Integrated Analysis of Multiple Microarrays Based on Raw Data Identified Novel Gene Signatures in Recurrent Implantation Failure

Hong Zeng[1,2], Yu Fu[2], Lang Shen[2]* and Song Quan[2]*

[1] Department of Reproductive Medicine Center, Foshan Maternal and Child Health Care Hospital, Southern Medical University, Foshan, China, [2] Department of Gynecology and Obstetrics, NanFang Hospital, Southern Medical University, Guangzhou, China

Background: Recurrent implantation failure (RIF) is an intricate complication following IVF-ET, which refers to the situation that good-quality embryos repeatedly fail to implant following two or more IVF cycles. Intrinsic molecular mechanisms underlying RIF have not yet been fully elucidated. With enormous improvement in high-throughput technologies, researchers screened biomarkers for RIF using microarray. However, the findings of published studies are inconsistent. An integrated study on the endometrial molecular determinants of implantation will help to improve pregnancy outcomes.

Objective: To identify robust differentially expressed genes (DEGs) and hub genes in endometrium associated with RIF, and to investigate the diagnostic role of hub genes in RIF.

Methods: Raw data from five GEO microarrays regarding RIF were analyzed. Integrated genetic expression analyses were performed using the Robust Rank Aggregation method to identify robust DEGs. Enrichment analysis and protein-protein interaction (PPI) analysis were further performed with the robust DEGs. Cytohubba was used to screen hub genes based on the PPI network. GSE111974 was used to validate the expression and diagnostic role of hub genes in RIF.

Results: 1532 Robust DEGs were identified by integrating four GEO datasets. Enrichment analysis showed that the robust DEGs were mainly enriched in processes associated with extracellular matrix remodeling, adhesion, coagulation, and immunity. A total of 18 hub genes (HMGCS1, SQLE, ESR1, LAMC1, HOXB4, PIP5K1B, GNG11, GPX3, PAX2, TF, ALDH6A1, IDH1, SALL1, EYA1, TAGLN, TPD52L1, ST6GALNAC1, NNMT) were identified. 10 of the 18 hub genes were significantly differentially expressed in RIF patients as validated by GSE111974. The 10 hub genes (SQLE, LAMC1, HOXB4, PIP5K1B, PAX2, ALDH6A1, SALL1, EYA1, TAGLN, ST6GALNAC1) were effective in predicting RIF with an accuracy rate of 85%, specificity rate of 100%, and sensitivity rate of 88.9%.

Conclusions: Our integrated analysis identified novel robust DEGs and hub genes in RIF. The hub genes were effective in predicting RIF and will contribute to the understanding of comprehensive molecular mechanisms in RIF pathogenesis.

INTRODUCTION

Implantation failure is the major limiting step in *in-vitro* fertilization and embryo transfer (IVF-ET) success. Approximately 15% of patients experience recurrent implantation failure (RIF) following IVF-ET. RIF refers to the situation that good-quality embryos repeatedly fail to implant following two or more ET cycles (1). Multifactorial pathogenesis plays a role in RIF, including embryo factors, endocrine factors, immunological problems, thrombotic conditions, uterine anomalies, genetic disorders, and endometrial factors (2). While the endometrial factor is one of the leading factors that account for RIF (3). However, intrinsic endometrial molecular mechanisms underlying RIF have not yet been fully elucidated. Endometrial gene expression profile may be disrupted in patients experiencing RIF (4). The value of identifying genetic biomarkers predictive of RIF is important as this would guide prognosis and inform potential therapeutic intervention for RIF. Increasing studies are investigating the endometrial gene expression profiles by microarray or RNA sequencing to identify biomarkers of prediction or diagnosis for RIF. Biomarkers including both coding genes and non-coding RNAs which are involved in implantation failure have been discovered and help to promote our understanding of pathogenesis underlying RIF (4–12). Nowadays, a tremendous amount of high-throughput data is piling up in public databases. However, results from the studies are inconsistent partly due to the small sample size in most studies or different platforms, or different bioinformatic methods. Most of these studies does not subject to validation on an independent cohort. Thus, there is an urgent need for timely identification and validation of more robust biomarkers for RIF, to improve pregnancy outcomes following IVF.

Biomarkers that are identified from a single study often appear to be biologically irrelevant or false positives. Meta-analysis allows integrating data from multiple studies to identify biomarkers across multiple conditions. Its main advantage is to boost power by increasing sample size and being able to catch signals that are small but consistent. Finding robust biomarkers for RIF is important for RIF prediction, diagnosis, and treatment. Researchers have characterized genetic cooperation and regulation through the menstrual cycle progression and characterized the genetic profiles for the acquisition of endometrial receptivity for a successful pregnancy (13). Recently, a panel of endometrial biomarkers acquired by endometrial receptivity test (ERT) was developed to accurately predict the window of implantation (WOI) and significantly improve the pregnancy outcomes of patients with RIF. These studies indicating the clinical potential of finding RIF related genetic profiles (14).

The purpose of the present study is to collect the available transcriptional microarray datasets from inconsistent measurements among various studies and perform an integrated analysis by robust rank aggregation (RRA) method (15) to generate more stable and robust DEGs. We perform protein-protein interaction analysis based on the robust DEGs to identify hub genes that may contribute to RIF, then validate the hub genes for prediction of RIF. Our study may help understand the mechanisms underlying RIF pathogenesis and promote the development of effective therapeutic targets of RIF.

MATERIALS AND METHODS

Data Collection

To identify gene expression data regarding recurrent implantation failure, we use the search term "implantation failure" to search gene expression datasets in Gene Expression Omnibus (GEO) database-series (https://www.ncbi.nlm.nih.gov/geo/browse/), The species are limited to humans. The searching keywords were "implantation homo". Datasets that met the following inclusion criteria were included: (1) Gene expression profile by array; (2) The sample is the endometrium during the window of implantation; (3) Recurrent implantation patients and fertile controls were contained in one experiment; (4) The sample size is at least ten, with at least five patients in each group. (5) Raw data were available in GEO; (6) Chip platforms were from "Agilent" or "Affymetrix" or "Illumina". By searching the GEO database, we identified 42 records in total. 24 records were excluded on reading titles and summaries, seven records were excluded on reading study designs. 11 GSEs were assessed for eligibility. Six GSEs were further excluded for the following reasons: duplicate samples and data (n=1, GSE71835); miRNA arrays (n=2, GSE71332, GSE108966); cirRNA array (n=1, GSE147442); blood samples (n=1, GSE106307); sample size < 10 (n=1, GSE103465). Finally, five GSEs (GSE26787, GSE92324, GSE111974, GSE58144, GSE71331) were included for statistical analysis. Among the five GSEs, four were included for RRA analysis, one was included in the validation and prediction process. The diagram of selected studies was shown in **Figure 1**. The following information were extracted from each identified GSE: GEO accession number, platform, sample size, array type, tissue, biopsy time, year, and country.

Quality Control

The raw data were retrieved from the GEO database. Quality controls (QC) were conducted using the R-package ArrayQualityMetrics (16). Raw intensity signals (*.CEL files,

FIGURE 1 | The diagram of selecting GSE datasets.
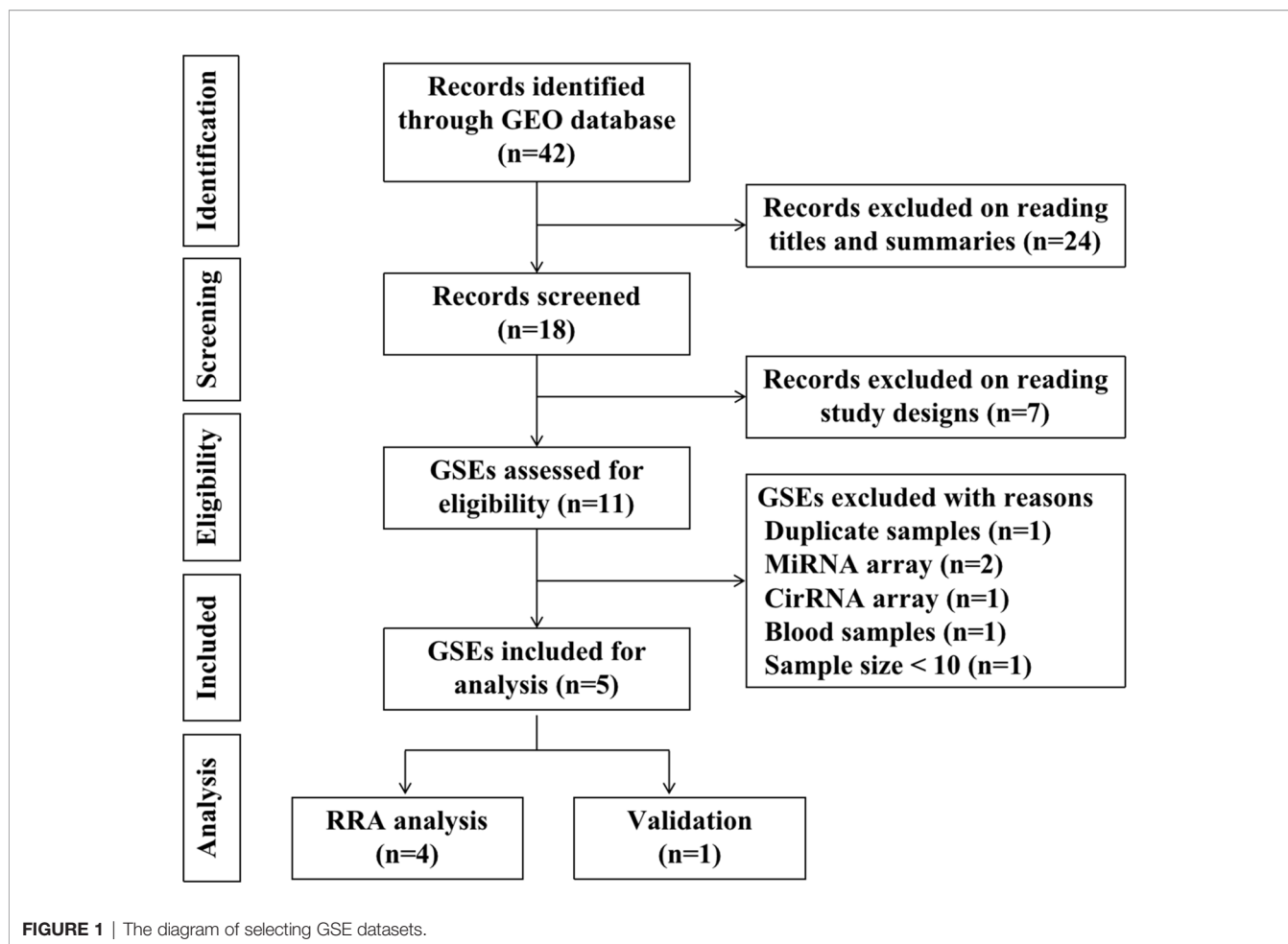
*.txt files, *.idat files) were extracted and normalized using the R-package affy and vsn (17, 18). The datasets were annotated with the R Bioconductor packages hgu133plus2.db, hugene10sttranscriptcluster.db, and hgu133a.db, depending on the platform. We excluded non-annotated probes. The batch effect was removed by removeBatchEffect function in the limma package if the batch effect was significant (19). When multiple probes were mapped to the same gene, those probes were averaged. All codes run under the R environment 4.1.0

Differentially Expressed Genes (DEG) Screening

The "limma" R package (19) was used to screen the DEGs between the RIF patients and the fertile control patients. The significantly DEGs in each GSEs were identified by p-value < 0.05 and |log2 fold change (FC)| >1. DEGs were ranked by p-value and |log2 FC|.

RRA Analysis of DEGs From Different GSEs

The "RobustRankAggregation" R package (15) was used to integrate the ranked gene lists of DEGs to find the robust DEGs. The genes with adjusted p-value < 0.05 in the RRA analysis were considered as

robust DEGs. For RRA analysis, we first analyzed the DEGs of each GSE dataset using the limma R package (19). Each DEG list was divided into the up-regulated gene list ($log_2FC > 0$) and the down-regulated gene list ($log_2FC < 0$). DEGs (both up-regulated and down-regulated) of each dataset were ranked according to the p-value and $|log_2FC|$. RRA for up-regulated genes and down-regulated genes was performed respectively. RRA is an effective tool to integrate multiple arrays outcomes. However, considering different array platforms, the input ranked gene lists only include the intersection of genes from the five GSEs. Then, all the DEGs were scored according to the ranked list and aggregately analyzed using the "Robust Rank Aggregation" R package. The final adjusted p-value in this method reflects the probability of the highly ranked genes in the datasets were identified as robust DEGs. We use the aggregateRanks() function to perform the RRA analysis. The parameters were set as the following: method="RRA", full=T, exact=T, topCutoff=NA. Details of the Robust Rank Aggregation (15) is as the following. For instance. let n be the number of experiments or GSEs and m be the number of genes studies. Assume that in DEG analysis of each GSEs genes are ordered according to their impact so that the genes locate in the beginning of the list are most likely to behave biologically functional. The rank of a gene is just the position in this ordering. If we divide ranks by the

maximal rank value m, we obtained normalized ranks with the maximal value of 1. for each gene let the corresponding rank vector $r = (r_1, …, r_n)$ be such that r_j denotes the normalized rank of the gene in the j-th preference list. To find the genes that are highly ranked in m studied ranked lists. We assume that all that all informative normalized ranks come from a distribution, which is strongly skewed toward zero and our task is to detect these distributions. For any normalized rank vector r, let $r_1…, r_n$ be a reordering of r such that $r_{(1)} \leq … \leq r_{(n)}$ Then we can ask how probable it is to obtain $\hat{r}_{(k)} \leq \cdots \leq \hat{r}_{(k)}$ when the rank vector $\hat{r}$ is generated by the null model, i.e. all ranks $\hat{r}_j$ are sampled from uniform distribution. Let $\beta_{k,n}(r)$ denote the probability that $\hat{r}_{(k)} \leq \hat{r}_{(k)}$. Then under the null model the probability that the order statistic $\hat{r}_{(k)}$ is smaller or equal to x can be expressed as a binomial probability

$$\beta_{k,n}(x) := \Sigma_{\ell=k}^{n} \binom{n}{\ell} x^\ell (1-x)^{n-\ell}, \qquad (1)$$

since at least k normalized rankings must be in the range $[0,x]$. Alternatively, $\beta_{k,n}(x)$ can be expressed through a beta distribution, as $\hat{r}_{(k)}$ is the order statistic of n independent random variables uniformly distributed over the range $[0,1]$. Since the number of informative ranks is not known, we define the final score for the rank vector r as the minimum of P-values.

$$\rho(r) = \min_{k=1,…,n} \beta_{k,n}(r) \qquad (2)$$

The ρ score is a minimum of $\beta_{n,k}(r)$ scores, each of which is a p-value measuring deviance of the k-th order statistic $\hat{r}_{(k)}$ from its expected distribution. If the null hypothesis holds, then all the $\beta_{n,k}(r)$ values follow uniform distribution. If these values would be independent then the distribution of p scores would be $Beta(1,n)$. We calculate the exact p-values based on p scores distribution. We want to calculate probability in the form

$$Pr\,[X \leq \rho] \qquad (3)$$

Using the definition of rho we can write

$$Pr\,[X \leq \rho] = 1 - Pr\,[X \geq \rho]$$
$$= 1 - Pr[\beta_{n,1}(\hat{r}) \geq \rho, …, \beta_{n,n}(\hat{r}) \geq \rho]$$
$$= 1 - Pr\,[\hat{r}_{(1)} \geq \beta_{n,1}^{-1}(\rho), …, \hat{r}_{(n)} \geq \beta_{n,n}^{-1}(\rho)]$$
$$= 1 - Pr\,[1 - \hat{r}_{(1)} \leq 1 - \beta_{n,1}^{-1}(\rho), …, 1 - \hat{r}_{(n)} \leq 1 - \beta_{n,n}^{-1}(\rho)],$$

where $\hat{r}$ is sample from uniform distribution with size n and $\beta_{n,k}^{-1}(\rho)$ is a quantile of $Beta(k, n - k + 1)$ distribution. Therefore, we can write

$$Pr\,[X \leq \rho] = 1 - Pr[\hat{r}_{(1)} \leq 1 - \beta_{n,n}^{-1}(\rho), …, \hat{r}_{(n)}$$
$$\leq 1 - \beta_{n,1}^{-1}(\rho)] \qquad (4)$$

The probability in (4) is the exact p-value.

Gene Ontology (GO) and KEGG Pathway Enrichment Analysis

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis were performed by the "clusterProfiler" R package (20). GO terms or KEGG pathways were visualized by the "ggplot2" R packages.

Protein-Protein Interaction (PPI) Network Analysis

The RRA identified robust DEGs (exact p-value < 0.01) was employed to construct the PPI network. We uploaded the robust DEG list into the STRING database (http://www.string-db.org/), the confidence interaction score was set as 0.7 in the PPI analysis. The network was imported into the Cytoscape (version 3.8.2) (21) for visualization and further identification of hub genes. Hub genes are usually of functional importance and highly interconnected with other genes. We use the Cytohubba plug-in (22) to screen candidate hub genes. The scores of the genes were calculated by MMC, DMNC, EPC, and Degree methods. The genes were then ranked by scores according to MCC, DMNC, EPC, and Degree, respectively. The overlap of the top 100 ranked genes identified by Cytohubba following four methods and the top 100 robust DEGs were determined as the hub genes.

Hub Genes Validation

We further conducted validation of hub genes by using the data from an Illumina microarray dataset (GSE111974) which consists of 24 RF patients and 24 fertile controls (8). The raw data were analyzed using the limma package (19). The background was corrected and the data was normalized. P-value <0.05 were considered to be significantly different.

The Diagnostic Role of Hub Genes in RIF

We use the dataset GSE111974 (8) to investigate the diagnostic role of hub genes in RIF. The dataset was divided into the training set (60%) and the validation set (40%) by randomization with the seed number set as 1234. To verify the diagnostic role of hub genes identified by RRA and PPI analysis, we construct a prediction model using the differentially expressed hub genes by the generalized multivariate regression with the training set. First, we supply the training set into algorithm to construct the model. Second, we predict the markers of our validation set. Third, we calculate the number of correct and incorrect predictions on the validation dataset to assess the model's prediction precision and calculate the accuracy rate, sensitivity rate, and specificity rate. Fourth. We performed the Receiver Operating Characteristic (ROC) analysis to detect the Area Under the Curve (AUC).

Statistical Analysis

Continuous variables are presented as mean ± standard deviation (SD) for normally distributed data, or as median and interquartile range. Normally distributed data were compared using the Student's t-test and non-normally distributed data using the Mann-Whitney U test. A p-value < 0.05 was considered to be statistically significant. All analyses were performed using R software (version 4.1.0). The R scripts and the corresponding Rdata file were deposited in our Github repositories (https://github.com/minizenghong/Genetic-meta-analysis-on-RIF).

RESULTS

Characteristics of the Included GSE Datasets From the GEO Database

A total of five microarray datasets (GSE26787, GSE92324, GSE111974, GSE58144, GSE71331) met the inclusion criteria were included for subsequent analysis. There are 91 RIF patients and 114 control patients in the five datasets. The characteristics of the datasets included in this study were listed in **Table 1**. GSE26787, GSE92324, GSE58144, and GSE71331 were included in the RRA analysis. GSE111974 was selected for subsequent validation of hub genes and to investigate the diagnostic role of hub genes in RIF. The diagram of the study was shown in **Figure 1**.

Identification of DEGs of Each GSE

We performed quality control, background correction, and normalization with the raw data from each GSE data set. Then DEGs of each GSE included in the RRA analysis were screened using the limma R package. The significantly DEGs were selected by p-value < 0.05 and |log2 FC| > 1. DEGs were ranked by p-value and |log2 FC|. The volcano plots showing DEGs of each GSE were shown in **Figure 2A**. The DEG list of each GSE dataset was shown in **Supplementary Table 1**. In summary, 121, 343, 1045 and 128 genes were significantly up-regulated in GSE26787, GSE92324, GSE58144, GSE71331, respectively. 156, 179, 1168 and 46 genes were significantly down-regulated in GSE26787, GSE92324, GSE58144, GSE71331, respectively.

Integrative Analysis of DEGs From Different GSEs

GSE26787, GSE92324, GSE58144, GSE71331 were included in the RRA analysis. 1532 significantly robust DEGs were identified by the RRA analysis with the criteria of adjusted p-value<0.05. 438 significantly robust DEGs were identified by the RRA analysis with the criteria of adjusted p-value<0.01. The heatmap of the top 25 up-regulated and 25 down-regulated genes were shown in **Figure 2B**. The result of the RRA analysis was shown in **Supplementary Table 2**.

Gene Ontology (GO) and KEGG Pathway Enrichment Analysis

We uploaded the 1532 robust DEGs to perform the GO and KEGG enrichment analysis. GO enrichment analysis showed that extracellular matrix organization, extracellular structure organization, external encapsulating structure organization, positive regulation of cell adhesion, cell-substrate adhesion,

substrate adhesion-dependent cell spreading, blood coagulation, hemostasis, cellular modified amino acid metabolic process, and coagulation were the top 10 enriched GO terms in biological process (BP) (**Figure 3**). Extracellular matrix structural constituent, heparin binding, actin binding, protease binding, phospholipid binding, extracellular matrix binding, protein tyrosine kinase activity, glycosaminoglycan binding, phospholipase activity, DNA-binding transcription activator activity were the top 10 enriched GO terms in molecular function (MF) (**Figure 3**). Collagen-containing extracellular matrix, endoplasmic reticulum lumen, platelet alpha granule, cell cortex, vesicle lumen, cytoplasmic vesicle lumen, high-density lipoprotein particle, secretory granule lumen, cell-cell junction, platelet alpha granule lumen were the top 10 enriched GO terms in cell component (CC) (**Figure 3**). KEGG enrichment analysis showed that complement and coagulation cascades, Human papillomavirus infection, NF-kappa B signaling pathway, Toxoplasmosis, PI3K-Akt signaling pathway, Osteoclast differentiation, Rap1 signaling pathway, ECM-receptor interaction, TNF signaling pathway, Human T-cell leukemia virus 1 infection were the top 10 enriched KEGG pathways (**Figure 3**). In summary, the robust DEGs were mainly enriched in processes associated with extracellular matrix remodeling, adhesion, coagulation, and immunity. The results of GO and KEGG pathway enrichment analysis was shown in the **Supplementary Table 3**.
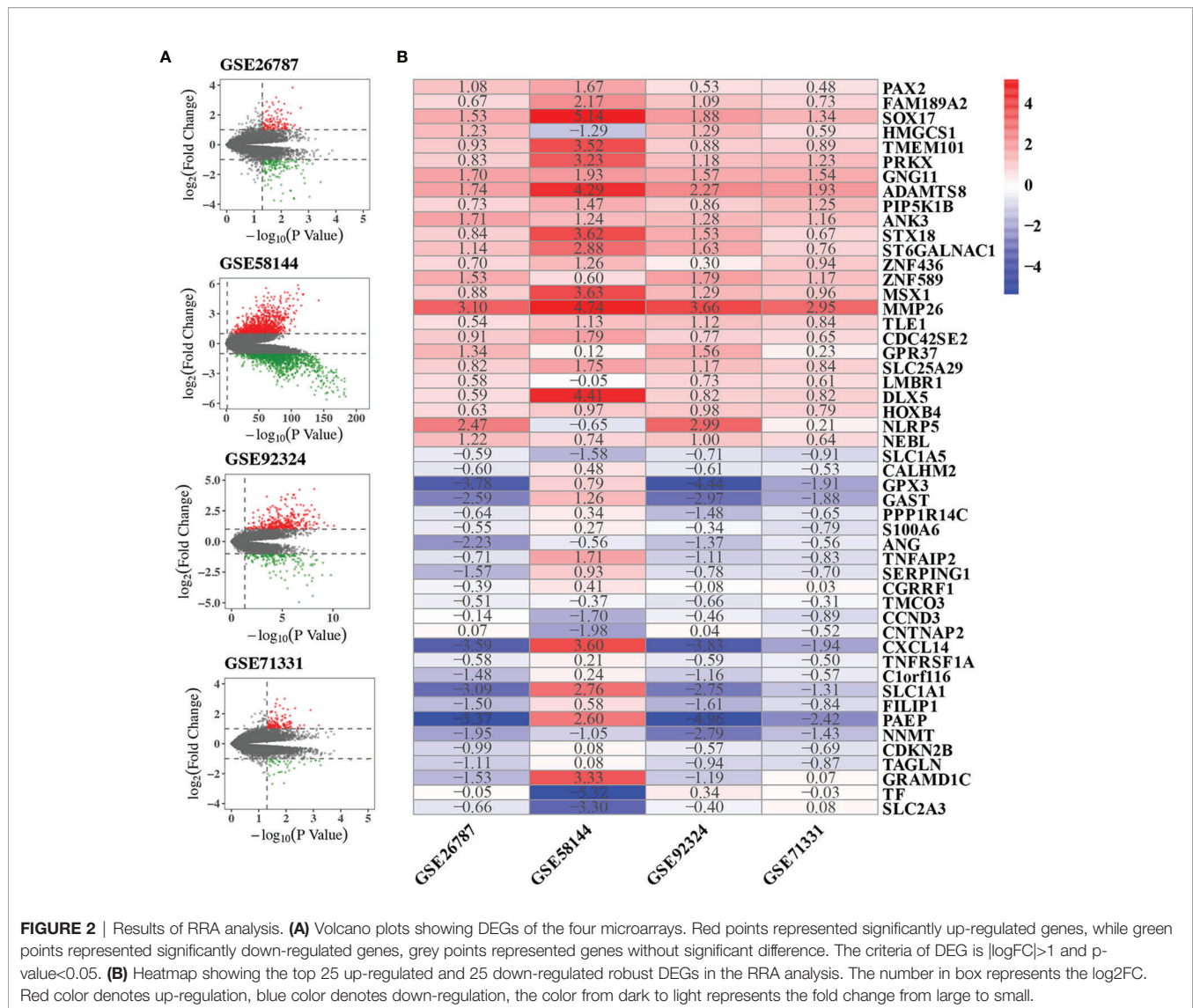
Protein-Protein Interaction (PPI) Analysis and Identification of Hub Genes

The PPI analysis of the 438 robust DEGs identified from the RRA analysis (p-value < 0.01) was constructed using the String website and visualized by Cytoscape. The PPI network is comprised of 160 nodes and 174 edges (**Figure 4**). The interactions of the proteins were shown in **Supplementary Table 4**. The network was then imported into the Cytoscape for subsequent analysis. The CytoHubba plugin was used to identify the candidate hub genes in the PPI network. The nodes were ranked by scores following MCC, DMNC, EPC, and Degree methods, respectively. Results of the CytoHubba are listed in **Supplementary Table 5**. The top 100 ranked nodes were candidate hub genes. The overlap genes of the top 100 genes identified by Cytohubba following four methods and the top 100 robust DEGs in the RRA analysis were determined as the hub genes. Finally, 18 genes (HMGCS1, SQLE, ESR1, LAMC1, HOXB4, PIP5K1B, GNG11, GPX3, PAX2, TF, ALDH6A1, IDH1, SALL1, EYA1, TAGLN, TPD52L1, ST6GALNAC1, NNMT) were determined as the hub genes.

TABLE 1 | Characteristics of the included microarray datasets.

GSE ID	Platform	Participants (Control/RIF)	Array type	Tissue	Biopsy time	Year	Country
GSE26787	GPL570	5/5	Affymetrix	Endometrium	WOI	2011	France
GSE92324	GPL10558	8/12	Illumina	Endometrium	WOI	2016	India
GSE111974	GPL17077	24/24	Agilent	Endometrium	WOI	2018	Turkey
GSE58144	GPL15789	72/43	Agilent	Endometrium	WOI	2014	Netherlands
GSE71331	GPL19072	5/7	Agilent	Endometrium	WOI	2018	China

RIF, recurrent implantation failure; WOI, window of implantation.

FIGURE 2 | Results of RRA analysis. **(A)** Volcano plots showing DEGs of the four microarrays. Red points represented significantly up-regulated genes, while green points represented significantly down-regulated genes, grey points represented genes without significant difference. The criteria of DEG is |logFC|>1 and p-value<0.05. **(B)** Heatmap showing the top 25 up-regulated and 25 down-regulated robust DEGs in the RRA analysis. The number in box represents the log2FC. Red color denotes up-regulation, blue color denotes down-regulation, the color from dark to light represents the fold change from large to small.

The intersection of the top 100 genes identified by Cytohubba following four methods and the top 100 robust DEGs was shown by the Venn diagram in **Figure 5**.

Hub Genes Validation

The expression of the 18 hub genes was then validated in GSE111974. As the violin plots showed that ALDH6A1, EYA1, HOXB4, PAX2, PIP5K1B, SALL1 SQLE, ST6GALNAC1 were significantly increased in the RIF patients compared to control patients; while LAMC1, TAGLN were significantly decreased in the RIF patients compared to control patients (**Figure 6**). The other hub genes that were not significantly different between the RIF group and the control group were not shown in the violin plot.
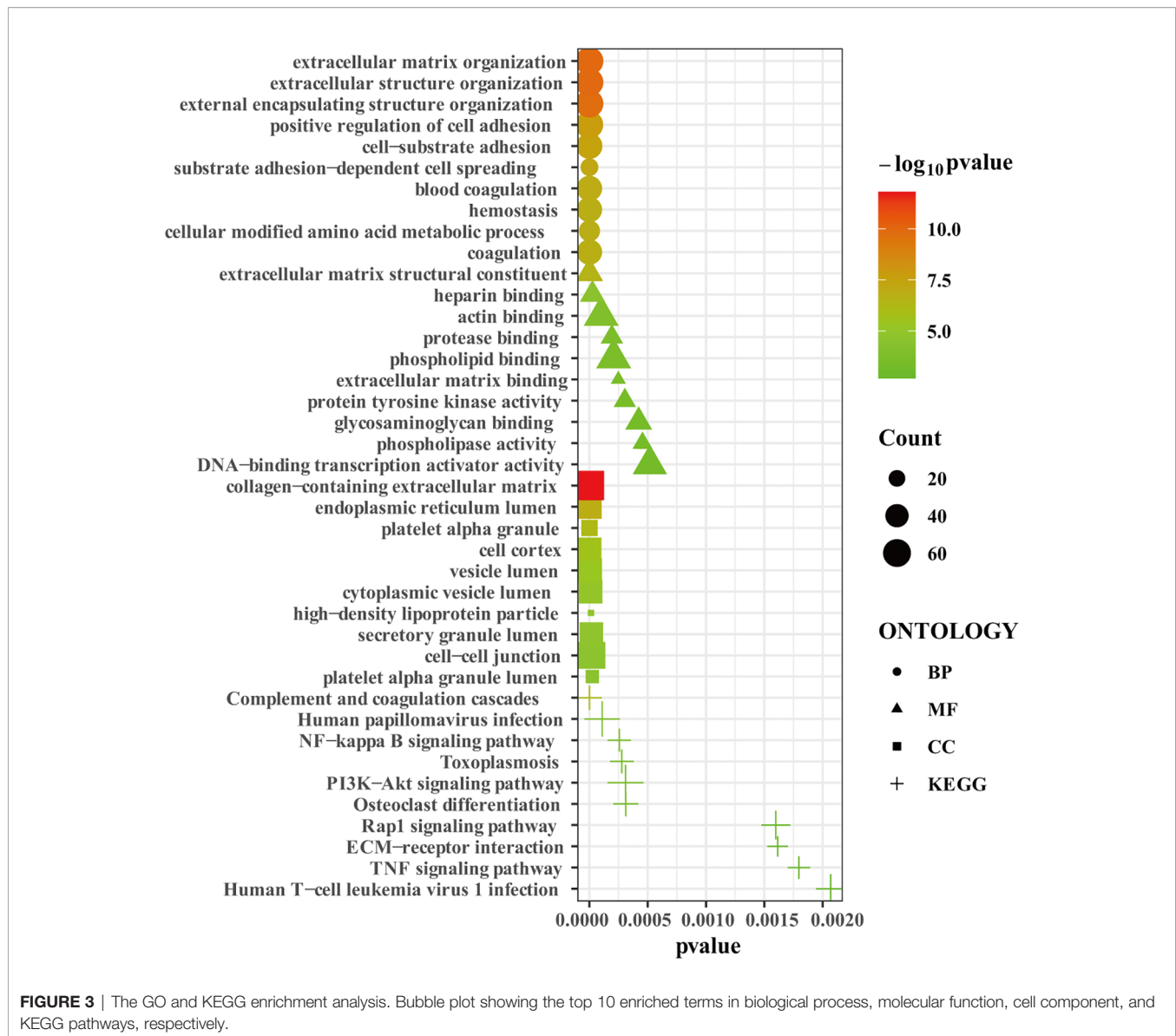
The Diagnostic Role of Hub Genes in RIF

We use the dataset GSE111974 to investigate the predictive effect of hub genes for RIF. 28 patients (13 controls and 15 RIF patients) were randomized divided into the training set, and 20

patients (11 controls and 9 RIF patients) were randomized divided into the validation set. In the validation set, the prediction model showed an accuracy rate of 85%, specificity rate of 100%, and sensitivity rate of 88.9%. The ROC analysis showed that the AUC is 0.980 (**Figure 7**).
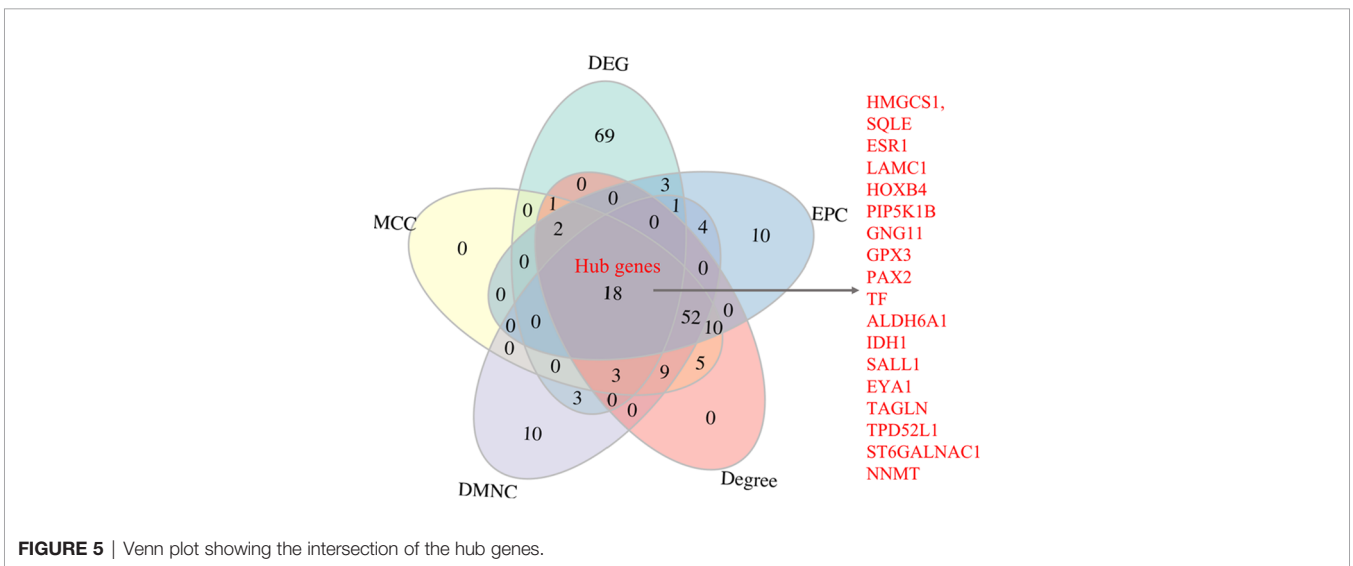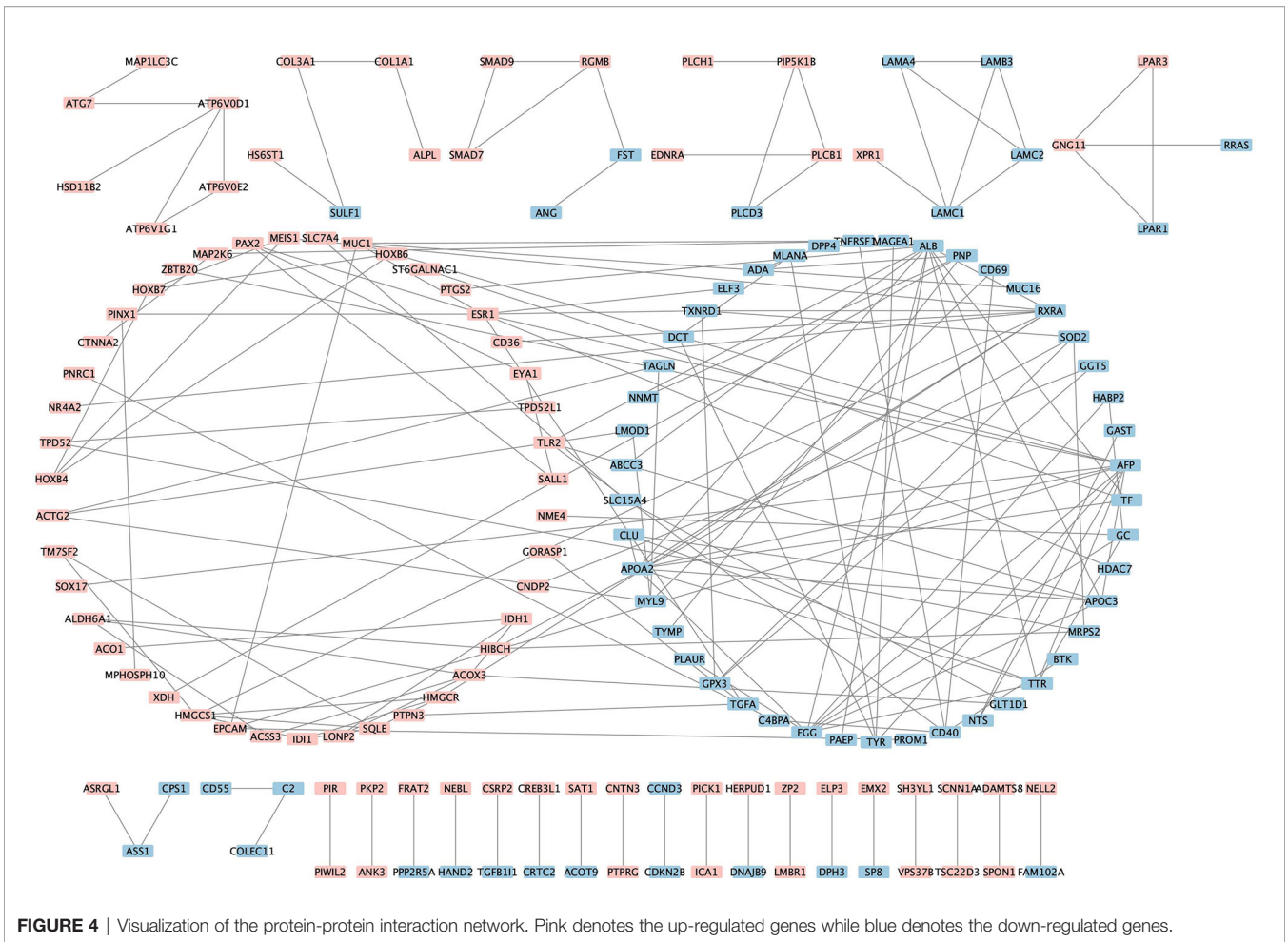
DISCUSSION

Embryo implantation begins as the blastocyst hatching from the zona pellucida, followed by blastocyst apposition, adhesion, and invasion. Recurrent implantation failure (RIF) refers to the situation that good-quality embryos repeatedly fail to implant following two or more IVF cycles (1). The endometrial factor is one of the leading factors account for RIF (3). Therefore, identifying dysregulated genes in the endometrium of RIF is very important for understanding the pathogenesis of RIF and is clinically useful for RIF prediction. To the best of our knowledge,

FIGURE 3 | The GO and KEGG enrichment analysis. Bubble plot showing the top 10 enriched terms in biological process, molecular function, cell component, and KEGG pathways, respectively.

this is the first study to explore novel robust DEGs and hub genes in the endometrium associated with RIF by the RRA method combined with PPI and other bioinformatic tools. Robust DEGs were identified by integrating four array datasets. Consistent with published data, the enrichment of these robust DEGs in several terms such as extracellular matrix organization, adhesion, and coagulation were closely related to embryo implantation (23, 24). In addition, enrichment of these robust DEGs in some KEGG pathways, such as complement and coagulation cascades, Rap1 signaling pathway, ECM-receptor interaction, also suggest their relevance in implantation failure pathogenesis. From the enrichment analysis, we can conclude that the dysregulated genes in RIF were mainly involved in the processes associated with cell-cell and cell-matrix adhesion, ECM remodeling, coagulation, and immune response. All these processes are essential for embryo implantation (23, 24).

After the analysis of RRA, PPI, and Cytohubba, we eventually obtained 18 hub genes (including HMGCS1, SQLE, ESR1, LAMC1, HOXB4, PIP5K1B, GNG11, GPX3, PAX2, TF, ALDH6A1, IDH1, SALL1, EYA1, TAGLN, TPD52L1, ST6GALNAC1, NNMT). Some of them were demonstrated to play a role in regulating endometrial function and may be linked to embryo implantation. For example, LAMC1 is produced by the decidualized cells and serves as a key factor that controls decidual cell architecture and differentiation. Down-regulation of LAMC1 can prevent the formation of the basal matrix and lead to decreased trophoblast outgrowth (25). ESR1 can mediate estrogen effects and endometrial preparation for implantation, the expression or polymorphism of ESR1 was reported to be related to RIF (26). GPX3 plays a major role in reducing ROS during decidualization with its expression peaks during the implantation window. While inhibition of GPX3 is associated

FIGURE 4 | Visualization of the protein-protein interaction network. Pink denotes the up-regulated genes while blue denotes the down-regulated genes.

**FIGURE 5** | Venn plot showing the intersection of the hub genes.

with a reduced pregnancy rate (27–29). TAGLN expression is significantly upregulated in the receptive endometrium compared to the pre-receptive endometrium (28) and is

involved in regulating cell invasion, migration, and differentiation (30). One member of the TAGLN family, TAGLN2, is required for embryo implantation by promoting

FIGURE 6 | The normalized expression of hub genes validated in GSE111974.
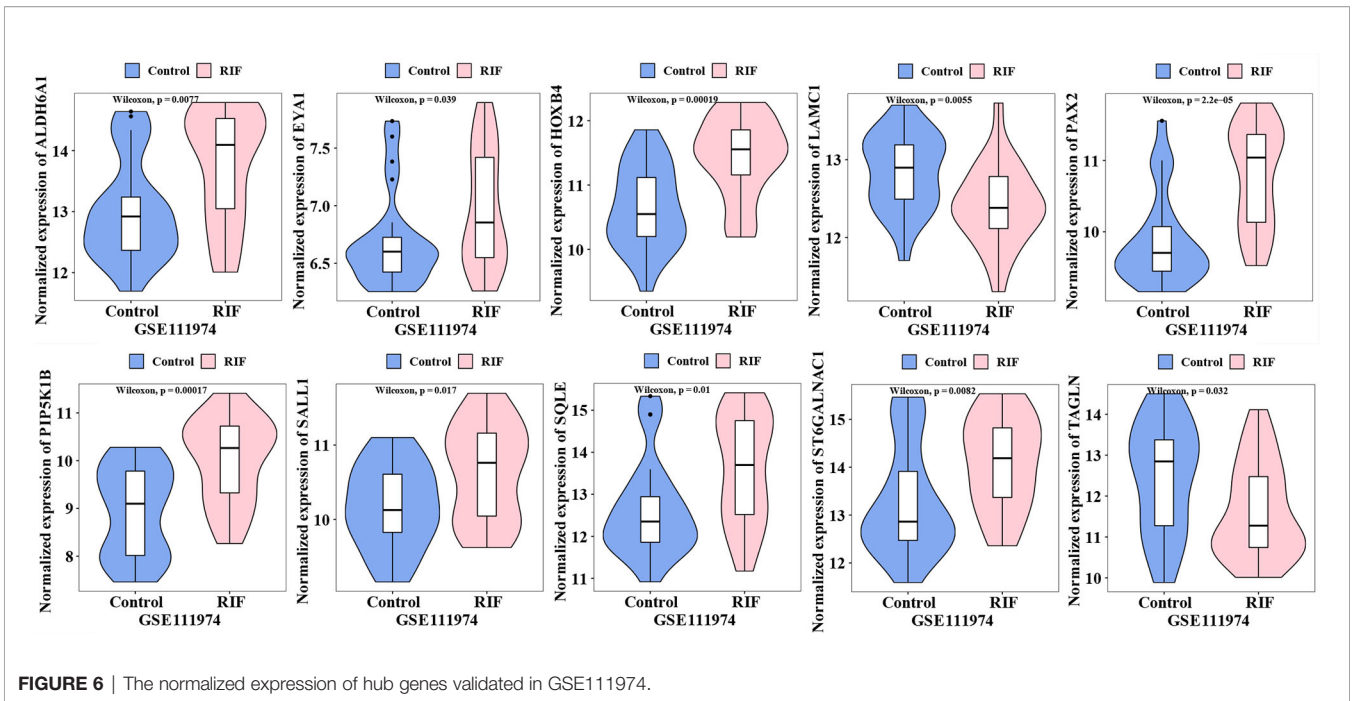
actin polymerization (31). NNMT in the uterine fluid is effective for the assessment of endometrial receptivity, NNMT protein in the uterine fluid was significantly decreased in RIF patients compared to fertile controls (32). IDH1 was expressed predominantly in the endometrial epithelia and involved in cellular defense against oxidative damage. IDH1 is required for a short time up to pregnancy recognition (33, 34). PAX2 is a known oncogenic gene in endometrial cancer, it is an estrogen-induced target gene (35). Abnormal increased PAX2 expression

indicates over activation of the estrogen pathway which may impede embryo implantation. Consistently, our study confirmed that the expression of PAX2 was higher in RIF patients. Besides, some of the top25 up-regulated and down-regulated robust DEGs are reported to be associated with embryo implantation. ANG is an essential angiogenesis factor that plays important role in embryo implantation (36). SOX17 plays a key role in endometrial receptivity and embryo implantation by regulating embryo adhesion (37). MSX1 is critical for conferring uterine

FIGURE 7 | The Roc curve of the combined ten hub genes in predicting RIF.
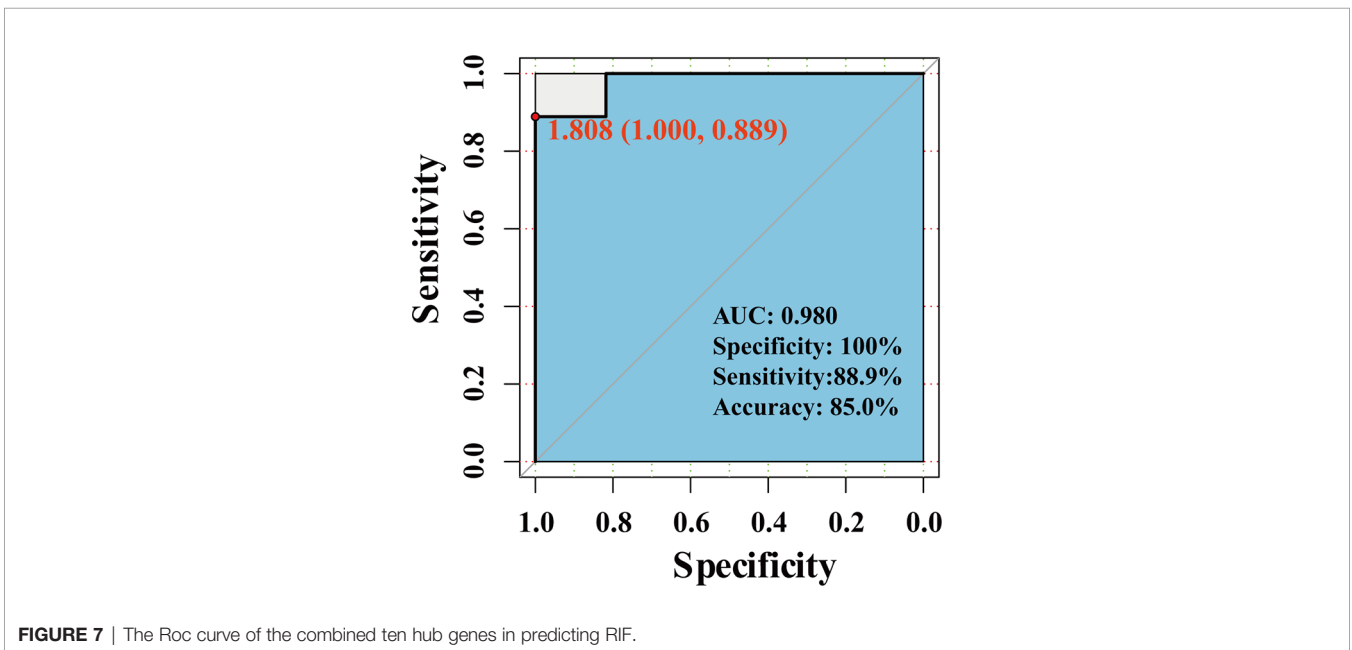
receptivity and readiness to implantation, it is significantly down-regulated in the receptive endometrium compared to the pre-receptive endometrium and reduced MSX1 in the endometrium is linked to infertility (28, 38–41). PAEP is also a marker for endometrial receptivity, PAEP was differentially expressed between RIF patients and fertile controls (7, 42). It is reported that receptive endometrial status can be determined based on the evaluation of mRNA expression levels of PAEP, DPP4, MSX1, and HLA-DOB genes (28). MMP-26 is localized in the epithelial cells of the human endometrium and its expression peaks in the early secretory phase. MMP-26 is reported to be an estrogen-sensitive gene with significantly higher expression at WOI in HRT cycles compare to natural cycles (43). Increased MMP-26 expression suggests overstimulation with E2. Consistently, our study indicates that the expression of MMP-26 was higher in RIF patients. SERPING1 mRNA is mainly in luminal and glandular epithelial cells and is significantly down-regulated in patients with recurrent miscarriage (44). SERPING1 is associated with decidualization and is involved in endometrial receptivity and immune regulation at the fetal-maternal interface (45–47). CXCL14 expression peaks at the embryo's implantation site during WOI (48). CXCL14 is necessary to recruit natural killer cells (49) and is associated with a normal epithelial/stromal gene expression pattern (50). Cxcl14 knockout mice are infertile (51). It is reported that intrauterine hCG co-cultured with PBMCs administration induced expression of CXCL14 (52), while intrauterine injection hCG is reported to increase implantation rate of RIF patients (53). CCND3 expression increases upon decidualization progression and peaks at WOI (54), CCND3 is reported to regulate decidualization of uterine stromal cells during implantation (55, 56). SLC1A5 is likely responsible for increases in amounts of neutral and acidic amino acids in the uterine lumen to support conceptus growth, development, and survival (57).

In summary, by combining RRA, PPI, and Cytohubba analysis, we have successfully identified 18 hub genes associated with RIF and may provide deeper insight into the comprehensive molecular changes in RIF. Among the hub genes, ALDH6A1, EYA1, HOXB4, PAX2, PIP5K1B, SALL1 SQLE, ST6GALNAC1 were significantly increased in the RIF patients compared to control patients; while LAMC1, TAGLN were significantly decreased in the RIF patients compared to control patients in GSE111974 (**Figure 6**). A combination of the ten hub genes was effective in the prediction of RIF with an accuracy rate of 85%, specificity rate of 100%, sensitivity rate of 88.9%, and an AUC of 0.980 (**Figure 7**). Though novel robust DEGs and hub genes were identified, however, the underlying molecular mechanisms have not yet been fully elucidated. In the future, large cohorts are needed to validate the actual prediction power of the hub genes in the real-world population.

The strength of the study is that we obtained a larger dataset by combining data from four GEO datasets, which increased the sample size and ensured the stability and relative reliability of the conclusions. On the other hand, the RRA method was used to reduce the influences of the measurement platform, the sample size of datasets, the experimental design, and other factors on the final results. However, the potential limitations should also be underlined: (1) The validity of our conclusions mainly rests on the reliability of the original microarray datasets. (2) We applied GSE111974 to validate the expression of hub genes and to determine the diagnostic role of hub genes in RIF, however, the results were limited since the sample size is 48. More experiments are needed to validate the expression and function of hub genes in the future.

In conclusion, our integrated analysis identified novel robust DEGs and hub genes in RIF. The hub genes were effective in predicting RIF and will contribute to the understanding of comprehensive molecular mechanisms in RIF pathogenesis.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

ETHICS STATEMENT

The study is an integrated analysis based on published datasets from Gene Expression Omnibus (GEO) database. Therefore, ethics approval is not applicable.

AUTHOR CONTRIBUTIONS

HZ designed the study, downloaded the raw data, performed the statistical analyses, draft the manuscript, tables and figures. LS and YF checked the tables and figures, revised the manuscript. SQ designed the study and revised the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fendo.2022.785462/full#supplementary-material

Supplementary Table 1 | DEGs in each microarray dataset.

Supplementary Table 2 | Robust DEGs identified by the RRA analysis.

Supplementary Table 3 | Results of GO and KEGG pathway enrichment analysis.

Supplementary Table 4 | PPI interactions.

Supplementary Table 5 | Results of the CytoHubba.

REFERENCES

1. Cimadomo D, Craciunas L, Vermeulen N, Vomstein K, Toth B. Definition, Diagnostic and Therapeutic Options in Recurrent Implantation Failure: An International Survey of Clinicians and Embryologists. *Hum Reprod* (2021) 36 (2):305–17. doi: 10.1093/humrep/deaa317

2. Coughlan C, Ledger W, Wang Q, Liu F, Demirol A, Gurgan T, et al. Recurrent Implantation Failure: Definition and Management. *Reprod BioMed Online* (2014) 28(1):14–38. doi: 10.1016/j.rbmo.2013.08.011

3. Sebastian-Leon P, Garrido N, Remohí J, Pellicer A, Diaz-Gimeno P. Asynchronous and Pathological Windows of Implantation: Two Causes of Recurrent Implantation Failure. *Hum Reprod* (2018) 33(4):626–35. doi: 10.1093/humrep/dey023

4. Koot YE, van Hooff SR, Boomsma CM, van Leenen D, Groot Koerkamp MJ, Goddijn M, et al. An Endometrial Gene Expression Signature Accurately Predicts Recurrent Implantation Failure After IVF. *Sci Rep* (2016) 6:19411. doi: 10.1038/srep19411

5. Luo J, Zhu L, Zhou N, Zhang Y, Zhang L, Zhang R. Construction of Circular RNA-MicroRNA-Messenger RNA Regulatory Network of Recurrent Implantation Failure to Explore Its Potential Pathogenesis. *Front Genet* (2020) 11:627459. doi: 10.3389/fgene.2020.627459

6. Liu C, Wang M, Zhang H, Sui C. Altered microRNA Profiles of Extracellular Vesicles Secreted by Endometrial Cells From Women With Recurrent Implantation Failure. *Reprod Sci* (2021) 28(7):1945–55. doi: 10.1007/s43032-020-00440-y

7. Pathare ADS, Zaveri K, Hinduja I. Downregulation of Genes Related to Immune and Inflammatory Response in IVF Implantation Failure Cases Under Controlled Ovarian Stimulation. *Am J Reprod Immunol* (2017) 78(1). doi: 10.1111/aji.12679

8. Bastu E, Demiral I, Gunel T, Ulgen E, Gumusoglu E, Hosseini MK, et al. Potential Marker Pathways in the Endometrium That May Cause Recurrent Implantation Failure. *Reprod Sci* (2019) 26(7):879–90. doi: 10.1177/1933719118792104

9. Lédée N, Munaut C, Aubert J, Sérazin V, Rahmati M, Chaouat G, et al. Specific and Extensive Endometrial Deregulation Is Present Before Conception in IVF/ICSI Repeated Implantation Failures (IF) or Recurrent Miscarriages. *J Pathol* (2011) 225(4):554–64. doi: 10.1002/path.2948

10. Guo F, Si C, Zhou M, Wang J, Zhang D, Leung PCK, et al. Decreased PECAM1-Mediated TGF-β1 Expression in the Mid-Secretory Endometrium in Women With Recurrent Implantation Failure. *Hum Reprod* (2018) 33(5):832–43. doi: 10.1093/humrep/dey022

11. Shi C, Han HJ, Fan LJ, Guan J, Zheng XB, Chen X, et al. Diverse Endometrial mRNA Signatures During the Window of Implantation in Patients With Repeated Implantation Failure. *Hum Fertil (Camb)* (2018) 21(3):183–94. doi: 10.1080/14647273.2017.1324180

12. Fan LJ, Han HJ, Guan J, Zhang XW, Cui QH, Shen H, et al. Aberrantly Expressed Long Noncoding RNAs in Recurrent Implantation Failure: A Microarray Related Study. *Syst Biol Reprod Med* (2017) 63(4):269–78. doi: 10.1080/19396368.2017.1310329

13. Sebastian-Leon P, Devesa-Peiro A, Aleman A, Parraga-Leo A, Arnau V, Pellicer A, et al. Transcriptional Changes Through Menstrual Cycle Reveal a Global Transcriptional Derepression Underlying the Molecular Mechanism Involved in the Window of Implantation. *Mol Hum Reprod* (2021) 27(5):gaab027. doi: 10.1093/molehr/gaab027

14. He A, Zou Y, Wan C, Zhao J, Zhang Q, Yao Z, et al. The Role of Transcriptomic Biomarkers of Endometrial Receptivity in Personalized Embryo Transfer for Patients With Repeated Implantation Failure. *J Transl Med* (2021) 19(1):176. doi: 10.1186/s12967-021-02837-y

15. Kolde R, Laur S. RobustRankAggreg: Methods for Robust Rank Aggregation. *Bioinformatics* (2013) 28(4):573–80. doi: 10.1093/bioinformatics/btr709

16. Kauffmann A, Gentleman R, Huber W. arrayQualityMetrics–A Bioconductor Package for Quality Assessment of Microarray Data. *Bioinformatics* (2009) 25 (3):415–6. doi: 10.1093/bioinformatics/btn647

17. Gautier L, Cope L, Bolstad BM, Irizarry RA. affy—Analysis of Affymetrix GeneChip Data at the Probe Level. *Bioinformatics* (2004) 20(3):307–15. doi: 10.1093/bioinformatics/btg405

18. Huber W, von Heydebreck A, Sültmann H, Poustka A, Vingron M. Variance Stabilization Applied to Microarray Data Calibration and to the Quantification of Differential Expression. *Bioinformatics* (2002) 18(Suppl 1):S96–104. doi: 10.1093/bioinformatics/18.suppl_1.S96

19. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma Powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies. *Nucleic Acids Res* (2015) 43(7):e47. doi: 10.1093/nar/gkv007

20. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: A Universal Enrichment Tool for Interpreting Omics Data. *Innovation* (2021) 2 (3):100141. doi: 10.1016/j.xinn.2021.100141.

21. Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T. Cytoscape 2.8: New Features for Data Integration and Network Visualization. *Bioinformatics* (2011) 27(3):431–2. doi: 10.1093/bioinformatics/btq675

22. Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. cytoHubba: Identifying Hub Objects and Sub-Networks From Complex Interactome. *BMC Syst Biol* (2014) 8(Suppl 4):S11. doi: 10.1186/1752-0509-8-S4-S11

23. Imakawa K, Bai R, Kusama K. Integration of Molecules to Construct the Processes of Conceptus Implantation to the Maternal Endometrium. *J Anim Sci* (2018) 96(7):3009–21. doi: 10.1093/jas/sky103

24. Guo X, Yi H, Li TC, Wang Y, Wang H, Chen X. Role of Vascular Endothelial Growth Factor (VEGF) in Human Embryo Implantation: Clinical Implications. *Biomolecules* (2021) 11(2):253. doi: 10.3390/biom11020253

25. Ramathal C, Wang W, Hunt E, Bagchi IC, Bagchi MK. Transcription Factor CCAAT Enhancer-Binding Protein Beta (C/EBPbeta) Regulates the Formation of a Unique Extracellular Matrix That Controls Uterine Stromal Differentiation and Embryo Implantation. *J Biol Chem* (2011) 286(22):19860–71. doi: 10.1074/jbc.M110.191759

26. Vagnini LD, Renzi A, Petersen B, Canas M, Petersen CG, Mauri AL, et al. Association Between Estrogen Receptor 1 (ESR1) and Leukemia Inhibitory Factor (LIF) Polymorphisms can Help in the Prediction of Recurrent Implantation Failure. *Fertil Steril* (2019) 111(3):527–34. doi: 10.1016/j.fertnstert.2018.11.016

27. Farimani Sanoee M, Alizamir T, Faramarzi S, Saidijam M, Yadegarazari R, Shabab N, et al. Effect of Myomectomy on Endometrial Glutathione Peroxidase 3 (GPx3) and Glycodelin mRNA Expression at the Time of the Implantation Window. *Iran BioMed J* (2014) 18(2):60–6. doi: 10.6091/ibj.1222.2013

28. Burmenskaya OV, Bozhenko VK, Smolnikova VY, Kalinina EA, Korneeva IE, Donnikov AE, et al. Transcription Profile Analysis of the Endometrium Revealed Molecular Markers of the Personalized 'Window of Implantation' During In Vitro Fertilization. *Gynecol Endocrinol* (2017) 33(sup1):22–7. doi: 10.1080/09513590.2017.1404236

29. Xu X, Leng JY, Gao F, Zhao ZA, Deng WB, Liang XH, et al. Differential Expression and Anti-Oxidant Function of Glutathione Peroxidase 3 in Mouse Uterus During Decidualization. *FEBS Lett* (2014) 588(9):1580–9. doi: 10.1016/j.febslet.2014.02.043

30. Dos Santos Hidalgo G, Meola J, Rosa ESJC, Paro de Paz CC, Ferriani RA. TAGLN Expression Is Deregulated in Endometriosis and May Be Involved in Cell Invasion, Migration, and Differentiation. *Fertil Steril* (2011) 96(3):700–3. doi: 10.1016/j.fertnstert.2011.06.052

31. Liang X, Jin Y, Wang H, Meng X, Tan Z, Huang T, et al. Transgelin 2 Is Required for Embryo Implantation by Promoting Actin Polymerization. *FASEB J* (2019) 33(4):5667–75. doi: 10.1096/fj.201802158RRR

32. Kasvandik S, Saarma M, Kaart T, Rooda I, Velthut-Meikas A, Ehrenberg A, et al. Uterine Fluid Proteins for Minimally Invasive Assessment of Endometrial Receptivity. *J Clin Endocrinol Metab* (2020) 105(1):dgz019. doi: 10.1210/clinem/dgz019

33. Forde N, McGettigan PA, Mehta JP, O'Hara L, Mamo S, Bazer FW, et al. Proteomic Analysis of Uterine Fluid During the Pre-Implantation Period of Pregnancy in Cattle. *Reproduction* (2014) 147(5):575–87. doi: 10.1530/REP-13-0010

34. Hayashi K, Spencer TE. Estrogen Disruption of Neonatal Ovine Uterine Development: Effects on Gene Expression Assessed by Suppression Subtraction Hybridization. *Biol Reprod* (2005) 73(4):752–60. doi: 10.1095/biolreprod.105.042812

35. Zeng K, Wu Y, Wang C, Wang S, Sun H, Zou R, et al. ASH2L Is Involved in Promotion of Endometrial Cancer Progression via Upregulation of PAX2 Transcription. *Cancer Sci* (2020) 111(6):2062–77. doi: 10.1111/cas.14413

36. Lee YL, Liu Y, Ng PY, Lee KF, Au CL, Ng EH, et al. Aberrant Expression of Angiopoietins-1 and -2 and Vascular Endothelial Growth Factor-A in Peri-

Implantation Endometrium After Gonadotrophin Stimulation. *Hum Reprod* (2008) 23(4):894–903. doi: 10.1093/humrep/den004

37. Kinnear S, Salamonsen LA, Francois M, Harley V, Evans J. Uterine SOX17: A Key Player in Human Endometrial Receptivity and Embryo Implantation. *Sci Rep* (2019) 9(1):15495. doi: 10.1038/s41598-019-51751-3

38. Daikoku T, Cha J, Sun X, Tranguch S, Xie H, Fujita T, et al. Conditional Deletion of Msx Homeobox Genes in the Uterus Inhibits Blastocyst Implantation by Altering Uterine Receptivity. *Dev Cell* (2011) 21(6):1014–25. doi: 10.1016/j.devcel.2011.09.010

39. Nallasamy S, Li Q, Bagchi MK, Bagchi IC. Msx Homeobox Genes Critically Regulate Embryo Implantation by Controlling Paracrine Signaling Between Uterine Stroma and Epithelium. *PloS Genet* (2012) 8(2):e1002500. doi: 10.1371/journal.pgen.1002500

40. Sun X, Park CB, Deng W, Potter SS, Dey SK. Uterine Inactivation of Muscle Segment Homeobox (Msx) Genes Alters Epithelial Cell Junction Proteins During Embryo Implantation. *FASEB J* (2016) 30(4):1425–35. doi: 10.1096/fj.15-282798

41. Bolnick AD, Bolnick JM, Kilburn BA, Stewart T, Oakes J, Rodriguez-Kovacs J, et al. Reduced Homeobox Protein MSX1 in Human Endometrial Tissue Is Linked to Infertility. *Hum Reprod* (2016) 31(9):2042–50. doi: 10.1093/humrep/dew143

42. Tapia A, Gangi LM, Zegers-Hochschild F, Balmaceda J, Pommer R, Trejo L, et al. Differences in the Endometrial Transcript Profile During the Receptive Period Between Women Who Were Refractory to Implantation and Those Who Achieved Pregnancy. *Hum Reprod* (2008) 23(2):340–51. doi: 10.1093/humrep/dem319

43. Pilka R, Oborna I, Lichnovsky V, Havelka P, Fingerova H, Eriksson P, et al. Endometrial Expression of the Estrogen-Sensitive Genes MMP-26 and TIMP-4 Is Altered by a Substitution Protocol Without Down-Regulation in IVF Patients. *Hum Reprod* (2006) 21(12):3146–56. doi: 10.1093/humrep/del180

44. Huang J, Qin H, Yang Y, Chen X, Zhang J, Laird S, et al. A Comparison of Transcriptomic Profiles in Endometrium During Window of Implantation Between Women With Unexplained Recurrent Implantation Failure and Recurrent Miscarriage. *Reproduction* (2017) 153(6):749–58. doi: 10.1530/REP-16-0574

45. Gurung S, Greening DW, Rai A, Poh QH, Evans J, Salamonsen LA. The Proteomes of Endometrial Stromal Cell-Derived Extracellular Vesicles Following a Decidualizing Stimulus Define the Cells' Potential for Decidualization Success. *Mol Hum Reprod* (2021) 27(10):gaab057. doi: 10.1093/molehr/gaab057

46. Rai A, Poh QH, Fatmous M, Fang H, Gurung S, Vollenhoven B, et al. Proteomic Profiling of Human Uterine Extracellular Vesicles Reveal Dynamic Regulation of Key Players of Embryo Implantation and Fertility During Menstrual Cycle. *Proteomics* (2021) 21(13-14):e2000211. doi: 10.1002/pmic.202000211

47. Bauersachs S, Mitko K, Ulbrich SE, Blum H, Wolf E. Transcriptome Studies of Bovine Endometrium Reveal Molecular Profiles Characteristic for Specific Stages of Estrous Cycle and Early Pregnancy. *Exp Clin Endocrinol Diabetes* (2008) 116(7):371–84. doi: 10.1055/s-2008-1076714

48. Kuang H, Chen Q, Fan X, Zhang Y, Zhang L, Peng H, et al. CXCL14 Inhibits Trophoblast Outgrowth via a Paracrine/Autocrine Manner During Early Pregnancy in Mice. *J Cell Physiol* (2009) 221(2):448–57. doi: 10.1002/jcp.21877

49. Cao Q, Chen H, Deng Z, Yue J, Chen Q, Cao Y, et al. Genetic Deletion of Cxcl14 in Mice Alters Uterine NK Cells. *Biochem Biophys Res Commun* (2013) 435(4):664–70. doi: 10.1016/j.bbrc.2013.04.106

50. Mesa AM, Mao J, Medrano TI, Bivens NJ, Jurkevich A, Tuteja G, et al. Spatial Transcriptomics Analysis of Uterine Gene Expression in Enhancer of Zeste Homolog 2 Conditional Knockout Mice†. *Biol Reprod* (2021) 105(5):1126–39. doi: 10.1093/biolre/ioab147

51. Meuter S, Schaerli P, Roos RS, Brandau O, Bösl MR, von Andrian UH, et al. Murine CXCL14 Is Dispensable for Dendritic Cell Function and Localization Within Peripheral Tissues. *Mol Cell Biol* (2007) 27(3):983–92. doi: 10.1128/MCB.01648-06

52. Chen J, Zhao X, Ao L, Yin T, Yang J. Transcriptomic Changes and Potential Regulatory Mechanism of Intrauterine Human Chorionic Gonadotropin Co-Cultured With Peripheral Blood Mononuclear Cells Infusion in Mice With Embryonic Implantation Dysfunction. *Ann Transl Med* (2020) 8(4):99. doi: 10.21037/atm.2019.12.109

53. Xie H, Zeng H, He D, Liu N. Effect of Intrauterine Perfusion of Human Chorionic Gonadotropin Before Embryo Transfer After Two or More Implantation Failures: A Systematic Review and Meta-Analysis. *Eur J Obstet Gynecol Reprod Biol* (2019) 243:133–8. doi: 10.1016/j.ejogrb.2019.10.039

54. Mestre-Citrinovitz AC, Kleff V, Vallejo G, Winterhager E, Saragüeta P. A Suppressive Antagonism Evidences Progesterone and Estrogen Receptor Pathway Interaction With Concomitant Regulation of Hand2, Bmp2 and ERK During Early Decidualization. *PloS One* (2015) 10(4):e0124756. doi: 10.1371/journal.pone.0124756

55. Tan J, Raja S, Davis MK, Tawfik O, Dey SK, Das SK. Evidence for Coordinated Interaction of Cyclin D3 With P21 and Cdk6 in Directing the Development of Uterine Stromal Cell Decidualization and Polyploidy During Implantation. *Mech Dev* (2002) 111(1-2):99–113. doi: 10.1016/S0925-4773(01)00614-1

56. Li DD, Zhao SY, Yang ZQ, Duan CC, Guo CH, Zhang HL, et al. Hmgn5 Functions Downstream of Hoxa10 to Regulate Uterine Decidualization in Mice. *Cell Cycle* (2016) 15(20):2792–805. doi: 10.1080/15384101.2016.1220459

57. Gao H, Wu G, Spencer TE, Johnson GA, Bazer FW. Select Nutrients in the Ovine Uterine Lumen. IV. Expression of Neutral and Acidic Amino Acid Transporters in Ovine Uteri and Peri-Implantation Conceptuses. *Biol Reprod* (2009) 80(6):1196–208. doi: 10.1095/biolreprod.108.075440