

Facing the challenges of non-standardized data: a curated meta-analysis on skin microbiome in aging

Fumagalli S.¹, Bruno A.¹, Soletta G.¹, Armanni A.¹, Labra M.¹, Casiraghi M.¹

E-mail: s.fumagalli66@campus.unimib.it

¹ Biotechnology and Biosciences Department, University of Milano-Bicocca, Italy.

Keywords: human microbiome, skin aging, data curation, machine learning

Abstract:

Years-long exponential increase of microbiome-related studies has resulted in a huge amount of data currently stored in public repositories, and more continue to be added. These data represent a crucial resource: pharmaceutical and cosmetic industries are showing a growing interest in human microbiome whose entire diversity can be captured solely through multiple datasets integration. Although research is thus being pushed towards meta-analyses, extracting the valuable information hidden in several datasets remains challenging.

Implementing SkiomeProject, we collected, harmonized, and analyzed 16S rRNA sequences obtained from skin microbiome-related projects deposited on international public databases. We focused on healthy skin samples, considering the overall person's well-being across all the stages of his life.

Here, we present the results of our meta-analysis, aiming at the age-related variations in healthy skin face microbiome. The lack of standardization across projects hinders biological information detection as biases introduced by the heterogeneous protocols' adoption are inbuilt into the stored data, hampering datasets comparison. However, by coupling multiple manual curated steps with advanced statistical analyses (i.e. supervised machine learning approach) we buffered the initial technical biases. Moreover, we identified the significant microbial taxa characterizing young and old skin, correctly predicting aging from microbiome composition.

In this study, we provide new insight into rebalancing the skin microbiome during aging. Finally, and most importantly, our results highlight the urgency of a shared coordinated effort in i) methods standardization, ii) metadata compilation consistency, and iii) raw data fairness to drive the future of skin science.