EULAC PERMED TECHNICAL WORKSHOP

Innovative methodologies for data use and management in PerMed Research

Speaker Biographies

Jorge Samayoa – Universidad Galileo (Guatemala)

Jorge A. Samayoa holds a B.S. in electronics and computer science, a M.S. in operations research, a M.S. in applied mathematics, and a Ph.D. in Industrial Engineering. In 2003, he started teaching undergraduate courses of mathematics and statistics at the Engineering School of Galileo University. He has taught in several universities in Guatemala and USA. In 2007 he was awarded the prestigious Fulbright-LASPAU scholarship to pursue his master's degree at Texas A&M University and later his Ph.D. at Purdue University, where he developed the Calculus for Decision Systems (a generalized systems theory). Currently he is the director of the Operations Research Institute at Galileo University, where he and his team promote research and applications of Data Science to a wide variety of areas.

Adriano Barbosa da Silva – Queen Mary University (UK/Brazil)

Adriano Barbosa is an UKRI (MRC) Rutherford Fellow at the Health Data Research UK London site (at QMUL) with expertise in computational biology, data curation and integration of multi-dimensional translational bioinformatics datasets (including OMICS, imaging and clinical data). He is using machine learning models to stratify patients suffering of cardiovascular diseases using UK Biobank and other EU-wide health datasets.

Preng Biba – Universidad Galileo (Guatemala)

Preng Biba holds a B.S. in Mechatronics engineering, a M.S. in Operation Research, and a M.S. in Computer Science and Electrical Engineering. He has been instructor and teaching assistant of statistics, optimization and data science in the school of engineering at Galileo University. He has several publications in computer networks and data science. Currently he is the academic coordinator of the Operations Research Institute at Galileo University, overseeing the distance learning programs in Data Science and Operations Research. He has work as consultant in business intelligence and data science. His current research interests include engineering education, mathematical programing and optimization models implementation.

Josep Maria Haro – Parc Sanitairi Saint Joan de Deu (Spain)

Josep Maria Haro, psychiatrist and Ph.D. in Public Health, is the Research and Innovation Director of Saint John of God Health Park in Barcelona, Spain and associate professor of medicine at the University of Barcelona. During the past twenty-five years he has worked both in clinical medicine and in public

health research and has published more than five hundred scientific papers. He has been included in the list of Clarivate Highly Cited Researchers in 2017 and 2018.

His areas of investigation have been epidemiology of mental disorders and the analysis of treatment outcomes. His last studies focus on the determinants of healthy ageing, analyzing both the impact of physical and mental co-morbidity and the effect of societal and environmental aspects. Dr. Haro is principal investigator of one of the groups of the CIBERSAM network. In 2011 he received the award of best researcher from the Spanish Society of Biological Psychiatry. In 2018 he received the award of professional excellence of the Barcelona Medical Association. He is currently the European coordinator of the EU funded project ATHLOS and SYNCHROS and was coordinator of the Roadmap for mental health and wellbeing research in Europe (ROAMER).

Fabio Porto - Laboratório Nacional de Computação Científica (Brazil)

Fabio Porto holds a BSc degree in Mathematics-Informatics from Rio de Janeiro State University, and a MSc and DSc in Informatics from PUC-Rio, in 1996 and 2001, respectively. During his PhD he spent 2 years at INRIA, in France, as part of a sandwich program. He was a Post-doc at Ecole Polytechnique Fédérale de Lausanne (EPFL), in Switzerland, from 2004 to 2006, and a Senior Researcher at EPFL, from 2007 to 2009. Since 2009, he has been a technologist at the National Laboratory of Scientific Computing (LNCC), where he founded the Data Extreme Lab (DEXL), focusing on research and development activities associated to Big Data Management and Machine Learning. Dr. Porto has published more than 90 papers, including conferences and journals. He was the General Chair of the 44th International Conference on Very Large Data Bases, in Rio de Janeiro, and of the Brazilian Symposium on Databases (SBBD), in 2015. He serves at the SBBD steering committee and is a reviewer of journals and top conferences in his area of expertise. Dr Porto is a member of the Association for Computing Machinery (ACM) and the Brazilian Computing Society (SBC).

Rosa Figueroa – Universidad de Concepcion / CENS (Chile)

Rosa L. Figueroa received the B.Eng. Degree from the University of Concepción in 2004, and her Ph.D. degree in Electrical Engineering from the same institution, in 2012. Her Ph.D. Thesis explored different methods to obtain useful information from free text. She is currently a faculty member and researcher in the Biomedical Engineering degree part of the Electrical Engineering Department, University of Concepcion and a Technical Board Member in the National Center on Health Information Systems. She has scientific publications in journals and conference proceedings. Her research interest is within the medical informatics area mainly machine learning and text mining. She is currently working in research projects related to secondary use of medical data and text classification.

Helder Nakaya - Universidad de Sao Paolo (Brazil)

Helder Nakaya is an associate professor at the University of Sao Paulo, Brazil, in the Department of Clinical Analyses and Toxicology, School of Pharmaceutical Sciences. He has a PhD in Molecular Biology with extensive training in Bioinformatics. He is an expert in Systems Vaccinology, an interdisciplinary field that combines systems-wide measurements, networks, and predictive modeling in the context of vaccines and infectious disease. Dr. Nakaya has developed systems biology approaches to understand and predict the mechanisms of vaccine induced-immunity for several vaccines. His lab is focused on investigating the basis of infectious diseases using computational systems biology. Additionally, Dr. Nakaya is an adjunct professor at Emory University School of Medicine in the Department of Pathology.

Adrian Turjanski – Universidad de Buenos Aires (Argentina)

Adrián Turjanski obtained his Master degree in Chemistry from the University of Buenos Aires in Argentina in 1999, and then his PhD in Biophysics in 2003. He conducted postdoctoral studies in the area of molecular modeling from 2003 to 2005 in the Department of Physiology and Molecular Biology, School of Sciences, University of Buenos Aires. He then conducted postdoctoral work in Bioinformatics as a 2005 Pew Latin American Fellow in the National Institute of Dental and Craniofacial Research at the National Institutes of Health. Bethesda, MD,USA. In 2008 he returned to Argentina where he is in charge of the Structural Bioinformatics Lab, in the school of sciences at the University of Buenos Aires. He is part of the research staff of the National Research and Technology Council (CONICET) as Independent Investigator. He has also been Full Professor of Bioinformatics at the University of Buenos Aires since 2008. In 2013 he became director of the Argentinian Bioinformatics Platform and in 2015, the Director of the Center for Interdisciplinary Sciences. His research is focused in the development and application of Bioinformatics tools for genomics, going from the genome to structural characterizations of proteins. In 2015 he co-founded BITGENIA, a precision medicine company. He is also one of the leaders of the Argentinian precision Medicine Initiative.

Elmer Fernandez – Universidad de Cordoba (Argentina)

Dr. Elmer Fernandez is a Bioengineer (National University of Entre Rios - Argentina). He got his a PhD in Computer Science and Artificial Intelligence at University of Santiago de Compostela (Spain) and he is currently leading the Bioscience Data Mining Group at CIDIE - UCC - CONICET in Córdoba, Argentina. He has more than ten years of experience in data science problems. His main interest are focused on multi-omics data integration and the development of tools and methods to dig into high-throughput molecular technologies data with the main aim of improving the characterization, identification and association of patients and their appropriate therapy.

Anton Ussi – EATRIS (EU – Netherlands)

Anton Ussi is the Operations & Finance Director for EATRIS (The European Infrastructure for Translational Medicine). He has a background in engineering, and is a specialist in the establishment and execution of strategic public-private and public- public collaborations based on the deployment of a high value translational research infrastructure for medicine. Anton has been co-responsible for the development of several ongoing public private partnerships and spin-out companies. He is currently leading the EATRIS representation in the European Joint Program on Rare Diseases and the EU Patient Centric Clinical Trials Platform, an international initiative that aims to use patient-centric data and knowledge sharing to accelerate the development of new treatments and reduce the operational costs of clinical trials.

Mary Furlan Feitosa – Washington University (USA/Brazil)

Dr. Mary F Feitosa is an Associate Professor of Division of Statistical Genomics, Department of Genetics at Washington University School of Medicine in St Louis, USA. Her investigation has covered epidemiologic and genetic aspects of complex human traits and diseases integrating genomic, statistical, biometrical, gene-lifestyle interaction, and longitudinal modeling approaches. Her work has mainly focused on subclinical atherosclerosis markers, cardiovascular disease and their risk factors. Her research interests also center on genetic and epidemiologic aspects of healthy aging in families enriched for exceptional longevity. In addition, she has worked on genetic epidemiology of human leishmaniasis, leprosy, malaria, and Chagas disease.

Horacio Botti – Universidad de la Republica (Uruguay)

Horacio Botti is an Adjunct Professor of Biophysics in the Facultad de Medicina de la Universidad de la República (Uruguay), where he has been a faculty member since 2016. In 2017 he founded the Integrative Biophysics Laboratory (L@Bi). He is a co-founder and the coordinator of *Semillero de Actividades Interdisciplinarias Ciencias de Datos y de la Información Aplicadas a la Salud Humana* (CIDASH, https://ei.udelar.edu.uy/grupos-financiados).

Horacio completed his Ph.D. at the Universidad de la República in 2010 and his MD studies in 2001. His current research interests lie in the area of the inference of models for complex systems behavior. He is particularly interested in closing the gap between model inference, knowledge advancement and personalized high performance medicine. He has collaborated actively with researchers in several other disciplines, particularly protein biophysics, biochemistry, redox medicine and clinical research, authoring 31 full articles and two edited book chapters. He has served on the University co-government as well as on organizations that foster biomedical science, like Programa para el Desarrollo de las Ciencias Básicas (PEDECIBA) and Fundación Manuel Perez. He is the advisor of one master degree student working on the ubiquitin proteasome system and co-advices a master degree student working on learning assessment using automatic learning techniques.

Stefan Klein – Erasmus MC (Netherlands)

Stefan Klein is Associate-Professor in Medical Image Analysis and is affiliated with the Biomedical Imaging Group Rotterdam (BIGR), Department of Radiology and Nuclear Medicine, Erasmus MC, Rotterdam/NL. In 2002, Stefan received his MSc degree from the faculty of Mechanical Engineering at the University of Twente, Enschede/NL. In 2008 he obtained his PhD degree at the Image Sciences Institute, UMC Utrecht/NL, for his research on optimisation methods for medical image registration. In 2010, He received an NWO Physical Sciences VENI fellowship. He is co-principal developer of a widely used open source software package for medical image registration, called Elastix (article cited >1700x), was co-organiser of the CADDementia and TADPOLE grand challenges on dementia prediction, and was general chair of the WBIR2018 conference. His current research interests include image reconstruction, radiomics, machine learning, and disease progression modelling. He is also active in setting up infrastructures for research and has for instance initiated a national research archive for medical imaging data, currently used by numerous multi-centre imaging studies in the Netherlands.

Justo Lorenzo Bermejo – University of Heidelberg (Germany)

Justo Lorenzo Bermejo is a professor and head of the Statistical Genetics group at the Institute of Medical Biometry and Informatics, University of Heidelberg, Germany with background in mathematics and statistical genetics. His current research focuses on the development and application of robust and Bayesian methods in statistical genetics and the identification of molecular markers for the personalized prevention and treatment of cancer and infectious diseases.

Abstracts

Jorge Samaoya and Preng Biba
Title of the talk: Data Science at work: an overview around PerMed

In this talk, we introduce the audience to the different application of data science to personalized medicine. There are hundreds of algorithms being developed in order to extract information from data. This information is used to predict the best treatment for a patient. For example, we can use customer stratification techniques to classify severity of a disses depending on patients' blood pressure. We will give a brief explanation of how this type of algorithms work and provide some of the most relevant applications in PerMed. Also, we will discuss why data science has become relevant nowadays, and how "modern" statistics is helping us to solve problems that were impossible to describe in the past.

Adriano Barbosa da Silva
Title of the talk: Multimodal patient data integration to foster machine-learning aided personalised medicine

This talk will present the importance of big data management in personalised medicine. Exploring how we used different datasets from the EU/EFPIA IMI project AETIONOMY for defining a new taxonomy for Alzheimer's disease; the translational capabilities of open-source software for health data integration and management (tranSMART) and finally how we use population genomics collections (ADNI and UK Biobank) for the training of machine learning algorithms for the classification of patients in different cardiac diseases sub-groups. It will also will shed some light on how to use transnational datasets for translational research and personalised medicine.

Josep Maria Haro
Title of the talk: Integrating cohorts across Europe

Many countries are very well served by a rich variety of population, patient and clinical trial cohorts, harmonised with International studies. However, the enormous potential benefits of the rich information contained in cohort studies have not been exploited. The optimisation of health-related data to support personalised medicine, and efficient knowledge transfer from research to clinical practice, has been an international priority. We will present examples of integration of population and clinical cohorts, examine the methodological advances that have facilitated this integration and discuss the challenges that still exist.

Fabio Porto
Title of the talk: Improving the training of High Performance Athletes based on multidisciplinary Data Analysis

In this talk, Dr Porto will present the work we developed to support the individual monitoring and comparative analyses of high performance athletes. The work has been implemented in the SAHA system that enables the integration of data coming from different instruments and used by various disciplines, providing a holistic view on athlete conditions

during different stages of training. The system is designed taking into account the longitudinal perspective of the follow-up activity of multi-disciplinary data. The integrated longitudinal data enable temporal analysis from different perspectives. Moreover, it is being used as the basis for training a machine learning model that intends to emit alerts about undesired events, such as injuries.

Rosa Figueroa
Title of the talk: Challenges for Secondary use of patient healthcare information in Chile: methods, data quality, results, and future perspectives

Every day we create vast amounts of data through the use of technology. According to the IDC by 2025, we are expected to have a DataSphere of 175 zettabytes of data. Healthcare is expected to grow the fastest of all industries. Secondary use of healthcare data through machine learning or artificial intelligence offers an opportunity to extract knowledge from clinical data to improve the delivery of care to patients. However, real-world patient data is not commonly used due to the inaccessibility, quality of the report, or segregated data. In this talk, we will show you our experience with secondary use of data in Chile, together with the challenges and barriers, and future opportunities.

Helder Nakaya
Title of the talk: Network Medicine of Infectious and Inflammatory Diseases

Diseases are mostly a consequence of an abnormality in multiple genes. Network Medicine investigates how genes interact to each other in complex intracellular and intercellular networks. The talk will show the recent advances on this emerging field of research and its impact on precision medicine and drug repositioning in infectious and inflammatory diseases. We will cover the integration of omics data and clinical/immunological data, and also the application of machine learning approaches to understand and predict human diseases.

Adrian Turjanski
Title of the talk: Translational Bioinformatics: Genomics data for clinicians and patients

Recent advances in genome sequencing, delivering data cheaper and faster every year, together with new bioinformatics algorithms, artificial intelligence and big data promise a new era of personalized medicine. As the DNA of more and more people is being obtained the bottleneck for a truly genomics era in healthcare is that data need to be transformed into valuable information. Translational bioinformatics is a key field that needs to be develop and implemented to have reliable storage, analytic and interpretive methods to optimize the transformation of these increasingly voluminous genomic data, into proactive, predictive, preventive, and participatory health. In this talk Dr Turjanski will describe the development of a bioinformatics platform to analyze patients' genomics information in the clinic. We have implemented a fully web based solution that goes from the Fastq sequences to the prediction of Single Nucleotide Polymorphisms (SNP) effect, as understanding the functional effects of SNPs and their relation to disease development is a major issue in clinical genomics. Dr Turjanski will explain the recenlty developed VarQ Web Server that automatically analyzes SNPs for their effect on protein activity, folding, aggregation and protein interactions among

others, and will comment on the application to several cases of rares disease diagnosis that performed in the context of a large collaboration between his group, public hospitals, and private companies in the context of the Argentinian precision medicine initiative.

Elmer Fernandez
Title of the talk: Who is there? The new paradigm of molecular signatures.

RNA sequencing has proved to be an efficient high-throughput technique to robustly characterize the presence and quantity of RNA in a patient biopsy allowing to analyze the continuously changing tissue transcriptome. More importantly, it can be used to computationally unravel the tumor immune infiltrate content and infer the immunological phenotypes of those cells (i.e digital cytometry). Given the significant impact of anti-cancer immunotherapy and the role of the associated immune tumor microenvironment (ITME) on its prognosis and therapy response, the estimation of the immune cell-type content in the tumor is crucial for designing effective strategies to understand and treat cancer. This talk will show MIXTURE, a new Data Science based approach that infers cell-type proportions from transcriptome data of bulk tumor samples. This approach provides a renewal for gene expression molecular signatures.

Anton Ussi
Title of the talk: Quality and reproducibility of data at preclinical level for the development of Personalized Medicine, a multiomic perspective

A major factor for efficient adoption of PM is the robustness of results from preclinical studies, which requires the use of validated patient-targeted biomarkers that support the process of patient stratification. In this context, quality and reproducibility of preclinical research represents one of the most important challenges for the development of PM strategies and a critical element for an efficient adoption of bigdata at the clinical decision level. The implementation of international quality initiatives (e.g. multisite comparative studies) are a powerful tool for addressing best practices for reproducibility, standards, and reference materials. The development of Genomic medicine together with other omic technologies (proteomic, transcriptomic, metabolomic) has opened the possibility of developing a strong pipeline for multiomic development of preclinical research for Personalized Medicine improving dramatically the identification of diagnostic and predictive biomarkers in the context of patient stratification. This talk will focus on how these challenges on the quality and reproducibility of data can be identified and what can be done to overcome them.

Mary Furlan Feitosa
Title of the talk: Disentangling the genetics of complex phenotypes integrating multiple data sources in meta-analysis of genome-wide association studies

Genome-wide association studies (GWAS) have had success in identifying genetic variants predisposing to complex diseases and providing insights into their biology. GWAS allowed to discover associations between hundreds of thousands to millions of genetic variants across the genomes of many individuals with a trait (biomarker or disease). GWAS have primarily

used single nucleotide polymorphisms (SNPs) via genetic chip technology or sequencing and mapping of genetic variation. The GWAS statistical power to identify a variant associated with a trait (disease) depends mainly on: (1) experimental sample size; (2) the effect size of variant that segregates in the data and its allele frequency; (3) the linkage disequilibrium (squared correlation, r^2) between observed genotyped variant and "unknown" causal variant; (4) and the precision with which the trait is measured or the disease is diagnosed.
The increase of cohort samples, large scales of biobank data, and efforts from international collaborative genetic study consortia have enabled GWAS to discover thousands of genetic variants across a wide range of complex traits, biomarkers, and diseases. Widespread use of whole-genome sequencing has contributed to improve imputation reference panels, with both common (minor allele frequency (MAF) >5%) and rare (MAF <5%) genetic variation, and the inference of missing genotypes from more globally human genetic diversity. In addition, new analytical methods, molecular studies, and pathway-based analyses have collaborated to connect the association between a trait (disease) with genetic variant at a genomic locus to the target gene(s) or mechanisms.

Although the proportion of individuals of non-European ancestry represented in GWAS has increased, the number of participants in GWAS is still largely from European Ancestry. It is of great importance that a broader representation of diverse ethnic ancestry populations across the globe be included in genomic research.

Several GWAS Consortia have studied large-scale multi-ancestry populations, including: the CKDGen Consortium for kidney function from analyses of a million individuals (PMID: 31152163), and the Gene-Lifestyle Interactions Consortium studies for multiple blood pressure and lipids using smoking or alcohol consumption as lifestyles from up to 610,091 individuals from 124 cohorts (*e.g.*, PMID: 29912962, PMID: 31127295, PMID: 30926973). These studies have identified several loci associated with these cardiovascular disease risk factors. Some of these loci were identified in African ancestry but not in European, Asian, Hispanic, or Brazilian ancestries. The minor allele frequencies of some index variants were higher in African ancestry than in other ancestries, which highlights the importance of genetic studies in diverse populations. Overall, the large-scale multi-ancestries GWAS have contributed to the understanding of the genetic architecture of biomarkers and disease outcomes among diverse populations, which may benefit the development of therapeutic treatment, preventive and personalized medicine, and consequently mitigate the social-public health disparities.

Horacio Botti
Title of the talk: PerMed Uruguay: Preliminary research experiences and developments planes associated with C2DASH

Uruguay faces health challenges commonly associated with both high and low income countries. Despite a low investment in research and innovation, we have a modest tradition in basic research, which overcame the terrible effects on science and culture of our most recent dictatorial period (1973-1984). Clinical and epidemiological research never accomplished such a catch-up process. For this to be accomplished, more and/or better

human resources, funds, infrastructure and organization are needed in our country. A few researchers are already moving towards PerMed and related paradigms of research and innovation, not only from the basic but also from the translational perspective. In 2018 we launched CIDASH, an interdisciplinary group of academics to promote the development of applications in Information and Data Sciences applied to human health. This academic space congregates approximately 50 researchers coming from relatively distant fields like Clinical Medicine, Biology (Biophysics, Genomics and Genetics), Epidemiology, Health Information, Information and Communication, Engineering (Signal Processing and Automatic Learning), Psychology, Biostatistics, Bioethics, etc. CIDASH is approaching a new stage: C2IDASH. Emphasis will be on Bioethics, Computation, Data and Information Science and Research Infrastructure as well as in making progresses towards PerMed, Precision Medicine and High Performance Medicine/Health more effective. This presentation has been developed in collaboration with other members of CIDASH: Lucía Spangenberg (Unidad de Bioinformática, Institut Pasteur de Montevideo), Víctor Raggio (Departamento de Genética, Facultad de Medicina, Universidad de la República), Nicolás Nin (Unidad de Cuidados Intensivos, Hospital Español Dr. Juan José Crottogini, Administración de Servicios de Salud del Estado), José Tort (Unidad de Cuidados Intensivos, Hospital Español Dr. Juan José Crottogini, Administración de Servicios de Salud del Estado), Hugo Naya (Departamento de Producción Animal y Pasturas, Facultad de Agronomía, Universidad de la República y Unidad de Bioinformática, Institut Pasteur de Montevideo).

Stefan Klein
Title of the talk: Medical Image Analysis for Personalized Medicine

Emerging artificial intelligence (AI) techniques based on machine learning, such as radiomics and deep learning, hold the promise to turn medical images (such as X-ray, CT, MRI, PET, Ultrasound) into objective and quantitative biomarkers, optimally guiding diagnosis and treatment. In the field of medical image analysis, machine learning has become the main workhorse in automating any image processing task, ranging from low-level operations like segmentation to high-level interpretations such as diagnosis, subtyping, and prediction. This presentation, will show several examples of the latter, illustrating how medical image analysis can play an important role in personalised medicine. Dr Klein will explain the underlying technical principles, and highlight some of the infrastructural challenges encountered in this type of research. Machine learning is a “data-hungry” approach: the computer needs vast amounts of clinically representative example data to learn models that generalise well to new data. Therefore, efficient workflows for data gathering, anonymisation, harmonisation, storage, sharing, annotation, processing and integration are necessary to accelerate personalized medicine research.

Justo Lorenzo Bermejo
Title of the talk: Collaborative research between Europe and Latin America in precision medicine for Chagas disease and gallbladder cancer

Chagas disease is a potentially fatal illness caused by the parasite Trypanosoma cruzi. There are more than 8 million people infected worldwide, mainly in endemic areas of Latin America,

but also increasingly in Europe due to migration and travel. Infected persons can be treated with benznidazole (BZ) or nifurtimox (NFX). BZ and NFX show close to 100% effectivity in killing the parasite shortly after infection, but efficacy sharply decreases with post infection time.
Using a three stage approach, we are conducting a collaborative research project in Bolivia, Chile, Germany and Italy to examine the impact of (epi)genetic variability on (1) BZ and NFX treatment efficiency and related toxicities, (2) the prognosis of Chagas disease (e.g. cardiomyopathy development), and (3) Chagas susceptibility and associated deaths. Human genetic and epigenetic variability (characterized by genome-wide genotypes and sncRNA profiles in pre-treatment serum), the genetic makeup of T. cruzi (characterized by DTUs), and possible interactions among them will be investigated. We aim to exploit the strengths of combining data from (i) 500 Chagas patients recruited in Chile and Italy with rich information on response to treatment (e.g. related toxicities), (ii) 1000 Chagas patients with clinical information (e.g. cardiomyopathy development), and (iii) 5000 individuals from a large geographic area (Colombia, Ecuador, Peru, Bolivia, Argentina and Chile).

Gallbladder cancer is a neglected disease with huge potential for prevention. In our Horizon2020-funded project “Establishment and Exploitation of a European–Latin American Research Consortium towards Eradication of Preventable Gallbladder Cancer” we aim at significantly improving the accuracy of risk estimation and early detection by identifying and adequately considering geographical, environmental, lifestyle, ethnic, gender and molecular differences. We plan to (1) build a unique European–Latin American biorepository integrated into a tailored IT platform, (2) identify, validate and functionally characterize novel GBC biomarkers, (3) develop a multifactorial risk score that integrates established and newly identified epidemiological and molecular risk factors, (4) improve the understanding of the causal mechanisms that link lifestyle, cultural and behavioural factors to GBC development and (5) exploit existing and newly generated epidemiological and multi-omics data to improve the accuracy of GBC risk prediction.
The planned biorepository will be unique regarding the number of collected biosamples. Samples and data collected before diagnosis from 200 European case–control pairs, and Chilean GBC cases (n=400), population-based healthy controls (n=3000) and gallbladder cancer families (n=18) are already available. They will be complemented by DNA, serum, plasma, saliva, urine, faeces and formalin-fixed paraffin-embedded (FFPE) gallbladder tumour tissues from 15,000 newly recruited Argentinean, Bolivian, Chilean and Peruvian gallstone disease and gallbladder cancer patients. Within the project, serum samples will be used for metabolomics assays and sncRNA sequencing, DNA samples will be used for genome-wide genotyping, targeted DNA sequencing and epigenome-wide methylation profiling. Unused aliquots, urine and faeces samples will be stored for future assays.