NC STATE UNIVERSITY

NC STATE College of Veterinary Medicine

BIOINFORMATICS RESEARCH CENTER

Meta-Analysis of the Vaginal Microbiome Reveals Cross-Study Signals of Preterm Birth Risk During Gestation

Caizhi Huang[1], Craig Gin[2], Betsy Foxman[3], Jennifer Fettweis[4], Bizu Gelaye[5],, David MacIntyre[6], Akila Subramaniam[7], William D Fraser[8], Negar Tabatabaei[8,9], Benjamin Callahan[1,2]

[1]NC State University, Bioinformatics Research Center, [2]NC State University, Department of Population Health and Pathobiology, [3]University of Michigan, Thomas Francis School of Public Health, [4]Virginia Commonwealth University, Department of Obstetrics and Gynecology, [5]Harvard University, Harvard T. H. Chan School of Public Health, [6]Imperial College London, Department of Metabolism, Digestion and Reproduction, [7]University of Alabama at Birmingham, Obstetrics & Gynecology and Maternal-Fetal Medicine, [8]University of Sherbrooke, Departments of Obstetrics and Gynecology, [9]Department of Pharmacology and Regenerative Medicine, University of Illinois College of Medicine

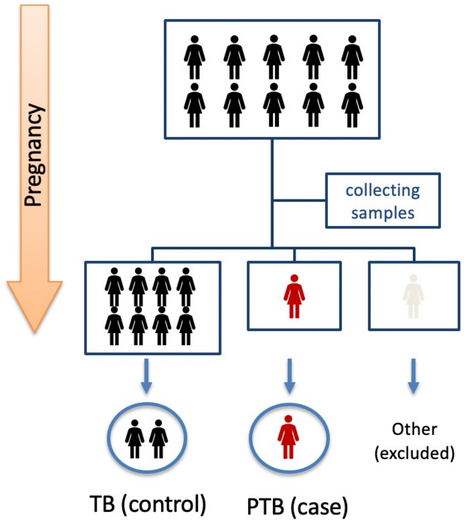
Abstract

Background: **Preterm birth** (PTB) is the primary cause of neonatal morbidity and mortality. Dozens of studies have been conducted to link the vaginal microbiome composition with PTB, resulting in mixed and sometimes inconsistent findings. **Heterogeneity** across studies along with generally low power has almost certainly contributed to the mixed findings. A cross-study comparison is needed to understand the **reproducibility** and the **consistent associations** across datasets.

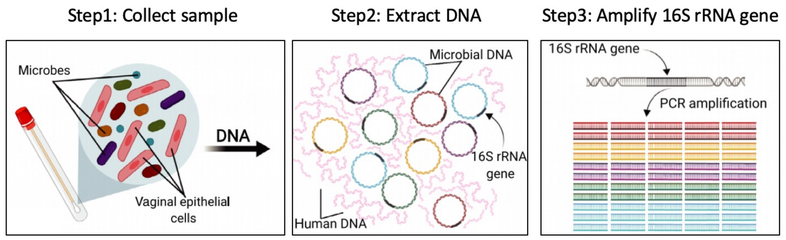
Methods: We performed a **meta-analysis** of 12 prospective case-control PTB studies using **16S rRNA gene sequencing** to measure the vaginal microbiome. All together, these 12 studies included 1926 pregnant women, 568 of whom went on to deliver preterm. We used a **machine learning** approach to investigate the predictability of PTB from the composition of the vaginal microbiome and further examined the cross-dataset reproducibility of these predictions and investigated the factors that affect the prediction accuracy.

Case-control design is used to study PTB

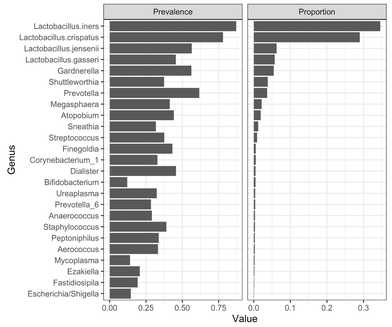
(a) Cohorts of pregnant women were recruited prospectively;
(b) Vaginal swabs were collected during the pregnancies;
(c) Birth outcomes were recorded;
(d) 16S rRNA gene sequencing was performed on a subset of women selected to meet pre-specified inclusion criteria.



Census of the vaginal microbiome - 16S rRNA gene sequencing: who is there and how much of each

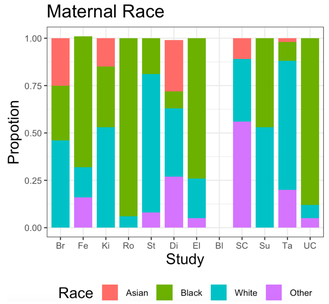


(Berman et al, 2019)

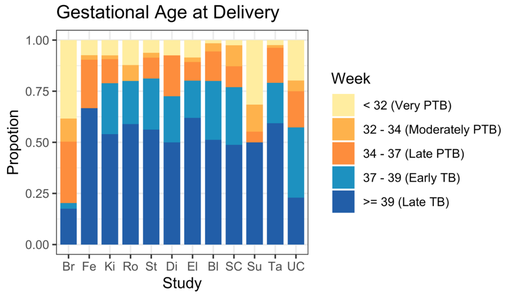


Published studies of the vaginal microbiome in term and preterm births are highly heterogeneous

Population



Definition of PTB/TB



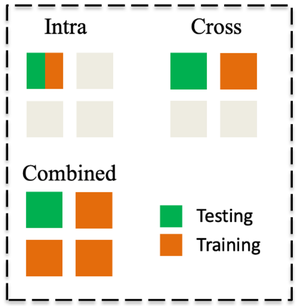
Study size/Technical choice

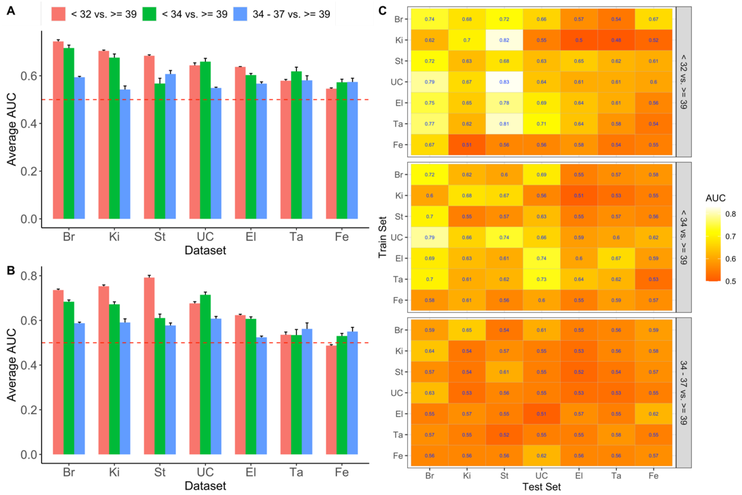
Study	Seq. region	No. of subject*	No. of sample
Br	V1-V2	141 | 36	165
Fe	V1-V3	45 | 90	546
Ki	V1-V3	34 | 127	161
Ro	V1-V3	18 | 72	349
St	V1-V3	24 | 104	158
Di	V3-V5	11 | 29	996
El	V3-V4	107 | 432	1505
Bl	V4	25 | 100	125
SC	V4	9 | 32	897
Su	V4	12 | 19	39
Ta	V4	94 | 356	450
UC	V4	41 | 55	1282

*No. of subject: preterm birth | term birth

Earlier preterm birth Is more predictable than late preterm birth

Schematic analytical strategy using machine learning





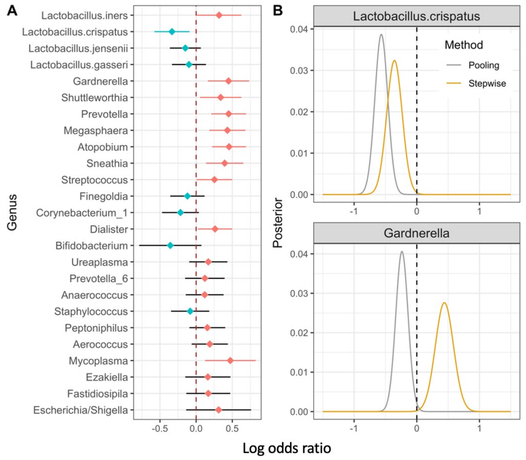
Assessment of prediction performance for different preterm birth groups using intra-dataset analysis (**A**), combined-dataset analysis (**B**) and cross-dataset analysis (**C**).

Risk of PTB is associated with microbiome

(a) Several aggregated risk factors of PTB including

- Decreasing Lactobacillus.crispatus
- Increasing Lactobacillus.iners, Gardnerella, Shuttleworthia, Prevotella, Megasphaera, Atopobium, et al.

(b) Accommodating dataset-specific effects is necessary to aggregate the effect of each microbe has on PTB across datasets.



A: point estimates and 95% confidence interval of log odds ratio between a genus present and absent using a generalized linear mixed model. **B**: posterior distribution of log odds ratio using pooling and stepwise methods for Lactobacillus.crispatus and Gardnerella.

Discussion and Conclusion

The association between the composition of the vaginal microbiome and preterm birth is obscured by differences in study population characteristics, study sizes, analysis methods and even the definitions of preterm and term birth. In this work, by using a machine learning framework to perform a meta-analysis of multiple high-throughput sequencing datasets of the vaginal microbiome, we found that the definitions of preterm and term birth affect the cross-dataset reproducibility and associations between some genera and preterm birth.

We were able to distinguish early PTB from late TB solely using microbial features with a modest to high prediction accuracy but could hardly distinguish late PTB from late TB. These findings might provide opportunity to do early detection of risk for preterm birth and develop new strategies for prevention and intervention.

Acknowledgement

We would like to thank all the collaborators who helped to collect all the sequence data and meta data. We also thank Dr. Craig Gin who contributes on the data analysis.