1. Title: Characterise robust biomarkers in early IVF cycle to predict final outcomes via metabonomics analysis to generate a database for constructing a predictive computer program.

2. Layman's summary

In vitro fertilisation (IVF), which regulates the relative hormonal levels and combined with laboratory fertilisation techniques to increase the successful pregnancy rate, is one of the major treatment considered by infertile couples. However, a standard IVF treatment would cost over 3000 pounds (1) with a lower than 50% successful pregnancy rate (2). Therefore, investigating the biochemical changes during different time-point in the IVF cycle would provide a better insight of what happening during the IVF cycle and improve the overall procedures to achieve a higher pregnancy rate. The pilot study has shown that the major biochemical variations occur during pituitary suppression, ovarian stimulation, and before egg collection. The aim of this proposal is to characterize several robust biomarkers in early IVF cycle to predict the final outcomes and construct an IVF database for building an IVF outcome prediction computer program in the future.

3. Background

Infertility is becoming a prevalent and global problem resulting from a variety of causes, such as physiological disease, bacterial infections, or abnormal nutritional conditions. Physiologically caused infertility is usually related to oocyte or anatomic abnormalities and poor semen quality. Three major infertility related pathogenic bacteria are *Mycoplasma tuberculosis*, *Chlamydia trachomatis*, and *Neisseria gonorrhoeae* (3). Additionally, it has been suggested that bacterial vaginosis – a condition in which the healthy *Lactobacillus*-dominated vaginal environment shifts to a poly-microbial dominated community – plays a major part in increasing the risk of infertility and preterm birth (3). Lastly, ovulation disorders associated with either an obese or an underweight state are predisposed to fertility-related problems.

IVF is one of the most common assisted reproduction techniques. The whole IVF cycle is usually consisted of six steps: 1. pituitary suppression, 2. ovarian stimulation, 3. final oocyte maturation, 4. oocyte retrieval (*Figure 1a*), 5. embryo preparation in laboratory (*Figure 1b*), 6. embryo transfer (*Figure 1c*).

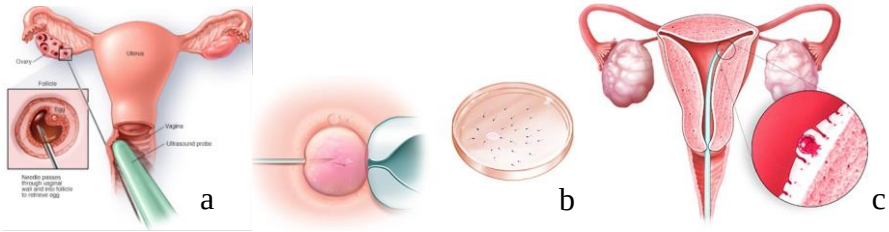

Figure 1: a: oocyte retrieval, b: embryo preparation in laboratory, c: embryo transfer (4).

a. Eggs are removed from mature follicles within an ovary via transvaginal ultrasound aspiration.

b. In vitro egg fertilization can be either done by injecting a sperm into the egg or mixing the egg with sperm in a petri dish.

c. Transfer the in vitro prepared embryo back into the uterus.

4. Scientific methodology

The preliminary research has done metabolic profiling of urine and plasma samples collected from women during the IVF cycle (6 steps) by nuclear magnetic resonance (NMR) spectroscopy. Principal component analysis (PCA) and orthogonal partial least squares – Discriminant Analysis (OPLS-DA) analysis were used to characterize the metabolic variations presented at each IVF step or between successful and unsuccessful outcomes within each step.

The starting metabolic status of the female subjects did not have influence on the IVF outcomes. The biomedical variations between positive and negative outcomes were observed during pituitary suppression or ovarian stimulation (*Figure 2*). Hippurate and lipoproteins which might related to obesity, were amongst the key metabolites prominently expressed in negative IVF outcomes. Additionally, compared to final successful pregnancy, the metabolic activities were significant different in pituitary suppression or egg collection (*Figure 3*).

For further metabolic profiling, the sample would mainly be collected after pituitary suppression, when ovarian stimulation, and before egg collection.

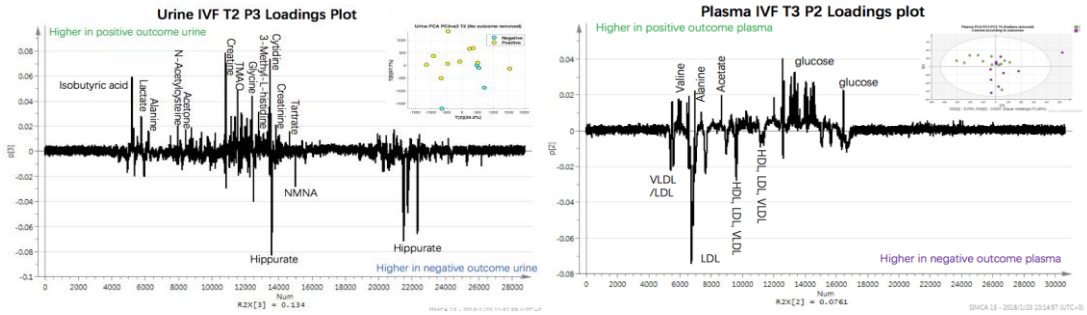

Figure 2: Loading plot showing metabolic variations between successful and unsuccessful outcomes within IVF cycle. T2: pituitary suppression. T3: ovarian stimulation. P3 (P2): principal component 3 (principal component 2).

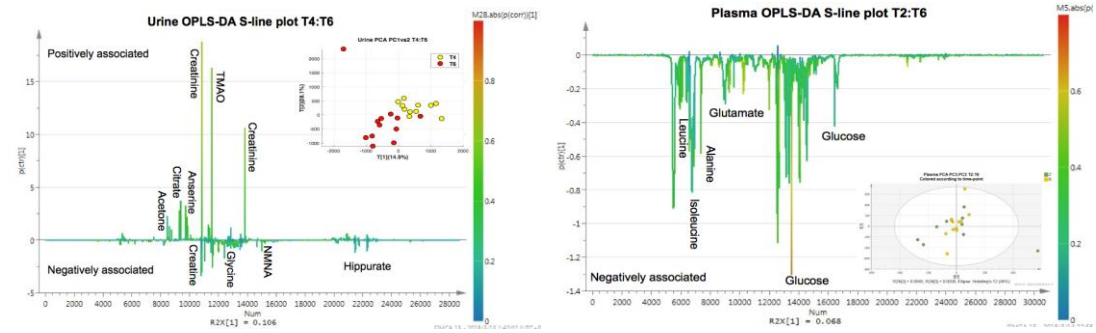

Figure 3: S-line plot showing metabolic differences between one IVF time-point and early successful pregnancy stage. Trying to identify metabolites that can indicate early pregnancy. T2: pituitary suppression. T4: before egg collection. T6: 6 weeks pregnancy.

Main Objective

- ***Develop a computer program based on metabonomics and microbiome information to predict IVF outcomes from early stage***

  The program would be used to monitor patient metabolic status and changes

during the IVF process in a simple and digitised way. In addition, the program can analysis the potential reason that would cause the negative results.

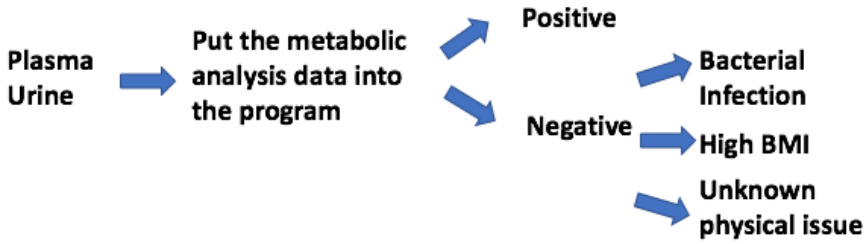

Figure 4: The flow chart of how the IVF outcome prediction computer program works.

Sub-objectives

- ***Investigate the appearance of microorganisms on embryo implantation rate*** There was a papers suggested that the presence of *Lactobacillus* in follicular fluid is related to successful embryo transfer and positive IVF results (5). 16s rRNA would be used in this project to analyse the in vitro embryo fertilisation media sample.

Methods

- Calculate the power of the study and the suitable number of samples.
- Collaborate with different reproductive clinics and Reproductive and Developmental Biology Department in Imperial College London to obtain the sample (urine and plasma, maybe in vitro embryo fertilisation media) from different IVF time-point. Use in conjunction with information on different ethnicities, ages, BMI, physical activity, and previous exposure to infectious vaginal disease to generate more representative metabolic profiles (Questionnaire).
- NMR metabolic profiling will be done in the Center of Computational and Systems Medicine, Sir Alexander Fleming Building, Imperial College London. Metabolic profiling will help to identify the metabolites in urine and plasma samples, and digitize the biochemical information into spectrums for analyzing. (Co-I)
- Multivariate data analysis and mixed effect models performed on MATLAB, R statistical software, and PRISM will be used for statistical validation. Other variables will also be considered, such as alanine/lactate ratio (6), BMI, excise level, and eating habit. (Co-I)
- Collaborate with and Life science department in Imperial College London for building an IVF metabonomics database.
- Code writing skills for IVF outcome predictive program development.

5. Contribution of the Co-Investigator

Computational spectral and statistical data pre-processing as well as statistical analysis and modelling of the datasets will be performed using R' and MATLAB analytical software by Miss Elena Alambriti (Co-I) at the Centre of Computational and Systems Medicine at Imperial College London (ICL). Untargeted profiling assays such as NMR generate extensively dense datasets making scientific analysis and biological

interpretation challenging. As a first step, spectral data pre-processing is required for the optimisation of the statistical analysis of the datasets. This includes the exclusion of spectral artefacts and unrelated variation as well as variable selection to make the data more comparable. An in-house pipeline developed by the centre at ICL will be used for the execution of spectral pre-processing in R' incorporating baseline correction, peak alignment, chemical shift metabolite assignment and data transformation and normalization.

As a result of the complexity of NMR datasets, multivariate statistical analysis will be used to characterize differential metabolic profiles in the variable IVF stages. This is due to the ability of this technique to account for multiple components incorporating into metabolic alterations as well as the detection of the main sources of variation to make the interpretability of the results easier. Statistical pre-processing approaches include data transformation and manipulation such as variable logarithmic transformations and data normalizations as well as data trimming and removal of duplicate information where required. Subsequently, multivariate metabolomic analysis using PCA, OPLS-DA and PLS-DA will be performed and mixed effect statistical models will be generated using variables such as BMI, ethnicity and age – obtained from questionnaires – to identify important and robust components of variation contributing to metabolic alterations that establish a relatively more effective IVF therapy.

6. Timeliness and novelty

This grant is aimed at developing an IVF outcome prediction model which can also monitor the metabolic changes during the cycle to increase the final pregnancy rate and avoid complications. The aim would be achieved step by step based on NMR performed metabolic profiling and multi-variate analysis. The project would collaborate with different IVF clinic and different variables such as BMI are also considered.

7. Impact

The successful pregnancy rate of IVF is low and the cycle requires several hormone regulations. Therefore, understanding and monitoring the biochemical changes during the IVF cycle will help to improve the successful pregnancy rate. Moreover, the computer program can also provide alternative information for doctor diagnosis and give suitable treatment for patients. Seminars will hold to introduce the program to clinicians who worked in gynaecology and obstetrics area.

8. Justification of resources

A clinician is needed to help set a standard sample collection procedure and help to assist the collaboration of clinics. A post-doc research associate will assign key metabolites to the PCA and OPLSDA analyzed NMR profile, and integrate the data into a database. In the final stage of the project, the post-doc would build a computer IVF outcome predictive program based on the database. NMR machine is used to metabolic profiling the sample of IVF patients. 16s rRNA analysis is used for sub-objectives which will investigate the influence of microorganisms existed in in vitro prepared embryo on implantation successful.

9. Programme of work

Timeline	0-12	13-14	15-20	21-24	25-30
1. urine and plasma sample collection during IVF cycle					
1* sampling in vitro embryo fertilization media					
2. Generating metabolic profiles of IVF samples via NMR					
2* 16s rRNA sequencing in vitro embryo fertilisation media sample					
3. Statistical analysis to find representative biomarkers and validation					
4. Building an IVF biomarker database					
5. Computer package development for IVF outcome prediction					

10. Costing

Category	*Breakdown by item heading*	*Total*
Staff salary	Clinician (12 month) = £3500*12 Post-doc research associate (18 month) = £4642*18	£125,556
Equipment	NMR analysis (1000*£20 per sample) = £20,000 16s rRNA extraction and sequencing (500*£35 per sample) = £17,500 Others (Eppendorf, filters, tips…) = £3,000	£40,500
Consumables	Sample transportation = £4,000 Software use (MATLAB, R statistical software…) = £3,000 A computer = £1,500 Travel expense = £3,000	£11,500
Overheads	=125,556 *1.3	£163,222.8
Total cost		£340,778.8

11. References

1. Clinic TLWs. Prices 2018 [cited 2018 05 May]. Available from: https://www.londonwomensclinic.com/about/prices/.
2. Clinic TLWs. Success Rates 2018 [cited 2018 05 May]. Available from: https://www.londonwomensclinic.com/about/success-rates/.
3. Sirota I, Zarek SM, Segars JH. Potential influence of the microbiome on infertility and assisted reproductive technology. Seminars in reproductive medicine. 2014;32(1):35-42.
4. Clinic M. In vitro fertilization (IVF) 2018 [cited 2018 05 may]. Available from: https://www.mayoclinic.org/tests-procedures/in-vitro-fertilization/about/pac-20384716.
5. Pelzer ES, Allan JA, Waterhouse MA, Ross T, Beagley KW, Knox CL. Microorganisms within Human Follicular Fluid: Effects on IVF. PLOS ONE. 2013;8(3):e59062.
6. Seli E, Botros L, Sakkas D, Burns DH. Noninvasive metabolomic profiling of embryo culture media using proton nuclear magnetic resonance correlates with reproductive potential of embryos in women undergoing in vitro fertilization. Fertil Steril. 2008;90(6):2183-9.