ISSN 2282-0930 • Epidemiology Biostatistics and Public Health - XIII^th SISMEC Congress - Vol. 20 Suppl. 1

Microbiota Data: A Statistical Analysis Workflow

Ciniselli Chiara Maura[1], Duroni Valeria[1], Blanda Adriana[1], Bernardo Giancarla[2], Licata Armando[3], De Cecco Loris[3], Sfondrini Lucia[2], Tagliabue Elda[2], Sozzi Gabriella[4], Verderio Paolo[1]

(1) Unit of Bioinformatics and Biostatistics, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy
(2) Unit of Microenvironment and Biomarkers of Solid Tumors, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy
(3) Unit of Integrated Biology of Rare Tumors, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy
(4) Unit of Epigenomics & Biomarkers of Solid Tumors, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy

CORRESPONDING AUTHOR: *Ciniselli Chiara Maura, chiara.ciniselli@istitutotumori.mi.it*

INTRODUCTION

The microbiota includes the collection of microorganisms that colonize our body, playing a pivotal role in maintaining homeostasis and modulating immune responses. An imbalance in the composition of these microbes is associated with various diseases, including oncological conditions [1]. From a methodological perspective, several aspects should be considered during microbiome analysis, such as the compositional nature of the data, its high dimensionality, overdispersion, and the prevalence of zero-inflated count data [2]. The institutional microbiome project currently active at the Fondazione IRCCS Istituto Nazionale dei Tumori was designed to assess the feasibility of investigating the microbiota in different tumors settings and biological matrices, with the final aim of identifying microbial communities able of distinguishing patients characterized by worse prognosis from those with less aggressive disease. The cancer settings included in the projects are lung, breast and prostate cancer, as well as pseudomyxoma peritonei.

OBJECTIVES

Given the unique characteristics of microbiome data, we propose a unified workflow designed for analysing data from various cancer settings to ensure consistency and comparability across the different cancer types.

METHODS

Biological samples analysed in this study were disease-specific and include, tumor tissues and matched normal counterpart for lung cancer, tumor tissues for breast and prostate cancer, matched tumor tissue, feces and mucin for pseudomyxoma peritonei. Samples were processed using the same standardized analytical pipeline across all pathologies. This process includes bacterial DNA extraction, enrichment, library preparation and sequencing of the variable regions of the bacterial 16S rRNA gene. Microbial communities were firstly characterized in terms of alpha- and beta-diversity: the first one quantifies the diversity within a given sample in terms of richness or eveness [3], whereas second one assesses diversity differences between sample-groups. Specifically, for alpha-diversity we used a set of different indices: Chao1, Hill and Observed for richness, Gini-Simpson and Shannon for eveness [3]; for beta-diversity the Principal Coordinate Analysis (PCoA) analysis [4,5] was adopted together with the PERmutational Multivariate ANalysis of VAriance (PERMANOVA) [4] and PERmutational Multivariate analysis of DISPersion (PERMDISP) tests [6] by using the Bray-Curtis distance metrics [5]. Then, the workflow incorporates tests and statistical models for both continuous and categorical data to identify bacterial taxa that are differentially expressed or present in distinct biological matrices or associated with clinical-pathological characteristics investigated in each cancer context. Test for paired or unpaired groups comparison were included for both continuous and categorical data analysis.

RESULTS

Using lung cancer as a model setting, matched tumor and normal counterpart of 155 lung cancer patients (stage I-III), were profiled using 16S rRNA gene sequencing for a total of 310 observation and 63 identified bacteria taxa (i.e. genus level). The analysis revealed distinct microbial compositions, in terms of beta-diversity, according to histology. Moreover, by modelling the count data and its presence/absence, a specific subset of bacteria was significantly associated with tumor progression and aggressiveness; specific bacteria were also found differentially expressed between tissue types (tumor or normal counterpart). We are now applying this workflow at the other settings, with the final aim of better understanding the bacteria that characterize tumor aggressiveness.

DOI: 10.54103/2282-0930/29363

CONCLUSION

The developed workflow allows: (i) the use of a shared pre-analytical and analytical workflow of analysis among the different tumor settings under investigation, (ii) the characterization of the microbial community within/between samples and (iii) the evaluation of associations between specific taxa bacteria and tumor characteristics.

REFERENCES

1. Derosa L, Iebba V, Silva CAC, et al., Custom scoring based on ecological topology of gut microbiota associated with cancer immunotherapy outcome. Cell., 2024 Jun;187(13):3373-3389.e16
2. Jonsson V, Österlund T, Nerman O, et al., Modelling of zero-inflation improves inference of metagenomic gene count data. Stat Methods Med Res., 2019 Dec;28(12):3712-3728
3. Koh H, Tuddenham S, Sears CL, et al., Meta-analysis methods for multiple related markers: Applications to microbiome studies with the results on multiple α-diversity indices. Stat Med., 2021 May;40(12):2859-2876
4. Goodrich JK, Di Rienzi SC, Poole AC, et al., Conducting a microbiome study. Cell., 2014 Jul;158(2):250-262
5. Peterson CB, Saha S, Do KA. Analysis of Microbiome Data. Annu Rev Stat Appl. 2024 Apr;11(1):483-504
6. Anderson MJ, Santana-Garcon J. Measures of precision for dissimilarity-based multivariate analysis of ecological communities. Ecol Lett. 2015 Jan;18(1):66-73

DOI: 10.54103/2282-0930/29363