Course information

Organizers: Léa Maitre and Augusto Anguita

Format: Online

Number of sessions: 12

Time: 30th September 2026- 16th December 2026 (16:00-17:30 CEST/CET, with some exceptions stated in the program)

Course overview:

We are excited to announce the second edition of the online workshop series, "Multi-Omics Data Integration in Human Exposome Studies," running from September to December 2026. While the first edition focused on foundational, introductory material, this second edition goes a step further into advanced analytical methodology and the open challenges of the field. To review the fundamentals or catch up on past sessions, you can access the recordings, slides, and code from the First Edition Workshop Materials.

The integration of multiple omics layers with environmental data represents a cornerstone for advancing precision environmental health and epidemiology. However, combining these complex datasets introduces significant computational and methodological challenges. To address these hurdles, this workshop introduces participants to state-of-the-art analytical frameworks and statistical workflows recently developed in the exposomics field. We begin with the foundations of multi-block dimensionality reduction, the workhorse for distilling high-dimensional, collinear exposomics layers into low-dimensional, biologically meaningful signatures. From there, our speakers will present methodological solutions for some of the field's most pressing challenges: relating exposures to biology through mediation and causal frameworks; disentangling the "mixture problem" of simultaneous environmental stressors; reconstructing the connectivity between molecular and exposure layers through network-based approaches; and incorporating emerging omic layers, such as the microbiome, that bridge the environment and the host. Participants will also explore specialised strategies to overcome recurrent data limitations, including missing and incomplete data, imbalanced outcomes, and longitudinal analyses that capture dynamic exposome-omic stability and time lags. Finally, the workshop turns to the AI frontier: the emerging role of Large Language Models (LLMs) and generative AI tools in multi-omics workflows, and the concept of XAI (eXplainable Artificial Intelligence), a growing trend that promotes the use of trustworthy, interpretable algorithms and post-hoc tools that allow researchers to understand the complex relationships between features and their contribution to a prediction. Together, these approaches transform overwhelming computational

complexity into a practical, actionable research toolkit for modern epidemiological and precision environmental health challenges.

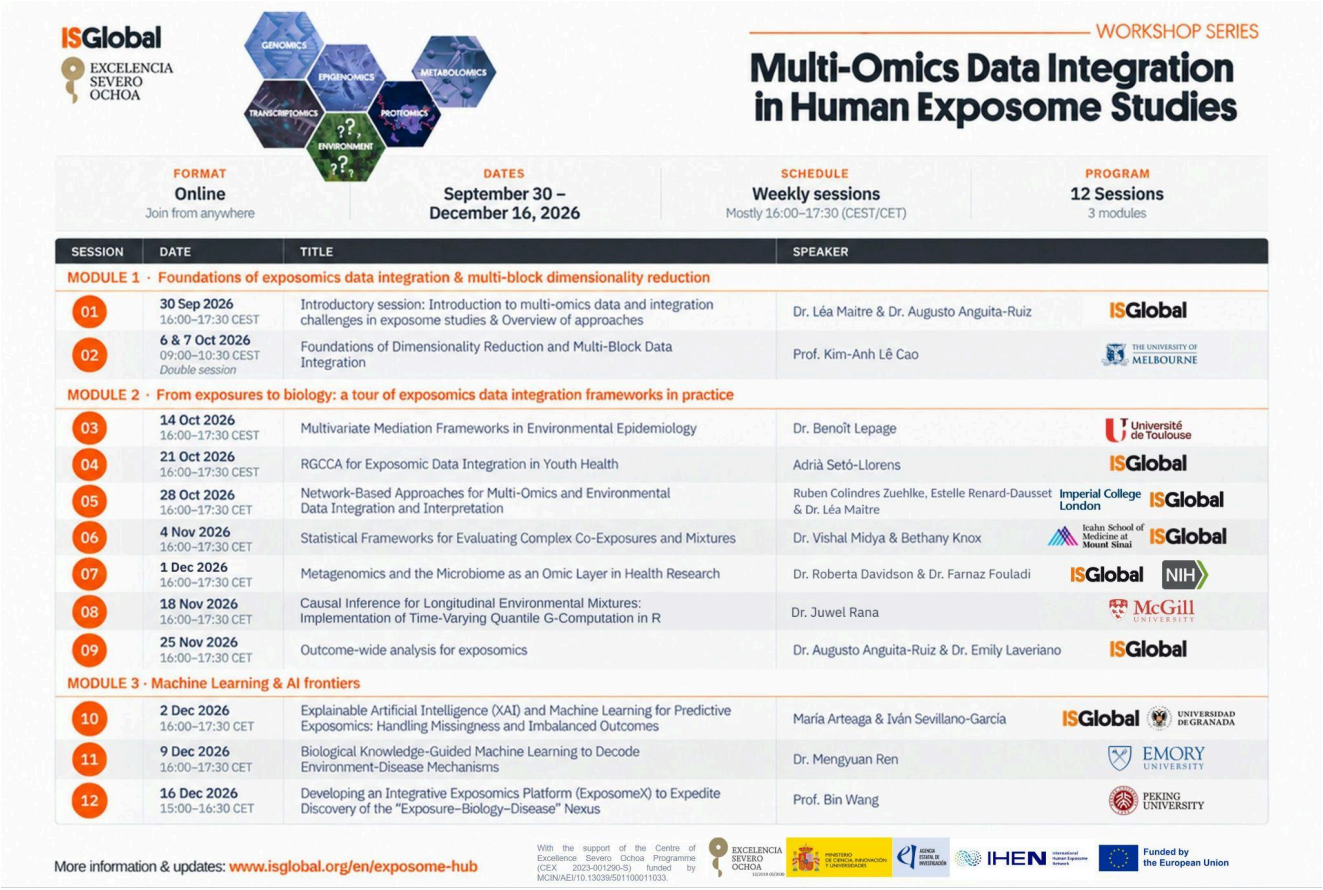
Code and materials:

Designed as a highly practical, hands-on experience, these seminars will provide participants not only with theoretical concepts of exposomic analyses but with code and resources for their implementation, empowering participants to apply the presented techniques in their own research endeavours. We will grant direct and free access to all session codes and materials during and after the workshop on our GitHub and YouTube channel.

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.

Summarized Program



Modules

Module 1: Foundations of exposomics data integration & multi-block dimensionality reduction

Session 1: Introductory session: Introduction to multi-omics data and integration challenges in exposome studies & Overview of approaches

Date: 30 September 2026 (16:00-17:30 CEST)

Speakers: Dr. Léa Maitre (ISGlobal) & Dr. Augusto Anguita-Ruiz (ISGlobal)

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 leadership 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.

Augusto Anguita-Ruiz is a biological data scientist and Junior Research Leader at ISGlobal, specializing in the analysis of complex epidemiological datasets such as those composed of clinical, omics, biochemical, and environmental data. 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: This talk will introduce a comprehensive taxonomy of integrative analytical approaches currently employed in exposomics research. Second, this taxonomy will serve as an organizing framework to examine recurrent methodological limitations across the field, with particular emphasis on five challenges: (1) interpretability and explainability, (2) mixture effects, (3) longitudinal and life-course modeling, (4) missing, unbalanced, and incomplete data, and (5) mediation and causal analysis. Finally, we highlight emerging and promising methodological developments that may help overcome these limitations, ultimately supporting the design of more reliable, transparent, and biologically informative exposome analyses.

Session 2 Foundations of Dimensionality Reduction and Multi-Block Data Integration

Date: Double Session 06 October (9:00-10:30 CEST) & 07 October 2026 (9:00-10:30 CEST)

Speaker: Prof. Kim-Anh Le Cao (University of Melbourne)

Kim-Anh Lê Cao is Professor of Statistical Genomics at the University of Melbourne and Director of Melbourne Integrative Genomics. She has secured 3 consecutive NHMRC fellowships since 2014, and numerous awards for her contributions to statistics applied to molecular biology, including the Moran Medal from the Australian Academy of Science. Her research focuses on 'omics data integration, and she leads the development of the popular R toolkit mixOmics.

Talk outline: Dimension reduction methods provide a powerful framework for exploring high-dimensional data and integrating multiple data sets measured on the same individuals. I will introduce the foundations of Principal Components Analysis and Partial Least Squares, including sparse and multiblock extensions implemented in the mixOmics package. Using data from the HELIX early-life exposome study, I will illustrate how these methods can identify major sources of variation, relate environmental exposures to molecular profiles, and reveal shared patterns across multiple omics layers. The seminar will provide the conceptual foundation for the practical session that follows.

Module 2: From exposures to biology: a tour of exposomics data integration frameworks in practice

Session 3: *Multivariate Mediation Frameworks in Environmental Epidemiology*

Date: 14 October 2026 (16:00-17:30 CEST)

Speaker: Dr. Benoît Lepage (University of Toulouse)

Benoît Lepage is a Professor of Public Health and Biostatistics at the University of Toulouse III – Paul Sabatier and a researcher at the Centre for Epidemiology and Research in Population Health (CERPOP, Inserm/Université de Toulouse). His research focuses on advanced biostatistical methods for life-course epidemiology, mediation analysis, and causal inference. He specializes in evaluating complex relationships between biological, environmental, and social determinants of health across the lifespan.

Talk outline: Understanding how complex environmental exposures impact human health requires robust analytical methods that can disentangle direct effects from biological pathways. This session provides an overview of multivariate mediation frameworks applied to environmental epidemiology and exposome studies. The talk will cover fundamental principles of causal inference, key methodologies for handling multiple and high-dimensional biological mediators (such as omics layers), and practical biostatistical strategies to decompose total exposure effects into direct and indirect mechanisms across the life course.

Session 4: RGCCA for Exposomic Data Integration in Youth Health

Date: 21 October 2026 (16:00-17:30 CEST)

Speaker: Adrià Setó-Llorens (ISGlobal)

Adrià Setó-Llorens is a doctoral researcher in Bioinformatics at ISGlobal. He has a background in Biology, Bioinformatics, and Biostatistics. His research focuses on understanding how early-life environmental exposures influence cardiometabolic health and the development of obesity-related insulin resistance during childhood and puberty. He integrates exposome, multi-omics, and explainable AI approaches to identify underlying molecular mechanisms and improve early risk prediction and prevention.

Talk outline: This talk will show how Regularized Generalized Canonical Correlation Analysis (RGCCA) integrates high-dimensional exposomics data to uncover low-dimensional, biologically meaningful signatures associated with health outcomes. To demonstrate its practical application, we will present a real-world use case using data from the HELIX multicenter European cohort study. Specifically, we will show how RGCCA integrates childhood plasma proteins, serum metabolites, and urine metabolites with blood pressure measured across childhood and adolescence. Finally, we will demonstrate how to couple this framework with exposome-wide analysis to pinpoint the prenatal environmental drivers behind the identified molecular signature

Session 5: Network-Based Approaches for Multi-Omics and Environmental Data Integration and Interpretation

Date: 28 October 2026 (16:00-17:30 CET)

Speaker: Ruben Colindres Zuehlke (Imperial College London), Estelle Renard-Dausset (ISGlobal) & Dr. Léa Maitre (ISGlobal)

Ruben Colindres is a PhD student and research assistant with a background in biochemistry and biostatistics, working with Prof. Marc Chadeau-Hyam at Imperial College London to develop and apply novel statistical methods for high-dimensional exposome and multi-omics data, including stability selection, network inference, and mediation analysis.

Estelle Renard-Dausset is a biostatistician at ISGlobal specializing in high-dimensional data analysis and computational approaches to exposome research. Together, they apply advanced statistical and network-based methods to investigate the relationships between environmental exposures, molecular signatures, and human health.

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 leadership 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: Using untargeted exposomics and metabolomics datasets generated within the ATHLETE project, we will present a case study mapping the chemical universe of adolescents in the HELIX cohort. We will demonstrate how Gaussian graphical models and community detection can be used to identify networks of interconnected chemical exposures and endogenous metabolites, revealing patterns of co-exposure and potential biological relationships.

Session 6: Statistical Frameworks for Evaluating Complex Co-Exposures and Mixtures

Date: 04 November 2026 (16:00-17:30 CET)

Speaker: Dr. Vishal Midya (Icahn School of Medicine, Mount Sinai) & Bethany Knox (ISGlobal)

Vishal Midya is an assistant professor in the Department of Environmental Medicine and Public Health at Icahn School of Medicine, Mount Sinai in New York. His research interests include developing and applying interpretable machine-learning models to understand how early-life environmental exposures and their dynamic interactions affect neurodevelopmental and metabolic outcomes later in life. He is the Director of Public Health Data Analytics of the MPH Program at Sinai. He is also a data science consultant at LinusBio, a startup from Mount Sinai Health System.

Bethany Knox is a predoctoral researcher in Environmental Epidemiology at ISGlobal, studying prenatal exposure to endocrine-disrupting chemicals and micro/nanoplastics with a focus on fetal growth and development, placental function, and Bayesian mixture models. She began her career as a pediatric nurse, later earning a BS and Master's in Public Health, which sparked her interest in research and prevention at the population level. Her expertise includes advanced statistical modeling, multi-cohort analysis, and translating findings into policy within large European Horizon 2020 projects. She aims to advance environmental epidemiology by incorporating clinical outcomes and addressing female health needs across the lifespan.

Talk outline:

This talk is divided into two main sections. First, it will introduce the methodological challenges of analyzing complex environmental mixtures in an exposomic landscape, discussing the necessary shift from traditional single-pollutant models to advanced mixture-based statistical frameworks designed to capture realistic, simultaneous high-dimensional exposure scenarios. The new method, penalized LASSO-type multigroup Bayesian Weighted Quantile Sum Regression (LBWQSR) will be presented as an extension of a Bayesian version of the Weighted Quantile Sum regression model that (1) considers the overall data-driven joint mixture effect and (2) allows concurrent inclusion of multiple groups of chemical compounds within the same mixture model while applying the elastic net-type penalties to prevent overfitting. Second, it will present a practical use case applying this method to an exposomic dataset from the Barcelona Life Study Cohort (BiSC). The use case will examine how prenatal exposure to a mixture of nonpersistent endocrine-disrupting chemicals (EDCs), including phthalates, phenols, parabens, and newer alternatives like DINCH, impacts angiogenic molecular biomarkers, fetoplacental hemodynamics, and fetal growth.

Session 7: Metagenomics and the Microbiome as an Omic Layer in Exposome Research

Date: 1 December 2026 (16:00-17:30 CET)

Speaker: Dr. Roberta Davidson (ISGlobal) & Dr. Farnaz Fouladi (National Institute of Environmental Health Sciences)

Roberta Davidson is a postdoctoral fellow at ISGlobal in the Environment and Health over the Lifecourse program, where she explores interactions between the human exposome and the microbiome. She obtained her PhD in genomics and bioinformatics from the University of Adelaide in 2024. Currently, actively contributes to major European collaborative initiatives, including the ATHLETE and ENDOMIX projects, bridging the fields of advanced bioinformatics and public health.

Farnaz's research is focused on analyzing microbiome data generated from high-throughput next-generation sequencing to study how the microbiome affects human health. She is interested in the application of statistical models and new algorithms in metagenomics. Farnaz earned her PharmD at Shahid Beheshti University in Iran and her PhD in Pharmaceutical Sciences at North Dakota State University, where she studied the relationship between the gut microbiome, bile acids, and gut peptides and their effects on weight outcomes following bariatric surgery.

Talk outline: This session will introduce core concepts regarding the different human microbiomes (gut, oral, vaginal, skin, etc.) and compare and contrast two main data types for microbiome research: metabarcoding (16S) and metagenomics. We will cover key factors

that are important for study design when sampling microbiomes and integrating microbiome data into human exposome studies.

Session 8: Causal Inference for Longitudinal Environmental Mixtures: Implementation of Time-Varying Quantile G-Computation in R

Date: 18 November 2026 (16:00-17:30 CET)

Speaker: Dr. Juwel Rana (McGill University)

Juwel Rana is Director of Research and Innovation at the South Asian Institute of Social Transformation (SAIST) in Bangladesh. He earned his PhD in Epidemiology from McGill University, Canada. His research investigates how complex environmental, nutritional, and social exposures jointly shape population health and health inequities across the life course within an exposome framework. Juwel leads multidisciplinary research initiatives, including the Bangladesh Longitudinal Child-Adolescent Development, Education, and Environment Study (BLADES; 2,500 households), a community-based exposome study involving children, adolescents, and their family members. He extended standard Quantile G-computation to Time-varying Quantile G-computation and implemented it in the R package *tvcQGComp*. He also adapted Weighted Quantile Sum (WQS) regression for mediation analysis to investigate pathways linking environmental mixtures to health outcomes. His empirical research draws on large-scale population cohorts across North America and Asia, including CanCHEC and MIREC in Canada, GUSTO in Singapore, and BEST and BLADES in Bangladesh.

Talk outline: This talk has two main parts. First, it will introduce the methodological challenges of estimating the health effects of longitudinal environmental mixtures, including correlated exposures measured repeatedly over time, time-varying confounding, and feedback between prior exposures and subsequent confounders. These challenges motivate moving beyond cross-sectional mixture models toward longitudinal causal inference frameworks. This part will introduce the causal structure, target estimand, identification assumptions, and methodological framework underlying Time-Varying Quantile G-computation.

Second, the talk will present a case study using Canadian environmental epidemiology data to estimate the joint causal effect of five time-varying $PM_{2.5}$ components, black carbon, sulfate, nitrate, ammonium, and organic matter, on nonaccidental mortality. Using the R package tvcQGComp, the hands-on implementation will demonstrate how to: (1) prepare longitudinal data and specify the required models; (2) estimate the cumulative joint effect of simultaneous quantile increases in multiple exposures across repeated time points; (3) adjust for time-varying confounders affected by prior exposure; (4) evaluate effect-measure modification by baseline characteristics; and (5) conduct model diagnostics by comparing

simulated natural-course trajectories with observed data and assessing positivity assumptions over time. Participants will also learn how to interpret and communicate the resulting joint-effect estimates and patterns of differential vulnerability across population groups.

The session will conclude with a brief discussion of future directions for causal inference in longitudinal environmental and multi-omics research, followed by an interactive Q&A.

Session 9: Outcome-wide analysis for exposomics

Date: 25 November 2026 (16:00-17:30 CET)

Speaker: Dr. Augusto Anguita-Ruiz (ISGlobal) & Dr. Emily Laveriano (ISGlobal)

Augusto Anguita-Ruiz is a biological data scientist and Junior Research Leader at ISGlobal, specializing in the analysis of complex epidemiological datasets such as those composed of clinical, omics, biochemical, and environmental data. 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.

Emily Laveriano is a postdoctoral researcher at ISGlobal and Adjunct Lecturer at the University of Barcelona, whose research focuses on the interplay between environmental and dietary exposures, metabolomics, and cardiometabolic health across the life course. She completed her PhD at the University of Barcelona on dietary (poly)phenols and cardiometabolic health in adolescents, developing omics-based approaches to assess dietary phenolic compounds. Her work includes collaborations with international institutions such as Johns Hopkins University and Imperial College London, as well as European research projects on phytochemicals and cardiometabolic health.

Talk outline:

In this talk, we will present the novel analytical framework of outcome-wide analysis and explore how it can be leveraged for exposomics data. We will survey state-of-the-art methodologies that allow researchers to simultaneously estimate the associations between environmental exposures and multiple health outcomes. Furthermore, we will discuss how omics can be modeled as outcomes, offering a powerful new strategy for integrating molecular and environmental data. We will introduce a cutting-edge technique called the mixed-responses sparse reduced-rank regression (msRRR) model, specifically designed to conduct outcome-wide exposomics analysis while handling high dimensionality, missingness, and mixed-type variables. Finally, we will demonstrate the practical application of this framework through a comprehensive use case. This hands-on example will illustrate

how to implement this methodology and interpret the resulting exposome-wide and multi-outcome associations.

Module 3: Machine Learning & AI frontiers

Session 10: Explainable Artificial Intelligence (XAI) and Machine Learning for Predictive Exposomics: Handling Missingness and Imbalanced Outcomes

Date: 02 December 2026 (16:00-17:30 CET)

Speaker: María Arteaga (ISGlobal & University of Granada) & Iván Sevillano-García (University of Granada)

María Arteaga Jover is a bioinformatician at the Barcelona Institute for Global Health (ISGlobal) and a doctoral researcher in Bioinformatics at the University of Granada. She has a background in Biochemistry, Bioinformatics, and Biostatistics. Her research focuses on developing and applying machine learning and explainable artificial intelligence methods to analyze clinical, exposomic, and multi-omics data, with a particular interest in obesity, insulin resistance, and metabolic health.

Iván Sevillano-García is a computer scientist and doctoral researcher at the DaSCI Institute (Data Science and Computational Intelligence) at the University of Granada. His research focuses on artificial intelligence, machine learning, and explainable AI, with a particular emphasis on developing interpretable and robust predictive models for complex biomedical and epidemiological datasets.

Talk outline: This session explores the application of machine learning and eXplainable Artificial Intelligence (XAI) in exposomics, bridging predictive modeling with model interpretability. Speakers will first demonstrate how to leverage environmental and multi-omics data to predict clinical outcomes across prospective cohorts using post-hoc explainers and interpretable algorithms. Building from a broader perspective, the session will highlight why interpretability in biomedical settings cannot be decoupled from data quality and performance. We will compare intrinsically interpretable models (such as rule-based systems) with post-hoc methods, examining how real-world data limitations, such as imbalanced outcomes and missing information, impact both model predictions and generated explanations. Finally, we will present existing approaches and multi-objective strategies that jointly optimize predictive performance, interpretability, and data imbalance in complex exposomic workflows.

Session 11: Biological Knowledge-Guided Machine Learning to Decode Environment-Disease Mechanisms

Date: 9 December 2026 (16:00-17:30 CET)

Speaker: Dr. Mengyuan Ren (Emory University)

Dr. Mengyuan Ren is a postdoctoral fellow in the Gangarosa Department of Environmental Health, Rollins School of Public Health, Emory University. He earned his PhD from Peking University, where he was trained in environmental epidemiology and environmental health. His research examines how environmental exposures shape maternal and child health through multi-omics approaches spanning the exposome, metabolome, and microbiome, with a focus on identifying disease-specific biomarkers and building interpretable machine learning prediction models. His recent work integrates biological knowledge into epidemiological studies within a machine learning framework.

Talk outline: Environmental exposures are increasingly implicated in adverse health outcomes, yet efforts to predict disease from these environmental contributors remain limited. Co-exposure patterns, mixture interactions, and underlying biological mechanisms further complicate the task. This talk presents a machine learning framework that combines environmental mixture exposures with a biological knowledge graph-based network (BKGN), which encodes how pollutants connect to disease outcomes through biological targets. Using early pregnancy loss as a case study, the talk shows how integrating this knowledge network improves predictive performance with considerable robustness, while also yielding biologically interpretable insights into how exposures may contribute to pregnancy loss. It closes by reflecting on how knowledge-guided models offer a scalable and cost-effective paradigm for precision environmental health, and on the potential of integrating exposomics, multi-omics, and AI-driven approaches going forward.

Session 12: Developing an Integrative Exposomics Platform (ExposomeX) to Expedite Discovery of the “Exposure–Biology–Disease” Nexus

Date: 16 December 2026 (15:00-16:30 CET)

Speaker: Prof. Bin Wang (Peking University)

Bin Wang is a tenured Associate Professor and Vice Dean at the Institute of Reproductive and Child Health, Peking University. His research primarily focuses on exposomics and AI-driven environmental health risk assessment. He has published over 70 papers as first or corresponding author in internationally renowned journals such as Environmental Health Perspectives, Environmental Science & Technology, The Lancet Regional Health-Western

Pacific, and The Innovation (H-index = 58, with over 6,500 citations). He developed ExposomeX, an integrative exposomics platform that accelerates research on the "Exposure–Biology–Disease" nexus. He serves as an Associate Editor for the journal Environmental Science & Technology, where he oversees topics related to machine learning and public health. He is also a pioneer in education, offering the course "Exposomics" to both undergraduate and graduate students, and teaching "Environment & Health" in the prestigious international master's program in global health and public health.

Talk outline: Human health is shaped by complex, lifelong environmental exposures, yet current risk assessment remains fragmented across exposure measurement, biological interpretation, disease relevance, and population burden estimation. To address this gap, we developed ExposomeX, an integrative exposomics platform designed to accelerate discovery of the exposure-biology-disease nexus and support mechanism-informed human health risk assessment. In this presentation, I will show how ExposomeX enables scalable and interpretable environmental health research through representative applications. First, knowledge graph-driven machine learning was used to integrate environmental mixture exposures with biological networks to predict early pregnancy loss, demonstrating the value of mechanistic pathways for risk prediction in small-sample epidemiological settings. Second, global air pollution-related lung cancer burden was assessed using benzo[a]pyrene as a mechanistic marker, revealing delayed and unequal risks across regions and birth cohorts. Third, ExposomeX-enabled frameworks were extended to global soil toxic metals and polycyclic aromatic hydrocarbon mixtures, integrating contamination intensity, pathway diversity, population exposure, and disease burden to generate health-oriented global risk atlases. Together, these studies show how ExposomeX advances environmental health research from exposure cataloging toward biologically informed, disease-specific, and policy-relevant risk prioritization.