Applying Precision Medicine Approach to Preterm Birth Prevention

PROTOCOL V.5 14/09/2017

Background & Rationale

We have recently carried out a systematic review of maternal genes implicated in spontaneous preterm birth (sPTB) and very early pre-labor preterm rupture of membranes (PPROM) using a systems biology approach (*Figure 1*).[1] We found that sPTB was related to autoimmune/hormonal gene regulation, whilst pathways implicated in the aetiology of PPROM included haematological/coagulation function disorder, collagen metabolism, matrix degradation and local inflammation. In addition, we have showed in a pilot study conducted in collaboration with Perkin Elmer Inc. that a combination of biochemical and clinical markers (cervical length) measured at 16 and 20 weeks gestation can predict preterm birth risk (ROC AUC 0.95) *(Figure 2).*

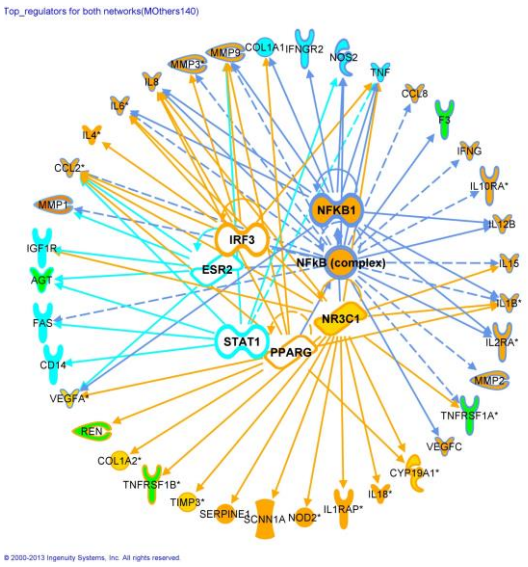


Figure 1. Top upstream regulators suggested by Ingenu Pathway Analysis for preterm birth. Targets for the t regulators are arranged in the outer circle. Functions of t main transcription regulators are presented with differ color lines. Blue lines represent both sPTB and PPR regulators (NFkB1, NFkB complex), orange lines repres mainly PPROM (PPARG, IRF3 and NR3C1) and turquoise li represent sPTB (STAT1 and ESR2) *(Capece et al. PLOS O*

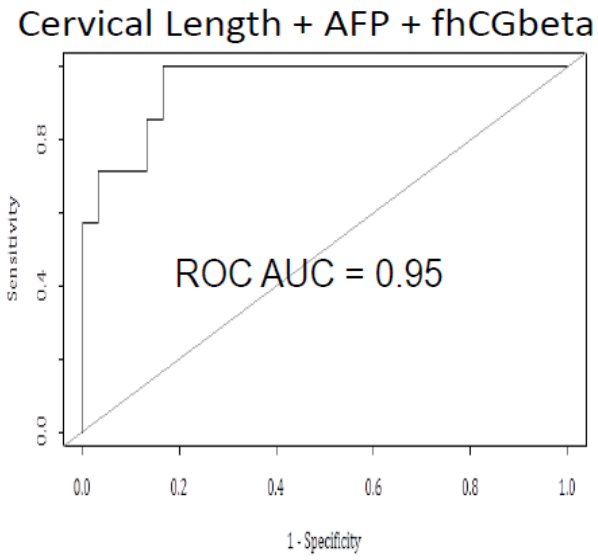


Figure 2. Combination of cervical length measured by transvaginal ultrasound, maternal serum α feto-protein and free β *human chorionic gonadotropin-beta measurements at 20 weeks gestation predicting recurrent spontaneous pre-term birth <34 weeks in 44 high risk women with at least one previous spontaneous preterm birth <34 weeks (unpublished data).*

Although valuable, such a reductionist approach to investigate the complexity of preterm birth may not be sufficient. Novel high-throughput technologies have now made possible simultaneous examination of millions of genes, transcripts, metabolites and microbiomes in an unbiased "hypothesis generating, multi layering" approach. Layers of data from genes through epigenetics, transcriptomics and functional proteomics should be made available for investigation of clinically well-phenotyped pregnancies linked to carefully selected pre-specified adverse outcomes. This approach may hold keys to understanding mechanisms of disease and identification of clinically useful prediction biomarkers and better targets for prevention therapy.

Currently, bioinformatics is capable of not only analysing very large sets of omics data, but also comparing cross-platform data and integrating the data into networks/pathways, which describe the structure and dynamics of the biological system in health and disease. It is crucial that the software used is able to interface with data from a large number of public databases. Such a system, named tranSMART is currently being established at the Centre for Genomic Research, University of Liverpool.

We will combine genetic and methylation data with metabolomics. Metabolomics has advantages over genomics, epigenetics and transcriptomics because the metabolic pathways are downstream of gene expression and therefore, may reflect functional cellular activity. Subclinical infection leading to local inflammation occurs in 1 in 4 preterm deliveries. Therefore, the analysis of the vaginal microbiome will be undertaken to determine both, eukaryotic and prokaryotic diversity of vaginal microbiome using sensitive next generation sequencing methods.

Protein quantification through incorporation of stable isotopes by iTRAQ (**i**sobaric **t**ags for **r**elative and **a**bsolute **q**uantification) is one of several techniques that our laboratory has successfully used to determine relative changes in protein and peptide abundance across diverse biological systems, enabling the comparative analysis of normal or disease and drug treated or un-treated conditions.

Our ultimate goal is to establish clinically useful personalized risk assessment with a combination of clinical and comprehensive molecular phenotyping. This approach will lead to better and safer use of currently available preventative therapies (drug repositioning) and development of novel, more effective therapies both in high and low resource settings.

Project Objectives

To investigate preterm birth phenotypes by using multiple omics and systems biology approaches in high risk population. Our goals are:

- To recruit a prospective cohort of women with spontaneous preterm birth at <34 weeks gestation (n=25) and controls who are women with spontaneous term delivery. All women will have well-characterized clinical phenotypes with biologic samples suitable for multi-platform systems biology analysis that are collected at a minimum of 2 time points during gestation and at delivery.

- To apply multi-omic high-throughput technologies to generate longitudinal datasets and integrate these into systems biology approaches to allow interpretation
- In collaboration with worldwide researchers, we aim to assess the feasibility and reproducibility of multi-platform systems biologic approaches using a limited number of samples from pregnancies that result in term and spontaneous preterm deliveries at < 34 weeks of gestation.

Methods

Women recruited prior to the ethical amendment in May 2015 (n=129) do not have urine, vaginal swabs or stool collected. These samples will be analysed together forming a pilot study and we will establish the necessary pipelines for data analysis of further samples.

Depending upon the results of the first study, the main cohort recruited after June 2015 will become a validation study for any significant findings from the pilot study or other significantly published works. New analysis will be based on the combination of vaginal microbiome results with other omics layers. Therefore the numbers to recruit have been clearly stated to include those recruited before July 2015, and recruitment will complete in December 2017.

This has been updated in Figure 3 to demonstrate both the pilot and validation study.

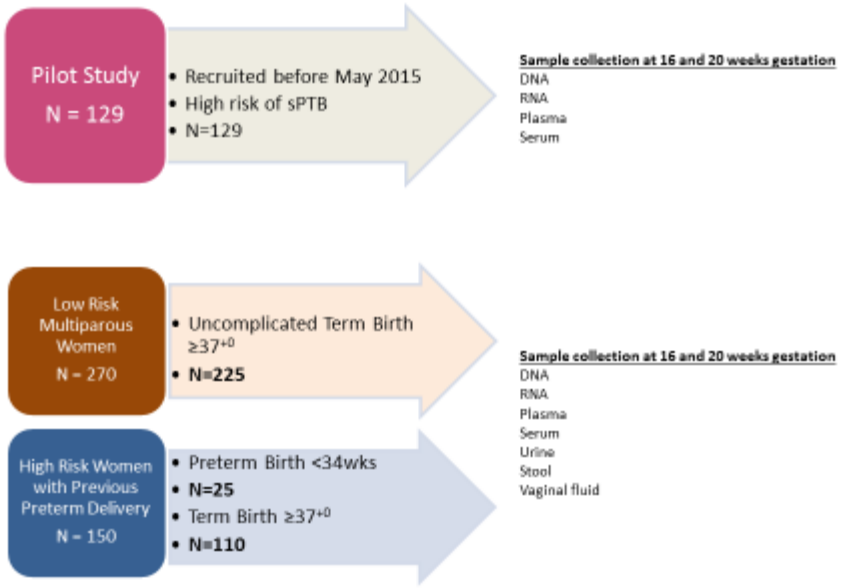
We aim to recruit 25 women with a preterm birth < 34 weeks and 225 low risk controls with a recruitment end date of 31 December 2017.

To achieve 25 paired samples from women who experience preterm births <34weeks we will recruit from our high risk patient clinic. Based on our clinical audit figures from 2010-2015 we know that our spontaneous preterm birth rate <24 weeks is 17%. We therefore need to recruit approximately 150 women to achieve 25 preterm births. As we have no way of foreseeing which women will have a preterm birth, there will be an additional 125 women recruited who will deliver at > 37 weeks (n = 110) or late preterm birth 34+1 – 36+6 (n = 15).

To achieve 225 normal deliveries we plan to recruit approximately 270 women who have previously had normal deliveries. We expect approximately 85% to have a subsequent normal pregnancy (i.e. no pathology such as gestational diabetes, pre-eclampsia, placental abruption, IUGR or spontaneous preterm birth), expecting 25 to be excluded due to unforeseen events of pregnancy and excluding those that are unable to attend both research appointments (unpaired data sets).

Therefore we will recruit two study cohorts with a total of approximately 420 participants:

- Pilot (pre May 2015)
- Main Study – Low risk 270 (approx.)
    - High risk 150 (approx.)



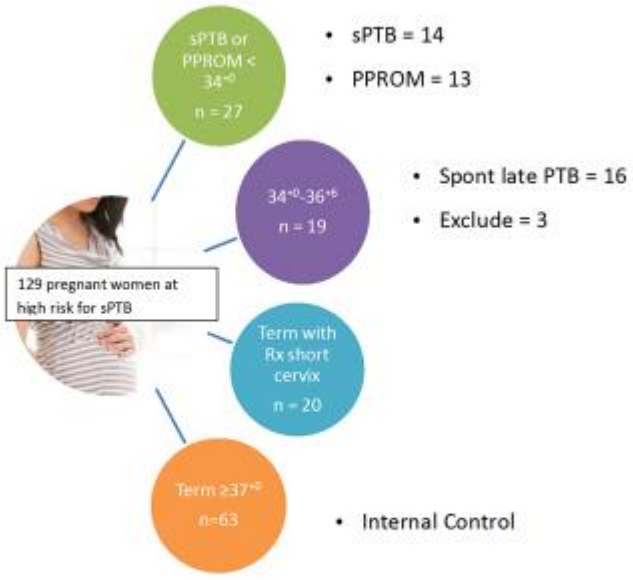


Figure 3. Patients included in the Pilot study

Approach all women who attend PTL clinic at 16 weeks due to the following criteria

- Previous PPROM >16 and <34 weeks
- Previous PTL >16 and <34 weeks

Inclusion criteria

- Singleton pregnancy
- Willing to undergo Transvaginal Ultrasound scan
- Age >18 years
- Understands English
- Understands and agrees to participate

Exclusion criteria

- Iatrogenic PTL or PPROM >16 and <34weeks

16 weeks attendance

- Vaginal swabs
- Urine sample
- Stool sample
- Blood for biomarkers, maternal genome and transcriptomics (plasma, serum, DNA, RNA)

NB. The participating woman may choose which samples to give if not able to give all. E.g. can give urine and vaginal swabs but not blood)

20 weeks attendance

- Vaginal swabs
- Urine Sample
- Stool Sample
- Blood for biomarkers, maternal genome and transcriptomics

NB. The participating woman may choose which samples to give if not willing to give all. E.g. can give urine and vaginal swabs but not blood, or time points 16 and 20 weeks samples.

<u>Blood taking</u>

Maternal Blood

Venepuncture of blood at 16, 20 weeks

- 6ml clot activator for biomarkers (Red)
- 6ml EDTA for maternal genome (Purple small)
- 10ml EDTA for plasma (Purple large)
- 2.5ml PAXgene RNA tube for RNA (PAXgene, red top)

Vaginal Swabs

At the time of screening patients will be sensitively asked about:

a. Any antibiotics during past two months – in pregnancy
b. Use of any vaginal preparations in past two months (eg. Progesterone)
c. Current symptoms of vaginal irritation or abnormal discharge
d. Intercourse in the past 48 hours

If current symptoms – an additional vaginal swab will be performed for assessment of active infection (BV/candidiasis) or STI screening if suggested through local hospital laboratories. The decision for clinical assessment will be made by the treating clinician. Any treatment required would be given as per local guidelines/clinician decision.

pH reading Collect vaginal fluid from the posterior fornix with a cotton swab. Press the swab on the **pH strip** (range 3.6-6.1). Discard the swab.

Collect vaginal fluid with sample collection swabs.

Urine sample

Fresh sample given at 16 and 20 weeks

- 10-20ml urine in sterelin pot

Stool sample

Sample given at home and posted to laboratory at 16 and 20 weeks, or alternatively brought to clinic and collected at appointment

- Stool sample (home collection kit containers)

Biomarker Processing

The samples will be labelled with study number, date and time of collection.

Maternal blood samples will be taken to haematology lab during or after each clinic, within 30 mins of them being taken.

Give the samples to the technician who will spin at 1500rpm for 10mins to generate plasma and serum which will then be placed in 10 aliquots and stored at -80 °C

Whole blood will be frozen without centrifugation.

PAXgene will be frozen as per manufacturer's guideline.

Serum

These blood samples will be dispatched to Perkin-Elmer laboratories, Finland for analysis at regular intervals determined by the research team. Contact for Perkin-Elmer is Mr Mikko Sairanen mikko.sairanen@perkinelmer.com

Plasma

We are working with the Centre for Structural Biology and the NMR team, University of Liverpool. Looking at NMR metabolomics analysis in plasma. Our point of contact is Dr Marie Phelan, mphelan@liverpool.ac.uk

Genomic testing

Maternal Genome

Samples of maternal blood will be stored for later analysis of maternal genotype and epigenetic environment in all women to attempt to better understand the mechanism of preterm labour and to look for a clinical biomarker or combination of biomarkers that may be used to assess preterm labour. This work will be performed by Dr Ana Alfirevic, Department of Molecular and Clinical Pharmacology, University of Liverpool, ana.alfirevic@liverpool.ac.uk

- Take 5ml of whole blood in EDTA tubes and store at -80 °C
- 2ml PAXgene whole blood RNA tube stored at -80 °C

Urine Sample

Samples of urine will be aliquoted and frozen at -80 °C for later analysis of metabolomics. This will be performed in the University of Liverpool Centre for Genomic Research facility. (CGR)

Stool samples

Samples delivered to the laboratory will be processed immediately and stored at -80 for metgenomic analysis at a later date at the University of Liverpool CGR.

Analysis performed at the CGR will be co-ordinated and supervised by Dr Christiane Hertz Fowler, chf@liverpool.ac.uk

Working in collaboration with Professor Chris Probert, chris.probert@liverpool.ac.uk to analyse stool samples using VOC analysis. **Vaginal Swabs**

Vaginal swabs taken for metagenomic analysis will be processed in the laboratory to ensure correct labelling and will be stored at -80 ^{o}C until DNA extraction.

Lead for analysis of vaginal microbiome email: j.vandewijgert@liverpool.ac.uk

Patient Data

Prior to each sample collection a case report form will be populated with demographic information, clinical/medical information and lifestyle information to allow for correct interpretation of biomarker findings.

We will analyse every sample during this project, if samples are left over then the sample will be transferred to the Liverpool Women's Tissue Bank after 5 years from the date of taking the sample.

Data Analysis

Our data layers will primarily be analysed in a univariate fashion wherever possible to minimize the probability of over-fitting. To this end we will be using mixed-models based approaches, which have the advantage of being very good at modelling (correcting) possible biases. We will accommodate longitudinal phenotypes by modelling the dependency structure between data measures obtained at different time points. In a secondary, more explorative layer we will look at bi-variate interactions on the phenotype (2,3). Finally, data will also be analysed using a modified random forest type of approach, a class of models which have the great advantage of avoiding over-fitting. This approach is not just robust with respect to over-fitting (similar to the univariate approaches mentioned above) in this implementation (RANGER, http://imbs-luebeck.de/imbs/de/taxonomy/term/1), but has the advantage of being very fast, allowing permutation for appropriate significance testing even on a multi-omics level. This is crucial when there are many possible predictors. We expect that, given the extreme phenotypes used, strong and thus clinically useful predictors will be present in the data. For a sample size of 75 individuals (25 cases <34 weeks versus 50 healthy controls) in a cross-sectional design we have a power of 80% to see an effect of 21% on the phenotypic variability at a significance level of 0.001. Given that a longitudinal analysis is generally more powerful, we are well powered to be able to identify clinically relevant predictors, if present. The data set will have a power of 80% to detect an effect size of 9% at a significance level of 0.001.

Integrative analysis of multiple data types will:

1) identify biomarkers and perform stratification analysis using information from multiple data types

2) reconstruct mechanisms of preterm birth starting from the biomarkers and subtypes identified in this research.

Data Management

Data is collected on a password protected spreadsheet that reflects the fields of the data collection form. This database is non-identifiable data. The password is only available to members of the research team. All data is stored on "Sharepoint", the University of Liverpool's web-based, collaborative platform that integrates with Microsoft Office. It is a secure document management and storage system.

Data analysis will be performed by existing research teams, each of which will be led by internationally recognized experts with expertise in each of the sequential phases of data processing, integration and analysis. Overall co-ordination of the data analysis pipeline will be performed by Prof. Sanderson, who leads the molecular networks research group and has past experience in the management of multi–centre Biomedical Systems Biology programs. Staff required to perform all stages of the data analysis pipeline are currently in place at the University of Liverpool and CETICS, with all salaries being guaranteed for the duration of the proposed research from existing funding.

In brief: the generation, processing and quality control of all DNA, RNA, miRNA and microbiome sequencing data will be performed by members of the Centre for Genomics Research, under the supervision of Dr Christiane Hertz Fowler (University of Liverpool). Archiving and integration of data from each multimodal analysis will be coordinated by Prof. Bertram Muller-Myhsok (University of Liverpool), who will act as overall coordinator of the data processing phase of the data analysis pipeline. Subsequent phases of data integration, leading to biomarker identification and predictive reconstruction of the causative mechanisms of preterm birth will be coordinated by Prof. Bertram Muller-Myhsok and communication between research leads for each phase of data processing and analysis will be ongoing throughout the program of research. Progress against proposed milestones will be monitored through quarterly meetings of coordinators from all stages of the data processing pipeline, together with Dr Care, Dr Alfirevic and a representative of the LCTU, in order to generate a quarterly progress report, which will then be presented at quarterly meetings of the main PMG. Training on data analysis and processing will be given to Juhi Gupta a PhD student with support from CGR and Computional Biology Facility (University of Liverpool).Collaboration between other research groups with the necessary expertise will be sought if required.

References

1) Capece A, Vasieva O, Meher S, Alfirevic Z, and Alfirevic A (2014). Pathway analysis of genetic factors associated with spontaneous preterm birth and pre-labor preterm rupture of membranes. PLOS One.2014 Sep 29;9(9):e108578.

2) Mehta D, Klengel T, Conneely KN, Smith AK, Altmann A, Pace TW, Rex-Haffner M, Loeschner A, Gonik M, Mercer KB, Bradley B, Müller-Myhsok B, Ressler KJ, Binder EB. Childhood maltreatment is associated with distinct genomic and epigenetic profiles in posttraumatic stress disorder. Proc Natl Acad Sci U S A. 2013 May 14;110(20):8302-7.

3) Kam-Thong T, Czamara D, Tsuda K, Borgwardt K, Lewis CM, Erhardt-Lehmann A, Hemmer B, Rieckmann P, Daake M, Weber F, Wolf C, Ziegler A, Pütz B, Holsboer F, Schölkopf B, Müller-Myhsok B. EPIBLASTER-fast exhaustive two-locus epistasis detection strategy using graphical processing units. Eur J Hum Genet. 2011 Apr;19(4):465-71.