Whole-transcriptome analysis of maternal blood for identification of RNA markers for predicting spontaneous preterm birth among preterm labour women: abridged secondary publication

SSC Chim *, TF Chan, TY Leung

KEY MESSAGES

1. Accurate prediction of spontaneous preterm birth (sPTB) before 37 weeks among women presenting with preterm labour may facilitate better patient care.
2. RNA-seq facilitates systematic search of markers in the transcriptome of maternal peripheral blood, which is relatively non-invasive, compared with amniotic fluid, chorionic villi, or foetal membrane.
3. We identified 68 differentially expressed transcripts between those preterm labour women ending in sPTB and those ending in term birth. Preterm labour women tested positive for two transcripts deliver significantly sooner than those tested negative.
4. The up-regulated transcripts in sPTB were over-represented with gene-ontology terms in inflammation and defence response to bacteria and other organisms.
5. Maternal peripheral blood with sPTB-associated transcripts are potentially useful for predicting and studying sPTB.

Hong Kong Med J 2020;26(Suppl 6):S20-3
HMRF project number: 01120066

[1] SSC Chim, [2] TF Chan, [1] TY Leung

The Chinese University of Hong Kong:
[1] *Department of Obstetrics & Gynaecology, Faculty of Medicine,*
[2] *School of Life Sciences*

* Principal applicant and corresponding author: sschim@cuhk.edu.hk

Introduction

Globally, over 13 million babies were born preterm (birth before 37 gestational weeks) each year. It is estimated that 10 neonates die per minute worldwide as a result of preterm birth. Only a certain proportion of women presenting with preterm labour (PTL) eventually end in spontaneous preterm birth (sPTB). Currently, the best predictive markers for sPTB are shortened cervical length and elevated cervicovaginal foetal fibronectin, but their sensitivity at high specificity is only moderate. Thus, new markers for sPTB are needed.

Discovering sPTB-associated markers is challenging, because it is unethical to obtain human placental or foetal gestational membranes, or myometrium before term from healthy pregnancies as control for comparison. We propose to use markers from maternal peripheral blood, which can be readily obtained before term.

Using strand-specific massively parallel cDNA sequencing (RNA-seq), we have systematically profiled the transcriptomes of maternal peripheral blood during the presentation of PTL. We compared the transcriptomes between sPTB and term birth (TB). We hypothesised that the blood concentration of certain RNA transcripts differed between PTL women ending in sPTB and those ending in TB. We also hypothesised that the maternal blood concentration of such RNA transcripts during PTL can be used to detect the imminent sPTB.

Methods

This case-control study was performed in two phases. With ethics approval from the institutional review board and informed consent from participants, blood samples were obtained from women during the PTL presentation. We followed up these pregnancies for delivery outcomes.

Inclusion criteria were women with uterine contractions more than once every 10 minutes before 34 weeks, an intact membrane, singleton pregnancy, and a Chinese or Korean ethnicity. Women were excluded if pregnancy was complicated with preterm prelabour rupture of membrane, multiple gestation, preeclampsia, foetal growth restriction, macrosomia, foetal distress, antepartum haemorrhage, foetal chromosomal or structural abnormalities, a history of uterine abnormality or cervical surgery, or indicated preterm births before 37 weeks (induction of labour, elective or emergency term caesarean deliveries) where delivery is iatrogenic, usually because of medical complications. Gestational age was established based on menstrual date confirmed by ultrasonographic examination prior to 20 weeks gestation.

We compared those women ending in sPTB

(<34 weeks) and those ending in TB (≥37 weeks). RNA-seq was performed on 20 women ending in sPTB (n=10) or TB (n=10). We then validated the differential expression of the 10 most-promising transcripts identified by RNA-seq using a different method on a different set of blood samples.

Results

A total of 129 PTL women were recruited. Upon quality check of the RNA using the Bioanalyzer (Agilent, Pico RNAchip), eight samples with an unacceptable low RNA integrity number were removed from analysis. Finally, 20 RNA samples of high quality were used for RNA-seq and the remaining 119 RNA samples were used for validation of the initial RNA-seq findings.

RNA-seq enabled summarisation of RNA levels at both gene levels and transcript levels or even exon levels. As one gene contains multiple transcripts and exons, multiple identifiers were generated to uniquely identify each transcript and exon in the whole genome. Each exon was uniquely identified by a 7-digit number preceded by 'e'. To account for technical variation in data (eg, different sequencing depth per sequencing library and intra- and inter-run variation), we normalised data using the established method[1] before differential expression testing.[2] After adjustment for multiple testing using the False Discovery Rate method,[3] we identified 68 differentially expressed RNA transcripts with ≥2-fold change (57 up-regulated and 11 down-regulated) between the sPTB and the TB groups (adjusted P<0.05, Fig. 1). Of these, 10 transcripts were selected for the validation study.

The validation set involved 119 RNA samples independent from the RNA-seq samples. To account for the difference in the varying amount of total RNA input, we divided the RNA level of the target transcript in a sample by the level of an empirically chosen reference transcript. We then $\log_2$-transformed that ratio and expressed it as the normalised RNA levels (Fig. 2). The normalised RNA levels in maternal peripheral blood of PTL women were significantly different between the sPTB and the TB groups in all 10 selected transcripts (Fig. 2), concordant with the RNA-seq data, suggesting that our RNA-seq dataset is valid.

The interquartile ranges of the sPTB and the TB groups did not overlap in four transcripts. The optimal cutoff in predicting sPTB was determined using the receiver operating characteristic (ROC) curve analysis. The area under the ROC curve ranged from 0.746 to 0.908. We designated the transcripts with the two greatest areas as the *PUT1* and *PUT2* mRNA, where *PUT* represents preterm up-regulated transcripts. Based on these optimal cutoffs, we achieved 92.6% (95% confidence interval (CI)=84.6%-97.2%) sensitivity and 69.6% (95% CI=57.3%-80.1%) specificity for the *PUT1* mRNA and 63.0% (95% CI=51.5%-73.4%) sensitivity and 85.5% (95% CI=75.0%-92.8%) specificity for the *PUT2* mRNA in predicting sPTB among PTL women.

To characterise the relationship of the blood levels of *PUT1* and *PUT2* mRNA transcripts and the timing of sPTB, we compared the Kaplan-Meier curve of women tested positive for these transcripts across the duration between blood sampling and delivery, with that of women tested negative. The median duration between blood taking and delivery was 5.09 days in the *PUT1*-positive women and 58.1 days in the *PUT1*-negative women (log-rank test, χ^2=43.32, degree of freedom=1, P<0.0001, hazard ratio=5.10 [95% CI=3.14-8.29], Fig. 3). Similarly, *PUT2*-positive women delivered sooner after blood sampling compared with *PUT2*-negative women (5.09 days vs 56.2 days, log-rank test, χ^2=44.66, degree of freedom=1, P<0.0001, hazard ratio=5.67 [95% CI=3.41-9.44], Fig. 3).

Discussion

PTL women with the *PUT1* and *PUT2* mRNA in their blood appear to deliver sooner. To explore whether our lists of up-regulated and down-regulated transcripts in sPTB bear any biological significance, we analysed the gene-ontology

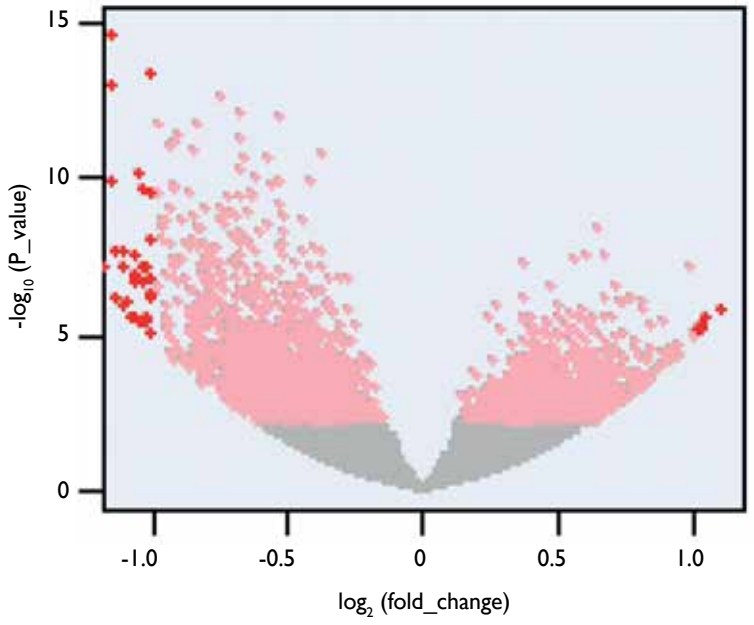

FIG 1. Volcano plot of the DESeq2 differential gene expression analysis of the RNA-seq experiment. Significantly changed transcripts between the spontaneous preterm birth (sPTB) and the term birth (TB) groups are highlighted in pink. Significantly up-regulated and down-regulated transcripts in sPTB with absolute fold-change of >2 are highlighted in red. These transcripts in red are used for validation in a different set of preterm labour women.

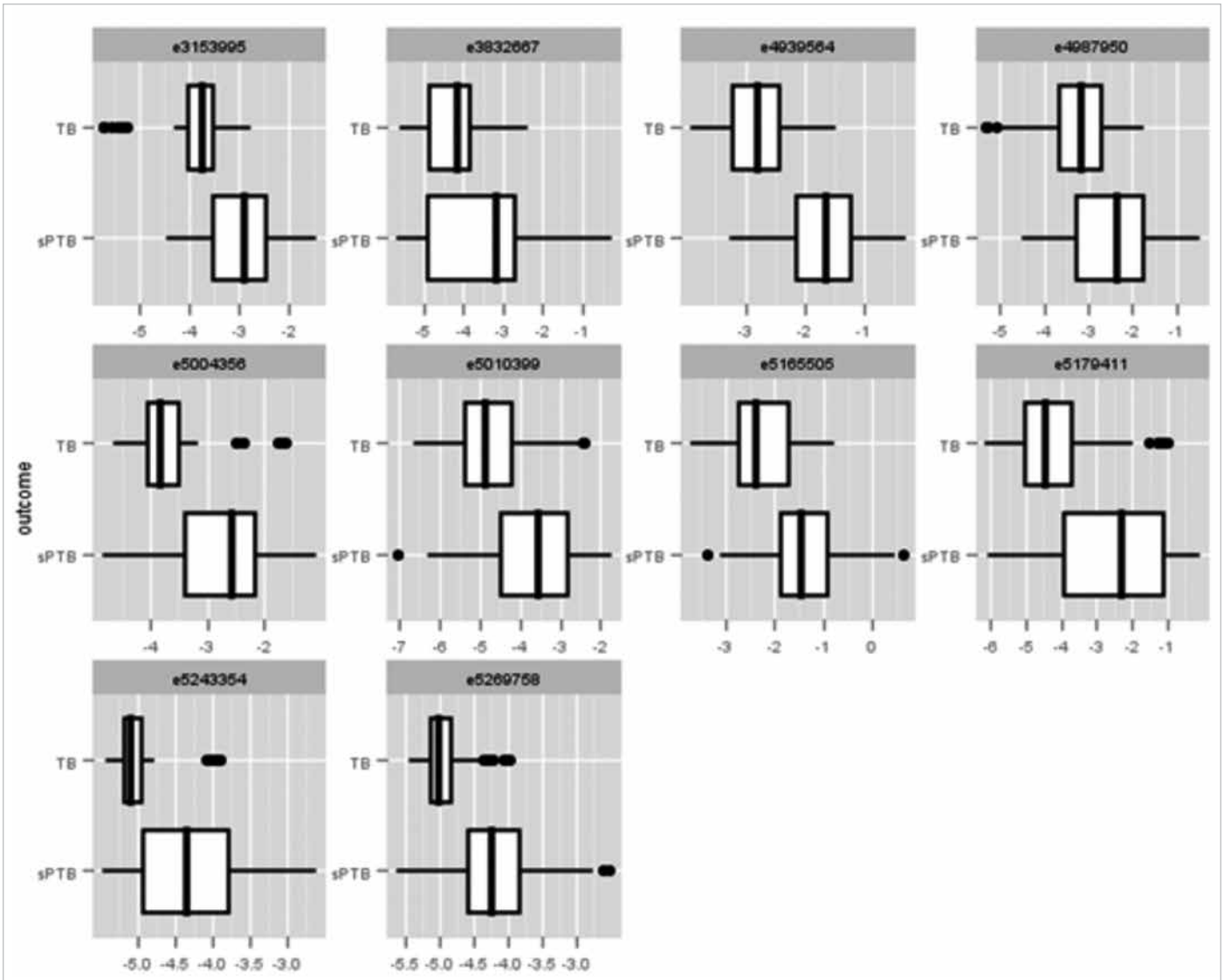

FIG 2. Box-plots of the normalised RNA levels (in log_2-scale) in the validation cohort of the 10 most promising transcripts identified using RNA-seq. Blood RNA levels are sampled in women with preterm labour and plotted according to eventual delivery outcome, either spontaneous preterm birth (sPTB) before 37 weeks or term birth (TB) on or after 37 weeks. Each assay is named by its exon ID of the transcript and normalised to one or more reference transcripts. The RNA transcripts with the greatest area under the receiver operating characteristic curve are the *PUT1* and *PUT2* mRNA.

terms annotating transcripts using the tools at the PANTHER[4] website and the gene-ontology terms in molecular function, biological process, and cellular components. Altogether, there are 329 up- and 239 down-regulated transcripts with ≥1.5-fold change in the sPTB group, compared with the TB group. Statistical over-representation tests showed that the sPTB-upregulated list comprised genes annotated more frequently with gene-ontology–biological process terms in acute inflammatory response (9.2-fold over-represented), positive regulation of defence response (4.0-fold), response to bacterium (3.6-fold), defence response to other organism (3.5-fold), innate immune response (3.1-fold), among other terms not associated with the immune system. In contrast, the sPTB-downregulated list comprised genes annotated more frequently with gene-ontology–biological process terms in complement activation (classical pathway) [16.6-fold], among a few other terms. For a more intuitive comparison, we expanded those gene-ontology–biological process terms related to the immune systems. The terms associated with innate immune response was over-represented among the sPTB-upregulated transcripts (3.1-fold), whereas those associated with humoral immune response mediated by circulating immunoglobulin (15.2-fold) was over-represented among the list of sPTB-downregulated transcripts. Apparently, there is a big difference in the immune system in the PTL women ending in sPTB, compared with those

ending in TB. Intriguingly, over-representation of terms associated with response to bacteria, defence response to other organism, and inflammation in the sPTB group are consistent with the literature on the potential pathogenesis of sPTB.

Our study is limited by a small sample size and recruitment of only East Asian women. Further studies are warranted to determine whether the predictive performance of the identified transcripts can be generalised to other ethnic groups.

Conclusion

PTL women tested positive for *PUT1* and *PUT2* in maternal peripheral whole blood are more likely to deliver sooner, compared with those tested negative. There is an over-representation of genes in inflammation and immune response against bacteria and other organisms in the list of up-regulated transcripts in PTL women ending in sPTB, compared with those ending in TB. Accurate identification of women at high risk for sPTB may facilitate appropriate management, including timely transfer to a hospital with neonatal intensive care ward and timely administration of antenatal corticosteroid to mature the foetal lungs.

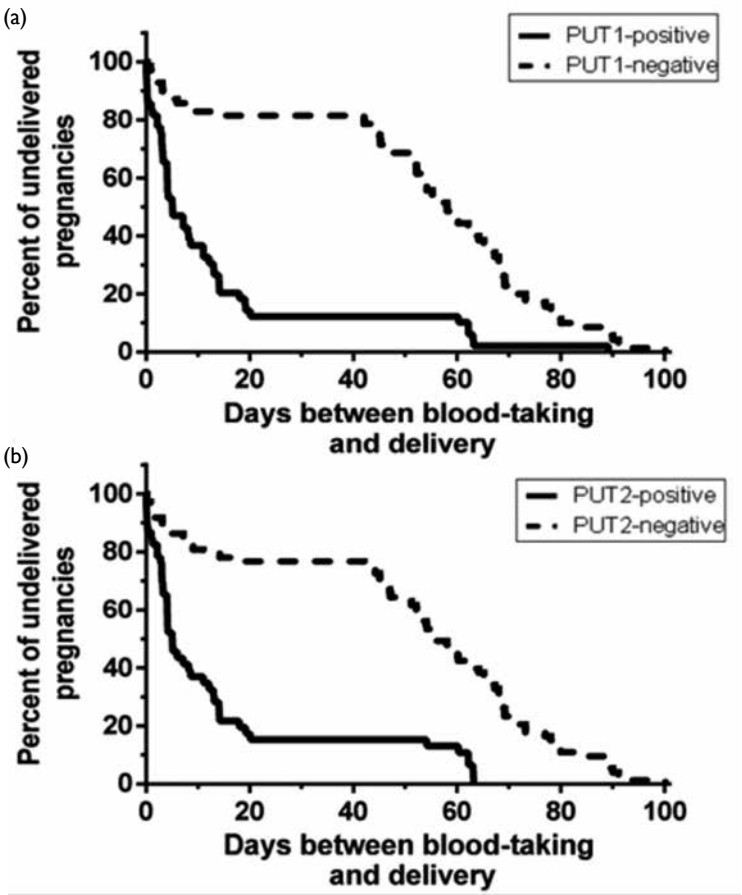

FIG 3. Percentage of undelivered pregnancies across gestational age in preterm labour women tested positive or negative for the *PUT1* and *PUT2* mRNA transcripts in maternal peripheral blood. (a) The median duration between blood taking and delivery is 5.09 days in the *PUT1*-positive women and 58.1 days in the *PUT1*-negative women (log-rank test, χ^2=43.32, degree of freedom=1, P<0.0001, hazard ratio=5.10 [95% CI=3.14-8.29]). (b) Similarly, *PUT2*-positive women delivered sooner than *PUT2*-negative women (5.09 days vs 56.2 days, log-rank test, χ^2=44.66, degree of freedom=1, P<0.0001, hazard ratio=5.67 [95% CI=3.41-9.44]).

Acknowledgements

We thank all participants for their contribution of blood samples, and local and overseas colleagues for their collection of clinical data and helpful discussion at different stages.

Funding

This study was supported by the Health and Medical Research Fund, Food and Health Bureau, Hong Kong SAR Government (#01120066). The full report is available from the Health and Medical Research Fund website (https://rfs1.fhb.gov.hk/index.html).

References

1. Risso D, Ngai J, Speed TP, Dudoit S. Normalization of RNA-seq data using factor analysis of control genes or samples. Nat Biotechnol 2014;32:896-902.
2. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014;15:550.
3. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Stat Methodol 1995;57:289-300.
4. Mi H, Muruganujan A, Casagrande JT, Thomas PD. Large-scale gene function analysis with the PANTHER classification system. Nat Protoc 2013;8:1551-66.