Deciphering the Molecular Complexity of Dyslipidemia: Multi-Omics Meta-Analysis of the Gut Microbiome and the Host Physiology

A Project Report

submitted by

SHIVANGI VERMA

in partial fulfilment of the requirements

for the award of the degree of

MASTER OF TECHNOLOGY

Computational Biology

INDRAPRASTHA INSTITUTE OF INFORMATION TECHNOLOGY DELHI,
NEW DELHI- 110020

August 2024

THESIS CERTIFICATE

This is to certify that the thesis titled *"Integrated multi-OMIC investigation of gut microbiota and metabolomics to identify gut-derived factors associated with Dyslipidemia"*, submitted by **Shivangi Verma**, to the Indraprastha Institute of Information Technology, Delhi, for the award of the degree of Master of Technology, is a bona fide record of the research work done by her under my supervision. This thesis has reached the standards fulfilling the requirements of the regulations relating to the degree.

The contents of this thesis, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

August, 2024

Advisor's Name: Dr. Tarini Shankar Ghosh
Department of Computational Biology
Indraprastha Institute of Information Technology
New Delhi 110020

ACKNOWLEDGEMENTS

I extend my heartfelt appreciation to Dr. Tarini Shankar Ghosh for their invaluable guidance and mentorship, which played a pivotal role in shaping this research endeavour. Additionally, I am grateful to the dedicated team at the Microbiome Informatics lab for their technical support and collaboration, without which this project would not have been possible. I also want to acknowledge my family and friends for their unwavering encouragement and support throughout this journey. Their belief in me has been a constant source of motivation and inspiration.

Abstract

The gut microbiome is intricately linked with various aspects of health and physiology, serving as both a potential diagnostic marker and therapeutic target. Understanding its mechanistic influences on systemic processes, particularly those affecting remote organs, is crucial. One key research area is the gut microbiome's interaction with the cardiometabolic system, particularly dyslipidemia. Dyslipidemia, characterized by abnormal levels of lipids in the blood, is a major risk factor for cardiovascular disease, Type II Diabetes, obesity, and metabolic syndrome. These conditions are characterized by disturbances in glucose metabolism, lipid metabolism, and blood pressure regulation, often leading to severe complications like heart attacks, strokes, Non-alcoholic fatty liver disease (NAFLD), and chronic kidney diseases.

Alterations in the gut microbiome are associated with dyslipidemia. Specific gut microbiome-derived metabolic products have been linked causatively to the onset or delay of these disorders. Despite this, our understanding remains incomplete, lacking a comprehensive catalogue of the microbiome-metabolite network-host-pathway links.

This study adopts a data-driven, multi-level analysis of human-derived gut microbiome and host metabolome datasets to investigate these connections. We performed a three-level interaction investigation: first, identifying gut-microbiome-associated serum metabolites linked with dyslipidemia health markers; second, deciphering links between the gut microbiome and serum metabolome; and third, validating these findings by associating specific gut microbial community members with dyslipidemia markers. To achieve this, we utilized 17 large, publicly available paired datasets, including gut microbiome and metabolome data, host metabolome and dyslipidemia marker profiles, and gut microbiome and host dyslipidemia marker profiles. This comprehensive approach has the potential to enhance our understanding of the gut microbiome's role in systemic health and dyslipidemia.

Contents

List of Figures

Chapter 1

Introduction

Dyslipidemia is characterized by abnormal lipid levels in the blood, including elevated triglycerides (TG), total cholesterol (TC), and low-density lipoprotein cholesterol (LDL-C), or decreased high-density lipoprotein cholesterol (HDL-C) (Mosca *et al.*, 2022). It is a major risk factor for cardiovascular diseases (CVD) such as coronary artery disease and stroke, with both genetic and environmental factors contributing to its development (Hedayatnia *et al.*, 2020). Despite established risk factors, a significant portion of dyslipidemia remains unexplained, prompting investigation into the gut microbiome's role.

Cardiovascular Disease and Dyslipidemia

Cardiovascular disease remains one of the leading causes of morbidity and mortality worldwide (Amini, Zayeri and Salehi, 2021). The pathological process often begins with the accumulation of LDL cholesterol in the arterial walls, leading to plaque formation and arterial narrowing. This process, known as atherosclerosis, can eventually result in reduced blood flow or complete blockage of the arteries (Kita *et al.*, 2001). Dyslipidemia accelerates this process by promoting inflammation, oxidative stress, and endothelial dysfunction, thereby increasing the risk of cardiovascular events (Hajjar and Gotto, 2013).

The Gut Microbiome and Its Influence on Dyslipidemia

Recent research has uncovered a profound connection between the gut microbiome—the diverse community of microorganisms residing in the gastrointestinal tract—and dyslipidemia (Jia *et al.*, 2021). The gut microbiome is known to influence lipid metabolism through various mechanisms, including:

1. **Bile Acid Metabolism**: Gut bacteria modulate the composition and function of bile acids, which are critical for lipid digestion and absorption. Altered bile acid metabolism can affect cholesterol homeostasis and lipid levels in the blood (Jia *et al.*, 2021).
2. **Short-Chain Fatty Acids (SCFAs)**: Produced by bacterial fermentation of dietary fibers, SCFAs like butyrate, propionate, and acetate have been shown to influence lipid metabolism and energy balance. SCFAs can regulate the expression of genes involved in lipid synthesis and breakdown (Jia *et al.*, 2021).

3. **Microbial Metabolites**: Certain gut bacteria produce metabolites that interact with host metabolic pathways. For instance, trimethylamine-N-oxide (TMAO), derived from dietary choline and carnitine by gut bacteria, has been linked to increased cardiovascular risk, including dyslipidemia (Zhen *et al.*, 2023).

Proven Associations Between the Gut Microbiome and Dyslipidemia

Numerous studies have demonstrated significant associations between the composition of the gut microbiome and lipid profiles (Ghosh and Valdes, 2023). Key findings include:

- **Microbial Diversity**: Reduced gut microbial diversity has been correlated with higher levels of LDL cholesterol and triglycerides.
- **Specific Microbes**: Certain bacterial taxa, such as *Bacteroides*, *Lactobacillus*, and *Akkermansia*, have been associated with favourable lipid profiles, whereas others, like *Firmicutes*, are linked to dyslipidemia.
- **Microbiome Modulation**: Interventions that modulate the gut microbiome, such as probiotics, prebiotics, and dietary changes, have shown promise in improving lipid profiles and reducing cardiovascular risk.

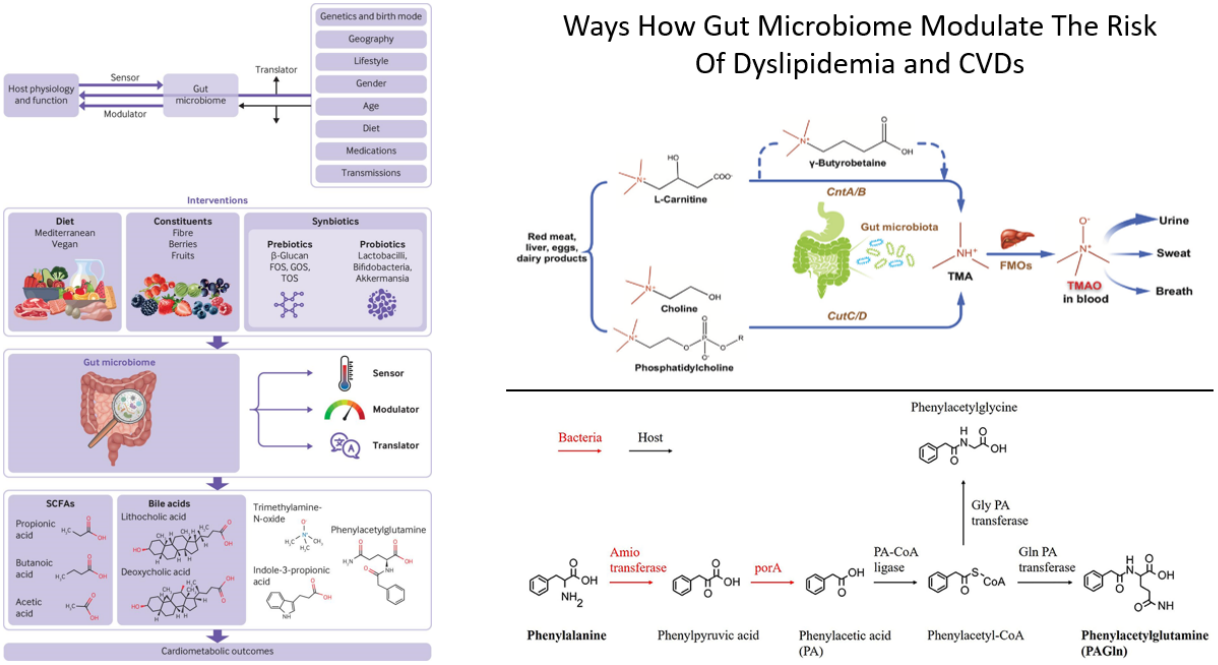

Fig 1. The Gut Microbiome as a Modulator of Host Physiology and Dyslipidemia: Impact of Dietary Interventions and External Factors (Ghosh and Valdes, 2023)

Chapter 2

METHODS

2.1 Dataset collection and pre-processing

In this work, we collected, processed, and analysed data from a few different multi-omic studies and the curatedMetagenomicData resource (R package curatedMetagenomicData (v3.8.0)). Below are details describing the data obtained from each source and the further processing we performed. We used a total of 14,133 gut microbiome – metadata matched samples collected from 17 data repositories, which had 4384 gut microbiome profiles. Three of the study cohorts from these 17 data repositories had matched metabolome-microbiome data available.

For our study on the association between dyslipidemia markers and the gut microbiome, some datasets were downloaded from the curatedMetagenomicData repository, for which the metaphlan3-generated taxonomic profiles and the sample metadata were already available (Beghini *et al.*, 2021). For the rest of the whole genome sequencing datasets, the same protocol adopted by the curatedMetagenomicData was used. All 16S-derived gut microbiome datasets were downloaded from the corresponding data sources (as denoted by the accession numbers) and processed using the SPINGO taxonomic classification tool (Allard *et al.*, 2015). The corresponding metadata for all datasets that were not a part of the curatedMetagenomicData were obtained by looking into the respective research papers. We thank the authors of the Diener cohort for providing us with the metadata of the corresponding participants.
The metadata for each study was filtered only to take the following dyslipidemia markers, ensuring they have the same units of measurement across each study cohort: Triglyceride, Cholesterol, HDL, and LDL levels in mmol/L.

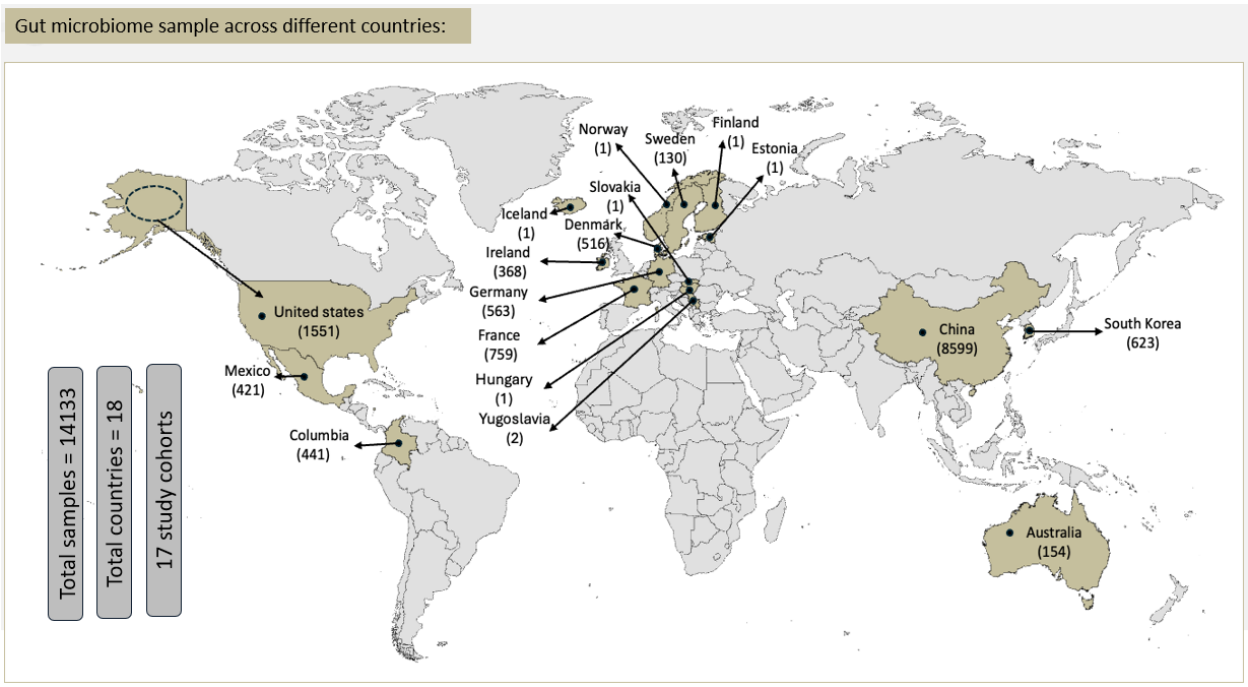

Figure 2.1: World map showing distribution of Gut Microbiome samples

In the data preprocessing phase, the raw species profile was row-sum normalized where rows represented each sample, and the columns consist of species names. To ensure consistency across the different dyslipidemia marker levels, they were converted into mmol\L to ensure uniformity across different datasets.

2.2 Methods Overview for filtering diagnostic features

A. Work Package 1

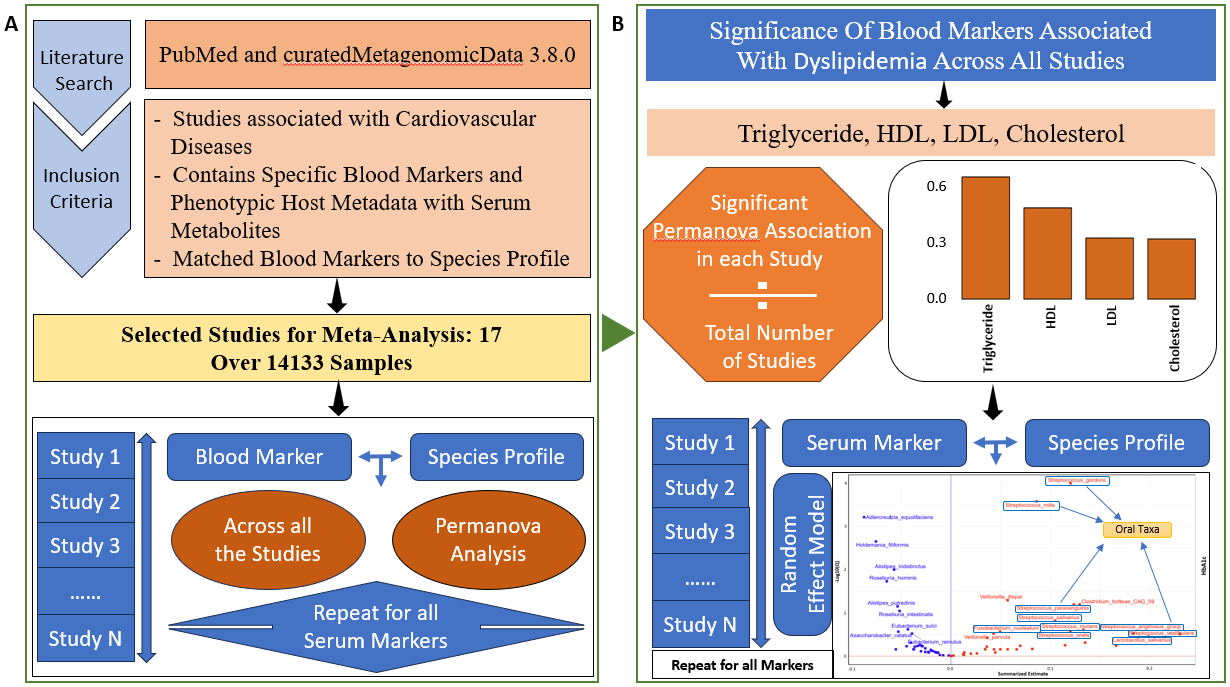

Figure 2.2.1: Method Schematic - Work Package 1 of Dataset processing

1. **Literature Search and Data Collection**:
   - Conducted a systematic search using PubMed and curatedMetagenomicData 3.8.0.
   - Applied inclusion criteria to select studies focused on dyslipidemia markers, phenotypic host metadata, and serum metabolites.
   - Selected studies included matched dyslipidemia markers and species profiles for integration.
2. **Study Selection for Meta-Analysis**:
   - A total of 17 studies were selected, comprising over 14,133 samples.
   - Selected studies met criteria for containing specific blood markers and associated species profiles.
3. **Meta-Analysis Workflow**
   - **Study-wise Analysis**:
     - For each study, blood markers and species profiles were analyzed.
     - PERMANOVA analysis was performed across all studies to assess associations between dyslipidemia markers and species profiles.

2. **Significance of dyslipidemia Markers Across Studies**:
    - Focused analysis was conducted on markers associated with dyslipidemia, including triglycerides, HDL, LDL, and cholesterol.
    - The significance of these associations was determined based on PERMANOVA results across studies (P value <0.05).
3. **Integration Using Random Effects Model**:
    - A random effects model was employed to synthesize results across all studies, considering the variability between studies.
    - Meta-analysis driven identification of microbial taxa associated with different dyslipidemia markers.

    This workflow ensured a robust meta-analysis approach by integrating dyslipidemia markers with microbial species profiles, thereby providing insights into microbiome community across diverse studies.

B. Work Package 2

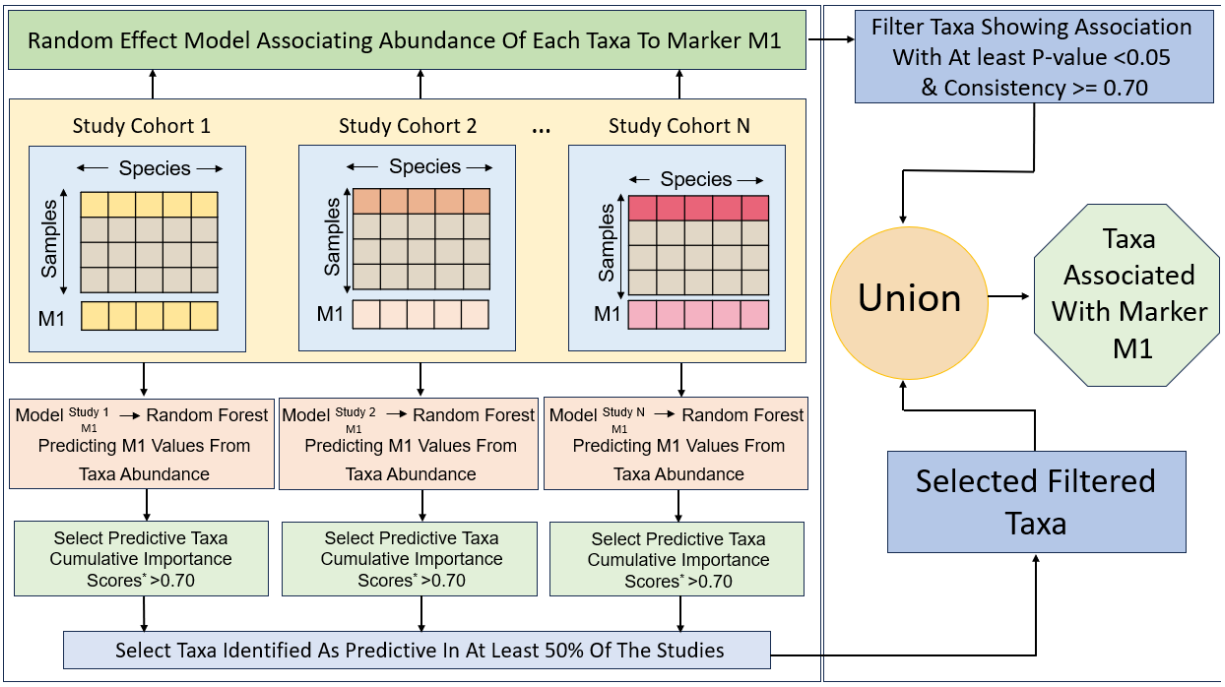

Figure 2.2.2: Work Package 2 of Dataset processing

Random Forest & Meta-Analysis Approach for filtering diagnostic features associating Taxa with each Blood Marker

Overview of Random Effect Model

The methodological approach illustrated in the figure involves the use of a Random Effect model to associate the abundance of each taxon with a specific dyslipidemia marker, denoted as M1. This process is carried out across multiple study cohorts, integrating the results to identify taxa significantly associated with the marker.

Steps Involved:

1. **Study Cohorts and Taxa Abundance**:
    - **Study Cohort 1 to N**: Each study cohort contains samples with species abundance data.
    - **Blood Marker M1**: The values for dyslipidemia marker M1 are included in each cohort.
2. **Random Forest Models**:
    - For each study cohort, a Random Forest model is constructed to predict the values of M1 based on the taxa abundance across all study cohorts.
3. **Selection of Predictive Taxa**:
    - Within each study, taxa are selected based on their predictive power.
    - **Cumulative Importance Scores**:
        - Taxa with cumulative importance scores greater than 0.70 are selected.
        - This score indicates the significance of each taxon's contribution to the model's predictions.
4. **Filtering Predictive Taxa**:
    - Taxa identified as predictive in at least 50% of the studies are selected for further analysis.
5. **Union of Taxa**:
    - The selected predictive taxa from all study cohorts and the meta-analysis taxa - Taxa showing an association with a p-value < 0.05 and a consistency ≥ 0.70 - are combined using a union operation.
    - This results in a list of taxa associated with marker M1.
6. **Final Selection of Taxa**:
    - The filtered taxa that meet the statistical criteria are then said to be associated with blood marker M1.
    - These selected taxa are then used for further analysis and validation.

Interpretation of Results

The methodology ensures that only taxa with a strong and consistent association with blood marker M1 across multiple studies are selected. This robust approach accounts for variability between studies and enhances the reliability of the identified associations.

- **Predictive Power**: By using Random Forest models, the approach effectively captures non-linear relationships and interactions between taxa and the blood marker.
- **Consistency**: The filtering criteria ensure that only taxa with consistent predictive power across multiple studies are selected.
- **Statistical Significance**: The final filtering based on p-value and consistency ensures that the selected taxa are statistically significant, reducing the likelihood of false positives.

Chapter 3

RESULTS

3.1 Investigation of the overall beta-diversity of gut microbiome with lipid markers

3.1.1 Statistical Analysis of Blood Markers and Microbial Species

We took 4 blood markers proven to be directly linked to dyslipidemia. The row-sum normalized species profiles were filtered based on the names of species detected in at least 40% of the studies, with a detection frequency of at least ten percent. These filtered species profiles with the blood marker metadata were run through the PERMANOVA test for statistical analysis to determine how significantly these blood markers were detected across all datasets. PERMANOVA was performed using the adonis2 function of the vegan R package.

PERMANOVA Analysis

- **Feature Selection**: Species containing specific keywords ("phage", "virus", "unclassified", or "satellite") were filtered out using the grep function.
- **Species Selection**: Identified species detected in at least 10% of samples and present in at least 40% of the studies.
- **Study-specific Analysis**: For each unique study associated with each blood marker variable, a Spearman correlation matrix for the top detected species was computed. This matrix was transformed into a dissimilarity matrix using the as.dist function. The PERMANOVA test was conducted with the matching samples of the dissimilarity matrix and the blood marker metadata as inputs. R^2 values and p-values from the test were stored.

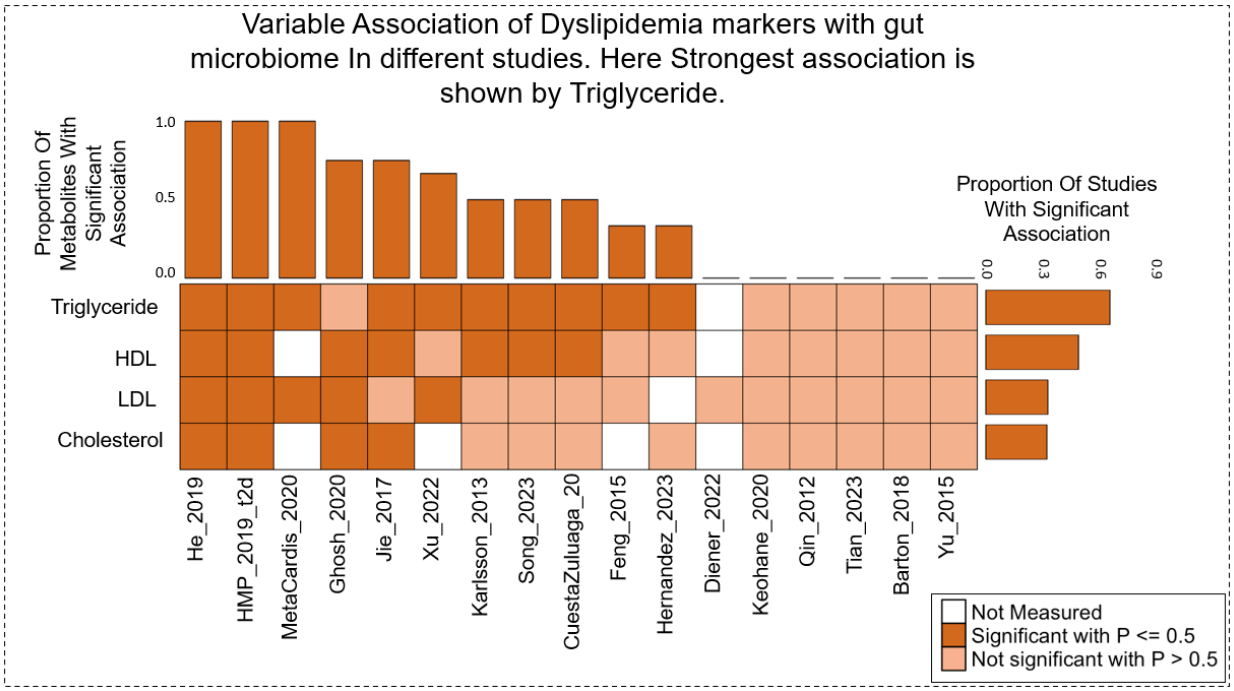

Figure 3.1.1: Variable association of Dyslipidemia markers with gut microbiome across different study cohorts

The PERMANOVA analysis assessed the association between microbial community composition and metadata variables across multiple studies. By iterating over each metadata variable and study, R^2 values and p-values were obtained, indicating the proportion of variance explained by the metadata and the statistical significance of the associations.

- **Helper Function for Directionality**: A helper function get_direction was defined to assign directionality based on p-values from the PERMANOVA results.
    - P-values ≤ 0.05 were assigned a value of 2 (indicating significance).
    - P-values > 0.05 were assigned a value of 1 (indicating non-significance).
    - Missing values were handled by assigning them a value of 0.

For each blood marker (row), the proportion of studies (columns) exhibiting significant associations was calculated by counting the occurrences of the value 2 and dividing it by the total occurrences of values 1 and 2. This approach quantified the prevalence of significant associations between blood markers and microbial species across multiple studies. A bar plot visualized the results, highlighting markers with the most consistent significant associations.

3.1.2 Meta-Analysis and Robust Regression

Meta-Analysis Using Random Effect Models

Subsequently, a meta-analysis using Random Effect models was performed to compute the association of the different taxa with blood markers across all study cohorts, treating the abundance as the response variable and the metadata as the predictor while considering the grouping variable. Estimation results, such as the beta coefficient, p-value, and consistency, were captured. P-values were adjusted using the false discovery rate (FDR) method to obtain q-values.

- **Directionality Score**: A directionality score was assigned based on the significance thresholds:
    - If q-value ≤ 0.10, indicating a high significance level, the sign of the estimate was multiplied by three.
    - If q-value > 0.10 but p-value ≤ 0.05, indicating a moderate significance level, the sign of the estimate was multiplied by two.
    - If p-value > 0.05, indicating a low significance level, the sign of the estimate was multiplied by one.

This approach ensured that the directionality of the estimates reflected both their statistical significance and the direction of the effect in the analysis.

Robust Linear Regression (MASS Package)

- The rlm function from the MASS package was used for robust linear regression to model the relationship between species abundance and the metadata variable within each study group.
- **Meta-Analysis (metafor package) R package version 4.2.0**: Computation of regression models for each group defined by the grouping variable was done, aggregating the results of individual regression models using meta-analysis techniques.
    - The escalc function computed effect sizes (ZPCOR) for each study group.
    - The rma function aggregated the results of individual regression models computed for each group using robust linear regression.

Directionality and Significance Testing

- The sign of the coefficient of the metadata variable was calculated to determine directionality. The coefficient represents the strength and direction of the relationship between species abundance and the metadata variable.
    - Positive coefficient: An increase in the metadata variable is associated with increased species abundance.

    - Negative coefficient: An increase in the metadata variable is related to a decrease in species abundance.
- P-values for the coefficients were computed, and a robust F-test for heteroscedasticity was performed using the f.robftest function of the sfsmisc R package 1.1.15.

Combining robust regression with meta-analysis techniques effectively analyzed the relationship between species abundance and metadata across different studies, accounting for potential heterogeneity and outliers within each study group.

Visualizing Results

- **Meta-analysis Driven Identification of Microbial Taxa Associated With Different Dyslipidemia Markers**: Volcano Plots were generated to visualize the impact of various metadata on microbial species abundances using the filtered dataset with a consistency greater than 0.70.
    - Created volcano plots with ggplot2 where each point represents a microbial species.
    - Points were coloured based on the significance and direction of the association.
- The **X-axis** represents the "Summarized Estimate," indicating the strength and direction of the association between specific microbial taxa and triglyceride levels. Taxa on the left side of the plot are associated with **decreasing** triglyceride levels, while those on the right are associated with **increasing** triglyceride levels.
- The **Y-axis** represents the significance of these associations as -log10(p-value), where higher values indicate stronger statistical significance.

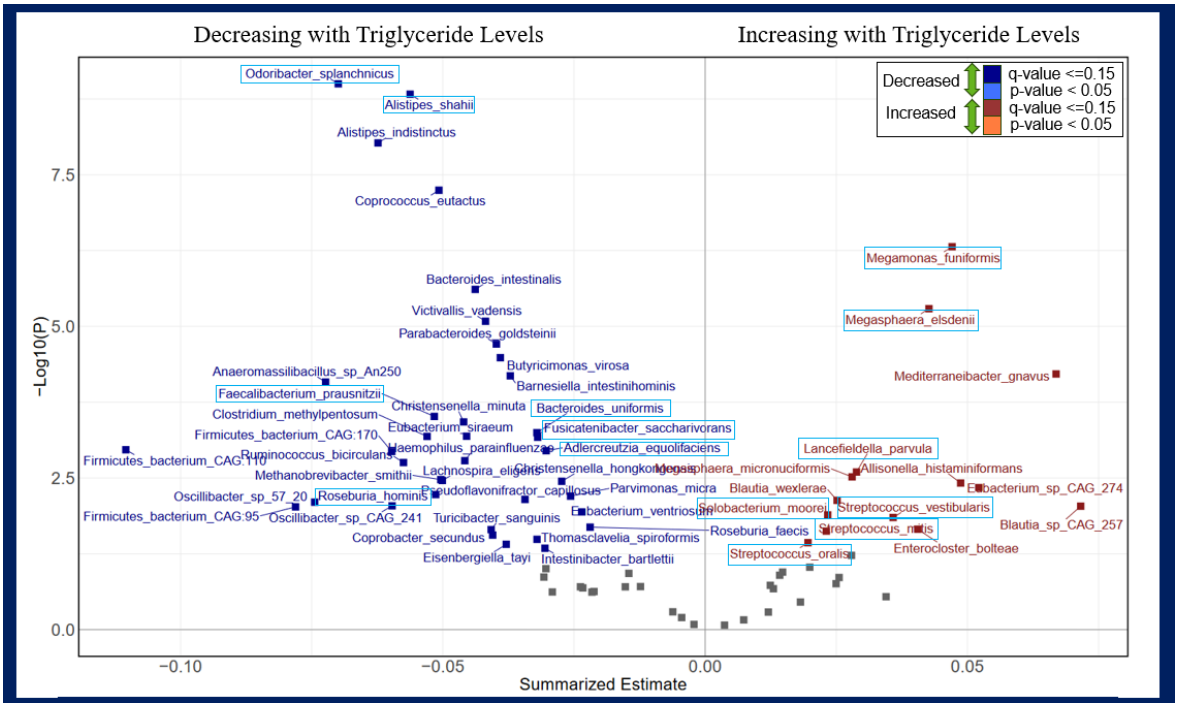

Fig. 3.1.2.1 Meta-analysis driven identification of microbial taxa associated with triglyceride marker

Key Features:

- Microbes are coloured based on their q-values, indicating adjusted statistical significance:
    - **Blue and light blue coloured taxa**: Microbes that are significantly associated with **decreased triglyceride levels** (q-value ≤ 0.15, 0.05, or 0.01) and negative summarized estimate value.
    - **Red and Orange coloured taxa**: Microbes that are significantly associated with **increased triglyceride levels** (q-value ≤ 0.15, 0.05, or 0.01) and positive summarized estimate value.
- **Highlighted Taxa**: Specific taxa which are generally referred to as oral microbiome have been highlighted, For example: *Streptococcus oralis*, and *streptococcus mitis* are showing positive association with increased Triglyceride levels. These bacteria generally also have been shown associated with multiple diseases ranging from Inflammatory Bowel Disease to cancer when transferred to gut.

  There are also good number of health associated bacteria which are negatively linked with triglyceride levels such as *Odoribacter splanchnicus*, *Alistipes shahii*, *Faecalibacterium prausnitzii* and *Roseburia hominis.*

The separation of taxa across the plot provides insights into potential microbial targets for managing dyslipidemia and associated cardiovascular risks.

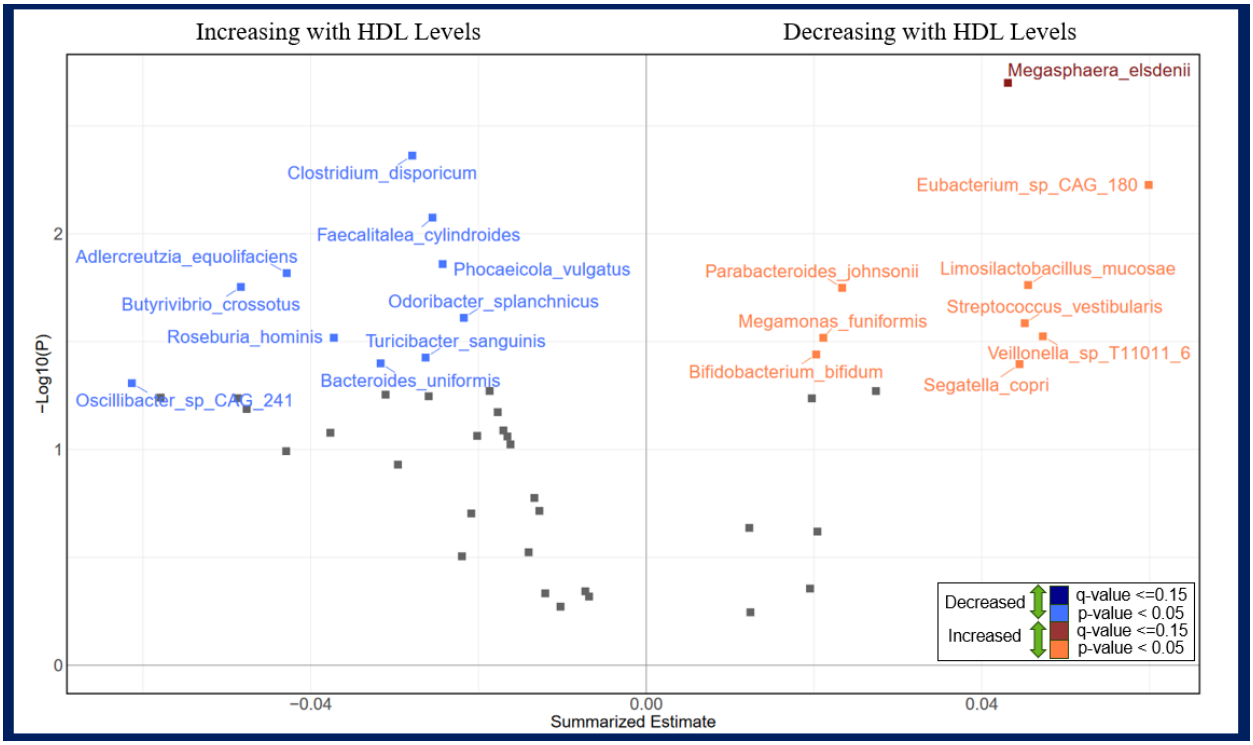

Fig. 3.1.2.2 Meta-analysis driven identification of microbial taxa associated with HDL marker

With Decreasing levels of HDL, we can see some of the following taxa showing positive association:

- ***Megasphaera elsdenii***: Involved in SCFA production; associated with gut health, obesity, and metabolic syndrome; potentially linked to dyslipidemia due to its role in lipid metabolism.
- ***Streptococcus vestibularis***: An oral commensal bacterium; linked to dental infections, endocarditis, and respiratory infections, especially in immunocompromised individuals.
- ***Megamonas funiformis****:* Involved in carbohydrate fermentation and SCFA production; associated with obesity, type 2 diabetes, and dyslipidemia due to its influence on energy metabolism and lipid profiles.

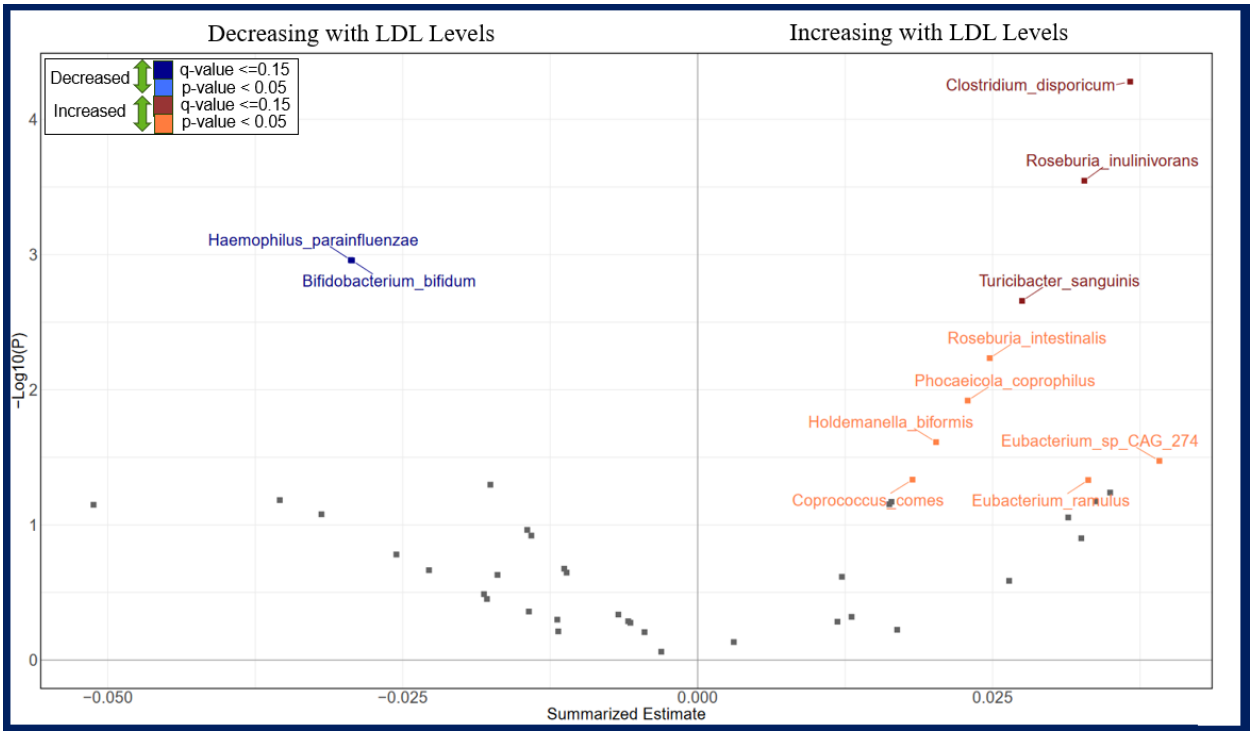

Fig. 3.1.2.3 Meta-analysis driven identification of microbial taxa associated with LDL marker

With Increasing levels of LDL, we see previously health associated beneficial bacteria, which is counterintuitive. This is a typical example of context dependent association of bacteria with health. Implying, we can not simply name a bacteria healthy or bad universally. Some of the bacteria, which are results show to be positively associated with LDL levels are -

- ***Clostridium disporicum**:* Typically found in the human gut microbiome and associated with gut health, it has shown to contribute to short-chain fatty acid (SCFA) production, which is important for maintaining gut integrity. Alterations in its abundance have been linked to inflammatory bowel disease (IBD) and metabolic disorders.
- ***Roseburia inulinivorans**:* A beneficial gut bacterium known for fermenting dietary fibers like inulin into butyrate, a key SCFA. Associated with anti-inflammatory effects and overall gut health. Reduced levels are linked to conditions like obesity, type 2 diabetes, and inflammatory bowel disease.
- ***Turicibacter sanguinis**:* A relatively lesser-known bacterium involved in gut microbial diversity. Associated with immune modulation and linked to gut health. Alterations in its levels have been noted in metabolic diseases like obesity and dyslipidemia.
- ***Roseburia intestinalis**:* A major butyrate-producing bacterium in the gut. It plays a protective role against inflammation and is associated with a healthy

gut barrier. Reduced abundance is linked to metabolic syndrome, IBD, and type 2 diabetes.

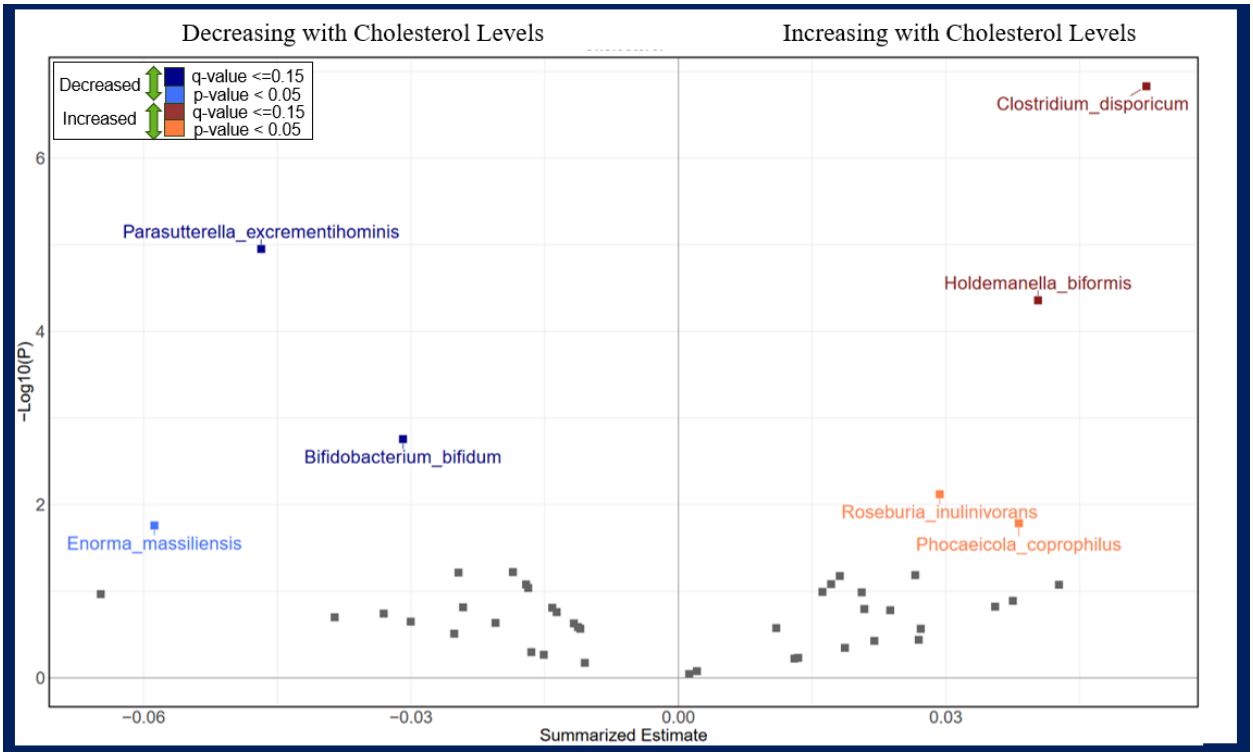

Fig. 3.1.2.4 Meta-analysis driven identification of microbial taxa associated with Cholesterol marker

Cholesterol gave us minimal association. We see taxa such as Holdemanella biformis and Roseburia which show positive association with increasing Cholesterol levels. However, one typical example is negative association of Bifidobacterium species with reduced cholesterol levels.

- ***Holdemanella biformis:*** Linked to better metabolic health, lower obesity risk, and anti-inflammatory effects.
- ***Clostridium disporicum***: Involved in gut health and SCFA production; linked to IBD and metabolic disorders.
- ***Bifidobacterium bifidum:*** Key probiotic, enhances digestion, gut barrier, and reduces inflammation.
- ***Parasutterella excrementihominis***: Affects bile acid metabolism and energy balance; associated with IBS and metabolic issues.

3.2 Results of the association of consistently associated gut microbial markers with different lipid risk factors

While Random Effect Model (Meta-Analysis) tells us about the strength of association and the directionality of the microbiota with the dyslipidemia markers, there might be a chance of missing out some consistent signatures. To find out which taxa are predictive of the markers but might not be strongly associated with markers across all studies, we trained models using Random Forest. These models helped us to track those predictive taxa which may not be strongly associated but may be linked to some other taxa which are strongly associated with dyslipidemia markers. The other aim was also to find out if the pattern of association of different taxa varies across study cohorts. That is why study specific Random Forest models are created.

Figure 3.2.1 shows heatmap of association of taxa with cumulative feature importance greater than 0.7 with study cohorts linked with Triglyceride. The white cells represent absence of that taxa in the corresponding study cohort. The y-axis lists microbial species, with distinct clusters based on how they correlate with triglyceride levels across different studies. The clustering indicates that certain microbial species are consistently associated with Triglyceride levels.

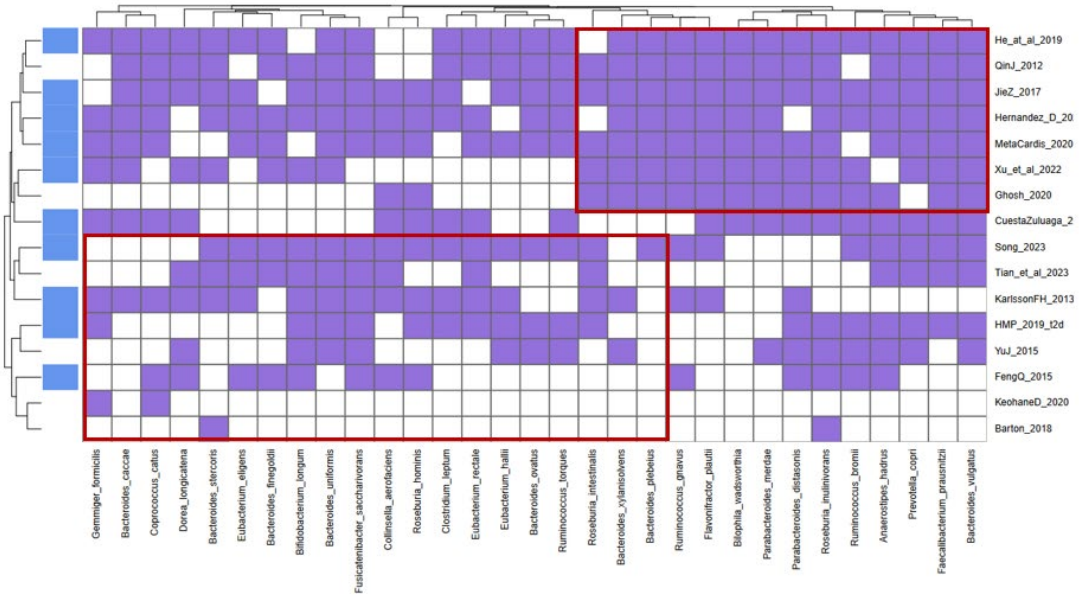

Fig. 3.2.1 Triglyceride associated microbial taxa through Random Forest (Machine Learning Based Approach)

Similar association of microbial markers with study specific linkage was seen for HDL levels. The consistent presence of some of these species in multiple studies strengthens their potential role in influencing or being influenced by HDL levels, highlighting their importance in dyslipidemia research.

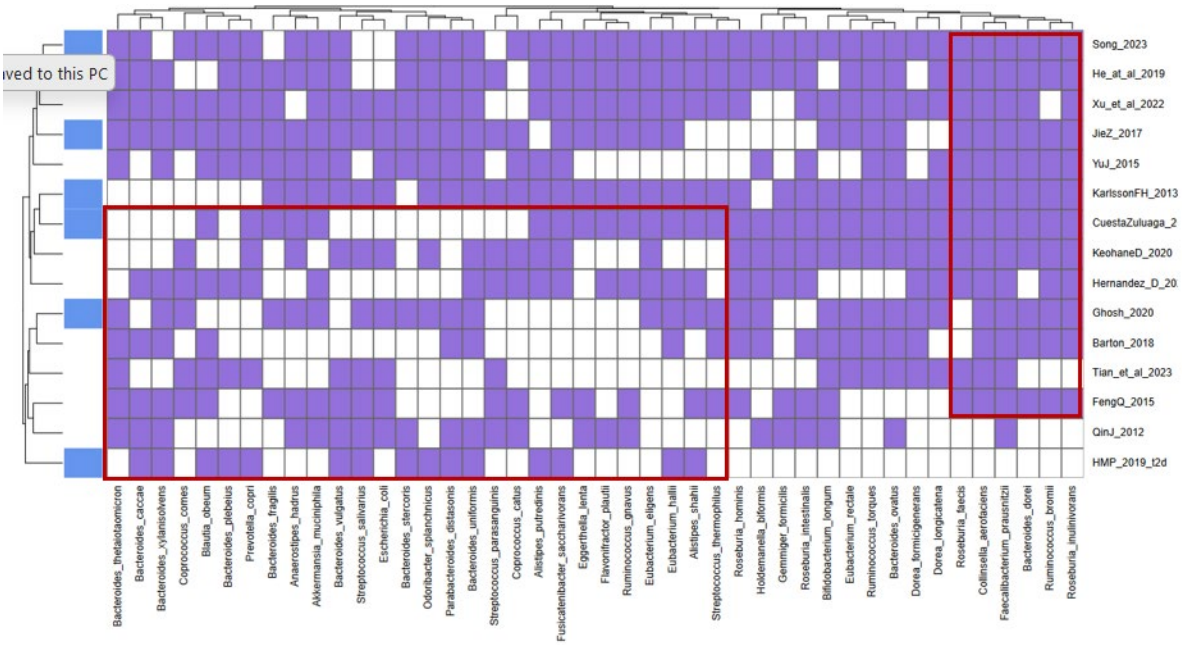

Fig. 3.2.2 HDL associated microbial taxa through Random Forest (Machine Learning Based Approach)

Finally, we conglomerated the two groups of taxa identified from meta-analysis and machine learning, and for every such taxa, for association with a given dyslipidemia marker, we computed the Marker Association scores. The formula is represented in the figure 3.2.4 where Marker score for a particular marker for a taxon was computed as – Number of studies in which it was significantly positively associated minus the number of studies in which it was significantly negatively associated divided by total number of studies in which it was profiled.

$$\text{Marker Score}^{\text{Marker}}_{\text{Taxa}} = \frac{N^{\text{Sig. Positive}}_{\text{Studies}} - N^{\text{Sig. Negative}}_{\text{Studies}}}{\text{Total Number of Studies}}$$

Fig. 3.2.3 Marker Association Scores Computation Formula

Here we identify a panel of sixty-two species that show strong link of dyslipidemia marker across all study cohorts. Here we see these taxa list was predominated by oral taxa members and inflammation associated taxa such as *streptococcus vestibularis* and *Clostridium disporicum*. The yellow shaded panel represents the cluster of disease risk associated members and the purple shaded members represent the health associated taxa such as Akkermansia and Alistipes shahii.

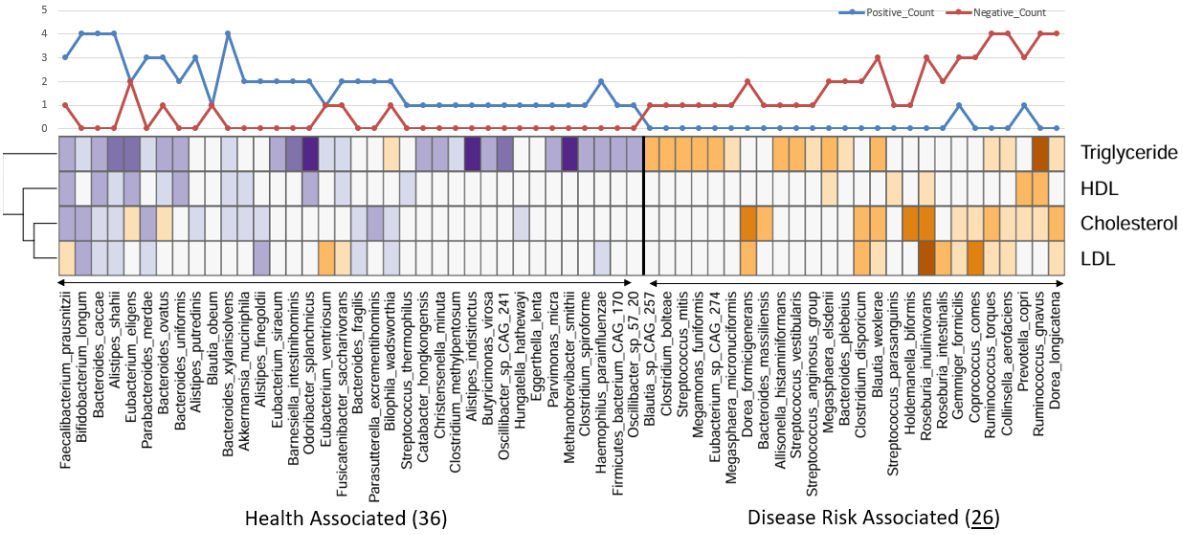

Fig. 3.2.4 Consistent Gut Microbial Markers Across all Dyslipidemia Markers

This investigation has its own importance because these results are showing us signatures which not only validate some of the typical health associated markers but also give us some contradictory context specific examples such as *Coprococcus comes*. *Coprococcus comes* in previous investigations have been shown to be linked with health, but here our findings conclude it to be associated with dyslipidemia risk.

We also computed the median of principle coordinates (PCA) for each study cohort based on the baseline microbiome composition which show us how different studies display different degree of relativeness based on their gut microbiome community (Figure 3.2.5).

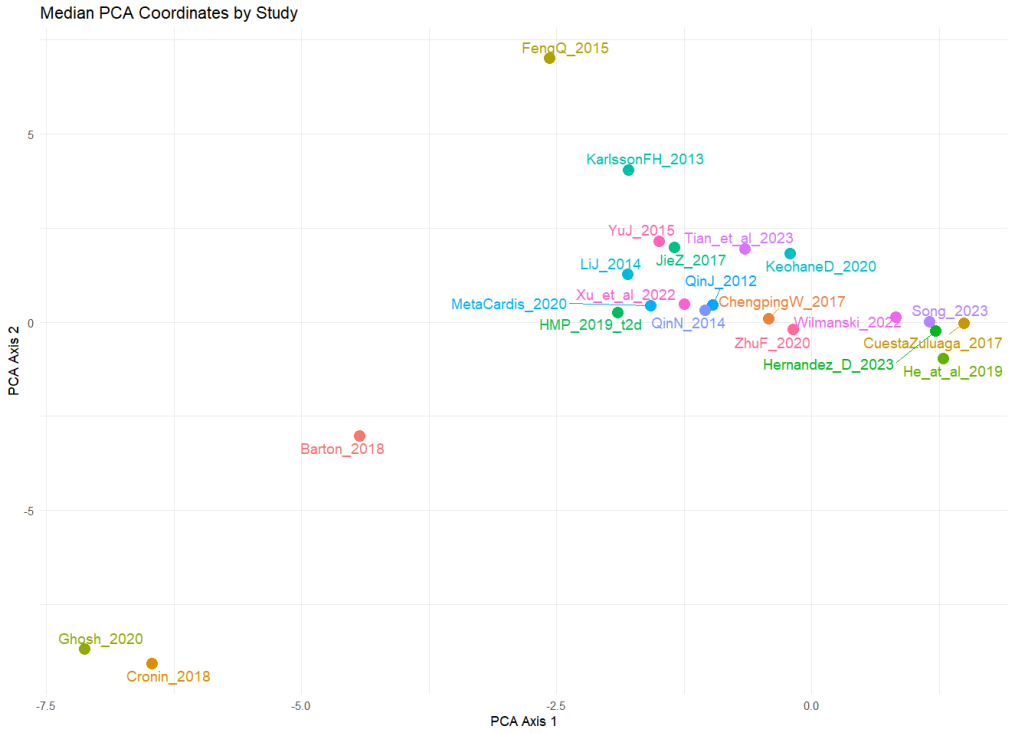

Fig. 3.2.5 Studies Show Different Degree of Relativeness Based on Their Gut Microbiome Community (Median PCA Plot of study cohorts)

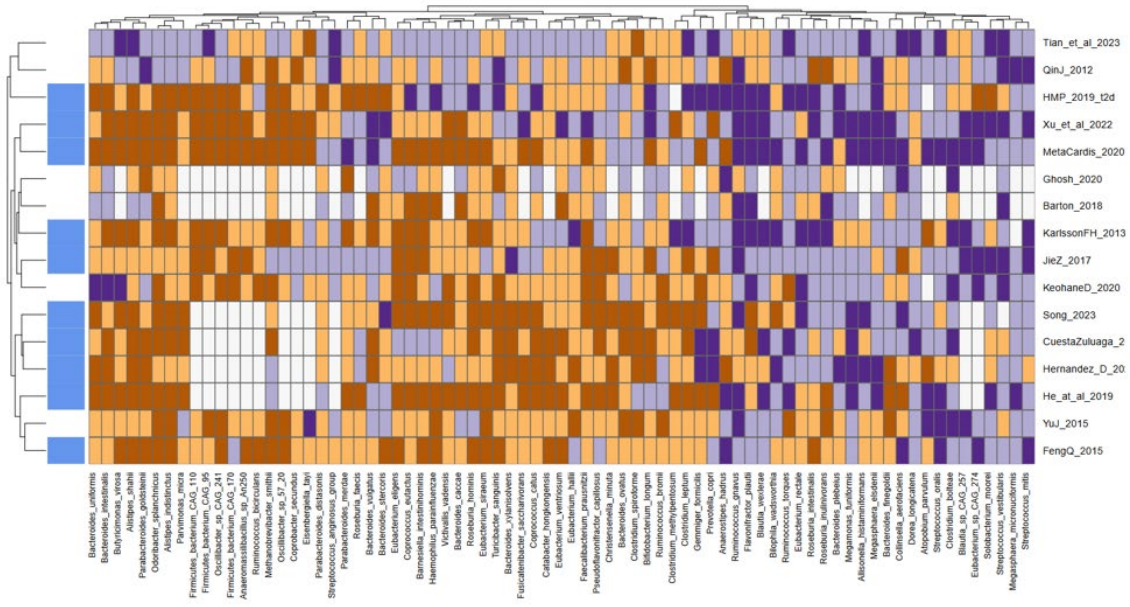

Fig. 3.2.6 Exploring Microbial Taxa Patterns Linked to Triglyceride Levels in Various Study Cohorts

The PCA analysis and the microbial taxa patterns linked with dyslipidemia markers show similarity in clustering of study cohorts. The results indicate that microbiome with similar communities tend to have similar bacterial markers associated with dyslipidemia marker levels. The study clusters which are closer to each other, show the same clustering in the microbial taxa pattern heatmaps shown for different dyslipidemia marker levels.

This shows us a new paradigm of computational models through the knowledge of baseline microbiome data, which will help us know the responsive microbes for a particular marker level.

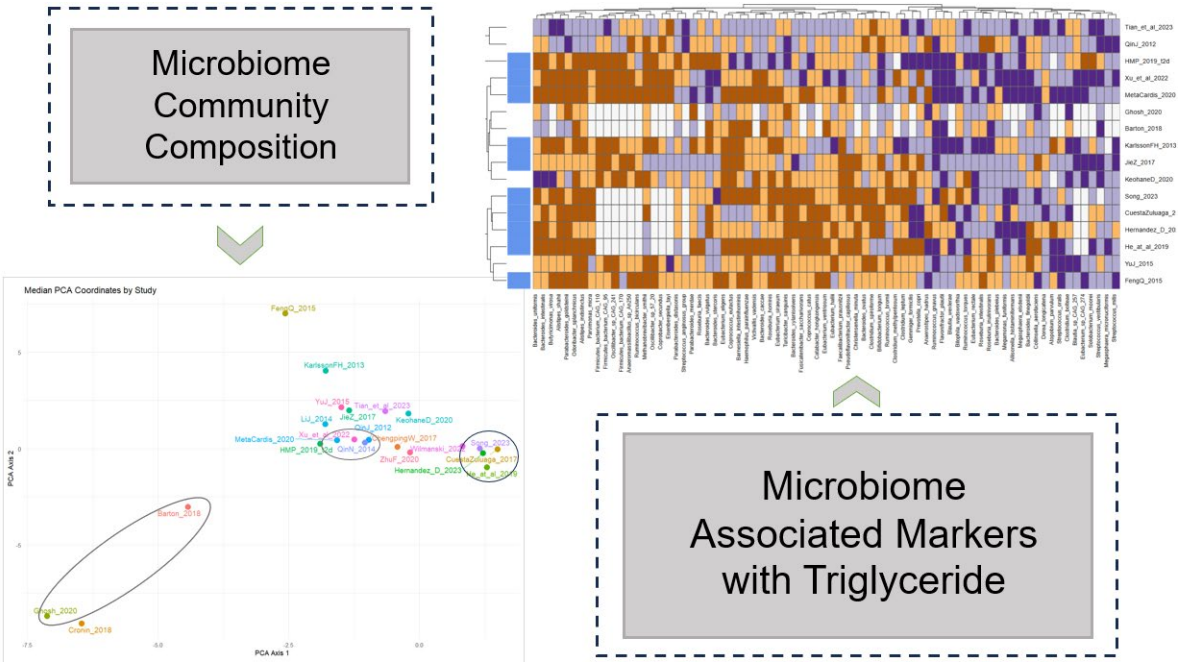

Fig. 3.2.7 Exploring Microbial Taxa Patterns Linked to baseline microbiome composition in Various Study Cohorts in Triglyceride

In the figure 3.2.7, we can see studies such as Song 2023, CuestaZuluaga 2017, He 2019 and Hernandez 2023 show similar microbial taxa pattern with Triglyceride values and present as a cluster as shown in PCA plot. This links to their inherent similarity in baseline microbiome composition.

This discovery points out the translational applicability of this study. The existing framework can be now utilized to predict Triglyceride levels and eventually all other marker levels based on the collated data we have built.

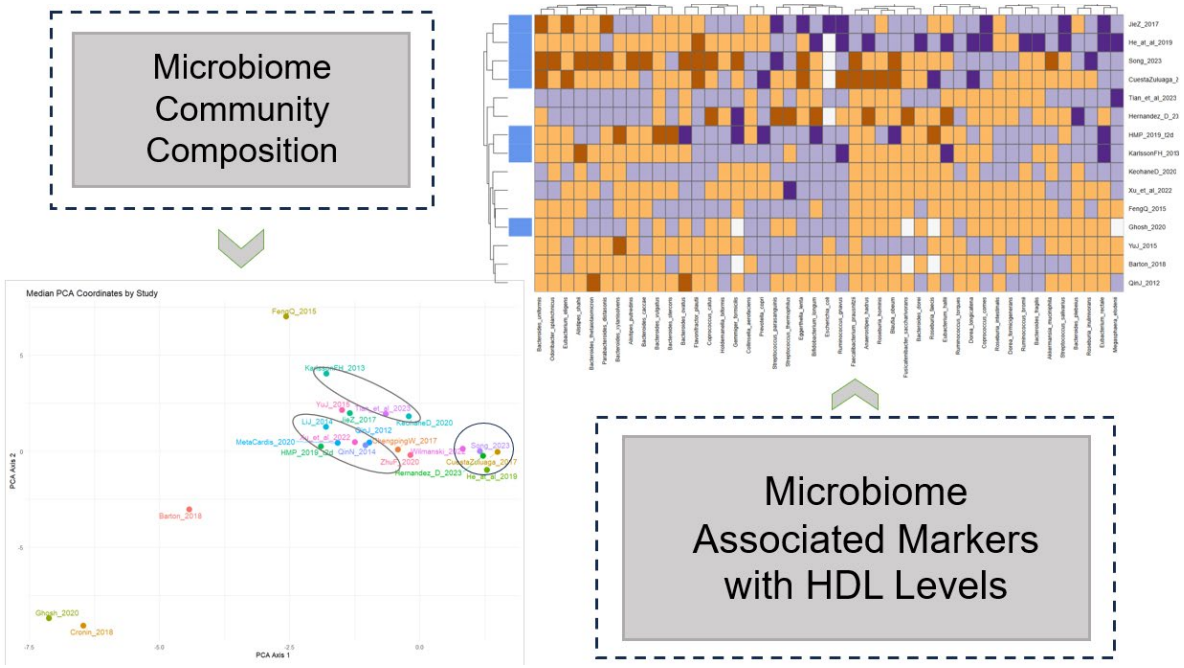

Fig. 3.2.8 Exploring Microbial Taxa Patterns Linked to baseline microbiome composition in Various Study Cohorts in HDL

Similar trend is being shown by HDL associated microbiome pattern and study cohorts linked to it. KarlssonFH 2013 and KeohaneD 2020 show similar microbial taxa pattern for HDL levels; Song 2023, CuestaZuluaga 2017, He 2019 and Hernandez 2023 show similar microbiome association pattern.

These findings illustrate that we can use the Baseline Community to identify the appropriate model or appropriate risk associated factors. This can be used to build microbiome-based models for future diagnostic purposes.

Chapter 4

4.1 Conclusion

Hierarchical Ensemble Models for Microbiome-Based Diagnostics

1. **Utilizing Baseline Community for Model Development**:
    - The baseline community of microbiome can serve as a critical foundation for identifying risk-associated factors and developing appropriate predictive models. The integration of baseline community data allows for the precise identification of risk factors, leading to the development of highly accurate and tailored hierarchical ensemble models for predicting cardiovascular disease risks.
2. **Novelty and Unprecedented Findings**:
    - The findings from our study are unique and novel to the best of our knowledge, presenting findings that highlight the critical role of the gut microbiome in cardiovascular health. These results pave the way for further research and validation, establishing a new paradigm in microbiome-based diagnostics and therapeutic strategies.
3. **Building Microbiome-Based Diagnostic Models**:
    - The development of microbiome-based diagnostic models, utilizing hierarchical ensemble methods, offers a promising avenue for future diagnostic applications. These models can enhance early detection, risk assessment, and personalized treatment strategies for cardiovascular diseases.

4.2 Future Perspectives

Integrative Multi-Omics Analysis: The depicted methodology outlines a comprehensive multi-omics approach that integrates data from the gut microbiome, metabolome, and host physiology to unravel the molecular complexity of cardiovascular disease (CVD). This integrative approach holds several promising future research directions and applications:

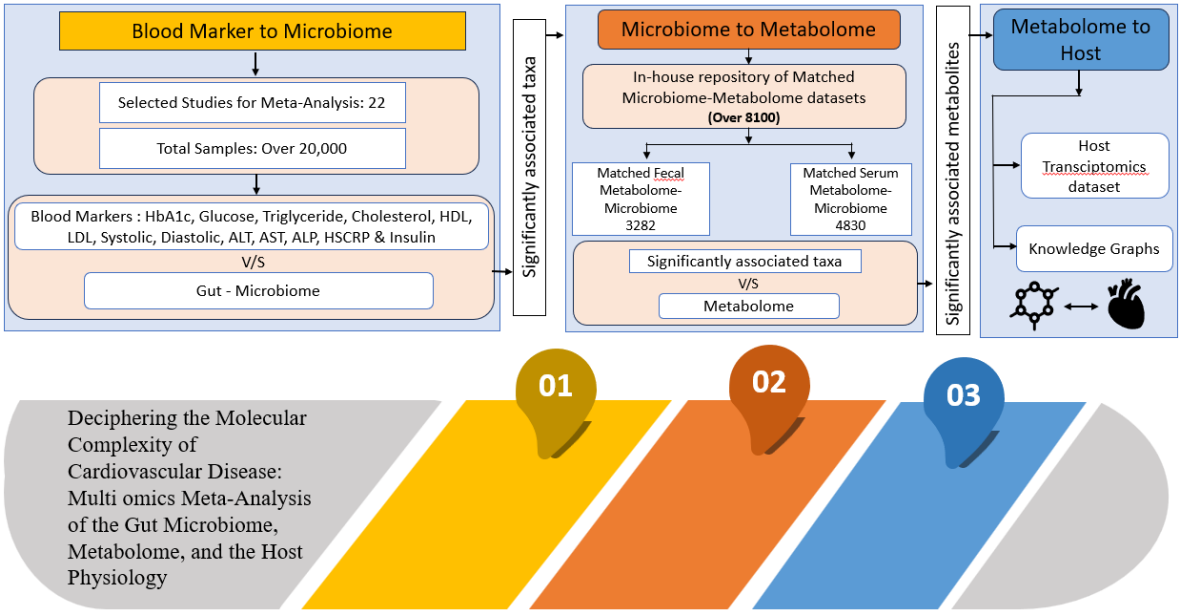

Fig. 4.2 Method Schematic of future perspective

4.2.1. **Enhanced Understanding of Disease Mechanisms:**

Blood Marker to Microbiome: By associating blood markers with gut microbiome composition, researchers can gain insights into how variations in microbiota may influence or reflect cardiovascular health. This can lead to the identification of microbial signatures indicative of disease states or progression.

Microbiome to Metabolome: Understanding the relationship between microbiome composition and metabolomic profiles allows for the identification of key metabolic pathways influenced by gut microbiota. This can highlight specific metabolites that play a role in disease mechanisms.

Metabolome to Host: Integrating metabolomic data with host transcriptomics and physiology can provide a holistic view of how

metabolic changes affect gene expression and overall host health. Knowledge graphs can be used to map complex interactions between metabolites, genes, and disease outcomes.

4.2.2. **Biomarker Discovery and Validation:**

The multi-omics approach facilitates the discovery of microbiome features for early detection, prognosis, and monitoring of CVD. By identifying taxa, metabolites, and gene expression patterns consistently associated with disease, researchers can develop robust biomarker panels for clinical use.

Bibliography

Allard, G. *et al.* (2015) 'SPINGO: A rapid species-classifier for microbial amplicon sequences', *BMC Bioinformatics*, 16. Available at: https://doi.org/10.1186/s12859-015-0747-1.

Amini, M., Zayeri, F. and Salehi, M. (2021) 'Trend analysis of cardiovascular disease mortality, incidence, and mortality-to-incidence ratio: results from global burden of disease study 2017', *BMC Public Health*, 21(1), p. 401. Available at: https://doi.org/10.1186/s12889-021-10429-0.

Beghini, F. *et al.* (2021) 'Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3', *eLife*. Edited by P. Turnbaugh, E. Franco, and C.T. Brown, 10, p. e65088. Available at: https://doi.org/10.7554/eLife.65088.

Ghosh, T.S. and Valdes, A.M. (2023) 'Evidence for clinical interventions targeting the gut microbiome in cardiometabolic disease', *BMJ*, 383, p. e075180. Available at: https://doi.org/10.1136/bmj-2023-075180.

Hajjar, D.P. and Gotto, A.M. (2013) 'Biological Relevance of Inflammation and Oxidative Stress in the Pathogenesis of Arterial Diseases', *The American Journal of Pathology*, 182(5), pp. 1474–1481. Available at: https://doi.org/10.1016/j.ajpath.2013.01.010.

Hedayatnia, M. *et al.* (2020) 'Dyslipidemia and cardiovascular disease risk among the MASHAD study population', *Lipids in Health and Disease*, 19, p. 42. Available at: https://doi.org/10.1186/s12944-020-01204-y.

Jia, X. *et al.* (2021) 'Impact of Gut Microbiota and Microbiota-Related Metabolites on Hyperlipidemia', *Frontiers in Cellular and Infection Microbiology*, 11, p. 634780. Available at: https://doi.org/10.3389/fcimb.2021.634780.

Kita, T. *et al.* (2001) 'Role of Oxidized LDL in Atherosclerosis', *Annals of the New York Academy of Sciences*, 947(1), pp. 199–206. Available at: https://doi.org/10.1111/j.1749-6632.2001.tb03941.x.

Mosca, S. *et al.* (2022) 'Dyslipidemia Diagnosis and Treatment: Risk Stratification in Children and Adolescents', *Journal of Nutrition and Metabolism*, 2022, p. 4782344. Available at: https://doi.org/10.1155/2022/4782344.

Zhen, J. *et al.* (2023) 'The gut microbial metabolite trimethylamine N-oxide and cardiovascular diseases', *Frontiers in Endocrinology*, 14, p. 1085041. Available at: https://doi.org/10.3389/fendo.2023.1085041.