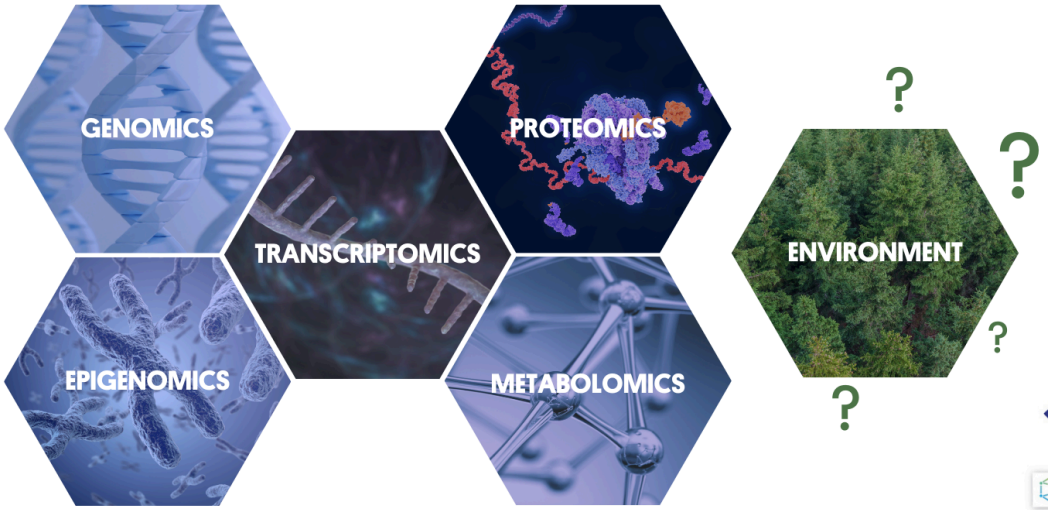
ISGlobal
EXCELENCIA SEVERO OCHOA

WORKSHOP SERIES

Multi-Omics Data Integration in Human Exposome Studies



ATHLETE The Exposome from Evidence to Translation

MEMBER OF THE European Human Exposome NETWORK

Organisers:
Léa Maitre and Augusto Anguita-Ruiz

Speakers:
Arthur Tenenhaus
Augusto Anguita
Congrong Wang
Chris Gennings
Inès Amine
Jesse A. Goodrich
Lea Maitre
Nikolaos Stratakis
Ville Pimenoff
Vishal Midya

Location: Online

N° Sessions: 10

Time: Sept 4th - Nov 20th 2024, Wednesdays. 4.30-5.30pm CET

Course overview:

Biological and biomedical research is increasingly driven by assays that measure multiple omics data types ("multi-omics"): transcriptome, proteome, genomic variants, epigenetic marks, metabolome, functional assays, imaging.The field of exposomics delves into the complex interplay of environmental, social, lifestyle, genetic, and other biological factors and warrants promise to advance precision environmental health. Beyond genetics, the integration of exposomics and multi-omics in epidemiology research can enable a more precise understanding of disease and promote health, especially for the most vulnerable population subgroups. Leveraging state-of-the-art approaches for exposomics in epidemiology research can accelerate our knowledge about complex exposure mixture effects, gene-environment interactions, disease mechanisms, and the translation of the exposome into public health and precision health strategies. As exposomics gains global recognition there is a pressing need to tackle barriers related to epidemiology study designs and analytical methods so that we can enhance the translation of exposomics into effective precision environmental health strategies.

In this workshop, we will introduce novel designs and analytical approaches for multi-omics integration that were recently developed and applied in the exposomics field to address critical data analysis challenges (dimensionality reduction approaches, mediation, predictive modelling, feature selection, networks, etc). The focus is on the practical application of these methods. Through practical sessions on relevant real-world examples from birth cohort exposome multi-omics studies, participants will gain familiarity with different analytical options for data integration, software implementations and learn about practical aspects, including quality assessment of inputs and outputs, method performance assessment, choice of tuning parameters, interpretation of the model fit and potential pitfalls.

Audience:

This workshop is aimed at Master/PhD students, researchers, clinicians, and other professionals working in the fields of exposomics, epidemiology, or toxicology with an interest in understanding how omics data, including the exposome, can be integrated in epidemiological research.

Objectives:

Comprehensive Understanding: Provide participants with a comprehensive understanding of the application of multi-omics analyses in unravelling the intricate connections between environmental exposures and human health.

Methodological Insights: Offer insights into state-of-art methods for multi-modal integration of omics data in human cohort studies, including epigenomics, transcriptomics, proteomics, metagenomics and metabolomics, enabling participants to grasp the tools at their disposal.

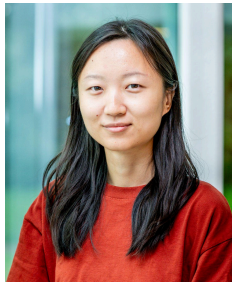
Collaboration: Foster interdisciplinary collaboration by bringing together researchers, scientists, and practitioners from diverse fields, creating a platform for knowledge exchange and the development of innovative research approaches.

Practical Applications: Facilitate hands-on learning experiences and practical applications of multi-omics analyses in exposome studies (providing participants with codes and resources for their implementation), empowering participants to apply these techniques in their own research endeavours.

Future Directions: Explore emerging trends and future directions in exposomics research, encouraging participants to envision novel possibilities and contribute to the evolving landscape of environmental health sciences.

Modules:

Date	Title	Speaker
Sept 4th 2024 *16:30h-18:00h Central European Time (CET)*	Introductory session: - Introduction to multi-omics data integration challenges in exposome studies - Overview of approaches	 **Dr. Léa Maitre**  **Dr. Augusto Anguita**
	- Multi-modal data integration through dimensionality reduction: Motivation, concepts and theory	 **Dr. Arthur Tenenhaus**

Sept 25th 2024 *16:30h-17:30h Central European Time (CET)*	- Practical aspects of multi-modal data integration through dimensionality reduction (MOFA, PLS, RGCCA, etc). - Case studies of multi-omics and exposome data integration through dimensionality reduction in HELIX cohorts	 **Inès Amine**  **Congrong Wang**
Oct 2nd 2024 *16:30h-17:30h Central European Time (CET)*	Strategies for Multivariate Mediation Frameworks in Environmental Epidemiology	 **Dr. Vishal Midya**

Oct 9th 2024 *16:30h-17:30h Central European Time (CET)*	Multi-Omic Analysis Framework For Precision Health: Omic mediation analysis in exposome research	 **Dr. Jesse A. Goodrich**
Oct 23th 2024 *16:30h-17:30h Central European Time (CET)*	Networks visualisation for multi-omics integration	 **Dr. Léa Maitre**

Oct 30th 2024 *16:30h-17:30h Central European Time (CET)*	Development of exposome mixture methods for high-dimensional omics data to investigate environmental risk factors of disease	 **Dr. Chris Gennings**  **Dr. Anna Young**
Nov 6th 2024 *16:30h-17:30h Central European Time (CET)*	Comprehensive biotic exposome estimation from wearable device metagenomes	 **Dr. Ville Pimenoff**

Nov 13th 2024 *16:30h-17:30h Central European Time (CET)*	Multi-omics and eXplainable Artificial Intelligence (XAI) for predictive purposes: a focus on interpretability and explainability	 **Dr. Augusto Anguita**
Nov 20th 2024 *16:30h-17:30h Central European Time (CET)*	Patient similarity networks for the identification of molecular endotypes of obesity	 **Dr. Nikolaos Stratakis**

Detailed program:

SESSION 1: Sept 4th 2024 / 16:30h-18:00h Central European Time (CET).

Title: *"Introductory session: Introduction to multi-omics data integration challenges in exposome studies & Overview of approaches"*

By Dr. Léa Maitre

Affiliation:

Barcelona Institute for Global Health (ISGlobal)

Short bio:

Léa Maitre is an Assistant Research Professor and director of the Exposome Hub at ISGlobal, promoting innovation, collaboration and communication about this new field of research. Her participation in European projects on this topic included the scientific coordination in the HELIX project (2013-2018), task leadership in the ATHLETE project (2020-2024), task leader in the ENDOMIX project and WP co-leadership in the International Human Exposome Network project (2024-2026). Her main lines of research include: Exposome in pregnancy, childhood and adolescence; Metabolomics integration in epidemiological studies; Molecular (omics) response to external stressors (chemical, physical, psycho-social) in observational human studies; Exposome, neurodevelopment and mental health.

Talk outline:

This talk will introduce the main challenges behind multi-omics data integration in environmental epidemiology (e.g., different data modalities, technical noise, confounders, etc.) and will present an overview of the different type of analytical approaches we can use for multi-omics data integration that will be covered in detail during the course.

Precourse reading: Maitre, L. (2022). State-of-the-art methods for exposure-health studies: Results from the exposome data challenge event. Environment international, 168, 107422. https://doi.org/10.1016/j.envint.2022.107422

SESSION 1: Sept 4th 2024 / 16:30h-18:00h Central European Time (CET).

Title: *"Multi-modal data integration through dimensionality reduction: Motivation, concepts and theory"*

By Dr. Arthur Tenenhaus

Affiliation:

Université Paris-Saclay, CentraleSupelec, CNRS

Short bio:

Arthur Tenenhaus is a professor at CentraleSupelec, a top engineering school in France, and a member of the Laboratoire des Signaux et Systèmes. He also co-holds the APHP-CentraleSupelec-INRIA chair. His primary research interest concerns the development of statistical methods for the joint analysis of heterogeneous and complex data.

Talk outline:

General overview of dimensionality reduction (DMR) techniques and how they can be useful for multi-omics data integration. This talk will cover the basic concept of dimensionality reduction, latent/orthogonal components, PCA. The suitability of DMR for data integration. Which DMR methods are available for multimodal data integration (PCA, PLS, CCA, RRR, factor analysis) and what are the main differences between these methods? The talk will specifically focus on Regularized Generalized Canonical Correlation Analysis (RGCCA), an unified and versatile framework consolidating over six decades of multiblock component methods. This presentation introduces the RGCCA package, which implements this framework. Beyond its implementation, the RGCCA package offers graphical outputs and statistics to assess the robustness and significance of the analysis. The usefulness of the RGCCA package is illustrated on real datasets. The RGCCA package is freely available on the Comprehensive R Archive Network (CRAN) at http://www.r-project.org/ and on GitHub at https://github.com/rgcca-factory/RGCCA.

SESSION 2: Sept 25th 2024 / 16:30h-17:30h Central European Time (CET)

Title: *"Practical aspects of multi-modal data integration through dimensionality reduction (MOFA, PLS, RGCCA, etc)"*

By Inès Amine

Affiliation:
Inserm Grenoble, France

Short bio:
Ines Amine is a data scientist and PhD student in environmental epidemiology at Inserm Grenoble, France. She is an investigator of the EU-H2020 ATHLETE exposome project, in which she is studying the broad effects of the exposome on multiple health outcomes in childhood. Following a recent visiting stay at ISGlobal with Léa Maitre, she is now investigating the multi-omic processes underlying the effect of the exposome on child health.

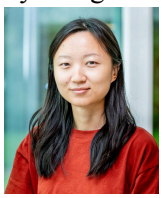
Talk outline:
This presentation, entitled "Multi-Omic Dimensionality Reduction Techniques", will present techniques for data integration of multi-block omic data through unsupervised (e.g. MOFA) and supervised (e.g. multi-block sPLS, RGCCA) dimensionality reduction models. These methods are of interest to identify molecular signatures using several interconnected omics layers, with potentially high-dimensionality (e.g. epigenomics). To illustrate this framework, we will examine two case studies involving mother-child pairs from the Human Early Life Exposome Project. The first study applied unsupervised multi-omics factor analysis to extract latent factors from both exposome and multiple molecular omics (epigenomics, transcriptomics, proteomics and metabolomics), facilitating mediation analysis to assess whether multi-omics mediated the effect of exposome on child BMI. The second study uses a supervised approach to identify immune signatures associated with a general health score in children, using epigenomic, proteomic and cell-type data.

Availability of material: All codes (in R) will be made available in a github repository.

SESSION 2: Sept 25th 2024 / 16:30h-17:30h Central European Time (CET)

Title: *"Case studies of multi-omics and exposome data integration through dimensionality reduction in HELIX cohorts"*

By Congron Wang



Affiliation:

Centre for Environmental Sciences,, Hasselt University, Belgium

Short bio:

Congrong Wang is a PhD candidate from the Centre of Environmental Sciences at Hasselt University in Belgium. She is a biostatistician and is currently using HELIX data to investigate the multi-omics signature of childhood telomere length. Together with her colleagues Dr. Rossella Alfano and Dr. Brigitte Reimann, she worked on a mediation analysis based on multi-omics integration to evaluate whether the effect of pre- and postnatal exposome on childhood anthropometrics is mediated by the changes in multi-omics.

Talk outline:

This presentation, entitled "Multi-Omic Dimensionality Reduction Techniques", will show techniques for data integration of multi-block omic data through unsupervised (e.g. MOFA) and supervised (e.g. multi-block sPLS, RGCCA) dimensionality reduction models. These methods are of interest to identify molecular signatures using several interconnected omics layers, with potentially high-dimensionality (e.g. epigenomics). To illustrate this framework, we will examine two case studies involving mother-child pairs from the Human Early Life Exposome Project. The first study applied unsupervised multi-omics factor analysis to extract latent factors from both exposome and multiple molecular omics (epigenomics, transcriptomics, proteomics and metabolomics), facilitating mediation analysis to assess whether multi-omics mediated the effect of exposome on child BMI. The second study uses a supervised approach to identify immune signatures associated with a general health score in children, using epigenomic, proteomic and cell-type data. **Availability of material:** All codes (in R) will be made available in a github repository.

SESSION 3: Oct 2nd 2024 / 16:30h-17:30h Central European Time (CET)

Title: *"Strategies for Multivariate Mediation Frameworks in Environmental Epidemiology"*

Vishal Midya

Affiliation:

Icahn School of Medicine at Mount Sinai, New York

Short bio:

Vishal Midya is a biostatistician and environmental epidemiologist at the Icahn School of Medicine at Mount Sinai. He is an assistant professor in the department of Environmental Medicine and Climate Science. His primary research focuses on understanding how prenatal and early-life exposure to environmental chemicals (especially metals) eventually leads to adverse neurodevelopmental outcomes later in life and whether childhood or adolescent gut microbiome can alter such detrimental effects. A major theme of his research involves incorporating exposure mixture models and interpretable machine learning tools to study aspects of precision environmental health.

Talk outline:

In this talk, I will discuss the strategies for multivariate mediation analyses with multiple chemical exposures, particularly aimed at studies in Environmental Epidemiology. I will delineate currently available tools to model simultaneous exposures to mixtures of chemicals and how to use them in standard mediation analyses framework with multiple mediators. I will discuss potential methods and explain what each strategy entails, how to interpret the results meaningfully, and the drawbacks of each tool. In the end, I will discuss what each method conveys in the context of chemical environmental health studies and the opportunities for future research involving machine learning techniques for precision environmental health. This talk will include multiple examples and illustrate efficient ways to present the results.

Availability of material:

All codes (in R), references, and illustrations will be available on the slides.

SESSION 4: Oct 9th 2024 / 16:30h-17:30h Central European Time (CET)

Title: *"Multi-Omic Analysis Framework For Precision Health: Omic mediation analysis in exposome research"*

By Jesse Goodrich

Affiliation:

University of Southern California Keck School of Medicine

Short bio:

Jesse Goodrich is an Assistant Professor in Environmental Health at the University of Southern California and the Associate Director of Data Science for the Center for Integrative and Translational Exposomics Research. Specializing in environmental epidemiology, data science, and physiology, his research focuses on uncovering the negative impacts of environmental chemicals on metabolic diseases and cancer. His work further aims to develop tools to facilitate the integration of environmental data with multiple omics layers to better understand the biology linking environmental factors with risk for disease.

Talk outline:

Precision Health aims to revolutionize disease prevention by leveraging information across multiple omic datasets. However, existing methods generally do not consider personalized environmental factors. Mediation analysis, which examines how intermediate variables link environmental factors with disease, can use multi-omics to inform precision health by identifying biomarkers, elucidating molecular mechanisms, and identifying susceptible individuals. This presentation, titled "Multi-Omic Analysis Framework For Precision Environmental Health: A Case Study on Maternal Mercury, Childhood Fatty Liver Disease, and Multiple Omic Layers", will introduce a novel precision environmental health framework that combines multi-omics integration approaches with mediation analysis techniques, including high dimensional mediation, mediation with latent factors, and integrated/quasi-mediation. To illustrate the practical application of this framework, we examine a case study involving 420 maternal-child pairs from the Human Early Life Exposome Project. This study focuses on the impacts of prenatal mercury exposure on children's risk of developing metabolic associated fatty liver disease, showcasing how integrating environmental data with multi-omic measurements can: uncover novel multi omic environmental disease associated biomarkers; elucidate biological mechanisms linking environment with disease; and identify of high-risk subgroups of individuals, characterized by unique combinations of exposure levels and omics profiles. **Availability of material:** All R code and resources are available through https://goodrich-lab.github.io/multiomics_book/.

SESSION 5: Oct 23th 2024 / 16:30h-17:30h Central European Time (CET)

Title: *"Networks visualisation for multi-omics integration"*

By Dr. Léa Maitre

Affiliation:

ISGlobal Barcelona Institute for Global Health

Short bio:

Léa Maitre is an Assistant Research Professor and director of the Exposome Hub at ISGlobal, promoting innovation, collaboration and communication about this new field of research. Her participation in European projects on this topic included the scientific coordination in the HELIX project (2013-2018), task leadership in the ATHLETE project (2020-2024), task leader in the ENDOMIX project and WP co-leadership in the International Human Exposome Network project (2024-2026). Her main lines of research include: Exposome in pregnancy, childhood and adolescence; Metabolomics integration in epidemiological studies; Molecular (omics) response to external stressors (chemical, physical, psycho-social) in observational human studies; Exposome, neurodevelopment and mental health.

Talk outline:

Her talk, entitled "Network visualisation and biological interpretation for exposure-omics associations" will go through the different challenges from implementation to data analysis and interpretation. I propose a session on multi-omic and exposure data visualisation based on the Nat Com. article public catalogue of results. I will go through the following points:

- Statistical methods to generate networks and analyse them
- How to create a network in Cytoscape
- Biological interpretation and triangulation of evidence in exposure-omics associations

Participants will have access to material to create their own network visualisation in Cytoscape and metabolic enrichment using the Exposome-explorer catalogue. Mainly allowing enrichment for sources of metabolites.

Precourse reading/ Lab materials for this workshop: Maitre, Léa, et al. Multi-omics signatures of the human early life exposome. *Nature Communications* 13.1 (2022): 7024. & Fabbri L, et al. Childhood exposure to non-persistent endocrine disrupting chemicals and multi-omic profiles: A panel study. *Environ Int*. 2023 Mar;173:107856. doi: 10.1016/j.envint.2023.107856

SESSION 6: Oct 30th 2024 / 16:30h-17:30h Central European Time (CET)

Title: *"Development of exposome mixture methods for high-dimensional omics data to investigate environmental risk factors of disease"*

By Dr. Chris Gennings

Affiliation:
Icahn School of Medicine at Mount Sinai

Short bio: Chris Gennings is professor and Director of the Biostatistics Division within the Department of Environmental Medicine and Climate Science at the Icahn School of Medicine at Mount Sinai in New York. Her research interests focus on design and analysis methodologies for studies of complex unintentional mixtures, with recent focus on the development and extensions of Weighted Quantile Sum (WQS) regression, development of a personalized nutrition index called My Nutrition Index (MNI) with a dietary tracking app that utilizes the MNI, and methods for risk assessment of environmental mixtures including a 'similar mixture approach' (SMACH), a strategy used in two recent EU Horizon 2020 research consortia, EDC-MixRisk and ENDpoiNTs.

And By Dr. Anna Young

Affiliation: Emory Rollins School of Public Health

Short bio: Anna Young is an associate research scientist in the Gangarosa Department of Environmental Health at Emory Rollins School of Public Health. With specialization in environmental epidemiology and data science, her research focuses on the reproductive and carcinogenic impacts of human exposures to complex mixtures of thousands of known and unknown chemicals (the exposome). Dr. Young previously received her MS and PhD degrees and postdoctoral training in the Department of Environmental Health at the Harvard T.H. Chan School of Public Health and her BA degrees in Computer Science and Environmental Studies from Yale.

Talk outline:

We will conclude with an integrated exposome–metabolome case study, in which we apply recent extensions of WQS to analyze untargeted data (generated by Emory's Walker Laboratory) using blood samples from the prospective CLUE cohort. We will demonstrate the power of this method to examine the cumulative impact of complex mixtures of untargeted chemical exposures on cancer risk, to prioritize 'bad actor' chemicals, and to determine potential intermediate metabolic pathways underlying the associations.

Availability of material: All examples and references will be available in the slides.

SESSION 7: <u>Nov 6th 2024 / 16:30h-17:30h Central European Time (CET)</u>

<u>Title:</u> *"Comprehensive biotic exposome estimation from wearable device metagenomes"*

By Dr. Ville Pimenoff

<u>Affiliation:</u>

University of Oulu, Finland

Short bio: Ville Pimenoff is an Adjunct Professor in Evolutionary Medicine at the University of Oulu, Finland. He is currently leading a '*Monitoring exposures during pregnancy with a wearable device´* exposome cohort study to estimate individual's biological and chemical exposures for the Horizon 2020 Human Exposome Assessment Platform (HEAP) -consortium. He is also a Co-PI for the Safe Biomonitoring by Bioict Integrated Wearables -project in Oulu University to apply machine learning methods for environmental exposure predictions and a co-investigator for the International Human Exposome Network consortium. His main lines of research include: Longitudinal population cohort epidemiology, population genetics of human microbiome, biotic and abiotic exposome in pregnancy, wearable device monitoring, metagenomics and phylodynamics of pathogens causing diseases.

Talk outline: In one aspect human health can be estimated from an individual's genome, microbiome, exposome and their complex interactions. This talk will introduce the concept of comprehensive personal microbial exposome estimation from wearable device metagenomes. Latest research suggest that we are exposed to a vast and dynamic array of biotic and abiotic environmental contaminants in our daily lives, far beyond what was previously quantified and understood. However, to identify an individual's seasonal and location -dependent patterns of biological exposure has several challenges from wearable device sampling to metagenome data curation and analysis for interpretable and reproducible results.

This lecture will present an overview of the technical and analytical approaches to systematically analyze individual's personalized biological exposome, which will eventually enable a comprehensive insight into the biotic environmental exposures likely affecting human health.

Precourse reading: Jiang, C. et al. Dynamic Human Environmental Exposome Revealed by Longitudinal Personal Monitoring. Cell 175, 277–291.e31 (2018).
<u>Availability of material:</u> This talk will include a pilot case study on monitoring exposures during pregnancy with a wearable device. R code employed for the generation of the results will be available through https://github.com/PimenoffV.

SESSION 8: Nov 13th 2024 / 16:30h-17:30h Central European Time (CET)

Title: *"Multi-omics and eXplainable Artificial Intelligence (XAI) for predictive purposes: a focus on interpretability and explainability"*

By Dr Augusto Anguita

Affiliation:

ISGlobal Barcelona Institute for Global Health

Short bio:

Augusto Anguita-Ruiz is a **biological data scientist** specialised in the analysis of complex epidemiological datasets such as those composed of clinical, omics, biochemical, and environmental data. At ISGlobal he is an investigator of the **EU-H2020 ATHLETE exposome project**. His main technical skills include a strong statistical, programming and data visualisation background, with special emphasis on the use of machine learning models for **multi-omics data analysis**.

Talk outline:

His talk, entitled "Multi-omics and eXplainable Artificial Intelligence (XAI) for predictive purposes: a focus on interpretability and explainability" will cover machine learning strategies for leveraging multi-omics data with predictive purposes in epidemiological studies. For that, a real showcase of clinical outcome prediction from multimodal molecular data in a prospective cohort will be presented. This talk will put special emphasis on the interpretability of generated predictive models, explaining how to use post-hoc explainers for the full exploitation of machine learning algorithms and multi-omic information. The talk will be structured as follows:

- Introduction to the real data showcase "Insulin Resistance Prediction in Paediatric Populations"
- Tips for multi-omics data processing and ML model selection and construction
- Use of SHAP values (SHapley Additive exPlanations) for the extraction of global and local explanations
- Use of SHAP values for the study of interactions

The topic will be adapted to a broad audience (including researchers in environmental epidemiology, molecular epidemiology, clinical field, both with more technical or clinical profiles) who are interested in building predictive models for clinical outcomes based on heterogeneous omic information. **Availability of material:** This talk will include a case study on childhood obesity. All R codes and resources employed for the generation of all presented results during the talk will be made available through the course Github account.

SESSION 9: Nov 20th 2024 /16:30h-17:30h Central European Time (CET)

Title: *"Patient similarity networks for the identification of molecular endotypes of obesity"*

 By Dr. Nikolaos Stratakis

Affiliation:

ISGlobal Barcelona Institute for Global Health

Short bio:

Nikos Stratakis is an Environmental Epidemiologist and Postdoctoral Fellow at the Environmental and Health over the Lifecourse programme at ISGlobal. His research lies at the crossroad of Epidemiology and Biology by combining exposure modelling methods with molecular (omics) data to disentangle the impact of environmental exposures on metabolic dysfunction and characterize underlying pathophysiological mechanisms. He has been awarded the prestigious Marie Skłodowska-Curie fellowship to apply exposomics approaches into the study of childhood liver disease in order to identify early-life environmental determinants.

Talk outline:

His talk will present a real application on the construction of patient similarity networks using extensive multi-omics profiling. Patient similarity networks, in which population samples are clustered or classified based on their similarities in various features, are considered an emerging parading for precision medicine. In this talk, Nikos will showcase how this paradigm can be combined with machine-learning tools (e.g., SHAP values) to derive highly-interpretable and clinically relevant molecular endotypes by using childhood obesity as example outcome. He will also touch upon the novel concept of "precision environmental health" and demonstrate how these similarity networks can be used in combination with exposome data to identify modifiable environmental determinants of disease.

Availability of material: All R codes used for this presentation will be made available through Github.