Regularized GLMs to find bacteria associated with colorectal cancer

Eliana Ibrahimi
Department of Biology, University of Tirana (AL)

Joint work with Marta B. Lopes
NOVA MATH, FCT NOVA, Lisbon (PT)

APC Microbiome Ireland & ML4Microbiome Conference

7-9 June 2023, University College Cork (UCC), Ireland

ML4 MICROBIOME

cost
EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

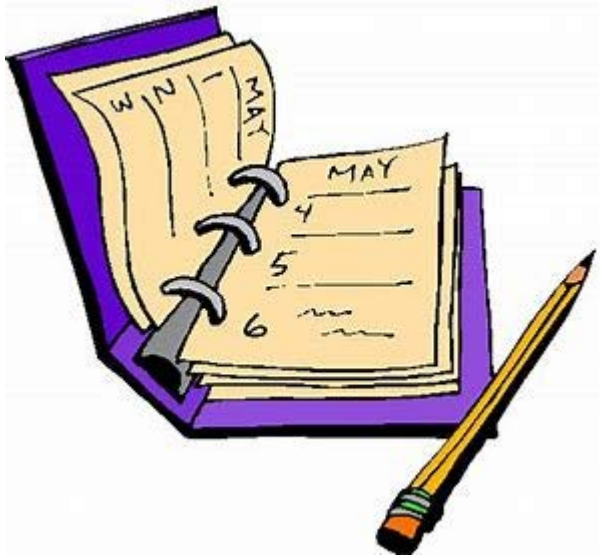
Talk outline

About the STSM

Research project

Benefits and post collaborations

Me & ML4Microbiome

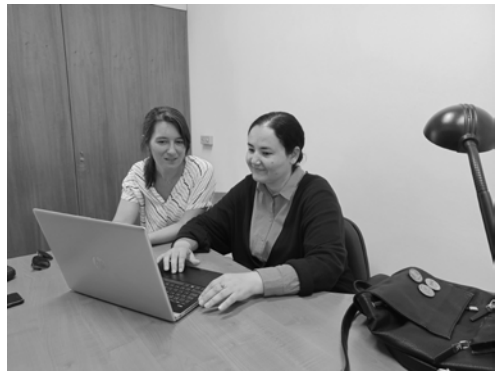


About the STSM

STSM at NOVA MATH, FCT NOVA, Lisbon, September 12 to October 4, 2022.

STSM host and supervisor: **Dr. Marta B. Lopes**

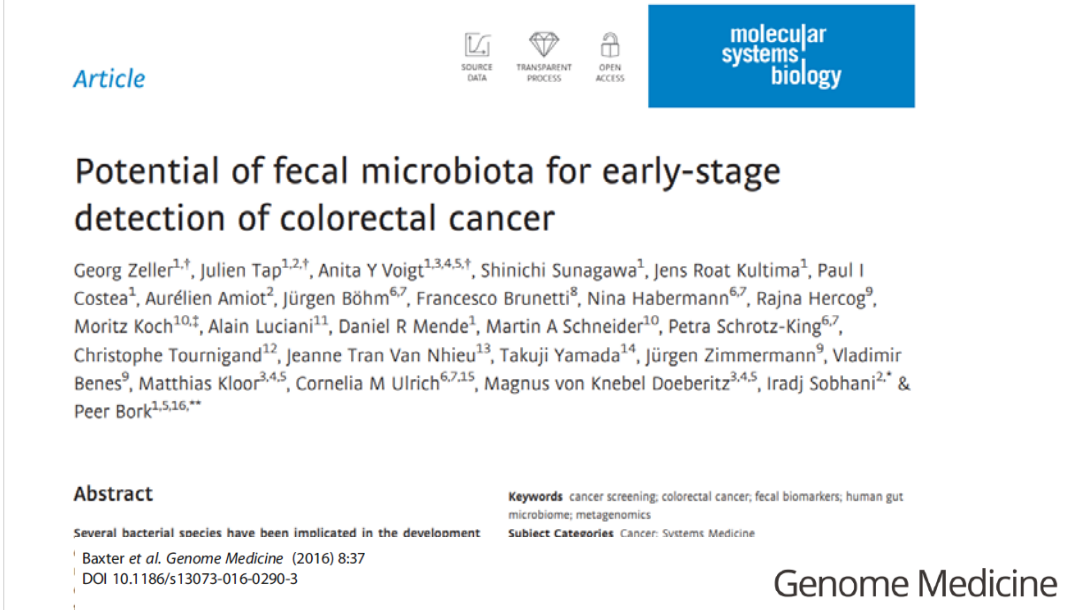
Topic:
Regularized GLMs to analyze the association of gut microbiome with colorectal cancer.



Research project

Aim: Optimize regularized GLMs to analyze the association of gut microbiome with colorectal cancer.

- ✓ CRC WG3 16S dataset
  - 219 samples
  - 503 features, including age, BMI, country, and genus counts.



Article

molecular systems biology

Potential of fecal microbiota for early-stage detection of colorectal cancer

Georg Zeller[1,†], Julien Tap[1,2,†], Anita Y Voigt[1,3,4,5,†], Shinichi Sunagawa[1], Jens Roat Kultima[1], Paul I Costea[1], Aurélien Amiot[2], Jürgen Böhm[6,7], Francesco Brunetti[8], Nina Habermann[6,7], Rajna Hercog[9], Moritz Koch[10,‡], Alain Luciani[11], Daniel R Mende[1], Martin A Schneider[10], Petra Schrotz-King[6,7], Christophe Tournigand[12], Jeanne Tran Van Nhieu[13], Takuji Yamada[14], Jürgen Zimmermann[9], Vladimir Benes[9], Matthias Kloor[3,4,5], Cornelia M Ulrich[6,7,15], Magnus von Knebel Doeberitz[3,4,5], Iradj Sobhani[2,*] & Peer Bork[1,5,16,**]

Abstract

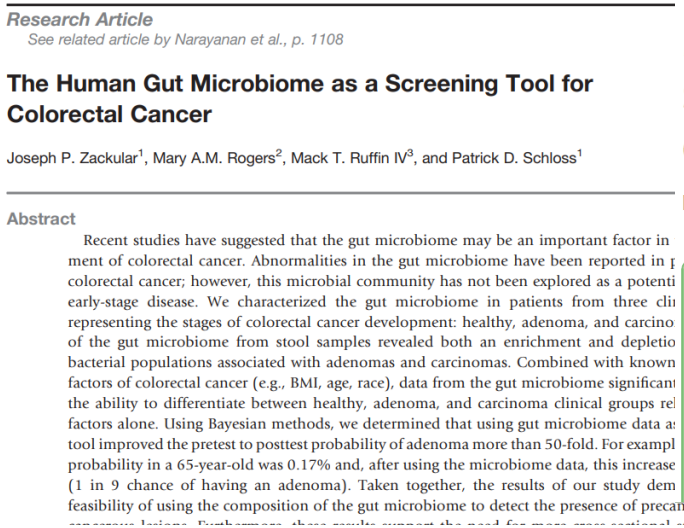
Several bacterial species have been implicated in the development

Keywords cancer screening; colorectal cancer; fecal biomarkers; human gut microbiome; metagenomics

Subject Categories Cancer; Systems Medicine

Baxter *et al. Genome Medicine* (2016) 8:37
DOI 10.1186/s13073-016-0290-3

Genome Medicine



Research Article

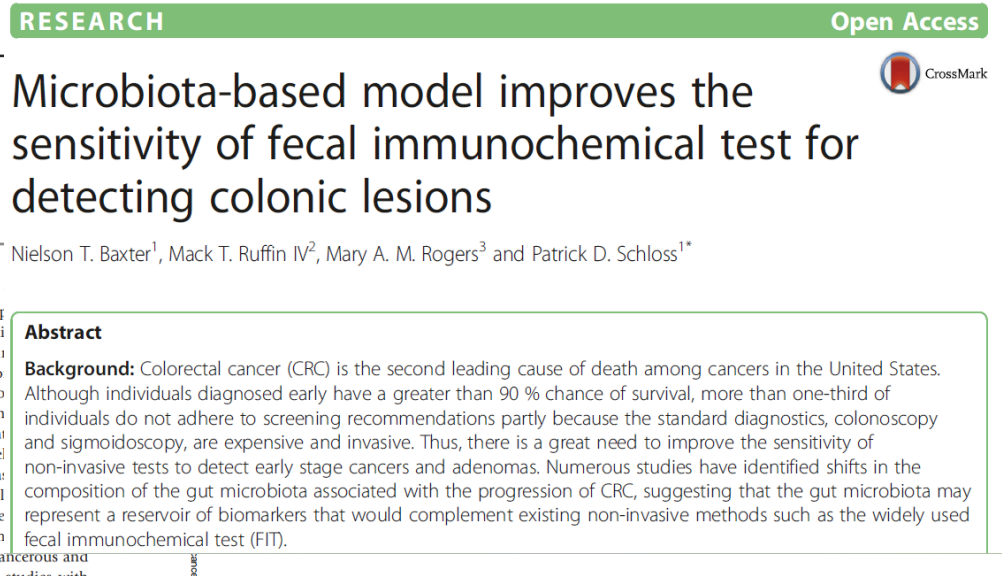
See related article by Narayanan et al., p. 1108

The Human Gut Microbiome as a Screening Tool for Colorectal Cancer

Joseph P. Zackular[1], Mary A.M. Rogers[2], Mack T. Ruffin IV[3], and Patrick D. Schloss[1]

Abstract

Recent studies have suggested that the gut microbiome may be an important factor in
ment of colorectal cancer. Abnormalities in the gut microbiome have been reported in
colorectal cancer; however, this microbial community has not been explored as a potenti
early-stage disease. We characterized the gut microbiome in patients from three cli
representing the stages of colorectal cancer development: healthy, adenoma, and carcino
of the gut microbiome from stool samples revealed both an enrichment and depletio
bacterial populations associated with adenomas and carcinomas. Combined with known
factors of colorectal cancer (e.g., BMI, age, race), data from the gut microbiome significant
the ability to differentiate between healthy, adenoma, and carcinoma clinical groups re
factors alone. Using Bayesian methods, we determined that using gut microbiome data a
tool improved the pretest to posttest probability of adenoma more than 50-fold. For exampl
probability in a 65-year-old was 0.17% and, after using the microbiome data, this increase
(1 in 9 chance of having an adenoma). Taken together, the results of our study dem
feasibility of using the composition of the gut microbiome to detect the presence of precancerous and



RESEARCH Open Access

CrossMark

Microbiota-based model improves the sensitivity of fecal immunochemical test for detecting colonic lesions

Nielson T. Baxter[1], Mack T. Ruffin IV[2], Mary A. M. Rogers[3] and Patrick D. Schloss[1*]

Abstract

Background: Colorectal cancer (CRC) is the second leading cause of death among cancers in the United States. Although individuals diagnosed early have a greater than 90 % chance of survival, more than one-third of individuals do not adhere to screening recommendations partly because the standard diagnostics, colonoscopy and sigmoidoscopy, are expensive and invasive. Thus, there is a great need to improve the sensitivity of non-invasive tests to detect early stage cancers and adenomas. Numerous studies have identified shifts in the composition of the gut microbiota associated with the progression of CRC, suggesting that the gut microbiota may represent a reservoir of biomarkers that would complement existing non-invasive methods such as the widely used fecal immunochemical test (FIT).

Data preprocessing and transformation

Filtering

- Low abundance filtering

Transformation

- Centered Log Ratio (CLR)
- Gaussian normalization (Z-score).

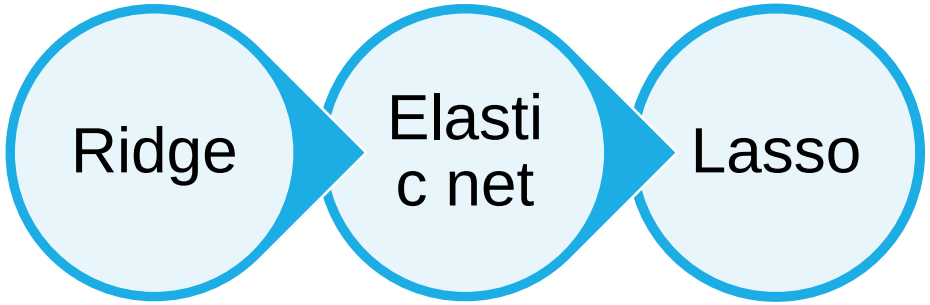
✓ Included in the analysis

- 219 samples
- 156 features, including age, BMI, and genus counts.

Regularized GLMs

- ✓ R packages glmnet, grplasso, SGL, and compositions.

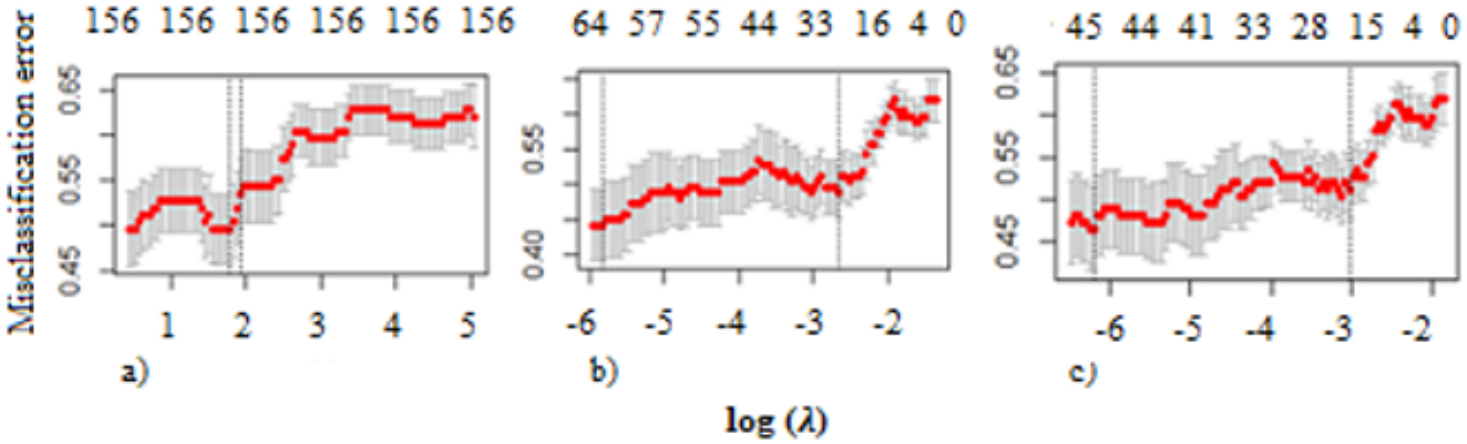
- ✓ Optimize regularized GLM for a **multiclass** (healthy/adenoma/cancer) classification task.



- ✓ Group Lasso and sparse group Lasso for a **binary** (healthy/cancer) classification task.

Parameter tuning

- Use a 10-fold cross-validation to select an optimal value for the tunning parameter, λ.
- For each λ, a predictive model is fitted in the training set (60% of the data) and then used to predict the outcome value of each sample in the test set (40%).

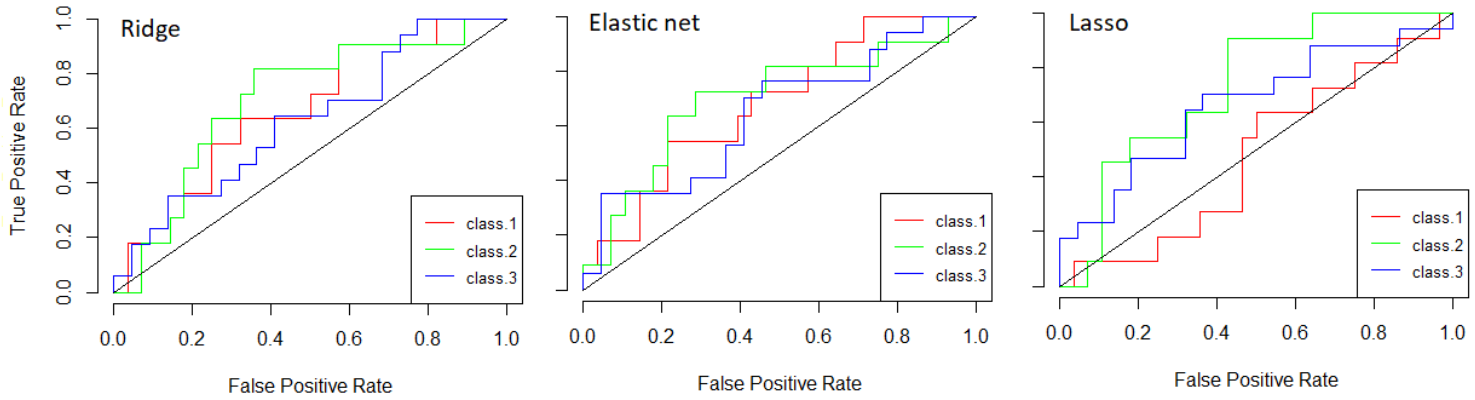


Plot of the misclassification error by log(λ) value for ridge (a), elastic net (α=0.6) (b), and lasso (c).

Findings

Multiclass models: Ridge and Elastic net applied on CLR transformed data showed higher accuracy.

The accuracy of the multinomial models when working with separate datasets was higher (i.e., 0.66 to 0.7) compared to that achieved from the merged dataset (i.e., 0.52 to 0.61).



class 1=adenoma; class 2=carcinoma; class3=healthy; Using data from Zeller et al. (2014).

Findings

- ✓ The group Lasso and sparse group Lasso analysis for a **binary** (healthy/cancer) classification task is not stabilized yet. Low accuracy is observed.

To summarize

- ✓ Factors such as the number of predictors, the sample size, and the desired sparsity level in the final model should be considered when selecting a regularization method.

- ➢ Extended findings will be presented at the ‘6th Conference on Statistics and Mathematics’ organized at Leipzig University of Applied Sciences (Germany) on July 14-16, 2023.

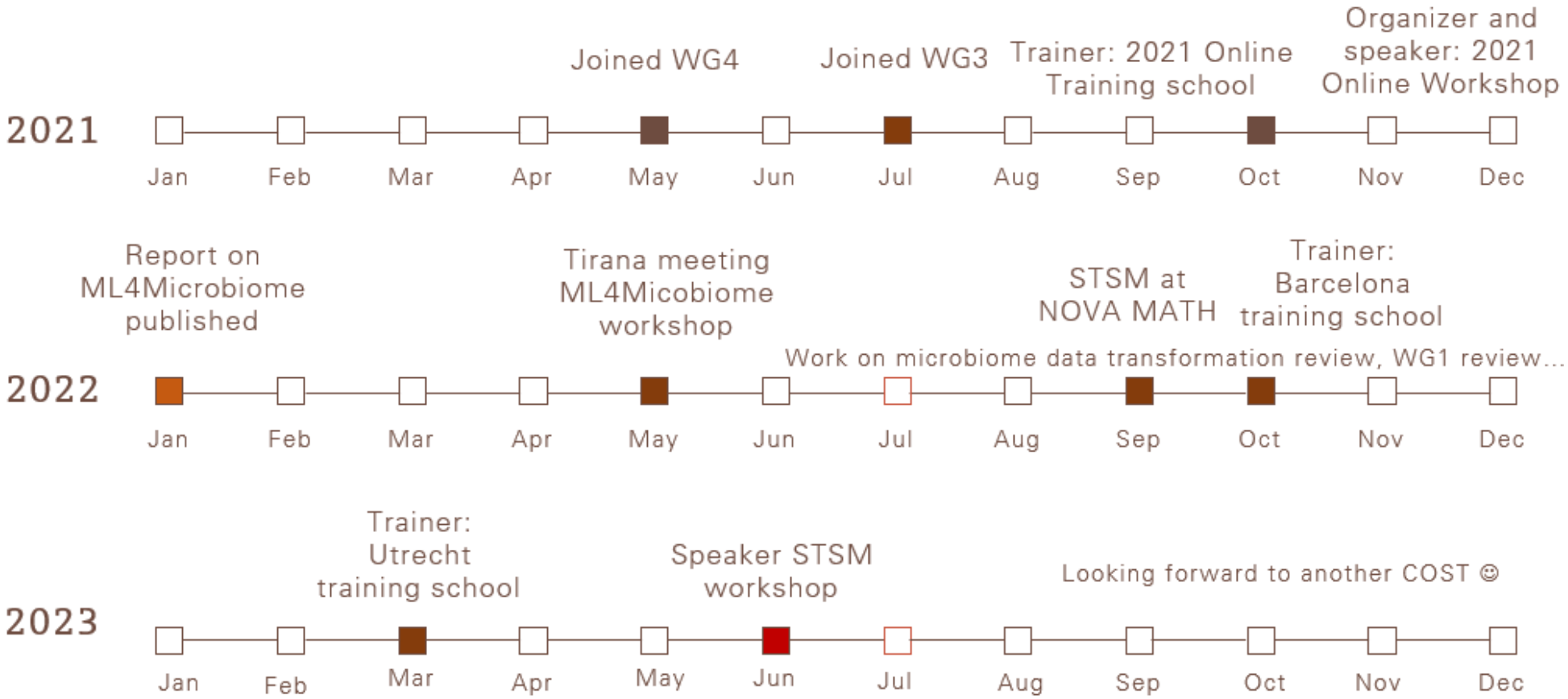
Other STSM activities

- Attended the ‘II Workshop in Statistics for Health and Public Health’.
- Gave a seminar on my current research and discussed it with interested researchers.
- Visited the research group of Prof. Susana Vinga at the Instituto Superior Técnico.

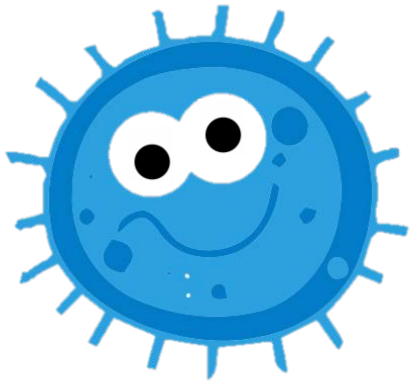
Post STSM collaborations

- Working on two papers on microbiome research.
- Co-supervising a master’s thesis at the Department of Biology, University of Tirana.
- Working on organizing a training school on statistical and machine learning modeling of biological data in October 2023 at the University of Tirana.

Me & ML4Microbiome



Many thanks



Marcus Claesson

Domenica D’Elia

Aldert Zomer

Leo Lahti

Tatjana Loncar-Turukalo

Magali Berland

Ilze Elbere

Christian Jansen

Marta B. Lopes

Alise Ponsero

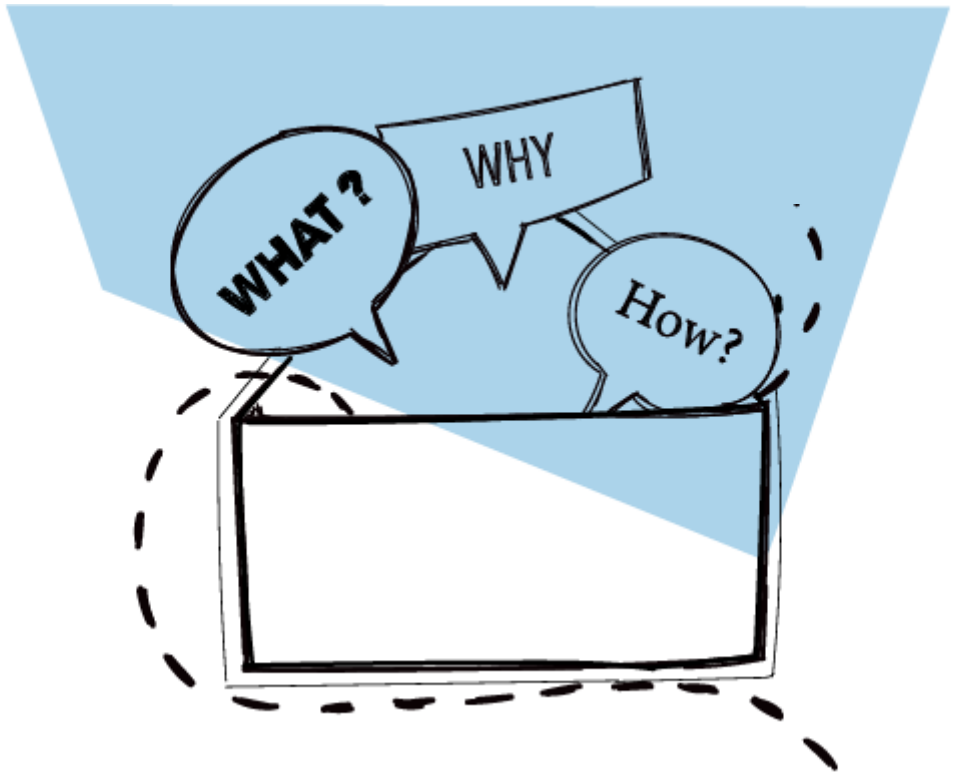
Laura Marcos Zambrano

Giorgos Papoutsoglou

ML4Microbiome community ...

cost
EUROPEAN COOPERATION
IN SCIENCE & TECHNOLOGY

ML4
MICROBIOME



Eliana Ibrahimi

Get in touch:

eliana.ibrahimi@fshn.edu.al

LinkedIn