

Testing Association between Cervical Electrical Impedance Spectroscopy Features and Metabolite Features

David Tian

1. Previous Work on Metabolite Analysis

In Stafford Graham P. et al. [1]:

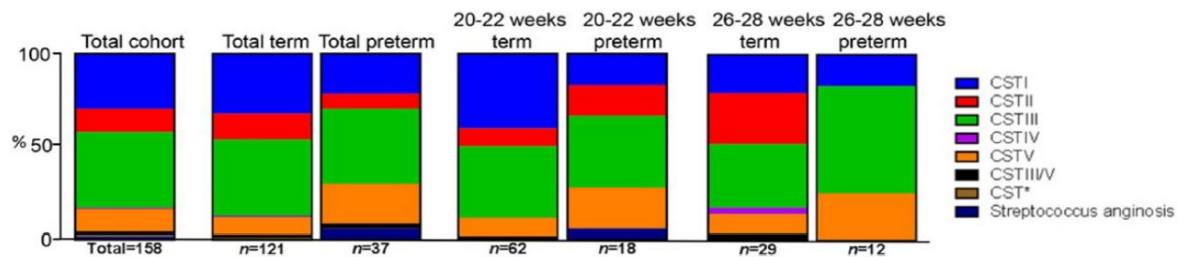


FIGURE 6 | Stacked bar-charts showing CST within total and sub-cohorts showing % of total [number of women in each group (*n*) shown] in each of the CSTs assigned where at least 60% of a sample is composed of a single species of *Lactobacillus* or at least 10% greater than the next most abundant, CSTI (*L. crispatus*, blue), CSTII (*L. gasseri*, red), CSTIII (*L. iners*, green), CSTIV (mixed anaerobes, purple), CSTV (*L. jensenii*, orange), CSTIII/V (50/50 mixed population), Group B (Milleri) *Streptococci* (navy blue) in relation to OTU within normalized trimmed sequencing samples. Total sample population, *n* = 158. CST* (*Lactobacillus delbrueckii* and *L. helveticus*).

Results show preterm women have more CST V (orange bar) than term women at 20-22 weeks. This implies CST V dominates the preterm class.

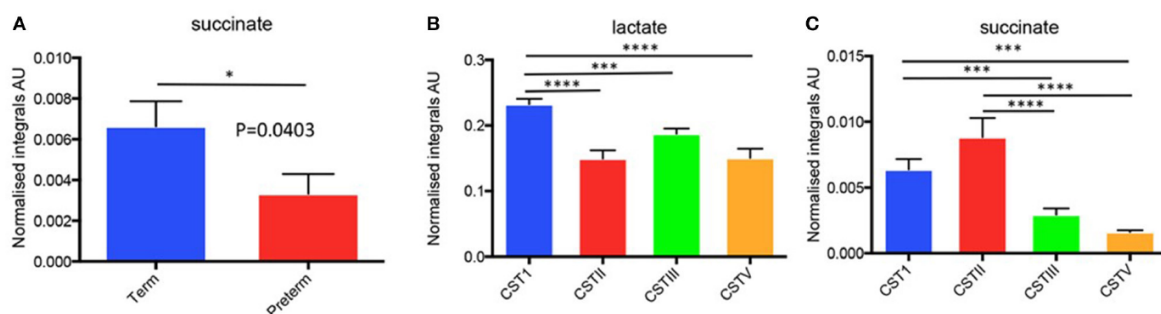


FIGURE 8 | (A) Normalized metabolite data for succinate for preterm vs. term in 26–28 week cohort and (B,C) Metabolite integrals plotted in relation to CST for the total study cohort (i.e., 20–22 and 26–28 week combined). **P* < 0.05, ****P* < 0.001, *****P* < 0.0001.

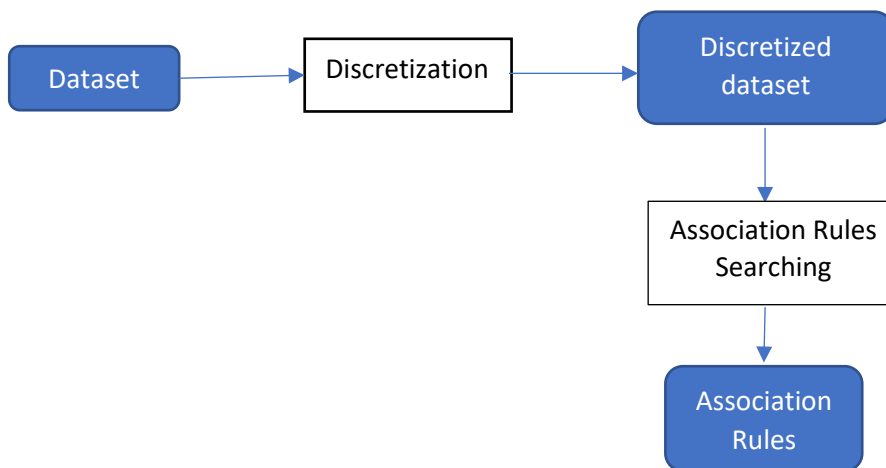
Results show CSTV has the smallest amount of Lactate and Succinate compared with the other CSTs.

Conclusion: CST V dominance and small amount of Lactate and/or Succinate in the vaginal fluid could increase the risk of PTB.

Reference:

[1] Stafford Graham P. et al. “Spontaneous Preterm Birth Is Associated with Differential Expression of Vaginal Metabolites by Lactobacilli-Dominated Microflora”, *Frontiers in Physiology*, vol 8, 2017.

2. Search of Association between Lactate/Succinate and EIS Features



2.1 Discretization of Numeric Features (Data Pre-processing Step)

1. Split the range of values of a numeric feature into smaller ranges (intervals).
2. Map the numeric values to corresponding intervals.
3. Represent the intervals using discrete values e.g. 'small', 'medium' or 'large'.

Equal width discretization method: Transform the range of values of a feature to K intervals of equal size.

The discretized LACTATE feature with 3 intervals (discrete labels):

LACTATE: $(-\infty, 26662145.666667]$, $(26662145.666667, 52352848.333333]$, $(52352848.333333, \infty)$
Discrete labels: 'Small' 'Medium' 'High'

Discrete data file: **asympt_22wks_filtered_data_28inputs_equal_width_discretized.csv**

References:

[Data Preparation \(Sections on Equal Width Binning and Discretization in Weka software\)](#)

[Weka 3 - Data Mining with Open Source Machine Learning Software in Java \(waikato.ac.nz\)](#)

2.2 Search of Association Rules using Apriori Algorithm

Use the Apriori algorithm of Weka machine learning software to find associations between EIS feature(s) and Metabolite feature(s). The association rules found represent associations between the itemsets. The association rules also represent possible associations between the features (the association exists if cramer's V > 0.2).

LACTATE Association Rules:

Fisher's Exact Test of Independence is used. Significance level 0.05 is used.

1. **Amplitude2='(-inf-78.483933]' ==> LACTATE='(-inf-26662145.666667]'**
Interpretation: Amplitude2 is associated with LACTATE

Amplitude2 and Lactate have no missing values.
 p-value = 0.02183, significant
 Cramer's V = 0.2505

2. Amplitude3='(-inf-72.047]' ==> LACTATE='(-inf-26662145.666667]'

Amplitude3 has no missing values.
 p-value = 0.02183, significant
 Cramer's V = 0.2505

3. Amplitude1='(-inf-99.386467]' and Amplitude2='(-inf-78.483933]' ==> LACTATE='(-inf-26662145.666667]'

Association of Amplitude1 and Lactate: p-value = 0.2568

4. Amplitude1='(-inf-99.386467]' and Amplitude3='(-inf-72.047]' ==> LACTATE='(-inf-26662145.666667]' not significant

5. Amplitude2='(-inf-78.483933]' and Amplitude3='(-inf-72.047]' ==> LACTATE='(-inf-26662145.666667]'

Step 1: Test Association between Amplitude2 and LACTATE: p-value=0.02183, strength: Cramer's V = 0.2505

Step 2: Test Association between Amplitude3 and LACTATE: p-value=0.02183, strength: Cramer's V = 0.2505

Step 3: Test Association between Amplitude2 and Amplitude3: p-value < 2.2e-16, strength: Cramer's V: 1

Therefore, the association between the pair: (Amplitude2, Amplitude3) and Lactate is significant and with strength approximately equals (0.2505+0.2505)/2.

6. Amplitude2='(-inf-78.483933]' and Amplitude4='(-inf-75.442567]' ==> LACTATE='(-inf-26662145.666667]' not significant

7. Amplitude2='(-inf-78.483933]' and Amplitude5='(-inf-54.804067]' ==> LACTATE='(-inf-26662145.666667]' not significant

8. Amplitude2='(-inf-78.483933]' and Amplitude6='(-inf-60.790133]' ==> LACTATE='(-inf-26662145.666667]' not significant

9. Amplitude2='(-inf-78.483933]' and Amplitude7='(-inf-71.501633]' ==> LACTATE='(-inf-26662145.666667]' not significant

10. Amplitude2='(-inf-78.483933]' and Amplitude8='(-inf-68.2614]' ==> LACTATE='(-inf-26662145.666667]' not significant

SUCCINATE Association Rules

1. Phase1='(-inf-6.209667]' ==> SUCCINATE='(-inf-1631213.333333]'
 p-value =1 when the observations with missing values '?' are removed

2. Amplitude7='(-inf-71.501633]' and Phase1='(-inf-6.209667]' ==> SUCCINATE='(-inf-1631213.333333]'
3. Amplitude8='(-inf-68.2614]' and Phase1='(-inf-6.209667]' ==> SUCCINATE='(-inf-1631213.333333]'
4. Amplitude9='(-inf-51.4239]' and Phase1='(-inf-6.209667]' ==> SUCCINATE='(-inf-1631213.333333]'
5. Amplitude10='(-inf-33.975233]' and Phase1='(-inf-6.209667]' ==> SUCCINATE='(-inf-1631213.333333]'
6. Amplitude11='(-inf-17.326567]' and Phase1='(-inf-6.209667]' ==> SUCCINATE='(-inf-1631213.333333]'
7. Amplitude12='(-inf-8.3441]' and Phase1='(-inf-6.209667]' ==> SUCCINATE='(-inf-1631213.333333]'
8. Amplitude13='(-inf-4.087067]' and Phase1='(-inf-6.209667]' ==> SUCCINATE='(-inf-1631213.333333]'
9. Amplitude7='(-inf-71.501633]' and Amplitude8='(-inf-68.2614]' and Phase1='(-inf-6.209667]' ==> SUCCINATE='(-inf-1631213.333333]'
10. Amplitude7='(-inf-71.501633]' and Amplitude9='(-inf-51.4239]' and Phase1='(-inf-6.209667]' ==> SUCCINATE='(-inf-1631213.333333]'

Succinate Association Rules of Hotspot Algorithm

Phase12='(-29.806567-inf)' and Phase2='(-11.059133-28.450833]' ==> SUCCINATE='(-inf-1631213.333333]'

Succinate has 1 missing value. The observation with missing value for Succinate is not counted in testing association and Cramer's V calculation.

Phase12: p-value = 0.07319 (Fisher Test)

Phase2: p-value = 0.5964 (Fisher Test)

Phase11='(-28.813-inf)' and Phase2='(-11.059133-28.450833]' ==> SUCCINATE='(-inf-1631213.333333]'

Phase12='(-29.806567-inf)' ==> SUCCINATE='(-inf-1631213.333333]'

Phase12: p-value = 0.07319 (Fisher Test)

Phase11='(-28.813-inf)' ==> SUCCINATE='(-inf-1631213.333333]'

Phase11: p-value = 0.08889

Reference:

[Weka - Association - Tutorialspoint](#)

[Cramér's V - Wikipedia](#)

Cramer V R library: [cramer: Calculate Cramer's V for categorical variables in greybox: Toolbox for Model Building and Forecasting \(rdrr.io\)](#)