Predicting changes in microbial metabolite productions using whole metagenomic sequencing data based on the microbial gene-metabolite network model

Jungyong Ji[1], Sungwon Jung[1,2,3*]

Department of Health Sciences and Technology, GAIHST, Gachon University, Incheon 21999, Republic of Korea[1]
Department of Genome Medicine and Science, Gachon University College of Medicine, Incheon 21565, Republic of Korea[2]
Gachon Institute of Genome Medicine and Science, Gachon University Gil Medical Center, Incheon 21565, Republic of Korea[3]

* : corresponding author

Abstract

Studies on crosstalk of microbiome-metabolome have been conducted to infer microbial or metabolic features that have importance on human health, through functional, mass spectrometry profiling. However, there have been few types of research identifying a change in the individual metabolite level based solely on the microbial genome. To the end, we propose the method that predicts individual metabolite change using the network built by utilizing known gene-metabolite interaction information. The network algorithm used here is reporter metabolite which identifies metabolites around which most significant microbial gene family abundance's changes occur, where it is constructed with individual interactions and gene families with differential abundances. For benchmark evaluation of the proposed method, we used a publicly available paired dataset of shotgun metagenomics and metabolomics from an inflammatory bowel diseases (IBDs) cohort with controls and evaluated the concordance between our predicted changes and the previously reported changes. The overlap between the predicted and the measured metabolic class was statistically significant, especially for the metabolites with large changes. This method of predicting individual metabolites based on microbial genomes will help narrow a number of metabolites down to their target range before performing costly and incomplete mass spectrometry and is expected to contribute to developing strategies for disease treatment and prevention by investigating candidate metabolites directly associated with disease pathogenesis.

Introduction

According to a recent study, metabolites of the gut microbiome play a central role in the interaction with humans, and they are also involved in disease phenotypes and disease pathogenesis in humans. Thus, We introduce the approach that predicts individual microbial metabolite's change using the network built by utilizing known gene-metabolite interaction information and differential gene based solely on shotgun metagenome sequencing data.

Methods

- Prediction of metabolic changes in microbial genome

First, for each of the two different samples, we calculate the abundance of gene family by aligning shotgun metagenomic reads to enzymatic gene family using *HUMAnN* tool. And, for the calculated gene family abundance, we perform differential analysis to detect gene family with differential abundances using *DESeq2* R package. Third, enzymatic metabolite networks are established with individual interactions and gene families with differential abundances using *Piano* R package, and the network constructed in this way is referred to as the reporter network. Here, the information on individual interactions was curated from KEGG database. Finally, the individual metabolite in the reporter network is linked to enzymatic gene families that contribute to the production of metabolite, where changes in this metabolite are determined by considering both the significance and direction of its gene families.

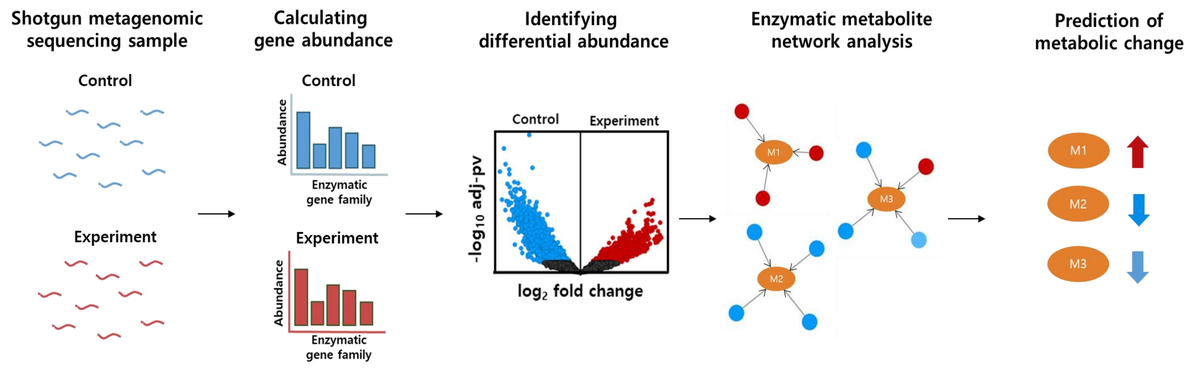

Figure 1. The conceptual outline of the proposed method

- Metabolic enrichment analysis in classes of molecules

Using statistics of individual metabolites as effect sizes, we perform enrichment analysis based on Wilcoxon signed-rank test to identify significantly enriched or depleted metabolic classes. Metabolic classes use the chemical classes level of Human Metabolome Database (HMDB).

- Statistical significance evaluation of the overlap using random permutation test

The null distribution is estimated by shuffling the label of enriched metabolic classes for the label of total metabolic classes in each predicted and measured group, overlapping between the predicted and measured group and recalculating many (e.g. 10,000) times. The p-value is calculated based on this null distribution.

Results

For benchmark evaluation of the proposed method, we used a publicly available paired dataset of shotgun metagenomics and metabolomics from an inflammatory bowel diseases (IBDs) cohort with controls, including 155 patients: 68 with CD, 53 with UC, and 34 non-IBD controls(Table 1).

IBD-subtype	Size
Crohn's disease (CD)	68
Ulcerative colitis (UC)	53
Non-IBD control	34

Table 1. Subtype and size of samples in IBD cohort

Through the prediction method described earlier, the identified metabolites are categorized under HMDB classes. 1,474 metabolites were assigned to 324 metabolic classes in CD (comparison group of CD versus non-IBD control), 1,056 metabolites were assigned to 264 metabolic classes in UC (comparison group of UC versus non-IBD control)(Figure 2).

We also evaluated the concordance between our predicted metabolic changes and the previously reported metabolic changes. In both CD and UC groups, the number of overlaps between the predicted and the measured metabolic class was 9 and 5, respectively, showing the statistical significance for the overlap in both groups as illustrated in Figure 3(p-value < 0.00005, random permutation test).

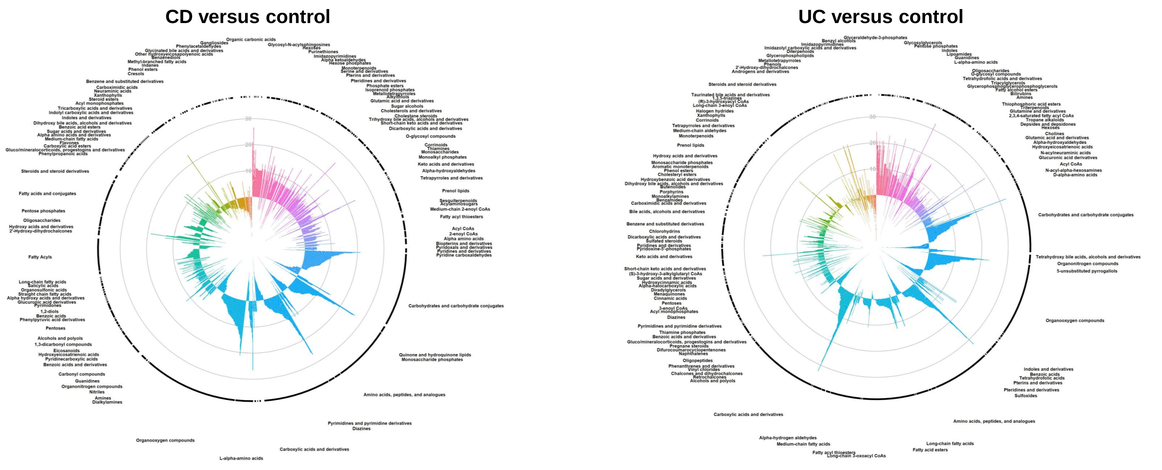

Figure 2. The circular plot of the overall changes and classes of predicted metabolites

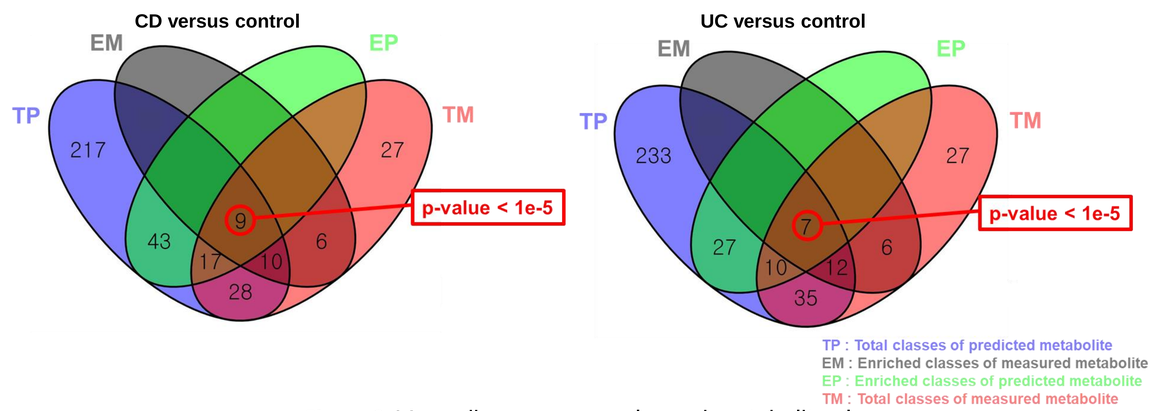

Figure 3. Venn diagram on overlapped metabolite classes

Among the overlapping between the predicted and the measured metabolic classes, we also checked that the direction of changes matched. In particular, it was shown that the predicted class with a large change coincided with the measured class in direction.

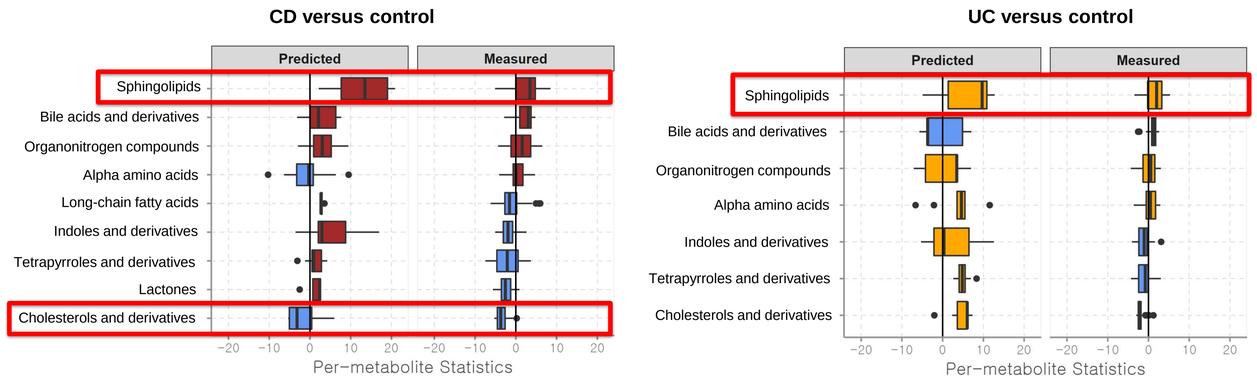

Figure 4. Comparison of the direction of changes between the predicted and measured metabolic class

Conclusion

In this study, the proposed method enables to directly estimate each metabolic potential of a given microbial genomic content by predicting the change of individual metabolites levels. The predicted metabolite with large changes shows statistically significant overlap with the measured metabolite and the concordance of the direction of its changes. This method will help narrow a number of metabolites down to their target range before performing costly and incomplete mass spectrometry.

References

1. Franzosa, E.A., et al., "Gut microbiome structure and metabolic activity in inflammatory bowel disease", *Nature Microbiology* 4, 293-305, 2019.
2. Väremo, L., et al., "Enriching the gene set analysis of genome-wide data by incorporating directionality of gene expression and combining statistical hypotheses and methods", *Nucleic acids research* 41, 4378–4391 2013.

Acknowledgement

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government Ministry of Science and Information & Communication Technology (MSIT) [2020M3C9A5085206].