Identifying gut microbes that affect human health

Confounders of Gut-Microbiome and Human Health

Sigal Leviatan, Eran Segal

The human microbiome is a complex ecology, consisting of tens of trillions of cells of a diverse community of microorganisms, mainly bacteria. Our gut microbiome, the largest and most diverse of these communities, is in constant interaction with the cells and systems of our body [1], and is both shaping and being shaped by our health status. The composition and diversity of the gut microbiome are known to be associated with multiple health disorders [2]. However, it is not known whether these associations are a result of the health conditions, or whether they partly cause them. Addressing this problem is highly challenging, due to the many physiology and lifestyle differences that exist between healthy and diseased individuals. Such confounders, variables that correlate with both the microbiome and the health status, may underlie the many discrepancies observed between different studies linking gut microbiome composition and human health [4].

On page xxx, Vujkovic-Cvijin et al. tackle this problem[3]. First, they identify physiological and lifestyle differences between diseased individuals and healthy individuals, which themselves may be associated with gut microbiome composition. Such differences can be a cause of variation in gut microbiome composition between diseased and health individuals. Without knowing about these differences, it is easy to misclassify an association between lifestyle (the confounder) and microbiome as an association between disease and microbiome. Then, they attempt to deal with such confounders by one-to-one matching [5] individuals from the disease cohort with healthy individuals who are similar to them in potential confounders (e.g., an individual of the same age, gender, and BMI). Such matching procedures are often used in observational studies, where one cannot assign individuals randomly to two groups and subject them to the two different scenarios being compared [6].

Vujkovic-Cvijin et al. find that gender, age, bowel movement quality, body-mass-index (BMI), and alcohol consumption are among the strongest such confounders to finding true associations between disease and gut-microbiome composition, as these characteristics are strongly associated both with microbiome and with disease status. When examining the differences between individuals with a disease such as Type-2-Diabetes (T2D) and healthy individuals, there appear to be many significant associations between the disease status and the abundances of different gut bacteria. In contrast, when matching individuals on the above phenotypes, many of these associations cease to be significant. This implies that some of the gut-microbiome changes previously attributed to disease may stem from different underlying causes related to these confounders. For example, if alcohol consumption causes gut-microbiome changes, and if

individuals suffering from a certain disease consume less alcohol (perhaps due to side effects of the drugs they take) then failing to match on alcohol intake may mislead a study into concluding that the changes are attributable to the disease itself when in fact they stem from reduced alcohol consumption.

A potential problem with the approach suggested by Vujkovic-Cvijin et al. is that some of the suggested confounders can be symptoms of a disease rather than lifestyle choices, and thus people having these symptoms may actually already be sick or on the path to being sick but they are currently undiagnosed. In such cases, the matching process might actually introduce bias [7]. For example, matching on alcohol intake makes no sense when studying alcoholic liver disease (ALD). Moreover, even if the potential confounders are not the defining symptoms of the disease in question or not unique symptoms of that disease, one should still worry that matching on them necessarily means that the resulting matched cohort would not be representative of a healthy cohort. For example, matching lung cancer patients to individuals with no cancer on the number of years of heavy smoking will not produce a truly healthy control cohort. With that in mind, one cannot match inflammatory bowel disease (IBD) patients to a healthy cohort on bowel movement quality, or match T2D patients to a healthy cohort on blood HbA1C levels (which the authors do not do). Moreover, one should be suspicious of matching T2D patients to a healthy cohort on BMI.

In an effort to address this issue, the authors repeated their analysis on a smaller cohort, with no individuals who self-reported any disease, and found similar results though less significant. Alas, removing individuals with any reported disease does not rule out matching with undiagnosed individuals or borderline diseased individuals, e.g., matching diabetics to pre-diabetics instead of a truly healthy match. This issue is of greater scope than merely for this study and merits discussion of what is considered to be a healthy cohort across all medical studies.

Finally, it is important to remember that identifying potential confounders between gut-microbiome composition and human health does not imply neither the existence not the lack of causality. For example, if alcohol consumption causes changes to the microbiome which in turn contribute to developing T2D, then a causal effect between the microbiome and disease exists, but will not be seen after matching on alcohol consumption. The same will be true if IBD causes the types of changes to the microbiome which cause diarrhea, and we match individuals on bowel movement quality. Thus, the results of Vujkovic-Cvijin et al. do not rule out causal effects through the microbiome. The question of causality between microbiome and disease is a major issue in microbiome human research and will certainly continue to fuel research in the field for years to

come, and Vujkovic-Cvijin et al. took an important step forward in advancing our thinking on this issue.

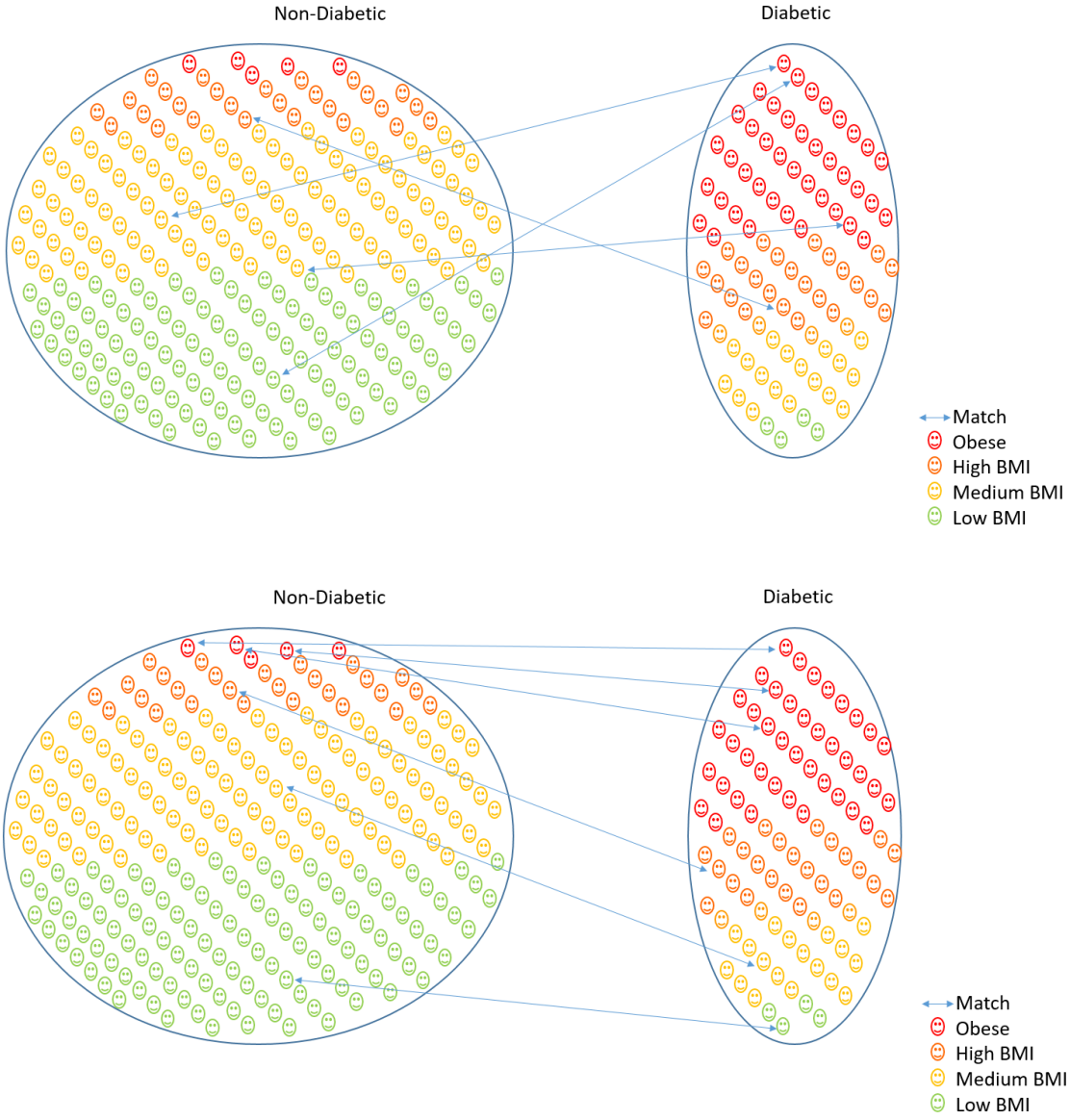

Figure1:

a. in random matching phenomena attributed to diabetes, may actually stem from BMI.
b. in matching on BMI, a non-representative sub-cohort of the non-diabetic is used, which might be defined by a non-healthy status, e.g. enriched for pre-diabetics.

1. DeSalle, R., & Perkins, S. L. Welcome to the Microbiome. Yale University Press (2015).
2. Honda, K., & Littman, D. The microbiota in adaptive immune homeostasis and disease. Nature 535, 75–84 (2016).
3. Vujkovic-Cvijin, I., Sklar, J., Jiang, L., Natarajan, L., Knight, R., Belkaid, Y. Host variables confound gut microbiota studies of human disease. Nature xxx, xx-xx (2020).
4. Clemente, J. C., Ursell, L. K., Parfrey, L. W. & Knight, R. The impact of the gut microbiota on human health: an integrative view. Cell 148, 1258–1270 (2012).
5. Schlesselman JJ. Case-Control Studies: Design, Conduct, Analysis. Oxford University Press; Oxford: 1982.
6. Rose, S., van der Laan, M.J. Why Match? Investigating Matched Case-Control Study Designs with Causal Effect Estimation. Int J Biostatistics 5:1 (2009).
7. Costanza MC. Matching. Preventive Medicine. 1995;24:425–433. doi: 10.1006/pmed.1995.1069.