

OPEN

# Trace biomarkers associated with spontaneous preterm birth from the maternal serum metabolome of asymptomatic nulliparous women – parallel case-control studies from the SCOPE cohort

Renato T. Souza<sup>1</sup>, Elizabeth J. McKenzie<sup>2</sup>, Beatrix Jones<sup>2</sup>, Jamie V. de Seymour<sup>2</sup>, Melinda M. Thomas<sup>2</sup>, Erica Zarate<sup>2</sup>, Ting Li Han<sup>2</sup>, Lesley McCowan<sup>2</sup>, Karolina Sulek<sup>2</sup>, Silas Villas-Boas<sup>2</sup>, Louise C. Kenny<sup>3</sup>, José G. Cecatti<sup>1</sup> & Philip N. Baker<sup>2,4</sup>

Prediction of spontaneous preterm birth (sPTB) in asymptomatic women remains a great challenge; accurate and reproducible screening tools are still not available in clinical practice. We aimed to investigate whether the maternal serum metabolome together with clinical factors could be used to identify asymptomatic women at risk of sPTB. We conducted two case-control studies using gas chromatography-mass spectrometry to analyse maternal serum samples collected at 15- and 20-weeks' gestation from 164 nulliparous women from Cork, and 157 from Auckland. Smoking and vaginal bleeding before 15 weeks were the only significant clinical predictors of sPTB for Auckland and Cork subsets, respectively. Decane, undecane, and dodecane were significantly associated with sPTB (FDR < 0.05) in the Cork subset. An odds ratio of 1.9 was associated with a one standard deviation increase in log (undecane) in a multiple logistic regression which also included vaginal bleeding as a predictor. In summary, elevated serum levels of the alkanes decane, undecane, and dodecane were associated with sPTB in asymptomatic nulliparous women from Cork, but not in the Auckland cohort. The association is not strong enough to be a useful clinical predictor, but suggests that further investigation of the association between oxidative stress processes and sPTB risk is warranted.

Spontaneous preterm birth (sPTB) due to spontaneous onset of labour or premature rupture of membranes (PROM) is a major cause of neonatal mortality and morbidity<sup>1-3</sup>. Although sPTB is prevalent in both high and low/middle-income countries, the major burden of sPTB is concentrated in Asian and African countries where about 85% of preterm births occur<sup>1,4,5</sup>. Short- and long-term consequences of preterm birth include bronchopulmonary dysplasia, neurodevelopment and cognitive impairment, retinopathy, as well as substantial impact on the functional, mental and social health of the infant and its family<sup>6-8</sup>. Despite an increase in the research conducted into sPTB and advancements in the implementation of management and prevention strategies, significant reductions of sPTB have not been achieved<sup>9,10</sup>.

Natural progesterone (oral or vaginal), 17 $\alpha$ -hydroxyprogesterone (intra-muscular), pessary and/or cervical cerclage have been variously recommended to prevent sPTB in some guidelines<sup>11-14</sup>, but their benefits are disputed. According to systematic reviews and meta-analyses, the use of such interventions in selected women can reduce the risk for sPTB, but they are less likely to reduce the risk for perinatal morbidity<sup>15-19</sup>. The identification

<sup>1</sup>Department of Obstetrics and Gynecology, University of Campinas, Campinas, Brazil. <sup>2</sup>The University of Auckland, Auckland, New Zealand. <sup>3</sup>The Department of Women's and Children's Health, Institute of Translational Medicine, Faculty of Health and Life Sciences, University of Liverpool, Liverpool, UK. <sup>4</sup>College of Life Sciences, University of Leicester, Leicester, United Kingdom. Correspondence and requests for materials should be addressed to R.T.S. (email: [renatotsouzasp@gmail.com](mailto:renatotsouzasp@gmail.com))

Received: 28 January 2019

Accepted: 9 September 2019

Published online: 23 September 2019

of a more accurate screening tool to predict which women are most at risk of sPTB could improve the selection of women that would benefit from interventions, increasing the likelihood of success and reducing the costs to the healthcare system. However, current screening tools used to identify women at risk of developing sPTB fail to accurately predict preterm birth in asymptomatic women.

Cervical length is possibly the most employed indicator for sPTB risk in clinical practice. While there is evidence for an increased risk of sPTB in women with a shortened cervix<sup>11,12</sup>, approximately 2/3 of women with short cervical length will have a term birth<sup>20,21</sup>. Chemical biomarkers for sPTB have also been sought, with fetal fibronectin (fFN), insulin-like growth factor-binding protein 1 (IGFBP-1), and interleukin-6 (IL-6)<sup>22–27</sup> among the most studied. A systematic review of sPTB biomarkers found more than 200 studies published between 1965 and 2008, reporting more than 100 potential biomarkers<sup>27</sup>. However, no single biomarker has proven to be a reliable predictor of sPTB. The authors of the systematic review concluded that there are many heterogeneities between the studies in terms of their experimental study design, timepoint of sample collection, and sample processing methods. One of the largest prospective cohort studies to date evaluated the performance of serial transvaginal cervical length measurements and quantitative vaginal fFN levels for predicting sPTB in a sample of approximately 10,000 nulliparous women with singleton pregnancies<sup>28</sup>. Despite determination of cervical length and fibronectin levels at three different gestations, the model still showed low predictive accuracy for sPTB.

sPTB is a condition with a multifactorial aetiology, a long pre-clinical phase, and adaptive mechanisms during pregnancy<sup>29</sup>. For instance, the cervical remodelling process, which involves softening and ripening of the cervix, invariably occurs in the spontaneous onset of labour and is considered its endpoint<sup>30</sup>. However, it does not occur in the same period of pregnancy for all women, which limits its potential as a marker for the prediction of sPTB<sup>31</sup>. The underlying complexity of the drivers of sPTB demands a robust analytical technique capable of taking into consideration the multiple pathways potentially affected in the development of sPTB.

Metabolomics, the study of low molecular weight compounds in a biological system, has previously been used to successfully investigate pregnancy complications such as fetal growth restriction, preeclampsia, and gestational diabetes mellitus<sup>32–35</sup>. Metabolomic measurements provide a biochemical snapshot of the physiological state of the organism<sup>36</sup>; their close relationship with the biological phenotype means they have the potential to reveal underlying mechanisms of disease. Due to the biochemical complexity of the metabolome, hyphenated analytical techniques are used. This involves subjecting sample extracts to a high-resolution separation technique such as gas chromatography (GC), high-pressure liquid chromatography (HPLC), or capillary electrophoresis (CE). Identification is then carried out using mass spectrometry (MS), or nuclear magnetic resonance (NMR).

The Preterm SAMBA study (Preterm Screening and Metabolomics in Brazil and Auckland) has been established to identify and validate metabolomic biomarkers for sPTB<sup>37</sup>, in collaboration with the SCOPE (Screening for Pregnancy Endpoints) study. SCOPE was an international multicentre prospective cohort study that collected standardized data and samples from more than 5,690 low-risk pregnant women<sup>38–40</sup>. The main objective of SCOPE was to develop a robust database and biobank to identify predictors and develop screening tests for preeclampsia, sPTB and small-for-gestational-age newborns. The current study describes the results of Preterm SAMBA phase 1, which used the SCOPE biobank to investigate the metabolomic profile, at 15- and 20-weeks' gestation, amongst SCOPE participants who went on to have sPTB, when compared to matched controls. Parallel analyses were performed using participants from the Irish and New Zealand's centres of the SCOPE study.

## Methods

**Study design and participants.** Two case-control studies were conducted, using a sub-set of women from the Auckland, New Zealand and Cork, Ireland centres of the Screening for Pregnancy Endpoints (SCOPE) study<sup>41</sup>.

The SCOPE study recruited 5,690 nulliparous low-risk women with a singleton pregnancy from New Zealand, Australia, Ireland and the United Kingdom between November 2004 and August 2008. Ethical approval was obtained from the local ethics committees in both Ireland and New Zealand (Cork: study number ECM5 (10) 05/02/08, Clinical Research Ethics Committee of the Cork Teaching Hospitals; New Zealand: study number AKX/02/00/364, Northern X Regional Ethics Committee) and all participants gave their written informed consent. Details on enrolment and inclusion criteria for the SCOPE study have been published elsewhere<sup>38–40</sup>. Collection of data and biological samples complied with standardized procedures in all participating centres and was conducted in accordance with the principles of the Declaration of Helsinki. Cases of sPTB were defined as those women who delivered before 37 weeks' gestation, and controls as those who delivered at or after 37 weeks' gestation.

### (a) Cork, Ireland.

The subset of SCOPE participants included in this study from the Cork, Ireland centre consisted of 55 cases of sPTB and 102 controls, matched to cases according to maternal age ( $\pm 3$  years) and maternal body mass index (BMI;  $\pm 5$  kg/m<sup>2</sup>)\*.

### (b) Auckland, New Zealand.

The subset of SCOPE participants included in this study from the Auckland, New Zealand centre consisted of 55 cases of sPTB and 109 controls, comprising 56 controls matched to cases according to maternal age ( $\pm 3$  years), and 53 controls matched to cases according to both maternal age ( $\pm 3$  years) and BMI ( $\pm 5$  kg/m<sup>2</sup>)\*.

(\*Our intention was to match each case to two controls, however, in Cork eight control subjects were excluded due to misclassification or lack of data; in Auckland one maternal age and BMI control was reclassified to a maternal age control and another was excluded due to technical difficulties).

**Data collection.** Demographic information including maternal age and BMI was collected at the time of recruitment ( $15 \pm 1$  weeks' gestation). History of vaginal bleeding and any infections during the pregnancy were also recorded via interview at the initial visit. The infant characteristics were recorded at birth and included birth weight, customised birth weight centile, and whether the infant was small, normal, or large for gestational age.

**Outcome.** Spontaneous preterm birth was defined as any birth that occurred before 37 weeks due to spontaneous onset of labour or premature rupture of membranes. Gestational age (GA) was estimated by last menstrual period (LMP) and confirmed by an ultrasound dating before 16 weeks of gestation. A discordance of seven or more days between LMP and ultrasound dating or an unsure LMP led to the estimation of GA exclusively by early ultrasound parameters. Term birth for the control group was determined as delivery after 37 weeks of gestation. Early sPTB before 34 weeks was considered for subgroup analysis.

**Sample collection and storage.** Maternal blood samples were collected in two 6 mL vacutainers by venepuncture at 15 and 20 weeks ( $\pm 1$  week). Following clot formation, the vacutainers were centrifuged at 3,000 rpm for 10 min at 4 °C, followed by a second centrifugation at 4,000 rpm for 10 min at 4 °C. After centrifugation, the resulting serum (supernatant) was pipetted into a sterile tube and aliquots of 250  $\mu$ L were dispensed into cryotubes. All samples were stored at  $-80$  °C. The study followed standard best practice procedures for repositories in all steps of sample collection and storage, registering all SPREC (Standard PREanalytical Coding) data accordingly in an online database<sup>42</sup>.

For this study, two 250  $\mu$ L maternal serum samples were obtained from the biobank for each participant, one from each time point; 15- and 20-weeks' gestation ( $\pm 1$  week).

**Sample preparation and extraction.** The serum samples underwent extraction and derivatization procedures based on the 2010 protocol of Smart *et al.*<sup>43</sup>. Briefly, samples were thawed on ice at 4 °C and transferred from cryotubes to 1.5 mL microcentrifuge tubes. An internal standard (IS; 20  $\mu$ L of 10 mM L-Alanine-2,3,3,4-d4) was added to all samples and the sample-IS mix was vortexed for 1 min. Samples were dried for 4 h at 0.8 HPa in a centrifugal vacuum concentrator with a  $-104$  °C refrigerated vapour trap (Thermo Fisher Scientific Savant SC250EXP SpeedVac Concentrator with Savant SP5121P Refrigerated Vapour trap). Metabolites were extracted using 50% and 80% cold methanol-water solution ( $-20$  °C, v/v). Specifically, 500  $\mu$ L of 50% cold methanol-water solution was added to all samples, followed by vortexing for 1 min and centrifugation at 3,000 rpm for 5 min at  $-4$  °C. After centrifugation, the supernatant was transferred to a fresh chilled microcentrifuge tube (kept on dry ice). Then, 500  $\mu$ L of 80% methanol-water solution was added to the pellet and this was centrifuged at 3,000 rpm for 5 min at  $-4$  °C. The supernatants obtained from both extraction steps were combined and dried in the centrifugal vacuum concentrator for 4 hours at 0.8 HPa, with a  $-104$  °C refrigerated vapour trap. Dried extracted samples were stored at  $-80$  °C prior to derivatisation.

Negative controls were produced by subjecting an empty microcentrifuge tube to the same processing as the samples. Pooled quality control samples (QC) were produced by pooling a small amount from every sample, mixing, and then making aliquots of the same volume as the samples.

Derivatization was carried out using methyl chloroformate (MCF). Samples were derivatised in batches of 18–24, on the same day that they were analysed on the GC-MS. Samples were re-suspended with 400  $\mu$ L of 1 M sodium hydroxide and were transferred to silanised glass tubes, followed by addition of 334  $\mu$ L of methanol and 68  $\mu$ L of pyridine. The sample was placed on a vortexer for the remainder of the derivatization process at  $\sim 1,500$  rpm. The rate limiting step began from the addition of 40  $\mu$ L of MCF. A second addition of 40  $\mu$ L of MCF was made, 30 sec later. After another 30 sec, 400  $\mu$ L of chloroform was added to extract the alkylated derivatives from the reaction mixture. After 10 sec, 400  $\mu$ L of sodium bicarbonate (50 mM) was added. Centrifugation was used to separate the aqueous layer from the chloroform layer. After centrifugation, the aqueous layer was removed and the remaining chloroform extract was dehydrated by the addition of sodium sulphate ( $\sim 0.3$  g). The remaining liquid was then transferred to an amber glass GC-MS vial with a glass 33  $\mu$ L insert.

**Gas chromatography – mass spectrometry (GC-MS) analysis.** The GC-MS instrument parameters were based on Smart *et al.*<sup>43</sup>, with modifications. One microliter of sample was injected for analysis. The injector was set to 290 °C in splitless mode. The column flow was maintained at 1.0 mL min<sup>-1</sup> in constant flow mode. The column was a fused silica ZB-1701 30 m long, 0.25 mm inside diameters, with a 0.15  $\mu$ m stationary phase constituting of 86% dimethylpolysiloxane and 14% cyanopropylphenyl (Phenomenex). Instrument grade helium ( $>99.99\%$ , BOC) was used as the carrier gas for the analysis. The detector was run in positive-ion, electron-impact ionisation mode, at 70 eV electron energy. Identification of compounds was carried out using mass spectra acquired in scan mode from 38 to 550 atomic mass units. The Cork samples were analysed on an Agilent 7890 A gas chromatograph coupled to an 5975 C inert mass spectrometer. The Auckland samples were analysed on a Thermo Scientific Trace GC Ultra gas chromatograph coupled to an ISQ mass spectrometer.

**Data extraction and compound identification.** Data processing was semi-automated. The raw files obtained from the GC-MS were converted into common data form (cdf) format for analysis and were deconvoluted and identified using the Automated Mass Spectral Deconvolution and Identification System (AMDIS - <http://www.amdis.net/>)<sup>44</sup> from an in-house mass spectral library for MCF-derivatised metabolites ( $\sim 210$  compounds) developed by Silas Villas-Boas. The library contained mass spectra predominantly from certified reference standards. In addition to the in-house library, the National Institute of Standards and Technology (NIST) mass spectral library (NIST14, 163,198 compounds) was also used to identify peaks in the raw chromatograms. Since AMDIS is not able to batch deconvolute with the entire NIST library, a NIST subset library was constructed employing a method developed by Elizabeth McKenzie using pooled quality controls.

The NIST subset was produced using the top five results for each feature from the Agilent Chemstation PBM (Probability Based Matching) deconvolution program. MassOmics in house software (version 2.3) was used to create the subset library. MassOmics is an R script based on the XCMS R package<sup>45</sup>, with a Windows graphic user interface (GUI) developed by Ting-Li Han. MassOmics used XCMS and the AMDIS report to integrate the peak areas for each of the identified metabolites. The summary report obtained from running the MassOmics script was then checked, and peaks with low ID hits, or with large retention time shift, and laboratory contaminants were removed.

Data was checked against negative controls to identify and remove background contaminants. Peaks that were not extracted correctly with XCMS were integrated separately using the Ion Extractor feature in MassOmics. Co-eluting peaks underwent manual integration. The relative abundance obtained for each metabolite was normalised to the internal standard (Alanine-d4). After internal standard normalisation, the remaining technical variation was corrected for by analytical batch median centering using the samples. An analytical batch was defined as ~25 injections, comprising ~18 samples, 4 QC's, one Alkane series, one negative control, and one standard mixture.

**Statistical analysis.** Statistical analysis of the normalised mass spectral data was performed using R 3.4.3 (<https://www.R-project.org>)<sup>46</sup>. Data was analysed separately for each site. Clinical predictors were analysed for univariate associations with preterm birth, and a multivariate logistic regression model predicting preterm birth with these predictors was selected using stepwise logistic regression with AIC, starting from an intercept only model, followed by backward elimination of variables with multivariate  $p > 0.05$ . Mann-Whitney tests were used to analyse the difference in metabolite levels between cases and controls. Ratios between the 20 and 15-week levels of each metabolite were also assessed in this way, to identify situations where changes from baseline rather than absolute metabolite levels were associated with case status. To adjust for multiple comparison testing, false discovery rates (FDR) were calculated for each comparison using the Benjamini-Hochberg procedure<sup>47</sup>. Metabolites with an estimated FDR  $< 0.05$  are reported.

Where potential metabolite predictors were identified, we then built logistic regression and random forest models using the selected clinical factors and metabolites. For logistic regression, we assessed whether one or more metabolites improved the area under the receiver operator characteristic curve compared to the model based on clinical predictors alone<sup>48,49</sup>. The average case-control difference in the linear predictor taken from the selected multivariate logistic regression model was also compared to average linear predictor differences from permuted data, to form a permutation test of model utility. Specifically, the model was refit to data where the response category was permuted to destroy any true association with the predictors. One thousand permutations were performed; the model was considered significantly better than random if the average linear predictor difference for the real data was better than 95% of the permuted replicates ( $p < 0.05$ ). The random forest was assessed via a permutation test based on the out-of-bag error rate.

To assess whether expanding the set of predictors might improve accuracy, the sparse partial least squares discriminant analysis (PLS-DA) method from the mixOmics package was employed<sup>50</sup>. The clinical variables, log transformed 15 and 20-week metabolite intensities, and 20 to 15-week ratios were used as candidate predictors. Up to three components with five predictors each were used. The number of components was selected based on the Mahalanobis distance balanced error rate using 10-fold cross validation, averaged over 10 repeats. This error rate was also used to assess the significance of this model via a permutation test.

The above procedures were repeated for comparing preterm birth  $< 34$  weeks to the entire control group.

## Results

**Participants.** Participant demographic and outcome characteristics are shown in Table 1. These show the success of the matching strategy: similar BMI and age in case and control groups. T-tests comparing case and control were insignificant (Cork  $p$ -values 0.68 for age, 0.47 for BMI; Auckland 0.66 and 0.54 respectively). The expected association between birthweight and gestational age category is also observed. Potential clinical predictors are shown in Table 2. For Auckland data, smoking was the only significant clinical predictor in univariate models, and the multivariate model selection also chose a model with smoking only (odds ratio = 4.9;  $p$ -value = 0.03). When considering preterm birth  $< 34$  weeks, smoking was no longer a significant predictor comparing preterm cases  $< 34$  weeks to the entire control group. For the Cork data, having vaginal bleeding before 15 weeks' was the only significant univariate predictor of sPTB ( $< 37$  weeks); multivariate model selection again chose a model with this variable only (odds ratio = 2.8;  $p$ -value = 0.006). The selected multivariate model for preterm birth  $< 34$  weeks included vaginal bleeding and being a current smoker, and history of miscarriage. However, the permutation test indicated overfitting ( $p = 0.27$ ) both of this model and the model including only current smoker and vaginal bleeding ( $p = 0.19$ ); accordingly, this model was reduced to include vaginal bleeding only, which was significantly better than random ( $p = 0.01$ ).

**Metabolites identified.** In the Cork samples, 176 compounds were detected. Of these, 77 were identified using an in-house library of reference standards and the remaining compounds were identified using mass spectrum alone (NIST 2014 mass spectral library). Of the in-house library matches, 61 were putatively identified (80–100% match to a reference standard) and 15 were tentatively identified (60–79% mass spectral match). Of the NIST14 library identifications, 25 were putatively identified (80–100% mass spectral match), 58 were tentatively identified (60–79% mass spectral match), and three were unknown ( $< 60\%$  mass spectral match).

In the Auckland samples, 142 compounds were detected. Of these, 50 were identified using an in-house library of reference standards and the remaining compounds were identified using mass spectrum alone (NIST 2014). Of the in-house library matches, 41 were identified, eight were tentatively identified, and one was unknown. Of

	Cork		Auckland	
	Term Birth N = 109	Spontaneous Preterm Birth N = 55	Term Birth N = 102	Spontaneous Preterm Birth N = 55
Maternal BMI*	23.4 (21.8, 25.8)	24.0 (21.6, 26.3)	23.5 (21.7, 26.0)	23.5 (21.5, 27.0)
Maternal Age (years)*	30 (27, 32)	30 (27, 34)	31 (28, 33)	32 (29, 34)
Birthweight (g)	3600 (3350, 3830)	2520 (2045, 2795)	3560 (3320, 3884)	2540 (2173, 2995)
Gestational age at delivery (weeks)	40.6 (39.9, 41.3)	35.3 (33.7, 36.4)	40.4 (39.3, 41.1)	35.7 (34, 36.4)
Gestational age at delivery				
<28 weeks	N/A	3 (6%)	N/A	3 (6%)
28–32 weeks		4 (7%)		2 (4%)
32–34 weeks		9 (16%)		8 (15%)
34–37 weeks		39 (71%)		42 (76%)
Preterm Premature Rupture of Membranes				
Yes	N/A	29 (53%)	N/A	25 (45%)
No		26 (47%)		30 (55%)

**Table 1.** Medians (Lower and Upper Quartiles) or Counts (%) for Matching and Outcome Variables. \*t-tests: Cork p-values 0.68 for age, 0.47 for BMI; Auckland 0.66 and 0.54 respectively.

the NIST14 library identifications, 51 were putatively identified, 34 were tentatively identified, and seven were unknown.

**Metabolites and performance of predictive models.** Table 3 shows the metabolites associated with preterm birth <37 weeks' and preterm birth <34 weeks', at each location and at each gestational time point (15 weeks or 20 weeks); ratios between the 15 and 20 week values were also assessed. Three metabolites detected at 20 weeks of gestation in the Cork subset were found to be significantly associated with sPTB (FDR < 0.05) when compared to term births: undecane, dodecane and decane. All three were found to have higher abundance in sPTB cases. Adding the natural log intensity of undecane to the clinical predictors in a logistic regression model estimates an odds ratio of 1.9 for a 1 standard deviation increase in log (undecane). This is significantly different from 1 ( $p = 0.0006$ ), and the model as a whole was significant via permutation test ( $p = 0.001$ ). The other metabolites were correlated with undecane (Pearson's correlation for log intensities  $r = 0.87$  decane,  $r = 0.89$  dodecane) (Suppl. Fig. a). Consequently, results were similar for the other metabolites added individually to the model but adding multiple metabolites did not improve the model further. Figure 1 shows the receiver operating characteristic (ROC) curves for the clinical model only (area under the curve 0.60), and the clinical model with the addition of undecane (area under the curve 0.73) predicting sPTB <37 weeks. Supplementary Fig. b shows the ROC curve for the models predicting sPTB <34 weeks (Suppl. Fig. b).

Similarly, the random forest model using log(undecane) and vaginal bleeding was significantly better than random ( $p = 0.002$ ) although the classification success was modest (70%). However, models also including log(decane) or log(dodecane), or both, had error rates similar to those based on permuted data ( $p = 0.29$ , 0.08 and 0.08, respectively).

Sparse PLS-DA also produced a 1-component model for the Cork preterm birth data with an error rate that was significantly better than error rates produced for random permutations ( $p = 0.01$ ). This model again included undecane, dodecane, decane, and vaginal bleeding. A fifth predictor, stearic acid measured at 15 weeks' gestation, was present; however, its loading was small, and it did not have an odds ratio significantly different from 1 ( $p = 0.07$ ) when incorporated into the logistic regression with vaginal bleeding and undecane.

No metabolites or 20–15 week ratios met the false discovery rate threshold for the Cork <34 weeks data, for either preterm <37 weeks or <34 weeks in the Auckland data, nor were the sparse PLS-DA models significant ( $p = 0.06$  preterm birth <34 weeks Cork;  $p = 0.07$ , preterm birth <37 weeks Auckland;  $p = 0.48$  preterm birth <34 weeks Auckland).

Log transformed Undecane did not meet the FDR threshold for association with preterm birth <34 weeks in the Cork data (compared to birth  $\geq 37$  weeks), but was associated with preterm birth <37 weeks. We therefore examined the sensitivity of our model to the threshold for preterm birth. The odds ratio associated with a 1 sd change in log(Undecane), in a model also including vaginal bleeding, when the divider between preterm birth and term birth was reduced to 36, 35 and 34 weeks were 2.1 ( $p = 0.0009$ ), 2.2 ( $p = 0.0014$ ), and 1.7 ( $p = 0.0571$ ) respectively.

## Discussion

We have analysed potential predictors of sPTB using serum samples and clinical data from Cork and Auckland participants of the SCOPE study, an international cohort of low-risk nulliparous women. An untargeted metabolomics approach was applied to serum samples collected at 15 and 20 weeks of gestation. More than one hundred metabolites were identified in each subset (Cork and Auckland). As expected for the metabolomics method employed, the most common classes of metabolites were fatty acids, followed by amino acids. Only three metabolites from the 20-week serum of Cork participants were found to be significantly associated with sPTB. Vaginal bleeding before 15 weeks, and smoking during pregnancy were the only clinical factors associated with

	Cork					Auckland				
	Term Birth N = 109	Spontaneous Preterm Birth <37w N = 55	p-value <37w vs term	Spontaneous Preterm Birth <34w N = 16	p-value <34w vs term	Term Birth N=102	Spontaneous Preterm Birth <37w N = 55	p-value <37w vs term	Spontaneous Preterm Birth <34w N = 13	p-value <34w vs term
Maternal height (cm)	165 (161, 168)	165 (160, 168)	0.99	167 (160, 169)	0.48	165 (161, 169)	165 (162, 168)	0.8	164 (162, 168)	0.53
<b>Infant sex</b>										
Female	47 (43%)	24 (44%)	>0.99	7 (44%)	1 *	47 (46%)	24 (44%)	0.9	4 (31%)	0.38*
Male	62 (57%)	31 (56%)		9 (56%)		55 (54%)	31 (56%)		9 (69%)	
<b>Smoking in pregnancy</b>										
no smoking	82 (75%)	40 (72%)	0.78 §	10 (62%)	0.33*	92 (90%)	44 (80%)	0.06* § 0.03*, †	12 (92%)	1* <sup>§</sup>
quit	17 (16%)	8 (15%)	0.66 <sup>†</sup>	3 (19%)		7 (7%)	4 (7%)		1 (8%)	
current smoker	10 (9%)	7 (13%)		3 (19%)		3 (3%)	7 (13%)		0	
Fertility Treatment	7 (6%)	2 (4%)	0.72*	2 (13%)	0.32*	11 (11%)	7 (13%)	0.92	2 (15%)	0.64*
Previous Miscarriage	15 (14%)	10 (18%)	0.61	4 (25%)	0.26*	15 (15%)	6 (11%)	0.67	2 (15%)	1*
<b>Gravidity</b>										
1	91 (83%)	43 (78%)	0.55*	12 (75%)	0.39*	73 (72%)	37 (67%)	0.36*	8 (62%)	0.35*
2	13 (12%)	10 (18%)		2 (13%)		23 (23%)	11 (20%)		3 (23%)	
3 or more	5 (5%)	2 (4%)		2 (13%)		6 (6%)	7 (13%)		2 (15%)	
Vaginal bleeding before 15 w visit	23 (21%)	23 (42%)	0.01	9 (56%)	0.01*	22 (22%)	14 (25%)	0.72	3 (23%)	1*
Any infection before 15 w visit <sup>a</sup>	26 (24%)	11 (20%)	0.72	5 (31%)	0.54*	45 (44%)	24 (44%)	>0.99	6 (46%)	1*

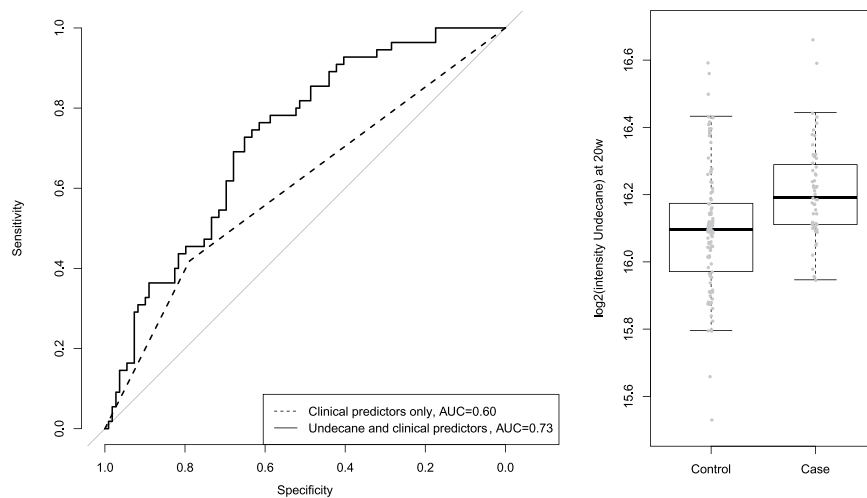
**Table 2.** Medians (Lower and Upper Quartiles) or Counts (%) for potential clinical predictors for sPTB <37 weeks and <34 weeks. P-values are from tests of case-control differences using the Mann-Whitney test (Continuous variables), Chi-squared test (categorical variables with adequate counts) or Fisher's exact test (\*). §p-value for the three categories. †p-value for current smoker vs no smoking in pregnancy. <sup>a</sup>Any infection during pregnancy means upper respiratory or urinary tract infection, pyelonephritis, gastrointestinal infection, vaginal candidiasis or other infections.

SCOPE site	Gestational age at sample collection	Metabolite	CAS number	P-Value*	FDR	Direction of association and PTB category
Auckland	15 weeks	None	—	—	—	—
Auckland	20 weeks	None	—	—	—	—
Auckland	Ratio between 15 and 20 weeks' gestation	None	—	—	—	—
Cork	15 weeks	None	—	—	—	—
Cork	20 weeks	Undecane	1120-21-4	3.73E-05	0.007	Higher in sPTB <37weeks
		Dodecane	112-40-3	8.11E-05	0.007	Higher in sPTB <37weeks
		Decane	124-18-5	7.06E-04	0.04	Higher in sPTB <37weeks
Cork	Ratio between 15 and 20 weeks' gestation	None	—	—	—	—

**Table 3.** Metabolites significantly associated with sPTB <37 weeks and <34 weeks. \*Mann Whitney U test.

sPTB, in Cork and Auckland subsets, respectively. In the Cork cohort, adding undecane to a multivariate logistic regression model for predicting sPTB improved its performance over a model with vaginal bleeding alone. This improvement is robust to the definition of preterm birth, persisting when the preterm threshold was decreased to 36 or 35 weeks; at 34 weeks the improvement is no longer significant. We note that, as depicted in Fig. 1, there are clear average differences in undecane levels between case and control groups, but also substantial overlap, limiting the utility of the metabolite measurements in clinical practice.

There is a biologically plausible explanation for observing elevated alkanes (decane, undecane, dodecane) in the serum of mothers who had sPTB. Oxidative stress may lead to degradation of cell membranes by lipid peroxidation, followed by conversion of polyunsaturated fatty acids to volatile alkanes. The association of several oxidative stress-associated processes and sPTB have been previously reported<sup>51–53</sup>, and oxidative stress has been associated with elevated alkane levels in gastroenteric disease, lung disease, and other chronic diseases of metabolism<sup>54–56</sup>. Glutathione, an important intracellular antioxidant, has been found to be decreased in maternal and umbilical cord blood of very low preterm neonates and their mothers<sup>57</sup>. Preterm birth seems to be associated with depletion of glutathione, reinforcing a possible increased oxidative status and lower antioxidant capacity. However, failure to demonstrate elevated alkanes among sPTB cases in the Auckland cohort limits our confidence, and there was no evidence these alkanes were associated with early sPTB (<34 weeks) at either study site. In addition, hypotheses describing the role of reactive oxygen species generation, metabolic and inflammatory imbalance and



**Figure 1.** Prediction model for sPTB <37 weeks: Comparison of ROC curves using clinical predictors only (vaginal bleeding) and clinical predictors with log undecane intensity for the Cork cohort; with comparison of  $\log_2$  intensity for undecane across case (Preterm Birth < 37w) and control (Term birth) individuals. Adapted from Souza RT<sup>63</sup>.

many other downstream mechanisms (telomerase reduction, cell apoptosis and senescence, etc.) caused by oxidative stress activation<sup>51</sup> do not clarify whether those mechanisms are trigger factors or consequences of underlying conditions/alterations resulting in preterm PROM and/or spontaneous onset of preterm labour.

The differences across study sites in the associations between sPTB and clinical factors such as smoking and vaginal bleeding in this case-control analysis also suggest that there are differences in the Cork and Auckland populations. The alkanes elevated in the Cork preterm birth group are present in outdoor and indoor air contaminants, which could potentially differ across study sites, providing an explanation for this observation. Association of environmental exposures with maternal and perinatal health have been reported for many years but are not yet well established<sup>58,59</sup>. Further investigation of individual pollutant exposure would be necessary to confirm whether elevated alkanes are associated with environmental exposure. It is also possible that technical rather than biological variability accounts for some of the differences observed across sites. Samples from the different sites, while analysed using the same protocols, were run on different platforms by different technical personnel. Thirty-four fewer metabolite species were detected in the Auckland samples, suggesting reduced sensitivity. In particular, two of the three metabolites elevated in cases in Cork (decane and dodecane) were not identified in the Auckland samples.

We also examined prediction of very preterm birth (<34 weeks). No clinical predictors were significant in the Auckland cohort, but vaginal bleeding before 15 weeks was associated with very preterm birth, as well as preterm birth, in the Cork cohort.

To the best of our knowledge, there are few studies in the literature applying metabolomics techniques to understand sPTB in asymptomatic pregnant women<sup>33,60–62</sup>. Previous studies have used between 20 and 70 samples of a variety of biofluids (amniotic and cervicovaginal fluid, as well as serum), equally divided between cases and controls. There is also a diversity of pregnancy stages and analytical techniques, so it is not surprising that there is little overlap in the specific compounds identified. However, Virgiliou *et al.* also suggested that the changes in amino acids and lipids they observed could be related to oxidative stress<sup>62</sup>.

Our study has strengths and limitations. Cases and controls were selected from a large cohort comprised of low-risk nulliparous women enrolled in early pregnancy and containing a high standard biobank. Several procedures were employed to assure data, sample, and analysis compliance and reliability according to Standard Operating Protocols. Shortcomings of our study include lack of data regarding cervical length, a previously reported risk factor, and assaying of the metabolome of the two cohorts at different times on different equipment. Genetic factors may also be contributors to PTB risk however these were not assessed in this study. In addition, we have not investigated predictive metabolites for the different preterm birth subtypes (spontaneous onset of preterm labour or preterm premature rupture of membranes), focusing instead on an early sPTB (<34 weeks) subgroup, due to the increased morbidity in this specific group. Using data and samples from women of different populations (Cork, Ireland; Auckland, New Zealand), enabled us to compare reproducibility of our technique and also to discuss possible local drivers for sPTB (Alkane pollutants in Cork, Ireland, for instance). Cork and Auckland samples were analysed by different laboratory experts and instruments and as such, the reproducibility and sensitivity differed across the sites.

While our finding of an association between elevated alkanes and sPTB at the Cork site is preliminary, it raises several interesting questions to pursue in the future. What differences are typical or expected across geographically distant study sites? What role might exposure to exogenous pollution sources, including alkanes, play in preterm birth? And finally, how might oxidative stress trigger, or be triggered by, processes leading to spontaneous preterm birth?

## Data Availability

The analysis in this study used the information from two centres from the SCOPE Consortium that has the ownership of data. Any requirement regarding this data can be directed to Prof. Philip N Baker.

## References

- Blencowe, H. *et al.* Born Too Soon: The global epidemiology of 15 million preterm births. *Reproductive health* **10**, S2–S2, <https://doi.org/10.1186/1742-4755-10-S1-S2> (2013).
- Blencowe, H. *et al.* Preterm birth-associated neurodevelopmental impairment estimates at regional and global levels for 2010. *Pediatric research* **74**(Suppl 1), 17–34, <https://doi.org/10.1038/pr.2013.204> (2013).
- Lawn, J. E., Cousens, S. & Zupan, J. 4 million neonatal deaths: when? Where? Why? *Lancet* **365**, 891–900, [https://doi.org/10.1016/S0140-6736\(05\)71048-5](https://doi.org/10.1016/S0140-6736(05)71048-5) (2005).
- Beck, S. *et al.* The worldwide incidence of preterm birth: a systematic review of maternal mortality and morbidity. *Bulletin of the World Health Organization* **88**, 31–38, <https://doi.org/10.2471/BLT.08.062554> (2010).
- Blencowe, H. *et al.* National, regional, and worldwide estimates of preterm birth rates in the year 2010 with time trends since 1990 for selected countries: a systematic analysis and implications. *Lancet* **379**, 2162–2172, [https://doi.org/10.1016/S0140-6736\(12\)60820-4](https://doi.org/10.1016/S0140-6736(12)60820-4) (2012).
- Patel, R. M. Short- and Long-Term Outcomes for Extremely Preterm Infants. *Am J Perinatol* **33**, 318–328, <https://doi.org/10.1055/s-0035-1571202> (2016).
- Natarajan, G. & Shankaran, S. Short- and Long-Term Outcomes of Moderate and Late Preterm Infants. *Am J Perinatol* **33**, 305–317, <https://doi.org/10.1055/s-0035-1571150> (2016).
- Saigal, S. & Doyle, L. W. An overview of mortality and sequelae of preterm birth from infancy to adulthood. *Lancet* **371**, 261–269, [https://doi.org/10.1016/S0140-6736\(08\)60136-1](https://doi.org/10.1016/S0140-6736(08)60136-1) (2008).
- Chang, H. H. *et al.* Preventing preterm births: analysis of trends and potential reductions with interventions in 39 countries with very high human development index. *Lancet* **381**, 223–234, [https://doi.org/10.1016/S0140-6736\(12\)61856-X](https://doi.org/10.1016/S0140-6736(12)61856-X) (2013).
- Zeitlin, J. *et al.* Preterm birth time trends in Europe: a study of 19 countries. *BJOG* **120**, 1356–1365, <https://doi.org/10.1111/1471-0528.12281> (2013).
- National Institute for Health and Care Excellence: Clinical Guidelines. Preterm Labour and Birth.* (2015 National Collaborating Centre for Women's and Children's Health, 2015).
- ACOG Practice Bulletin. Assessment of risk factors for preterm birth. Clinical management guidelines for obstetrician-gynecologists. Number 31, October 2001. *Obstet Gynecol* **98**, 709–716 (2001).
- Sentilhes, L. *et al.* Prevention of spontaneous preterm birth: Guidelines for clinical practice from the French College of Gynaecologists and Obstetricians (CNGOF). *Eur J Obstet Gynecol Reprod Biol* **210**, 217–224, <https://doi.org/10.1016/j.ejogrb.2016.12.035> (2017).
- Di Renzo, G. C. *et al.* Guidelines for the management of spontaneous preterm labor: identification of spontaneous preterm labor, diagnosis of preterm premature rupture of membranes, and preventive tools for preterm birth. *J Matern Fetal Neonatal Med* **24**, 659–667, <https://doi.org/10.3109/14767058.2011.553694> (2011).
- Saccone, G. *et al.* Cervical Pessary for Preventing Preterm Birth in Singleton Pregnancies With Short Cervical Length: A Systematic Review and Meta-analysis. *J Ultrasound Med* **36**, 1535–1543, <https://doi.org/10.7863/ultra.16.08054> (2017).
- Berghella, V. *et al.* Cerclage for sonographic short cervix in singleton gestations without prior spontaneous preterm birth: systematic review and meta-analysis of randomized controlled trials using individual patient-level data. *Ultrasound Obstet Gynecol* **50**, 569–577, <https://doi.org/10.1002/uog.17457> (2017).
- Jin, Z. *et al.* Cervical pessary for preventing preterm birth: a meta-analysis. *J Matern Fetal Neonatal Med*, 1–7, <https://doi.org/10.1080/14767058.2017.1401998> (2017).
- Romero, R. *et al.* Vaginal Progesterone for Preventing Preterm Birth and Adverse Perinatal Outcomes in Singleton Gestations with a Short Cervix: A Meta-Analysis of Individual Patient Data. *American Journal of Obstetrics & Gynecology*, <https://doi.org/10.1016/j.ajog.2017.11.576>.
- Jarde, A. *et al.* Effectiveness of progesterone, cerclage and pessary for preventing preterm birth in singleton pregnancies: a systematic review and network meta-analysis. *BJOG* **124**, 1176–1189, <https://doi.org/10.1111/1471-0528.14624> (2017).
- Berghella, V., Roman, A., Daskalakis, C., Ness, A. & Baxter, J. K. Gestational age at cervical length measurement and incidence of preterm birth. *Obstetrics and gynecology* **110**, 311–317, <https://doi.org/10.1097/01.AOG.0000270112.05025.1d> (2007).
- Iams, J. D. *et al.* The length of the cervix and the risk of spontaneous premature delivery. National Institute of Child Health and Human Development Maternal Fetal Medicine Unit Network. *The New England journal of medicine* **334**, 567–572, <https://doi.org/10.1056/NEJM199602293340904> (1996).
- Conde-Agudelo, A., Papageorghiou, A. T., Kennedy, S. H. & Villar, J. Novel biomarkers for the prediction of the spontaneous preterm birth phenotype: a systematic review and meta-analysis. *BJOG* **118**, 1042–1054, <https://doi.org/10.1111/j.1471-0528.2011.02923.x> (2011).
- Abbott, D. S. *et al.* Quantitative fetal fibronectin to predict preterm birth in asymptomatic women at high risk. *Obstetrics and gynecology* **125**, 1168–1176, <https://doi.org/10.1097/AOG.0000000000000754> (2015).
- Conde-Agudelo, A. & Romero, R. Cervical phosphorylated insulin-like growth factor binding protein-1 test for the prediction of preterm birth: a systematic review and metaanalysis. *American journal of obstetrics and gynecology* **214**, 57–73, <https://doi.org/10.1016/j.ajog.2015.06.060> (2016).
- van der Ven, J. *et al.* The capacity of mid-pregnancy cervical length to predict preterm birth in low-risk women: a national cohort study. *Acta Obstetrica et Gynecologica Scandinavica* **94**, 1223–1234, <https://doi.org/10.1111/aogs.12721> (2015).
- Celik, E. *et al.* Cervical length and obstetric history predict spontaneous preterm birth: development and validation of a model to provide individualized risk assessment. *Ultrasound in obstetrics & gynecology: the official journal of the International Society of Ultrasound in Obstetrics and Gynecology* **31**, 549–554, <https://doi.org/10.1002/uog.5333> (2008).
- Menon, R. *et al.* Biomarkers of Spontaneous Preterm Birth: An Overview of The Literature in the Last Four Decades. *Reproductive Sciences* **18**, 1046–1070, <https://doi.org/10.1177/1933719111415548> (2011).
- Esplin, M. S. *et al.* Predictive Accuracy of Serial Transvaginal Cervical Lengths and Quantitative Vaginal Fetal Fibronectin Levels for Spontaneous Preterm Birth Among Nulliparous Women. *JAMA* **317**, 1047–1056, <https://doi.org/10.1001/jama.2017.1373> (2017).
- Di Renzo, G. C. The great obstetrical syndromes. *The journal of maternal-fetal and neonatal medicine* **22**, 633–635, <https://doi.org/10.1080/14767050902866804> (2009).
- Preterm Birth: Causes, Consequences, and Prevention. National Academy of Sciences (2007).
- Word, R. A., Li, X. H., Hnat, M. & Carrick, K. Dynamics of cervical remodeling during pregnancy and parturition: mechanisms and current concepts. *Semin Reprod Med* **25**, 69–79, <https://doi.org/10.1055/s-2006-956777> (2007).
- Romero, R. *et al.* The use of high-dimensional biology (genomics, transcriptomics, proteomics, and metabolomics) to understand the preterm parturition syndrome. *BJOG* **113**, 118–135, <https://doi.org/10.1111/j.1471-0528.2006.01150.x> (2006).
- Horgan, R. P., Clancy, O. H., Myers, J. E. & Baker, P. N. An overview of proteomic and metabolomic technologies and their application to pregnancy research. *BJOG* **116**, 173–181, <https://doi.org/10.1111/j.1471-0528.2008.01997.x> (2009).
- Sulek, K. *et al.* Hair metabolomics: identification of fetal compromise provides proof of concept for biomarker discovery. *The ranostics* **4**, 953–959, <https://doi.org/10.7150/thno.9265> (2014).



35. He, X. *et al.* Maternal hair metabolome analysis identifies a potential marker of lipid peroxidation in gestational diabetes mellitus. *Acta diabetologica*, <https://doi.org/10.1007/s00592-015-0737-9> (2015).
36. Hollywood, K., Brison, D. R. & Goodacre, R. Metabolomics: current technologies and future trends. *Proteomics* **6**, 4716–4723, <https://doi.org/10.1002/pmic.200600106> (2006).
37. Cecatti, J. G. *et al.* Use of metabolomics for the identification and validation of clinical biomarkers for preterm birth: Preterm SAMBA. *BMC Pregnancy and Childbirth* **16**, 212–212, <https://doi.org/10.1186/s12884-016-1006-9> (2016).
38. McCowan, L. M. E. *et al.* Risk factors for small-for-gestational-age infants by customised birthweight centiles: data from an international prospective cohort study. *BJOG* **117**, 1599–1607, <https://doi.org/10.1111/j.1471-0528.2010.02737.x> (2010).
39. North, R. A. *et al.* Clinical risk prediction for pre-eclampsia in nulliparous women: development of model in international prospective cohort. *BMJ* **342**, d1875–d1875 (2011).
40. Australian New Zealand Clinical Trials Registry: ACTRN12607000551493, <https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=82254>. Accessed 27 Sep 2017 (2007).
41. Kenny, L. C. *et al.* Early pregnancy prediction of preeclampsia in nulliparous women, combining clinical risk and biomarkers: the Screening for Pregnancy Endpoints (SCOPE) international cohort study. *Hypertension* **64**, 644–652, <https://doi.org/10.1161/HYPERTENSIONAHA.114.03578> (2014).
42. Betsou, F. *et al.* Standard preanalytical coding for biospecimens: defining the sample PREanalytical code. *Cancer epidemiology, biomarkers & prevention* **19**, 1004–1011, <https://doi.org/10.1158/1055-9965.EPI-09-1268> (2010).
43. Smart, K. F., Aggio, R. B., Van Houtte, J. R. & Villas-Bôas, S. G. Analytical platform for metabolome analysis of microbial cells using methyl chloroformate derivatization followed by gas chromatography-mass spectrometry. *Nat Protoc* **5**, 1709–1729, <https://doi.org/10.1038/nprot.2010.108> (2010).
44. Automated Mass Spectral Deconvolution and Identification System - AMDIS. <http://www.amdis.net>. Accessed 20 May 2019.
45. Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R. & Siuzdak, G. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem* **78**, 779–787, <https://doi.org/10.1021/ac051437y> (2006).
46. R Development Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (2006).
47. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society* **57**, 289–300 (1995).
48. DeLong, E. R., DeLong, D. M. & Clarke-Pearson, D. L. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* **44**, 837–845 (1988).
49. Robin, X. *et al.* pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* **12**, 77, <https://doi.org/10.1186/1471-2105-12-77> (2011).
50. Le Cao, K. A., Boitard, S. & Besse, P. Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems. *BMC Bioinformatics* **12**, 253, <https://doi.org/10.1186/1471-2105-12-253> (2011).
51. Menon, R. Oxidative stress damage as a detrimental factor in preterm birth pathology. *Front Immunol* **5**, 567, <https://doi.org/10.3389/fimmu.2014.00567> (2014).
52. Venkatesh, K. K. *et al.* Inflammatory and oxidative stress markers associated with decreased cervical length in pregnancy. *Am J Reprod Immunol* **76**, 376–382, <https://doi.org/10.1111/aji.12545> (2016).
53. Moore, T. A., Ahmad, I. M. & Zimmerman, M. C. Oxidative Stress and Preterm Birth: An Integrative Review. *Biol Res Nurs* **20**, 497–512, <https://doi.org/10.1177/1099800418791028> (2018).
54. Phillips, M. *et al.* Volatile organic compounds in breath as markers of lung cancer: a cross-sectional study. *Lancet* **353**, 1930–1933, [https://doi.org/10.1016/s0140-6736\(98\)07552-7](https://doi.org/10.1016/s0140-6736(98)07552-7) (1999).
55. Phillips, M. *et al.* Detection of lung cancer with volatile markers in the breath. *Chest* **123**, 2115–2123 (2003).
56. Phillips, M. *et al.* Volatile biomarkers of pulmonary tuberculosis in the breath. *Tuberculosis* **87**, 44–52, <https://doi.org/10.1016/j.tube.2006.03.004> (2007).
57. Küster, A. *et al.* Cord blood glutathione depletion in preterm infants: correlation with maternal cysteine depletion. *PLoS One* **6**, e27626, <https://doi.org/10.1371/journal.pone.0027626> (2011).
58. Ferguson, K. K., O'Neill, M. S. & Meeker, J. D. Environmental contaminant exposures and preterm birth: a comprehensive review. *J Toxicol Environ Health B Crit Rev* **16**, 69–113, <https://doi.org/10.1080/10937404.2013.775048> (2013).
59. Royal College of Obstetricians and Gynaecologists. Chemical Exposures During Pregnancy: Dealing with Potential, but Unproven, Risks to Child Health (2016).
60. Parfieniuk, E., Zbucka-Kretowska, M., Ciborowski, M., Kretowski, A. & Barbas, C. Untargeted metabolomics: an overview of its usefulness and future potential in prenatal diagnosis. *Expert Rev Proteomics*, <https://doi.org/10.1080/14789450.2018.1526678> (2018).
61. Ghartey, J., Bastek, J. A., Brown, A. G., Anglim, L. & Elovitz, M. A. Women with preterm birth have a distinct cervicovaginal metabolome. *Am J Obstet Gynecol* **212**, 776.e771–776.e712, <https://doi.org/10.1016/j.ajog.2015.03.052> (2015).
62. Virgiliou, C. *et al.* Amniotic Fluid and Maternal Serum Metabolic Signatures in the Second Trimester Associated with Preterm Delivery. *J Proteome Res* **16**, 898–910, <https://doi.org/10.1021/acs.jproteome.6b00845> (2017).
63. Souza, R. T. A comprehensive integrative approach to investigate factors associated with preterm birth, related perinatal outcomes and its prediction using metabolomic markers (University of Campinas, 2019).

## Acknowledgements

We would like to acknowledge Rennae Taylor for assistance accessing the Auckland SCOPE biobank, and Emma Snapes, Ali Khashan and the SCOPE team at University College Cork for assistance with shipping samples. Thanks also to the staff at the Centre for Genomics, Proteomics & Metabolomics (University of Auckland), and at the Liggins Institute. We acknowledge that this manuscript is part of the RTS's PhD thesis – Post-Graduation Programme in Obstetrics and Gynaecology of Unicamp, Brazil<sup>63</sup>. This study was jointly supported by the Brazilian National Research Council (CNPq) (Award 401636/2013-5), and the Bill and Melinda Gates Foundation (Award OPP1107597- Grand Challenges Brazil: Reducing the burden of preterm birth). The SCOPE database is provided and maintained by MedSciNet AB (<http://medscinet.com>). The New Zealand SCOPE study was funded by the New Enterprise Research Fund, Foundation for Research Science and Technology; Health Research Council (04/198); Evelyn Bond Fund, Auckland District Health Board Charitable Trust. The Australian SCOPE study was funded by the Premier's Science and Research Fund, South Australian Government (<http://www.dfeest.sa.gov.au/science-research/premiers-research-and-industry-fund>). The Irish SCOPE study was funded by the Health Research Board of Ireland (CSA/2007/2; <http://www.hrb.ie>). The UK SCOPE study was funded by National Health Service NEAT Grant (Neat Grant FSD025), Biotechnology and Biological Sciences Research Council (www.bbsrc.ac.uk/funding; GT084) and University of Manchester Proof of Concept Funding (University of Manchester); Guy's and St. Thomas' Charity (King's College London) and Tommy's charity (<http://www.tommys.org/>); King's College London and University of Manchester); and Cerebra UK ([www.cerebra.org.uk](http://www.cerebra.org.uk));

University of Leeds). The current analyses were designed, sponsored and developed as part of the Preterm SAMBA study, a collaborative study group on Preterm Birth Research jointly issued by the Brazilian National Research Council (CNPq) and the Bill and Melinda Gates Foundation (Award 401636/2013-5 - Grand Challenges Brazil: Reducing the burden of preterm birth, number 05/2013). RTS has been awarded PhD scholarship from the CAPES Foundation, an agency under the Ministry of Education of Brazil, process 88881.134095/2016-01. The sponsors played no role on the study design or manuscript writing.

### Author Contributions

Baker P.N., Kenny L.C. and McCowan L. participated in the primary cohort study design and project implementation. Baker P.N., Kenny L.C. and Cecatti J.G. conceived and enabled the current study. Thomas M.M. carried out sample preparation and instrumental analysis and Sulek K. directed and supervised the analysis for the Auckland samples. Villas-Boas S. developed the analytical method for sample analysis. Zarate E. carried out sample preparation and instrumental analysis and McKenzie E.J. directed and supervised the analysis for the Cork samples. Han T.L. provided significant help in data acquisition and instrumental analysis, and produced the software used for data extraction. Thomas M.M. processed the Auckland data and McKenzie E.J. processed the Cork data; Jones B. and de Seymour J.V. carried out the statistical analysis for the study. de Seymour J.V. collated material written by Thomas M.M., McKenzie E.J., Jones B. and Souza R.T. for the first draft of the manuscript, and Souza R.T. coordinated the final draft of the manuscript. All the authors read, reviewed and approved the final version of the manuscript.

### Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-019-50252-7>.

**Competing Interests:** Baker P.N. and Kenny L.C. are minority shareholders of Metabolomics Diagnostics Ltd, a company dedicated to developing innovative screening tests using metabolomics technology.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019