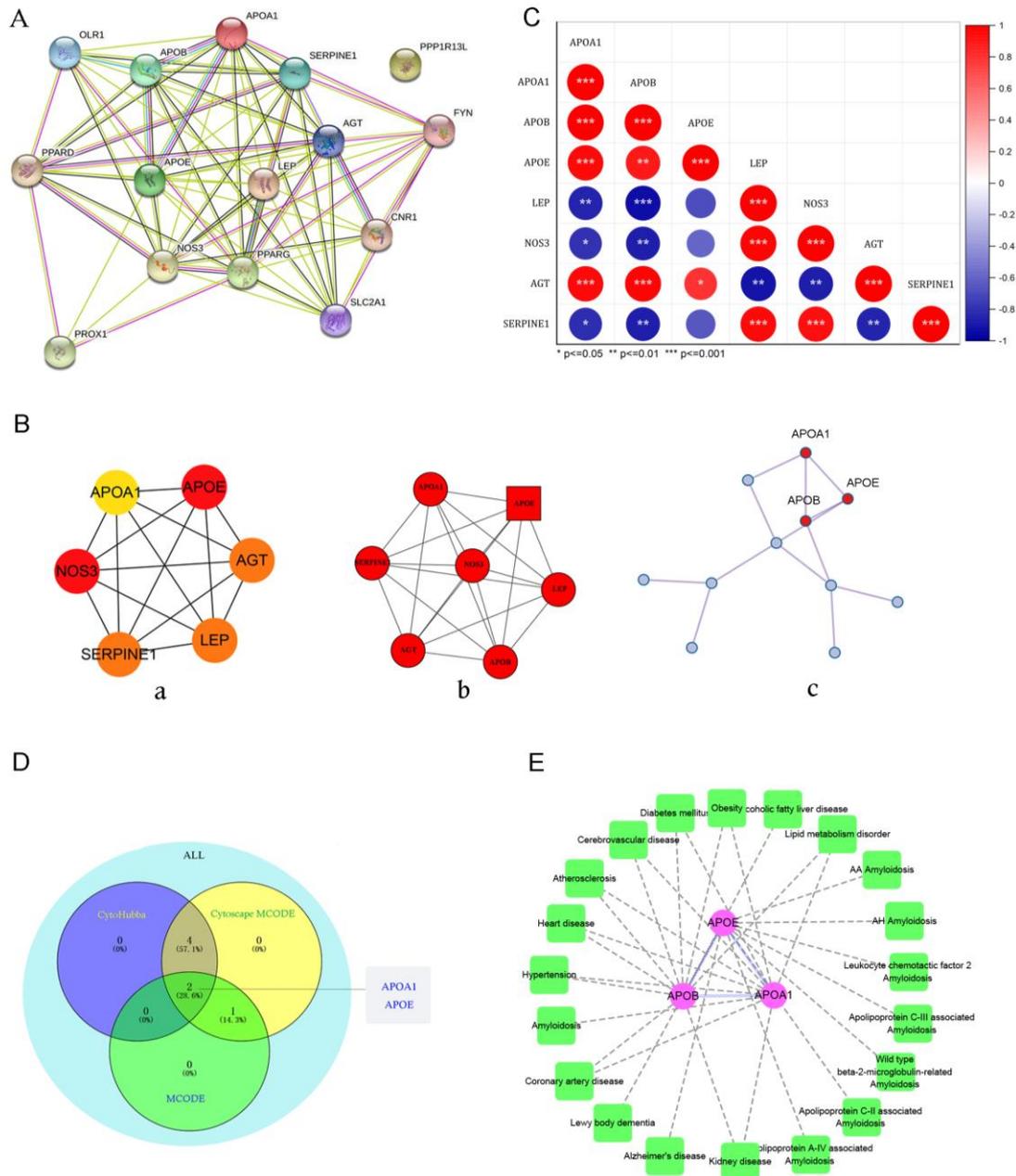


Figure 4. Key genes network analysis (A) The common gene protein-protein interaction network, each node represents a protein, and the lines represent interactions. (B) a. Analysis of key node network by cytohubba plugin. b. Analysis of key node network by MCODE cytohubba plugin. c. Metascape online analysis of key gene complexes in the network. (C) Correlation of gene expression, color is positively correlated with correlation, red (positive correlation), blue (negative

correlation). (D) Network analysis Venn diagram. (E) Gene-disease interaction networks.

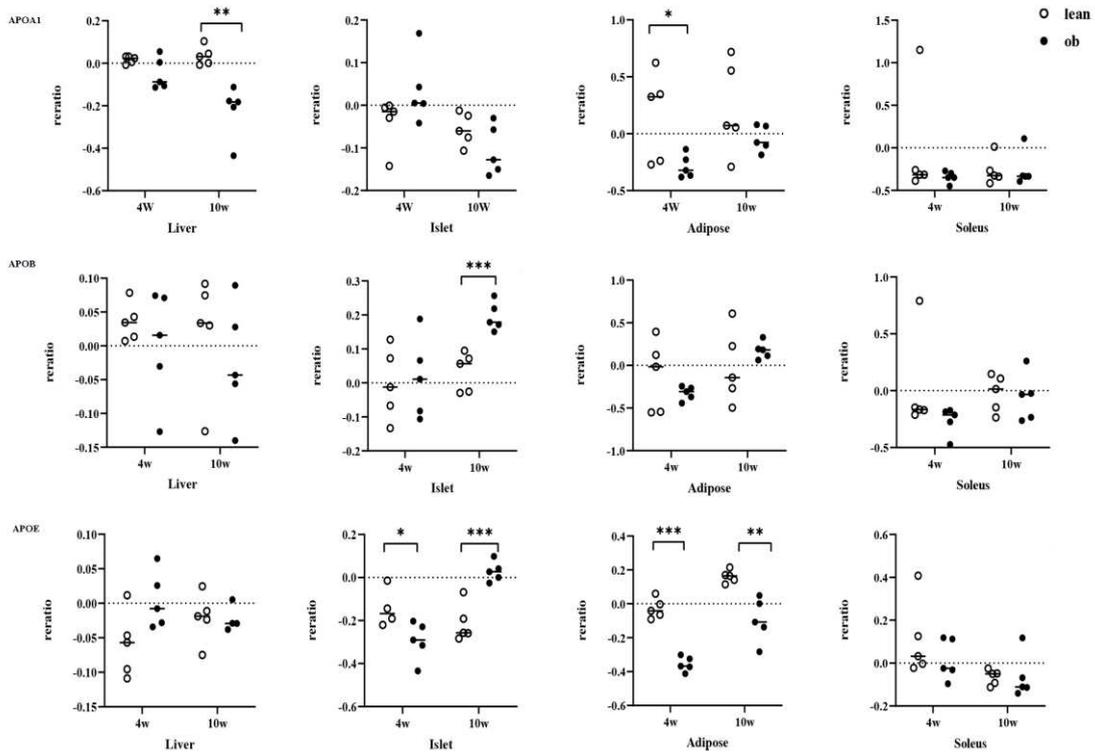
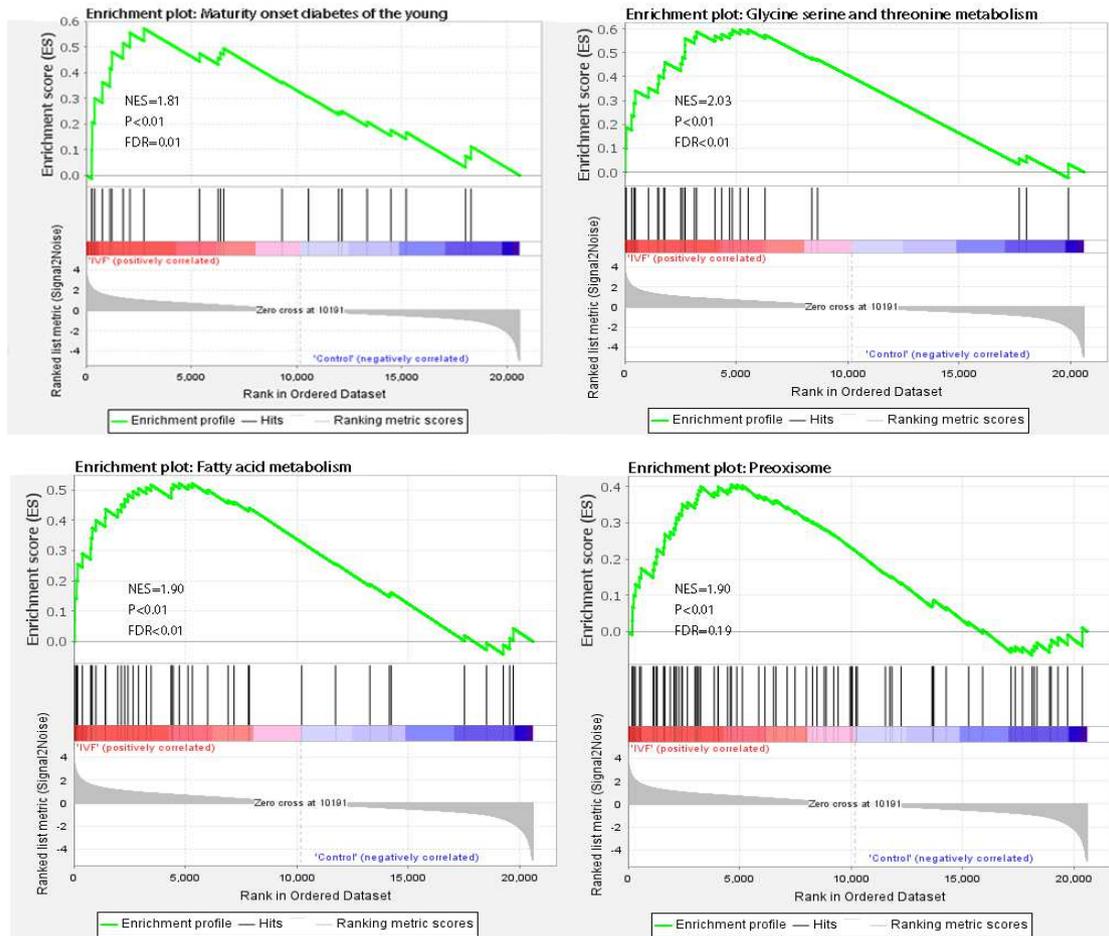


Figure 5. The key gene expression in the 4-weeks and 10-weeks BTBR obese diabetic mice different tissues (Liver, Islet, Adipose, and Soleus).



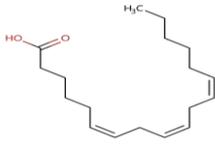
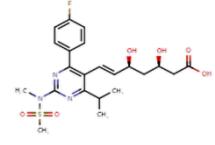
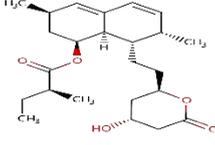
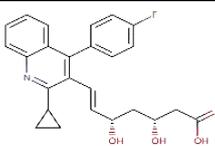
Figures 6. GSEA-based KEGG-enrichment plots of representative gene sets for activated pathway from GSE122214 data series: maturity-onset diabetes of the young; glycine serine and threonine metabolism; fatty acid metabolism; peroxisome.

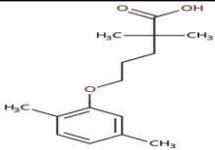
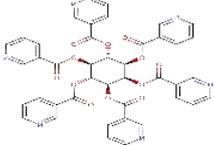
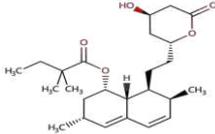
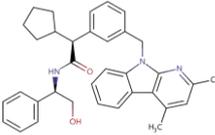
Table 1. The properties of common genes between IVF and T2DM (hg38)

Gene name	Chr.	Length (aa)	MW (kDa)	PI	Instability index	GRAVY	N-glycosylation site
AGT	Chr.01	485	53.15	5.87	Unstable 41.16	0.065	4(47,170,304,328)
PROX1	Chr.01	737	83.20	6.74	Unstable 64.99	-0.757	2(131,273)
APOA1	Chr.11	267	30.78	5.56	Unstable 40.83	-0.717	-
APOB	Chr.02	4563	515.60	6.58	Stable 35.84	-0.296	10(34,185,456,983 ...)

PPARG	Chr.03	505	57.62	5.61	Unstable 50.20	-0.308	1(24)
PPARD	Chr.06	441	49.90	7.53	Unstable 51.90	-0.351	1(24)
SLC2A1	Chr.01	492	54.08	8.93	Stable 36.57	0.534	2(45,411)
OLR1	Chr.12	273	30.96	6.94	Unstable 42.23	-0.399	4(73,139,183,210)
CNR1	Chr.06	472	52.86	8.47	Stable 39.74	0.329	4(77,83,112,389)
SERPINE1	Chr.07	402	45.06	6.68	Stable 35.33	-0.059	4(232,288,352,388)
LEP	Chr.07	167	18.64	5.88	Unstable 47.52	0.131	-
PPP1R13L	Chr.19	828	89.09	6.37	Unstable 72.87	-0.655	4(251,303,672,704)
NOS3	Chr.07	1203	67.86	8.58	Unstable 50.54	-0.311	-
APOE	Chr.19	317	36.15	5.65	Unstable 44.88	-0.596	-
FYN	Chr.06	537	60.76	6.23	Stable 36.41	-0.489	3(23,68,113)

Table 2. Drugs and structures for key targets in DrugBank database

DrugBank ID	Drug	Groups	Gene	Structure
DB13854	Gamolenic acid	Approved, Investigational	APOE	
DB01098	Rosuvastatin	Approved	APOB	
DB00227	Lovastatin	Approved, Investigational	APOB	
DB08860	Pitavastatin	Approved	APOB	

DB01241	Gemfibrozil	Approved	APOB	
DB05528	Mipomersen	Approved, Investigational	APOB	
DB08949	Inositol nicotinate	Approved, Withdrawn	APOB	
DB00641	Simvastatin	Approved	APOB	
DB04852	Implitapide	Investigational	APOB	
DB01094	Hesperetin	Experimental	APOB	