**R code tutorial: Assessing the effect of Vitamin D exposure on Fn biofilm formation**

By Leticia Testa

The gut microbiome is a vital component of human health. The gastrointestinal tract is home to an abundance of microorganisms such as bacteria [1]. However, even though bacteria are a crucial part of the microbiome, many of these bacteria are known to form biofilms, which can hinder human health. Studies have shown that biofilm formation is correlated with tumor progression [3]. Therefore, there is a growing need to introduce new agents to deal with the emergence of biofilm forming bacteria that are involved in infection and tumor evolution. New scientific research has emphasized the role of vitamin D in the immune system and inflammatory processes [3]. This tutorial describes the data analysis that wishes to answer the question: Does vitamin D change the ability of a colorectal carcinoma (CRC) associated Fusobacterium nucleatum strain to form biofilm?

Vitamin D, along with compounds that are known to increase biofilm-formation (used as control) will be tested in different concentrations (dilution factor). The absorbance was loaded in an excel file, and further manipulated so it would be plotted in R studio. It is worth noting that since this is unpublished data, the absorbances had to be modified. In excel, the data was first separated into three columns: treatment, absorbance, and absorbance (ABS). After data manipulation, I cleaned the global environment and installed all packages needed – tidyverse, readxl, ggpubr, and ggtext [4]. After uploading the data from my directory, I separated the data into different treatment groups to further perform statistical analysis [5,6]. I spent a good number of hours searching in literature good packages and functions for this, and found that ‘geom\_abline’, ‘geom\_smooth’, and ‘geom\_errorbar’ were quite helpful. Furthermore, the rest of the tutorial involved using ‘annotate’ to make the graph easy to understand [7]. Statistical correlations were a crucial part of my tutorial and the function ‘stat\_cor’ was very helpful. The plot was generated through the easy to manage ‘ggplot’. As we learned in class, that is an easy function used to create a scatterplot. My data was saved with ‘ggsave’ and my plot was ready for presentation (slides included in GitHub).

**References**

[1] Motta, JP., Wallace, J.L., Buret, A.G. *et al.* Gastrointestinal biofilms in health and disease. *Nat Rev Gastroenterol Hepatol* **18**, 314–334 (2021). <https://doi.org/10.1038/s41575-020-00397-y>

[2] Kherad Z, Yazdanpanah S, Saadat F, Pakshir K, Zomorodian K. Vitamin D3: A promising antifungal and antibiofilm agent against *Candida* species. Curr Med Mycol. 2023 Jun;9(2):17-22. doi: 10.18502/cmm.2023.345062.1416.

[3] Saputo S, Faustoferri RC, Quivey RG Jr,. Antimicrob Agents Chemother. 2017 Dec 21;62(1):e01675-17.

[4] Giorgi FM, Ceraolo C, Mercatelli D. The R Language: An Engine for Bioinformatics and Data Science. Life (Basel). 2022 Apr 27;12(5):648. doi: 10.3390/life12050648.

[5] Ito K, Murphy D. Application of ggplot2 to Pharmacometric Graphics. CPT Pharmacometrics Syst Pharmacol. 2013 Oct 16;2(10):e79. doi: 10.1038/psp.2013.56. PMID: 24132163; PMCID: PMC3817376.

[6] Gelfond J, Goros M, Hernandez B, Bokov A. A System for an Accountable Data Analysis Process in R. R J. 2018 Jul;10(1):6-21. Epub 2018 May 15.

[7] Giorgi FM, Ceraolo C, Mercatelli D. The R Language: An Engine for Bioinformatics and Data Science. Life (Basel). 2022 Apr 27;12(5):648. doi: 10.3390/life12050648.

[8] Xie, Yihui. 2015. *Dynamic Documents with R and Knitr*. 2nd ed. Boca Raton, Florida: Chapman; Hall/CRC

[9] Ito K, Murphy D. Application of ggplot2 to Pharmacometric Graphics. CPT Pharmacometrics Syst Pharmacol. 2013 Oct 16;2(10):e79. doi: 10.1038/psp.2013.56.

[10] Drăghici S. Statistics and data analysis for microarrays using R and Bioconductor, 2nd Edition. CRC Press; 2012.

[11] Stiglic G, Watson R, Cilar L. R you ready? Using the R programme for statistical analysis and graphics. Res Nurs Health. 2019 Dec;42(6):494-499. doi: 10.1002/nur.21990.