A comparison of transcriptomic profiles in endometrium during window of implantation between women with unexplained recurrent implantation failure and recurrent miscarriage

Jin Huang1,*, Hao Qin2,*, Yihua Yang1,3, Xiaoyan Chen1, Jiamiao Zhang1,3, Susan Laird4, Chi Chiu Wang1,5,6, Ting Fung Chan2 and Tin Chiu Li1

1Department of Obstetrics and Gynaecology, 2School of Life Sciences, The Chinese University of Hong Kong, Hong Kong SAR, China, 3Reproductive Medicine Centre of the Affiliated Hospital, Guilin Medical University, Guilin, Guangxi, China, 4Biomolecular Sciences Research Centre, Sheffield Hallam University, Sheffield, UK, 5Li Ka Shing Institute of Health Sciences, and 6School of Biomedical Sciences, The Chinese University of Hong Kong, Hong Kong SAR, China

Correspondence should be addressed to T C Li; Email: tinchiu.li@cuhk.edu.hk

*(J Huang and H Qin contributed equally to this work)

Abstract

The endometrium becomes receptive to the embryo only in the mid-luteal phase, but not in the other stages of the menstrual cycle. Endometrial factors play an important role in implantation. Women with recurrent miscarriage and recurrent implantation failure have both been reported to have altered expression of receptivity markers during the window of implantation. We aimed to compare the gene expression profiles of the endometrium in the window of implantation among women with unexplained recurrent implantation failures (RIF) and unexplained recurrent miscarriages (RM) by RNA sequencing (RNA-Seq). In total 20 patients (9 RIF and 11 RM) were recruited. In addition 4 fertile subjects were included as reference. Endometrium samples were precisely timed on the 7th day after luteal hormone surge (LH+7). All the 24 endometrium samples were extracted for total RNA. The transcriptome was determined by RNA-Seq in the first 14 RNA samples (5 RIF, 6 RM and 3 fertile). Differentially expressed genes between RM and RIF were validated by quantitative real-time PCR (qPCR) in all 24 RNA samples (9 RIF, 11 RM and 4 fertile). Transcriptomic profiles of RM and RIF, but not control samples, were separated from each other by principle component analysis (PCA) and support vector machine (SVM). Complementary and coagulation cascades pathway was significantly up-regulated in RIF while down-regulated in RM. Differentially expressed genes C3, C4, C4BP, DAF, DF and SERPING1 in complement and coagulation cascade pathway between RM and RIF were further validated by qPCR. This study compared endometrial transcriptome among patients with RIF and RM in the window of implantation; it identified differential molecular pathways in endometrium between RIF and RM, which potentially affect the implantation process.

Reproduction (2017) 153 749–758

Introduction

Implantation is a process when the embryo attaches to the endometrium, followed by migration and invasion into the deeper layer of the endometrium to become embedded, which involves a complex sequence of cellular and molecular changes. There are two well-defined categories of reproductive failure attributable to implantation disorders, namely recurrent miscarriage (RM) and recurrent implantation failure (RIF). Recurrent implantation failure refers to failure to achieve a clinical pregnancy after transfer of at least four good-quality embryos in a minimum of three cycles in a woman under the age of 40 years (Coughlan et al. 2014). The failure usually occurs at an earlier stage, resulting in complete failure to implant or failure to establish the pregnancy. On the other hand, recurrent miscarriage is defined as three or more consecutive clinical miscarriages (Saravelos & Li 2012). It commonly manifests as pregnancy loss later on in the pregnancy, often around 6–8 weeks gestation.

A number of earlier studies suggested that the endometrium in unexplained RM and RIF shared some common pathological changes. For example,
both uNK cell count and interleukin 15 expressions have been reported to be increased in the two conditions (Tuckerman et al. 2007, 2010). On the other hand, certain molecules have been found to be deranged in one condition but not in the other. A notable example is beta3 integrin, which is down-regulated in RM (Germeyer et al. 2014) but not in RIF (Coughlan et al. 2013); whereas leukemia inhibitory factor is down-regulated in RIF (Mariee et al. 2012), but not in RM (Xu et al. 2012, Karaer et al. 2014). In most of these earlier studies, only one specific marker was examined, with the exception of a few, which examined up to 3 specific markers at the same time (Xu et al. 2012). The study of a single or a few biomarkers has a limited value especially in the case of implantation disorder. A more comprehensive study of several markers would be needed.
All expressions were normalized by quantile normalization method using median (Risso et al. 2014). The differential expressed genes (DEGs) were determined by two criteria: (a) the fold change between the means of groups was higher than 1.5; and (b) the $P$-value calculated from pooled T-test was smaller than 0.05.

**Hierarchical clustering and principle component analysis (PCA)**

The expressions of each gene were firstly scaled as follow:

$$S_{ij} = \frac{R_{ij}}{\max(R_j)}$$

where $S_{ij}$ is the scaled expression of gene $i$ in sample $j$, $R_{ij}$ is the raw normalized expression of gene $i$ in sample $j$, $\max(R_j)$ is the largest value of gene $i$ among all samples from RM, RIF and fertile groups. Afterward the scaled expressions will be used for unsupervised hierarchical clustering and then PCA by R packages gplots (Warnes et al. 2009) and prcomp.

After PCA was done, the vector of each principle component was calculated. In the space constructed by any two principle components $V_p$ and $V_q$, the direction of classification which was vertical to the calculated boundary $ap + bq = c$ by SVM is

$$V_D = \begin{bmatrix} a \\ b \end{bmatrix},$$

where $a^2 + b^2 = 1$. Thus the contribution of each gene to the classification direction was calculated by:

$$aV_p + bV_q = \mathbf{X}$$

where $\mathbf{X}$ contains the contribution of the corresponding genes to the classification direction. Support vector machine (SVM) was performed by Python library sklearn 0.17.0 (Pedregosa 2011). Genes whose absolute values of the contribution scores were larger or equal to 0.01 were considered to have significant contribution.

**Gene ontology and pathway analysis**

Only genes with significant differential expression were retrieved for gene ontology (GO) and pathway analysis. Pathway enrichment was analyzed by DAVID 6.7 (the Database for Annotation, Visualization and Integrated Discovery) (Risso et al. 2014). The pathways whose correlated $P$-value (q-value) smaller than 0.05 were considered significantly enriched.

---

**Table 1** Demographic characteristics in women with recurrent miscarriage or fertile controls.

<table>
<thead>
<tr>
<th></th>
<th>RIF</th>
<th>RM</th>
<th>C</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (year), mean ± s.d.</td>
<td>36.6 ± 3.4</td>
<td>35.3 ± 3.0</td>
<td>31.8 ± 4.3</td>
<td>NS</td>
</tr>
<tr>
<td>Body mass index (kg/m$^2$)</td>
<td>21.8 ± 2.1</td>
<td>20.6 ± 2.2</td>
<td>21.1 ± 2.5</td>
<td>NS</td>
</tr>
<tr>
<td>Cycle length (day), mean ± s.d.</td>
<td>29.1 ± 4.0</td>
<td>28.5 ± 4.2</td>
<td>30.0 ± 3.1</td>
<td>NS</td>
</tr>
<tr>
<td>Endometrium thickness on the biopsy day, mean ± s.d.</td>
<td>10.2 ± 5.1</td>
<td>9.8 ± 4.7</td>
<td>10.5 ± 4.4</td>
<td>NS</td>
</tr>
<tr>
<td>Average times of live birth</td>
<td>0</td>
<td>0.1</td>
<td>2.0</td>
<td>–</td>
</tr>
<tr>
<td>Average times of miscarriage</td>
<td>0.1</td>
<td>3.5</td>
<td>0</td>
<td>–</td>
</tr>
<tr>
<td>Average times of implantation failure</td>
<td>5.8</td>
<td>0</td>
<td>0</td>
<td>–</td>
</tr>
</tbody>
</table>

C, fertile subjects; NS, not significant; RIF, recurrent implantation failures; RM, recurrent miscarriages.

**Quantitative RT-PCR**

In additional to the 14 sequenced samples, extra 10 independent samples (4 RIF, 5 RM and 1 fertile) were added to measure relative gene expression using quantitative real-time RT-PCR (qPCR) for validation. TATA-box binding protein (TBP) and ribosomal protein L13a (RPL13A) were used as reference genes for expression normalization. Total RNA was extracted and quality checked as above. One microgram of total RNA was
used for reverse transcription. Quantitative PCR was performed by using SYBR Green PCR Master Mix (Applied Biosystems) with Roche LightCycler 480 II. Primer sequences can be found in Supplementary Table 1 (see section on supplementary data given at the end of this article). Wilcoxon test was used to examine the statistical significance between RIF and RM.

**Ethics**

This study was approved by the Joint Chinese University of Hong Kong – New Territories East Cluster Clinical Research Ethics Committee. Written consent was obtained from all participants.

**Results**

The demographics of the recruited subjects are summarized in Table 1. There was no significant difference in age, BMI, cycle length and endometrium thickness at the time of biopsy among RIF, RM and fertile groups (Table 1).

The reads mapping of all 14 samples for RNA-Seq were satisfactory (Supplementary Fig. 1). All samples had over 80% reads mapped to the human genome hg38. The raw sequencing data was uploaded to NCBI with reference BioProject ID: PRJNA314429. Firstly, we explored whether RM, RIF and fertile samples could be separated in the transcriptome profiling. After normalization and scaling of gene expressions, unsupervised hierarchical clustering was performed to all samples (Fig. 1). Most RIF and RM samples were clustered to two sub-trees, while fertile samples could not be grouped and were clustered within RIF or RM samples. There were 661 genes significantly up-regulated in RIF compared with RM; while 301 genes up-regulated in RM compared with RIF. To further compare and contrast the differences between RIF and RM, fertile samples were excluded and principle component analysis (PCA) was performed. RIF and RM samples showed distinct spatial distribution in the three-dimensional space constructed by the first three components. In the space constructed by the first component and the third component, RIF and RM were linear-separated perfectly (Fig. 2). The boundary between RIF and RM could be further learned by SVM, where the classification direction which was vertical to the linear boundary gave the best resolution to distinguish RIF and RM. To identify the genes that contribute the most to the classification direction, the contribution score for each gene was calculated (see ‘Materials and methods’ section). The genes with positive contribution scores showed higher expressions in RM, while genes with negative contribution scores showed higher expressions in RIF (Fig. 3). Genes whose absolute values of contribution scores were larger or equal to 0.01 were considered to have significant contribution to the classification of RIF and RM, where 183 genes had significant positive contribution and 380 had significant negative contribution.

To investigate which biological and molecular pathway contributed the most to the differential transcriptomic pattern between RIF and RM, pathway enrichment analysis was applied on genes with significant contribution on both directions (Fig. 3).
Pathways with $-\log(q\text{-value}) > 2$ were considered as significant. The localization of proteins encoded by genes in both directions showed high enrichment in extracellular regions indicated by GO cellular component terms. However, the pathways where they were involved were distinct, which could be revealed by GO biological process terms and KEGG pathways (Fig. 4). On negative direction, many responses to wounding and inflammatory genes were predominately enriched, including the most enriched complement...
and coagulation cascades in KEGG (Fig. 5). Several central components of complement and coagulation cascades have significantly higher expressions in RIF than those in RM, with significant t-test P-values less than 0.05. We chose the top six over-expressed genes in RIF from complement cascade, namely C3, C4, C4BP, DAF, DF and SERPING1, for real-time PCR validation (Fig. 6). In contrast, on a positive direction, genes that were involved in extracellular structure organization and biological adhesion by GO biological process terms and neuroactive ligand–receptor interaction and calcium signaling pathways in KEGG were significantly enriched, but only few (2/78) significant differentially expressed genes were identified.

Real-time PCR was performed on all 24 samples to validate the differential expression of 6 genes in complement cascade (Fig. 6). All the gene expressions were significantly up-regulated in RIF group compared with that of RM with Wilcoxon test, P-values less than 0.05. However, the real-time PCR results of control group showed great variation, and most of the difference is not significant.
Discussion

In this study, we have found that the transcriptome profiles of the two groups of patients studied (RM and RIF) are distinctively different from one another. In addition, we have found significant amounts of differentially expressed genes (661 up-regulated in RIF and 301 up-regulated in RM) and one distinctly and validated pathway between women with RIF and RM.

Among all the enriched pathways, the complement and coagulation cascades pathway was the most significantly affected pathway with $-\log(q\text{-values}) > 4$ in KEGG pathway analysis in RIF. This particular pathway was up-regulated in RIF patients compared with RM patients, the genes involved in this pathway included C3, C4, C4BP, DAF, DF and SERPING1. All of these 6 up-regulated genes had been validated with qPCR, which confirmed that the up-regulation generally existed in patients with unexplained RIF. Reference to Fig. 6 showed that C3 expression in RIF was significantly higher than that of control, whereas the expression in RM was not different to that of control subjects. The complement system, represented by complement component 3 (C3), is a proteolytic cascade in plasma and an upstream mediator of innate immunity. It is known that human chorionic gonadotropin (hCG) has positive effects on endometrial C3 expression (Palomino et al. 2013). While the adverse effect of decreased C3 expression on placental development and fetal development has been shown in C3 deficient mice (Chow et al. 2009), and variants in FOXD1 that enhance the expression of C3 were associated with miscarriage in humans and mice (Laissue et al. 2016). The possible adverse effect of over expression of C3 has not been previously reported. It seems therefore a fine balance is necessary; both under expression as well as over expression of C3 may be detrimental. Our
observation that the C3 was over expressed in RIF but not in RM suggested that the two conditions affect uterine receptivity in different ways. One of the major immune functions of C3 pathway is to form the membrane attack complex, leading to cell lysis (Ricklin et al. 2010). While DAF (complement-protective protein decay-accelerating factor, also known as CD55) is considered as an inhibitor of increased complement activity (Young et al. 2002), and the expression of DAF was minimal in the proliferative and early secretory phase in endometrium, increasing to a maximum on LH+7, and decreasing until next cycle. Endometrial C3 and DAF expression was associated with human chorionic gonadotropin, indicating its roles in early embryo development (Palomino et al. 2013). While there has not been any study in the literature which reported on the expression of any of the genes involved in this pathway in the endometrium of RIF or RM at the time of implantation, previous genetic association studies found the loss of functional mutation of some genes in this pathway were associated with RM (Mohlin et al. 2013) or other adverse pregnancy outcome, such as preeclampsia (Salmon et al. 2011). Although both C3 and DAF were up-regulated in RIF when compared with RM, the increased C3 expression (3.5 folds) was higher than the increased DAF expression (2.8 folds), suggesting the inhibitory complement system in RIF may be more likely a reactive response. Further studies of its inhibitory mechanism and subsequent downstream innate immune response in RIF are needed.

On the other hand, though the neuroactive ligand-receptor interaction pathway and calcium signaling pathway were enriched in RM according to the positively contributed gene list from SVM analysis, we did not consider them as important as the complement and coagulation cascades pathway as discussed above. Firstly, most of the genes with positive contribution in these two pathways were expressed at very low expression level. Furthermore, almost no differentially expressed genes were identified in those two pathways. One explanation for the lack of significantly expressed genes in calcium pathway in our study could be that this activity could be prominent in endometrial epithelium cells (Thie & Denker 2002, Brosens et al. 2014, Ruan et al. 2014), which might be diluted if sequencing endometrium tissue as a whole. However, it has long been known that Ca\(^{2+}\) channels involve in a variety of implantation processes and increased Ca\(^{2+}\) mobilization can assist blastocyst-endometrium adhesion (Thie & Denker 2002, Brosens et al. 2014, Ruan et al. 2014). Brosen and coworkers found that competent and low-quality embryos elicited different Ca\(^{2+}\) channel responses in vitro, which indicate the active role of endometrial selective function of human embryos (Brosens et al. 2014). And this selection was impaired in the endometrium of RM subjects (Teklenburg et al. 2010). In this study, the genes with positive contribution in calcium signaling pathway may suggest the higher activity of the Ca\(^{2+}\) channel in the endometrium of RM compared with that of RIF. It might indicate that the endometrium during WOI is more favorable for implantation in RM compared with RIF, which would also be consistent with the in vitro study carried out by Brosen and coworkers that pattern of Ca\(^{2+}\) signals was associated with the implantation results.

According to the hypothesis put forward by Brosen and Macklon and coworkers (Teklenburg et al. 2010, Brosens et al. 2014, Macklon & Brosens 2014), women with unexplained RM would be superfertile because the endometrium is over-receptive, less able to discern and prevent the abnormal embryos from implantation, in contrast to that of women with RIF, in which the abnormality makes it difficult for even the normal embryo to implant. While our findings do not directly confirm or refute the Brosen hypothesis that the endometrium is over-receptive in women with RM, our finding regarding the differential regulation of the pathways between the two groups of women may provide insight into the molecular mechanism controlling the implantation process in the endometrium to make it under-receptive (as in women with RIF) or over-receptive (as in women with RM).

The transcriptome pattern of fertile women seems dispersedly distributed among RIF and RM subjects, but the sample size in our study indeed is very small to make any conclusion. The dispersed distribution may be due to the heterogeneity of endometrial receptivity status. Although they were classified as fertile controls as they had successful pregnancy in early years, unfortunately it does not necessarily imply they will still be able to achieve successful pregnancy if conceived. This is one of the limitations of our study.

A particular strength of our study is the precise timing of the endometrial specimen, all obtained on day LH+7. While some earlier studies did time the specimen precisely on a single day (Diaz-Gimeno et al. 2011, Hu et al. 2014) others obtained the specimen over two or more days (Ledee et al. 2011, Koot et al. 2016). Given that the endometrium changes very rapidly around the time of implantation, the inclusion of samples collected on different days after the LH surge could introduce a significant source of variance to the results. It may help to explain why, in a previous study by Ledee and coworkers which studied similar subjects groups as in our study but with biopsies obtained over a three-day period from days LH+7 to +9, they could find gene expression differences between RM and RIF, consistent with our findings, but not able to identify the pathways (Ledee et al. 2011).

Another strength of this study is that we used RNA-seq rather than micro-array to analyze the specimens. Earlier transcriptome studies of the endometrium used micro-array analysis (Ledee et al. 2011, Othman et al. 2012, Ruiz-Alonso et al. 2013, Koot et al. 2016) although 2
recent studies did use sequencing techniques (Hu et al. 2014, Kosova et al. 2015). It is now well accepted that sequencing technique is more comprehensive in coverage and precise in quantification of global gene expression profiles (McGee et al. 2013). Furthermore, in our study we have used more straightforward and more comprehensive methods to mine the features which contributed to the classification of the two groups of women, and thus identified genes and pathways that were differentially expressed between RIF and RM. The chosen testing platforms and analysis methods could also greatly contributed to the identification of significant pathways.

One possible limitation of our study is the relatively small sample size (RIF = 9, RM = 11 and control = 4) and so the conclusions reached in this study should be considered preliminary, especially in view of the potential heterogeneity of the study populations.

To conclude, we have identified that the complement and coagulation cascade pathways are significantly different between women with RM and RIF. The identified pathways provide an insight into how the process of implantation in these two types of implantation disorder differs from one other.

**Supplementary data**

This is linked to the online version of the paper at http://dx.doi.org/10.1530/REP-16-0574.

**Declaration of interest**

The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

**Funding**

Direct Grant for Research (Ref No. 2014.1.042) from the Chinese University of Hong Kong. H Q is supported by a General Research Fund (GRF14102014) from the Research Grant Committee, Hong Kong Special Administration region to T F C.

**Authors’ contribution statement**

J H and T C L designed the study, prepared the samples, interpreted the data and wrote the manuscript. H Q and T F C performed the bioinformatics analysis and prepared the figures. Y Y and J Z performed the qPCR validation. X C helped on patient recruitment and sample collection. C C W and S L contributed to the interpretation of data and the manuscript preparation. All of the authors contributed to finalizing of the manuscript.

---

**References**


