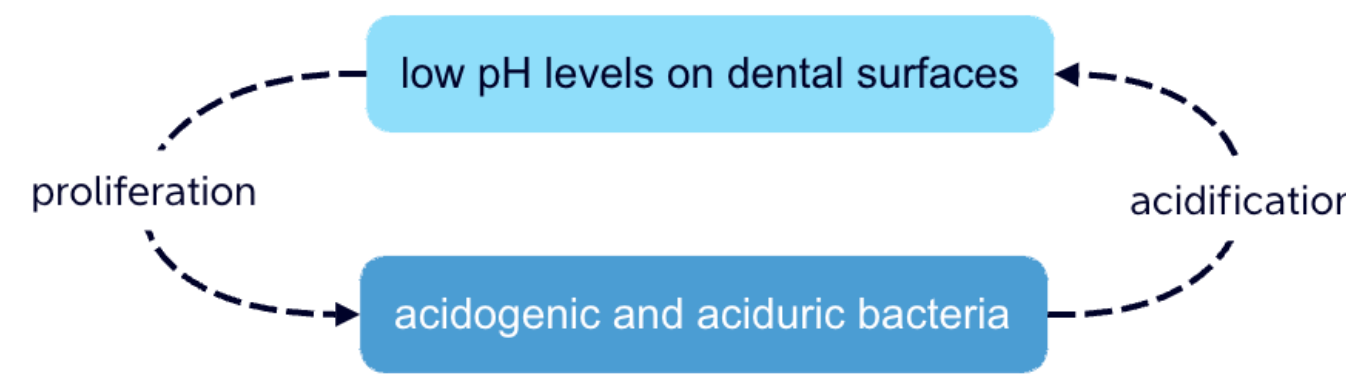


DCM Analysis of Microbiome and Metabolome in Oral Biofilm Reveals Key ECC-associated Interaction Modules

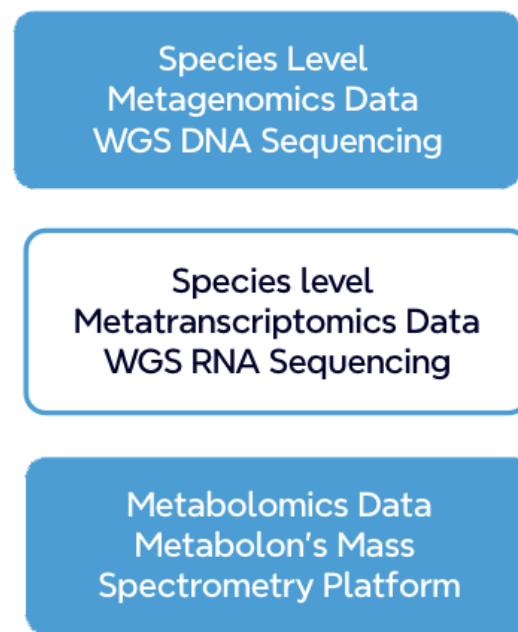
Yilan (Maya) Yin, Di Wu
UNC Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC

Introduction

- Early Childhood Caries (ECC) is a significant public health challenge with a global prevalence of approximately 50% among preschool children [1].
- ZOE 2.0 cohort [2, 3]:
 - public school children
 - aged 3 to 5 years
 - from North Carolina, United States



From dental plaque biofilm sample:



Goal: locate cliques of differentially correlated features/species across metagenomics, metatranscriptomics, and metabolomics between the group with and without Early Childhood Caries (ECC) using Differential Correlation Mining (DCM) analysis

Method and Materials

Differential Correlation Mining (DCM)

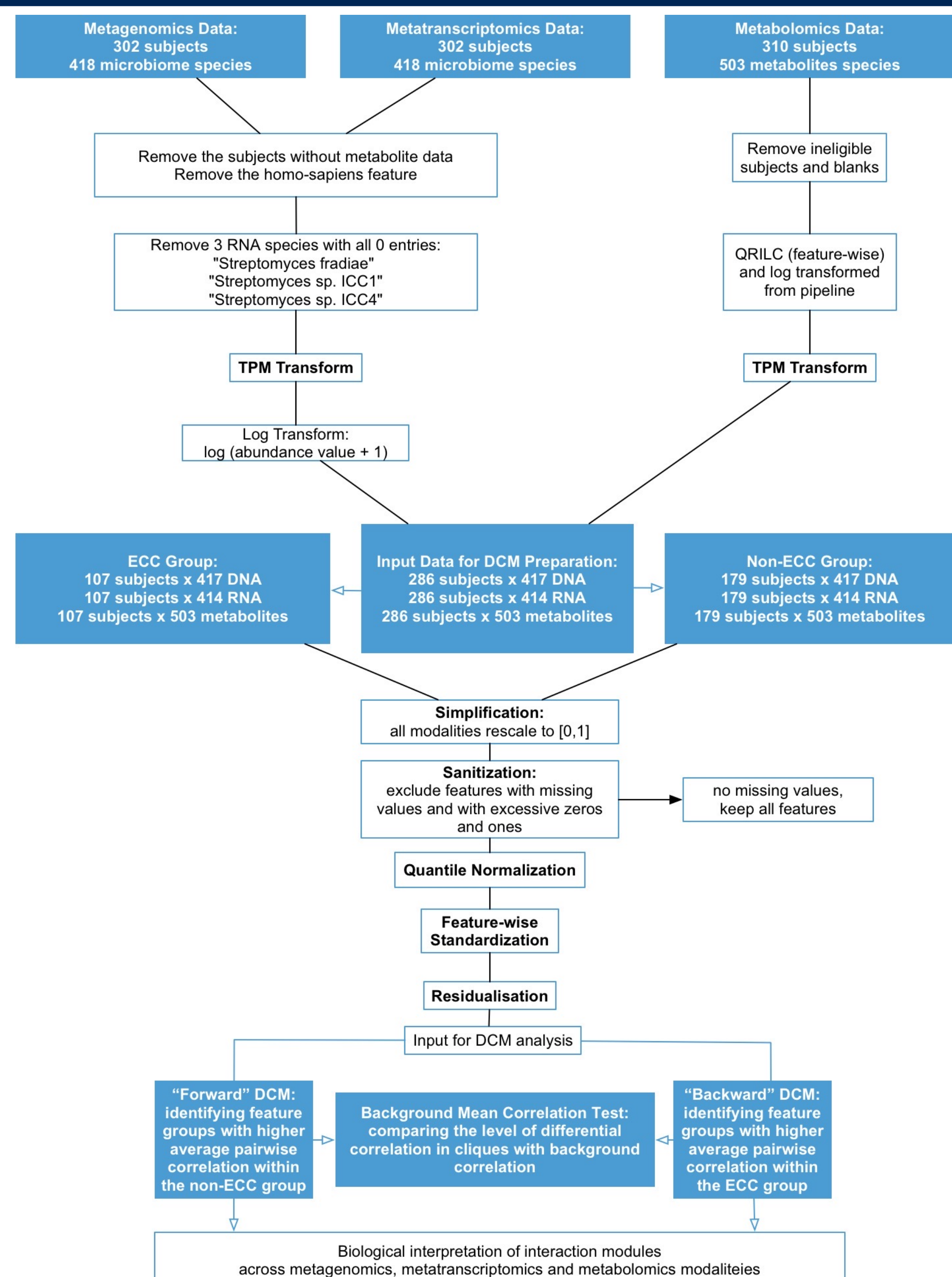
Locate groups of variables, as fully interconnected nodes, across two distributions that exhibit differential correlation based on the mean pairwise correlation [4].

DCM core search process:

Adjustments made for multi-modality microbiome-metabolome data:

- Transcripts-per-Million (TPM) transformation to normalize data within modality
- 90% termination criteria to avoid oscillation between two candidate cliques and control within the iteration limit
- Absolute value mining to highlight strong correlations in both positive and negative directions

Workflow



DCM identifies two stable cliques without applying the TPM transformation. The forward clique successfully converges within 50 iterations. The backward clique has consistent features selected across multiple runs.

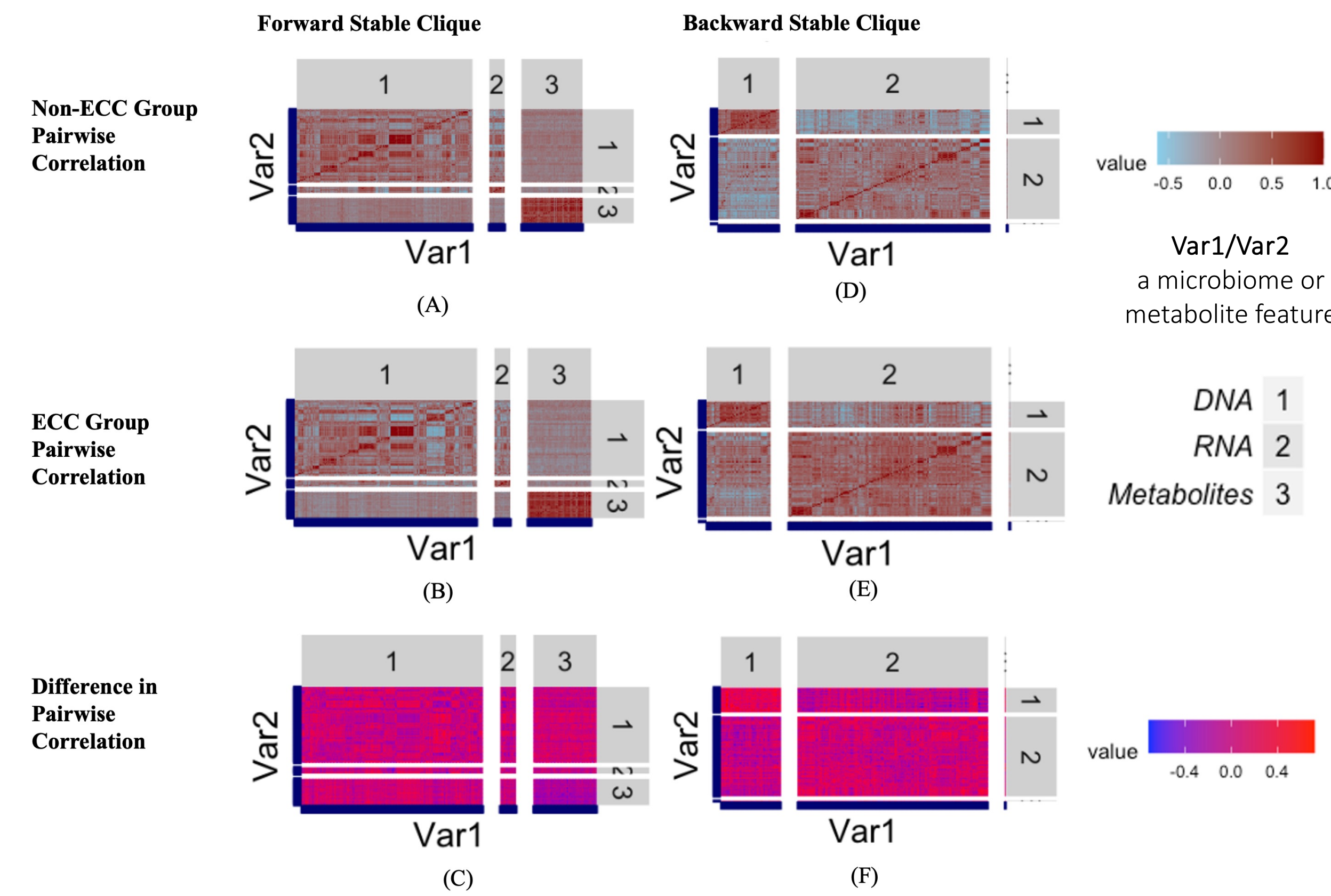


Figure 1 Pairwise correlation plots for stable clique in forward (A-B) and backward direction (D-E). The difference plot (C and F) calculated by the difference in each pairwise correlation between the non-ECC and ECC group provides a better way to visualize the differential correlation pattern. Metabolites such as N-acetylaspartate (NAA) and 3-(4-Hydroxyphenyl) lactate (HPLA), known to contribute to ECC classification, are in the forward clique [5, 6]. Half of the 16 species previously identified as significantly associated with dental caries experience are included in the forward clique and 11 of the 16 in the backward clique [2].

The termination criterion is changed to at least a 90% overlap in selected features with those in the candidate clique from the previous iteration. Initializing with 20 features while restricting the alpha level to 0.001, DCM can locate smaller cliques. A backward clique with features from all three modalities is shown below. The pathways enriched suggested by metabolites and microbiomes are highly aligned.

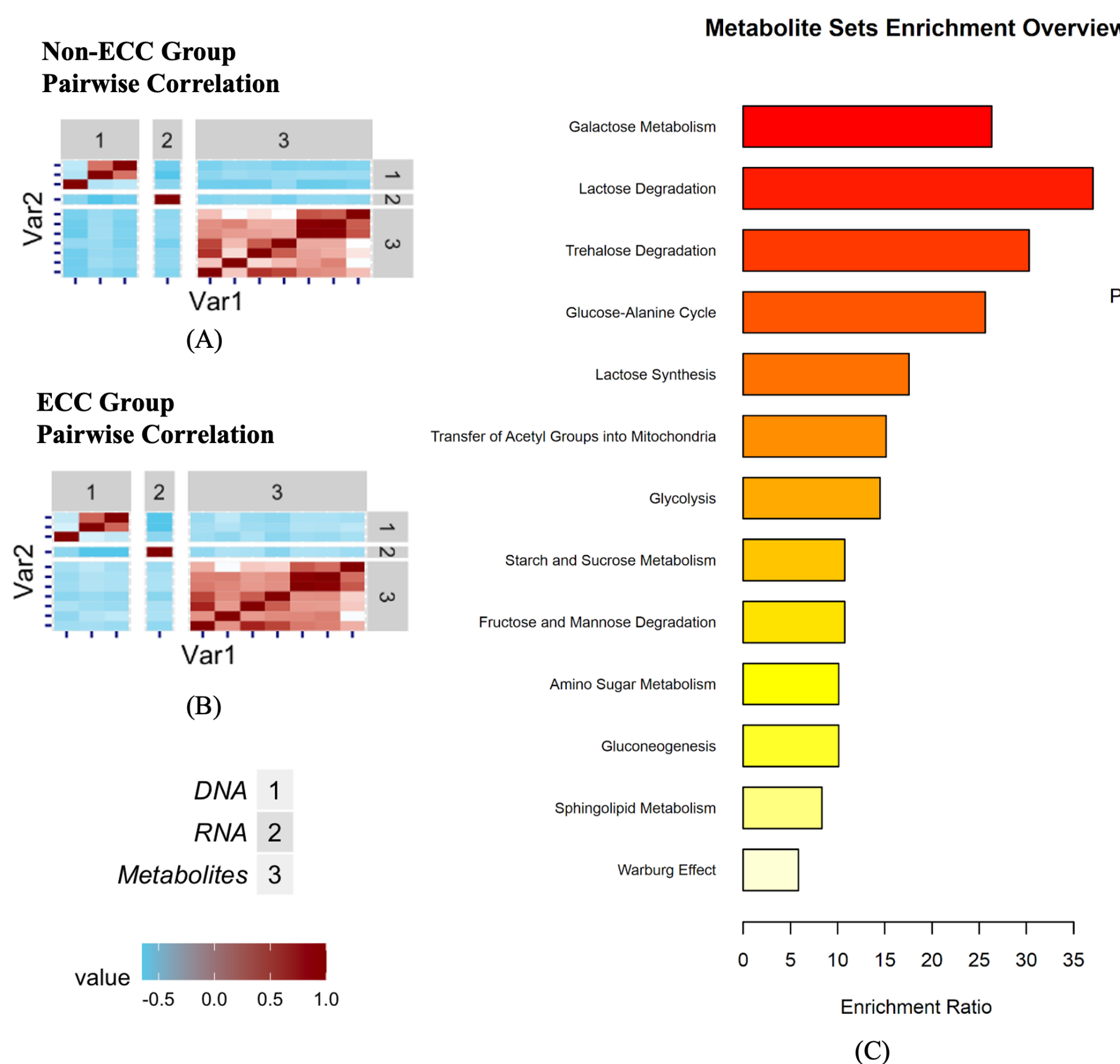


Figure 2 The pairwise correlation pattern for a backward clique using the 90% termination criteria in non-ECC individuals (A) and ECC individuals (B). (C) is generated using MetaboAnalyst [7], showed the enriched pathways suggested by metabolites included in this clique accompanied by p-values. Microbiome species in the clique – *Veillonella sp. oral taxon 780*, *Parvimonas microa*, and *Treponema socranskii* – are involved in pathways including galactose metabolism, gluconeogenesis, glycolysis, and fructose degradation [8].

Results

Table 1 Summary for differentially correlated cliques presented, including direction, size, the number of features coming from each modality, and mean feature-wise correlations in the non-ECC and ECC groups.

	Direction	Clique Size	Modality (#features)	Mean Correlation in Non-ECC Group	Mean Correlation in ECC Group		
Stable (Figure 1)	Forward (correlation higher among the non-ECC)	334	DNA (232)	0.189	0.044		
			RNA (21)				
			Metabolites (81)				
90% Termination Criteria (Figure 2)	Backward (correlation higher among the ECC)	189	DNA (45)	0.032	0.222		
			RNA (143)				
			Metabolites (1)				
Absolute Value (Figure 3)	Forward	75	DNA (3)	0.155	0.391		
			RNA (1)				
			Metabolites (7)				
			DNA (39)			0.697	0.660
			RNA (28)				
			Metabolites (8)				
62	DNA (33)	0.696	0.659				
	RNA (18)						
	Metabolites (11)						

Absolute values of correlation coefficients are used to find cliques with high positive and negative differential correlations. The top negative and positive correlation pairs driving the differential pattern are highlighted.

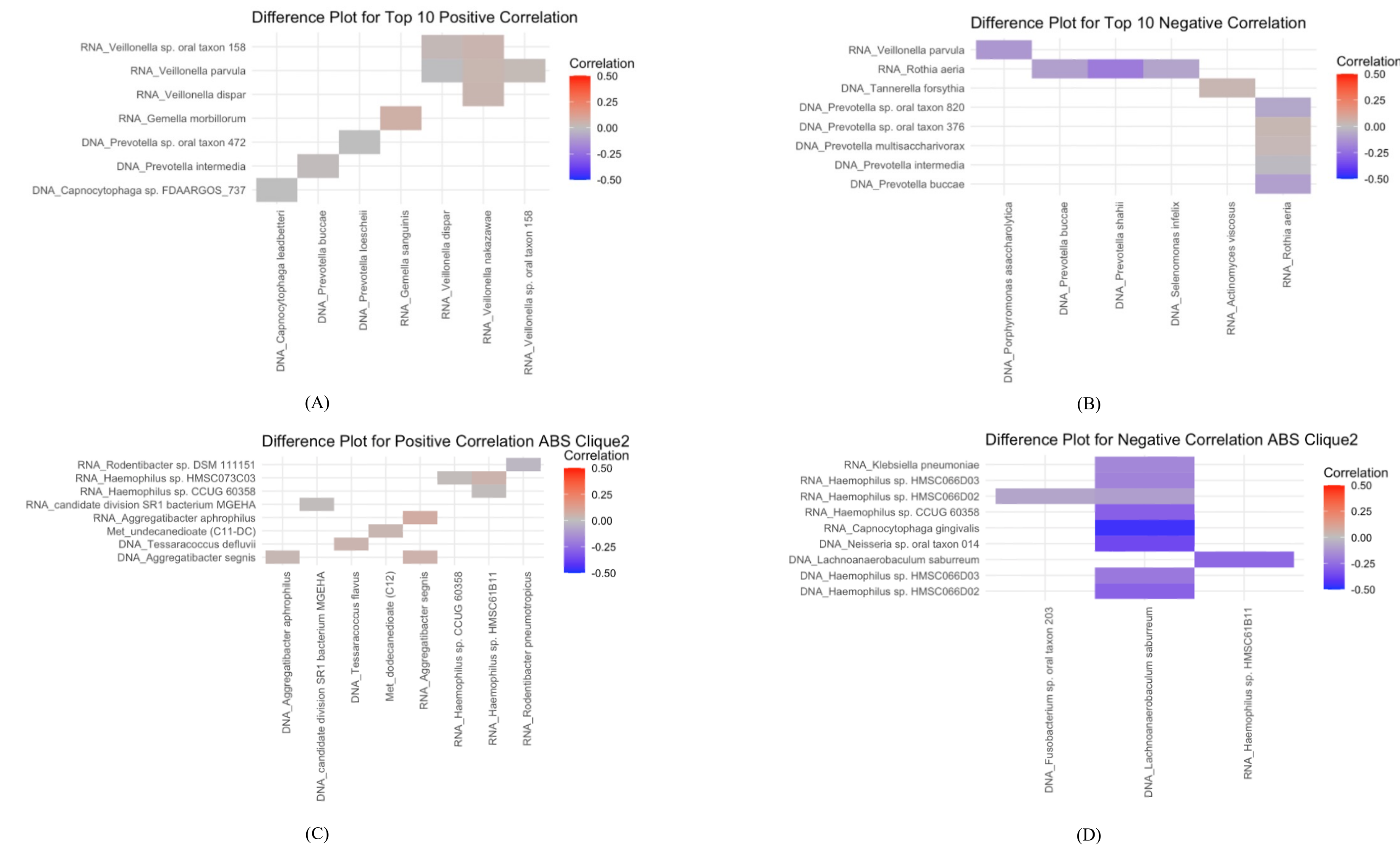


Figure 3 Difference plot for the top 10 positive correlations (A and C) and top 10 negative correlations (B and D) in the forward cliques identified using absolute correlation values in the core search step. *Rothia aerea*, a known dental infection agent, is negatively correlated with multiple species within the *Prevotella* genus, such as *P. intermedia*, *P. baroniae*, *P. multisaccharivorax*, and *P. buccae* [9, 10]. In the second clique, *Lachnoanaerobaculum saburreum* is a primary species associated with caries experience [2]. The correlation between *L. saburreum* and *Capnocytophaga gingivalis* shifts from positive in the ECC group to negative in the non-ECC group.

In addition to procedures required by DCM, we also perform the following tests:

- Test statistics visualization in the feature refinement process

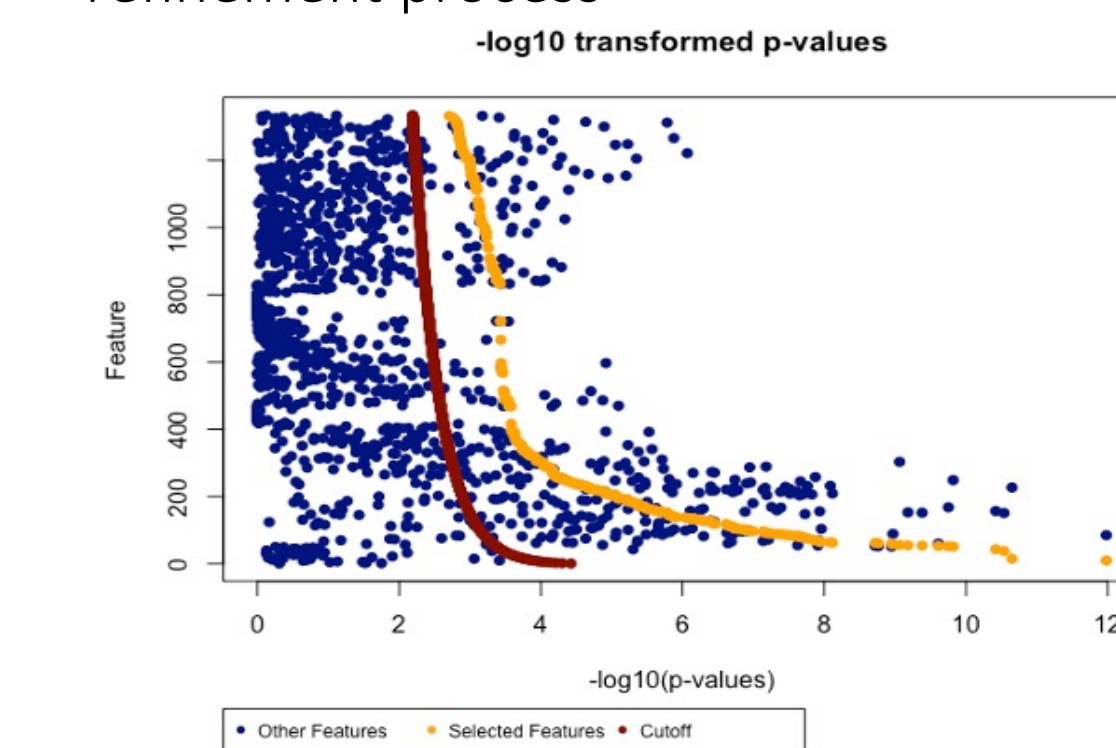


Figure 4 Test statistics curve from the forward stable clique. P-values for all features from the last iteration are shown as navy dots. The FDR-adjusted cutoff values calculated given the alpha level are in dark red. Selected features in the forward stable clique are in yellow.

2) Feature permutation test to monitor the difference in background correlation in the ECC and non-ECC group

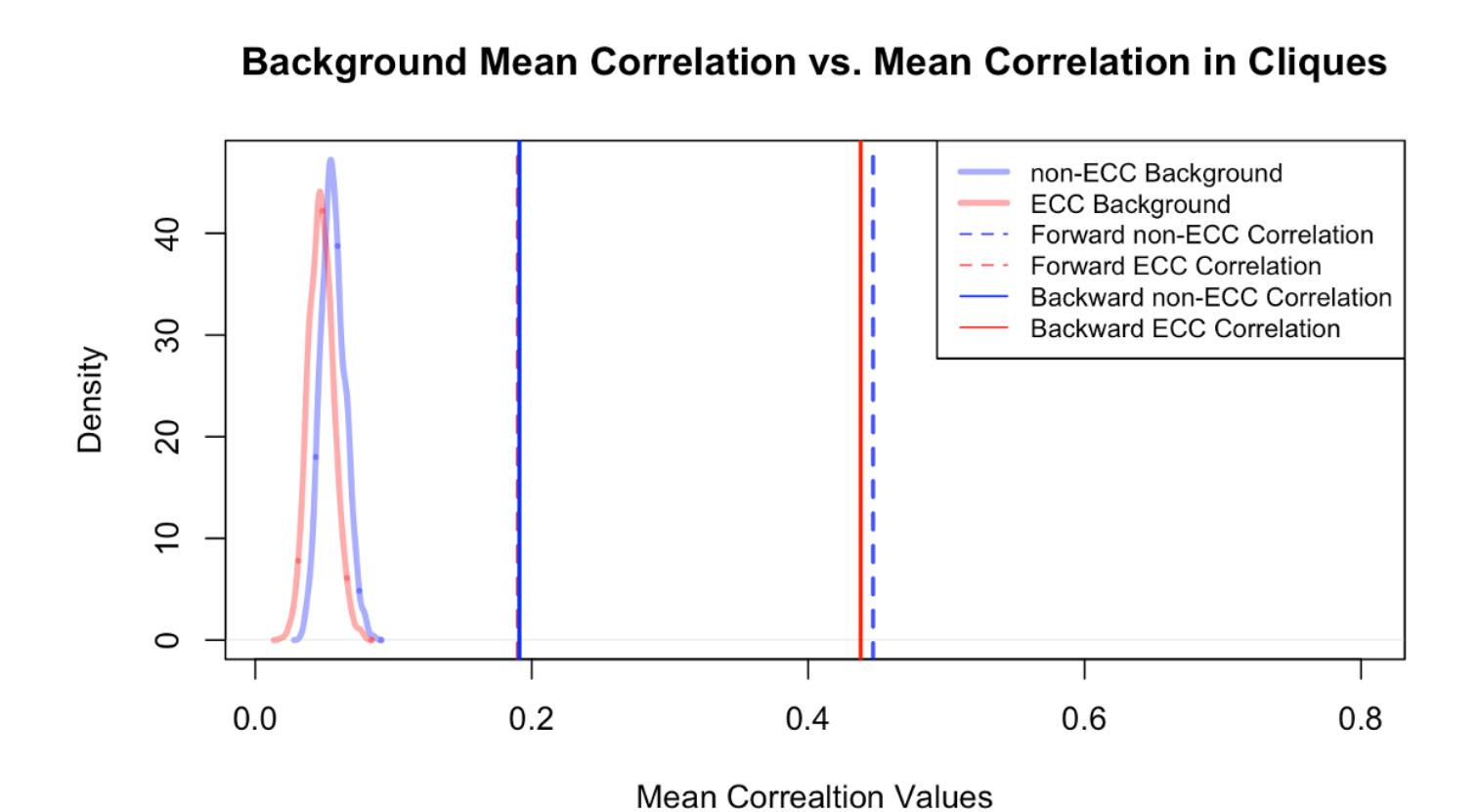


Figure 5 The distributions of mean correlation of randomly selected 50 features 1000 times from the non-ECC and ECC groups are compared to the mean correlation values for a forward clique containing 51 features and a backward clique with 40 features. While the distributions of background mean correlations between the two groups do not fully overlap, the magnitude of discrepancy is substantially smaller when compared to the differential correlation exhibited in the cliques identified.

References

- [1] Uribe, S.E., Innes, N., et al., "The global prevalence of early childhood caries: A systematic review with meta-analysis using the WHO diagnostic criteria," *International Journal of Paediatric Dentistry* 31(6), 817–830 (2021).
- [2] Cho, H., Ren, Z., et al., "Selenomonas sp. sp. acts as a pathobiont mediating spatial structure and biofilm virulence in early childhood caries," *Nature Communications* 14, 2919 (2023).
- [3] Dvaris, K., Slade, G.D., et al., "Cohort Profile: ZOE 2.0-A Community-Based Genetic Epidemiologic Study of Early Childhood Oral Health," *International Journal of Environmental Research and Public Health* 17(21), 8056 (2020).
- [4] Bodwin, B.K., Zhang, K., et al., "A TESTING BASED APPROACH TO THE DISCOVERY OF DIFFERENTIALLY CORRELATED VARIABLE SETS," *The Annals of Applied Statistics* 12(2), 1180–1203 (2018).
- [5] Lin, B.M., Cho, H., et al., "B2INB Model-Based Pathway Analysis and Module Identification Facilitates Integration of Microbiome and Metabolome Data," *Microorganisms* 11, 766 (2023).
- [6] Heimsdottir, L.H., Lin, B.M., et al., "Metabolomics Insights in Early Childhood Caries," *Journal of Dental Research* 100(6), 615–622 (2021).
- [7] Xia, J., Psychogios, N., et al., "MetaboAnalyst: a web server for metabolomic data analysis and interpretation," *Nucleic Acids Research* 37(suppl_2), W652–W660 (2009).
- [8] Karp, P.D., Billington, R., et al., "The BioCyc collection of microbial genomes and metabolic pathways," *Briefings in Bioinformatics* 20(4), 1085–1093 (2019).
- [9] Sewell, J., Sathish, R., et al., "Rothia aerea vertebral discitis/osteomyelitis in an immunocompetent adult: Case report and literature review," *IDCases* 27, e01459 (2022).
- [10] Känönen, E., Fletta, D., et al., "Prevotella species as oral residents and infectious agents with potential impact on systemic conditions," *Journal of Oral Microbiology* 14(1), 2079814 (2022).