# **Introduction**

The past years of research have unveiled a fundamental role of the human gut microbiome in health and disease. The gut microbiome provides the host with valuable resources, for example, nutrient bioconversion through metabolic capabilities, otherwise absent in the host, or protection against pathogens (Young, 2017) \*https ://doi.org/10.1136/bmj.j831.

A deeper understanding of the functional potential and the taxonomic composition of the microbiome may have major implications in the identification of signatures in health and disease across different regions (David et al., 2014; Lozupone et al., 2012; Sommer et al., 2017)\* [8, 10, 11]; could eventually lead into a better comprehension in the underlying mechanisms of the interactions between the host and the microbiome and even, into new treatments and therapies (Schupack et al., 2021).

Shotgun Metagenomics has constituted a central tool for the study of microbiomes without the need of cultivating them. A large number of studies in the las years had provided taxonomic and functional information about of normal and disease populations. These studies have covered a broad group of diseases across many geographic areas, most of them focused on the discovery of new uncultured microbes and some of them also on the functional potential of the gut microbiome (David et al., 2014; Jalanka-Tuovinen et al., 2011; Mehta et al., 2018)\* [7]–[9].

Huge amounts of sequence data from shotgun metagenomic studies have been deposited in public databases, and are steadily increasing, providing a unique opportunity to perform a large-scale integration analysis. Large-scale integration of microbiome would bring a broader point of view bringing the possibility of extracting new information, getting more general conclusion or bringing stronger support to previous observations about the microbiome and host physiology.

Public resources collecting and processing microbiome data exist, contributing to laborious and necessary task to standardize and make accessible this accumulated information. Some of them focused in particular in Human gut microbiome (gutMDisorder , GIMICA, and DISBIOME and GMRepo (Cheng et al., 2020; Dai et al., 2022; Janssens et al., 2018; Tang et al., 2021)). Herein, we integrated publicly available data from many studies studyng in particular the Human Gut Microbiome across different countries from healthy and diseased individuals taking advantage of metagenomic approaches which enabled us the identification of metagenome species and their pangenome content.

The current study provides a global map of the human gut microbiome in healthy individuals, the identification of species and functions enriched or depleted across a number of disease cohorts and machine learning models to classify healthy against disease samples. The analysis is presented in an open-access Human Gut Microbiome Atlas ([www.microbiomeatlas.org](http://www.microbiomeatlas.org)), allowing researchers to explore for the first time an integrative analysis on composition, functional, richness, diseases, and region signatures for the gut microbiota across 19 geographical regions and diseases.

Cheng, L., Qi, C., Zhuang, H., Fu, T., & Zhang, X. (2020). gutMDisorder: a comprehensive database for dysbiosis of the gut microbiota in disorders and interventions. *Nucleic Acids Research*, *48*(D1), D554–D560. <https://doi.org/10.1093/NAR/GKZ843>

Dai, D., Zhu, J., Sun, C., Li, M., Liu, J., Wu, S., Ning, K., He, L., Zhao, X.-M., & Chen, W.-H. (2022). GMrepo v2: a curated human gut microbiome database with special focus on disease markers and cross-dataset comparison. *Nucleic Acids Research*, *50*(D1), D777–D784. <https://doi.org/10.1093/NAR/GKAB1019>

David, L. A., Materna, A. C., Friedman, J., Campos-Baptista, M. I., Blackburn, M. C., Perrotta, A., Erdman, S. E., & Alm, E. J. (2014). Host lifestyle affects human microbiota on daily timescales. *Genome Biol*, *15*(7), R89. <https://doi.org/10.1186/gb-2014-15-7-r89>

Jalanka-Tuovinen, J., Salonen, A., Nikkila, J., Immonen, O., Kekkonen, R., Lahti, L., Palva, A., & de Vos, W. M. (2011). Intestinal microbiota in healthy adults: temporal analysis reveals individual and common core and relation to intestinal symptoms. *PLoS One*, *6*(7), e23035. <https://doi.org/10.1371/journal.pone.0023035>

Janssens, Y., Nielandt, J., Bronselaer, A., Debunne, N., Verbeke, F., Wynendaele, E., van Immerseel, F., Vandewynckel, Y. P., de Tré, G., & de Spiegeleer, B. (2018). Disbiome database: Linking the microbiome to disease. *BMC Microbiology*, *18*(1), 1–6. <https://doi.org/10.1186/S12866-018-1197-5/TABLES/1>

Lozupone, C. A., Stombaugh, J. I., Gordon, J. I., Jansson, J. K., & Knight, R. (2012). Diversity, stability and resilience of the human gut microbiota. *Nature*, *489*(7415), 220–230. <https://doi.org/10.1038/nature11550>

Mehta, R. S., Abu-Ali, G. S., Drew, D. A., Lloyd-Price, J., Subramanian, A., Lochhead, P., Joshi, A. D., Ivey, K. L., Khalili, H., Brown, G. T., DuLong, C., Song, M., Nguyen, L. H., Mallick, H., Rimm, E. B., Izard, J., Huttenhower, C., & Chan, A. T. (2018). Stability of the human faecal microbiome in a cohort of adult men. *Nat Microbiol*, *3*(3), 347–355. <https://doi.org/10.1038/s41564-017-0096-0>

Schupack, D. A., Mars, R. A. T., Voelker, D. H., Abeykoon, J. P., & Kashyap, P. C. (2021). The promise of the gut microbiome as part of individualized treatment strategies. *Nature Reviews Gastroenterology & Hepatology 2021 19:1*, *19*(1), 7–25. <https://doi.org/10.1038/s41575-021-00499-1>

Sommer, F., Anderson, J. M., Bharti, R., Raes, J., & Rosenstiel, P. (2017). The resilience of the intestinal microbiota influences health and disease. *Nat. Rev. Microbiol.*, *15*(10), 630–638. <https://doi.org/10.1038/nrmicro.2017.58>

Tang, J., Wu, X., Mou, M., Wang, C., Wang, L., Li, F., Guo, M., Yin, J., Xie, W., Wang, X., Wang, Y., Ding, Y., Xue, W., & Zhu, F. (2021). GIMICA: host genetic and immune factors shaping human microbiota. *Nucleic Acids Research*, *49*(D1), D715–D722. <https://doi.org/10.1093/NAR/GKAA851>

Young, V. B. (2017). The role of the microbiome in human health and disease: an introduction for clinicians. *BMJ*, *356*. <https://doi.org/10.1136/BMJ.J831>

*Introduction content.*

*1 Human gut microbiome, source of metabolic variabliity, role in health and disese.*

*2 Metagenomic approach, , studying healthy populations and broad range of diseases.*

*3 Large scale integration, data accumulation. some integration efforts seen. Allows accessibility novel knowledge extraction, require standardized analytical techniques*

*4 Human gut microbiome atlas*

*Supplementaty FIgure. Streptococcus effect size across all cohorts.* The figure above displays the branch containing all the MSPs from the Streptococcus genus and the estimated effect size across all disease cohorts included in this study. The bar plots on the leftmost side of the figure show the number of cohorts where an MSP‘s effect size was above 0.3. Red colour indicates depletion and blue colour, enrichment. The graph on the top shows the hierarchical clustering result using the Ward method on the Effect Size matrix.