

HORIZON  
2020

# Innovative Statistical modelling for a better Understanding of Longitudinal multivariate responses in relation to Omic datasets

## Informe

### Información del proyecto

#### ISULO

Identificador del acuerdo de subvención:  
840383

[Sitio web del proyecto](#)

#### DOI

[10.3030/840383](https://doi.org/10.3030/840383)

Proyecto cerrado

Fecha de la firma de la CE  
4 Abril 2019

Fecha de inicio  
1 Marzo 2020

Fecha de  
finalización  
3 Agosto 2023

#### Financiado con arreglo a

EXCELLENT SCIENCE - Marie Skłodowska-Curie  
Actions

#### Coste total

€ 275 619,84

#### Aportación de la UE

€ 275 619,84

#### Coordinado por

CENTRE DE COOPERATION  
INTERNATIONALE EN  
RECHERCHE AGRONOMIQUE  
POUR LEDEVELOPPEMENT -  
C.I.R.A.D. EPIC

 France

## Periodic Reporting for period 2 - ISULO (Innovative Statistical modelling for a better Understanding of Longitudinal multivariate responses in relation to Omic datasets)

## Resumen del contexto y de los objetivos generales del proyecto

In medicine or agronomy, the understanding of dynamic processes, such as disease progress or growth, is crucial. Such processes are controlled by complex biological mechanisms that involve markers from different molecular levels and by environmental factors. Disentangling these mechanisms may help in the development of more effective therapeutic strategies in human disease and may contribute to gaining insights into the adaptation of plants in the face of climate change.

During the last decades, advances of high-throughput technologies have led to the acquisition of various types of -omic data, including whole-genome sequencing, methylation, transcriptomic, glycomic, proteomic, and metabolomic. Such data offer the opportunity of gaining insights into complex biological mechanisms by making possible the study of relationships between features from various molecular levels and the study of their implication in biological processes. The first approaches have analyzed each dataset separately providing a snapshot of the molecular processes involved. Over the last years, integrative approaches that combine complementary information from each data type have been proposed. While they have led to significant results, they have also highlighted that the integration of several biological layers of information is still challenging. Indeed, statistical models need to efficiently deal with high-dimension while considering dependence between/within -omic data. In addition, there has been a growing interest in identifying biological mechanisms involved in the evolution of traits along time, for example by identifying active genes along plant growth, or biological pathways involved in fetus development. The associated data are characterized by measures along time on the same individuals and are called longitudinal data. The dependence among measurements on a same individual needs to be considered in the statistical models.

The joint analysis of traits evolving over time in relation to high-dimensional -omic data from various biological levels is a recent active area that gives new opportunities for an enhanced understanding of dynamic biological processes. But this also raises new statistical challenges related to, among others, high-dimensionality and dependence structure between variables. To this end, innovative statistical methods need to be developed. The ISULO project's main objective was to propose innovative methodologies to simultaneously analyze longitudinal data and high-dimensional -omic data. The research methodology has been subdivided into two major parts, one focusing on the analysis of one longitudinal outcome in relation to one -omic dataset, and a second part focusing on the integration of multiple high-dimensional -omic datasets for explaining non-longitudinal and/or longitudinal outcomes. In the two parts, the proposed methods aim at dealing with high-dimensional data for which prior knowledge about relationships between predictors and/or outcomes may be available.

The ISULO project has provided innovative statistical approaches addressing biological questions that arise both in medicine and agronomy. The results have highlighted the advantage of integrating known structures between covariates, between responses or between covariates and responses into the statistical models. They have also pointed out the difficulties/limitations that may be encountered

when jointly modelling multivariate outcomes with multiple -omic datasets. The interdisciplinary nature of the project has allowed exploration of various biological and statistical areas and high-collaboration with researchers in statistics, genetics, epidemiology, molecular biology, and oncology. The wide applicability of the proposed approaches as well as the expertise acquired in the analysis of data generated by cutting-edge technologies have led to new collaborations with national and international researchers working in different domains. By addressing challenging methodological questions, this action has also generated new interdisciplinary questions that will be investigated in future projects.

## Trabajo realizado desde el comienzo del proyecto hasta el final del período abarcado por el informe y los principales resultados hasta la fecha ▼

Throughout the project, innovative Bayesian variable selection methods have been developed or applied to handle high-dimensional datasets with structure between covariates, between responses or between covariates and responses. For all developed approaches, the associated R code are available on the web interface github. Whether in medicine or in agronomy, those approaches have allowed to disentangle genetic and/or epigenetic bases for phenotypes under investigation. The results have given new insights into the biological mechanisms governing complex processes. For example, in a study following patients with hepatocellular carcinoma (HCC), key miRNA-mRNA pairs and pathways that are potentially associated with HCC have been identified. In the meantime, new genomic regions associated with relatively small effects in oil palm have been discovered. All of these findings, which need to be experimentally validated, will provide, in medicine, a better selection of biomarkers that may serve as diagnostic, prognostic, or therapeutic candidates. In agronomy, it will help breeders select plants with a good capacity to deal with climate change. Moreover, the dissemination that has occurred throughout the action as well as the wide applicability of the proposed approaches has allowed to target audiences from various application domains and diverse research institutes and, so, to reach many potential users. New collaborations and working groups combining statisticians and biologists from national institutes have also been initiated.

## Avances que van más allá del estado de la técnica e impacto potencial esperado (incluida la repercusión socioeconómica y las implicaciones sociales más amplias del proyecto hasta la fecha) ▼

The developed statistical methods contribute to a better understanding of complex biological mechanisms between longitudinal or non-longitudinal outcomes and biological features from different biological levels.

They address biological and/or methodological questions that commonly arise from longitudinal and -omic data and for which there are no efficient methods available. The results on simulated and real datasets have shown better performance compared to existing approaches. They have also shown that the integration of dependence structure between and within predictors and outcomes leads to a better identification of discriminatory covariates and, so to a better understanding of complex

biological mechanisms. This brings new opportunities for identifying diagnostic, prognostic, or therapeutic biomarker candidates in medicine, and for identifying genotypes most adapted to climate change in agronomy.

As most of the proposed approaches have been developed to address biological questions encountered in medicine and agronomy, the plan is to widely apply these approaches on various real datasets from the two domains.

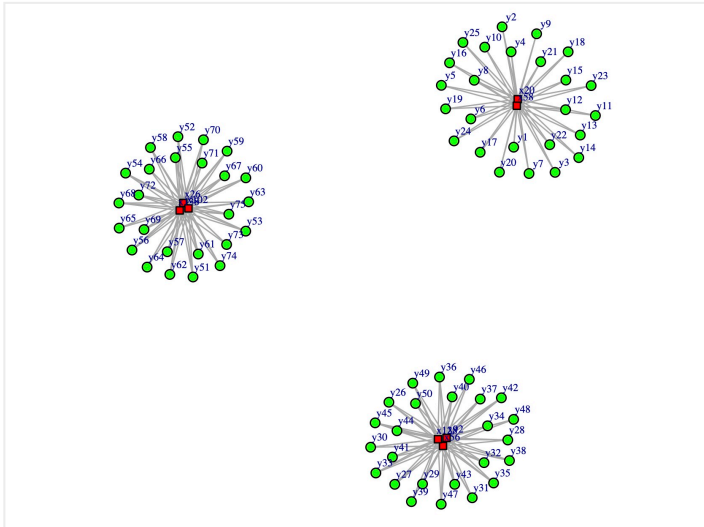
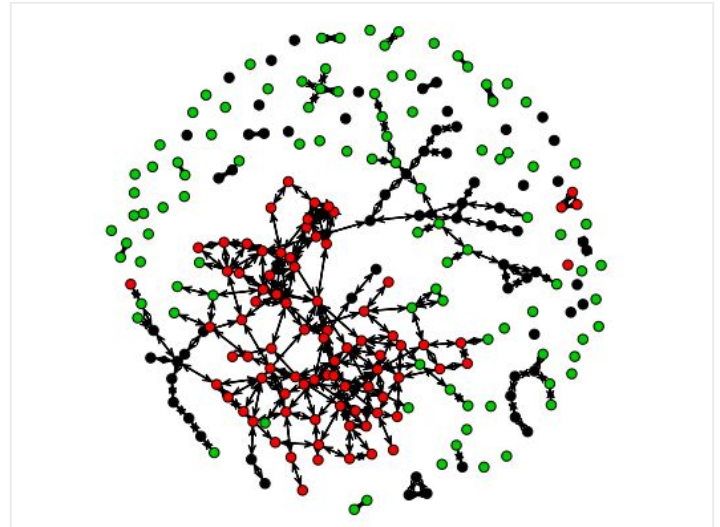


illustration2.jpg



a) Example of network structure of three -omic datasets (genes in red, methylations in black, metabo)

Última actualización: 7 Junio 2024

Permalink: <https://cordis.europa.eu/project/id/840383/reporting/es>

European Union, 2025