

rexposome

A bioinformatic tool for characterizing multiple environmental factors and its association with different omics biomarkers and disease

Carles Hernandez-Ferrer^{1,2,3,*}, Martine Vrijheid^{1,2,3}, Juan R González^{1,2,3}

¹ISGlobal, Centre for Research in Environmental Epidemiology (CREAL) / ² Universitat Pompeu Fabra (UPF) / ³ CIBER Epidemiología y Salud Pública (CIBERESP)
* Contact information: chernandez@creal.cat / www.creal.cat/brge

www.projecthelix.eu

@HELIXexposome

#exposome



INTRODUCTION

Exposome encompasses the totality of human environmental exposures from conception onwards, complementing the genome. The overall aim of **HELIX** is to exploit novel tools and methods to characterise early-life exposure to a wide range of environmental hazards, and to integrate and link these exposures with data on major child health outcomes in population studies, thus developing an Early-Life **Exposome** approach. For that, **HELIX** focuses on the most critical periods of life during pregnancy and the first years of childhood. One of the **HELIX** objectives is to determine molecular signatures associated with environmental exposures through analysis of metabolites, proteins, RNA transcripts and DNA methylation. To achieve this goal two different approaches will be followed: single-integration, that corresponds to integrate the **exposome** with each **single-omic**, and **cross-omics**, that corresponds to the integration of the **exposome** with all omics.

METHODS

In order to implement the required pipelines to analyse the **exposome** and its integration with omics data, three R packages have been developed: **MultiDataSet** and **rexposome** and **CTDq**. **MultiDataSet** is a package implementing a class to encapsulate datasets from multiple omics with no samples restriction. It aims to be the new master class in Bioconductor for multi omics analysis. **rexposome** includes tools to find associations between the exposome and transcriptome data, methylome data, preoteome data or phenotypes. It also includes tools to integrate the exposome with multiple omics data (genome, transcriptome, methylome, proteome) with and without supervision (figure 1). **CTDq** was developed to query “The Comparative Toxicogenomics Database” and “DisGeNET” to enrich and visualize the results obtained by **rexposome**.

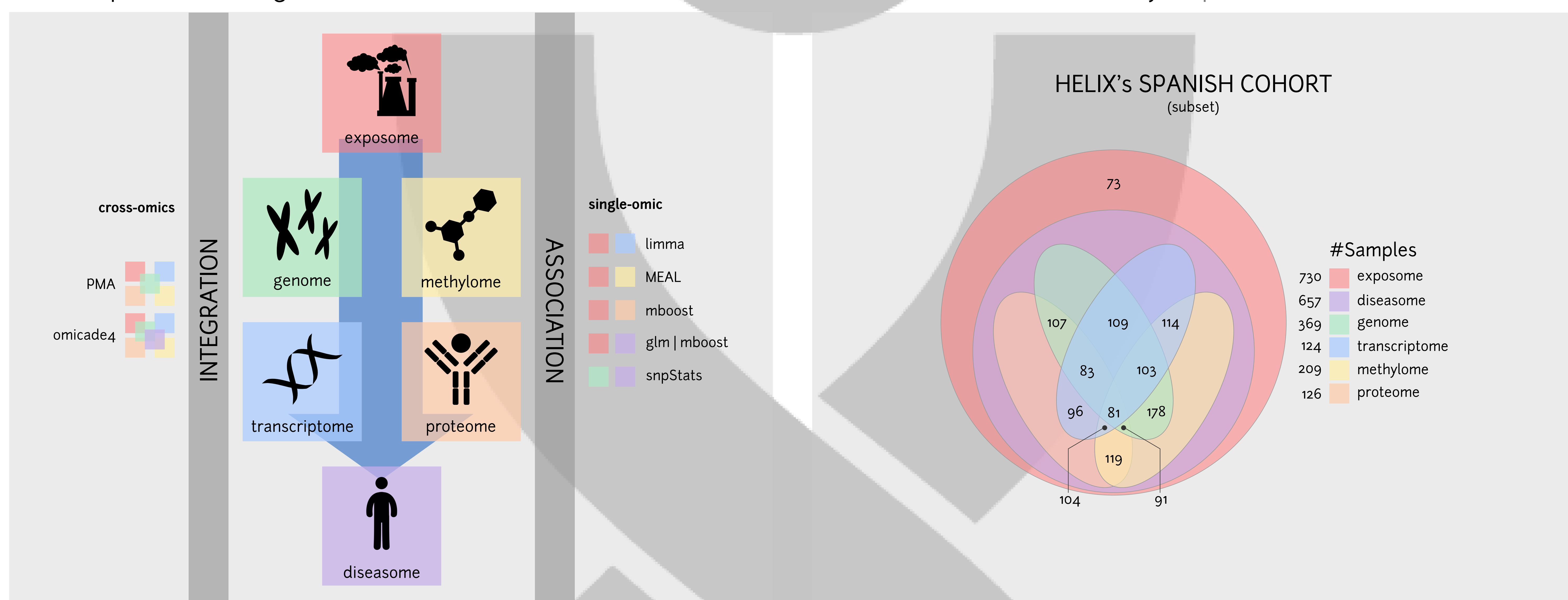


FIGURE 1 - rexposome is an R package conceived to perform massive association and integration analysis in HELIX framework. It uses standard and well known pipelines to discover associations between exposome and omics data, also with phenotypes (uni-variant and multi-variant). It also brings two strategies to deal with multi omics analysis (crossomics): multi canonical correlation for non supervised analysis and multiple co-inertia for phenotype-supervised analysis.

FIGURE 2 - The subset of the HELIX's Spanish cohort includes 88 exposures grouped in 13 families for 730 subjects. It also includes 4 phenotypes for 657 subjects, gene expression data for 124 subjects (Affymetrix HTA 2.0), methylation data for 209 subjects (Illumina 450k), SNP data for 396 subjects (Illumina HumanCore BeadChip) and proteome data including 47 analytes for 125 subjects (two R&S System Quantikine Human kits).

RESULTS

Package performance was tested by analysing a subset of the HELIX's Spanish cohort. This included a set of 4 phenotypes (sex, age, chest whistling and wheezing), 88 exposures, SNP data, gene expression data, methylation data and proteome data (figure 2). **ExWAS** (Exposome-Wide Association Study) showed association between Monoisobutyl phthalate, Vanadium in PM₁₀ and Cotinine with chest whistling (P-Value < 0.05). Vanadium in PM₁₀, Mono-(2-ethyl-5-oxohexyl), Mono(2-ethyl-5-carboxypentyl) and Mono-(2-ethyl-5-hydroxyhexyl) phthalates were linked to wheezing (P-Value < 0.05).

Single-omics association analyses found three BDEs (Polybrominated diphenyl ethers: BDE₁₅₃, BDE₂₀₉ and BDE₄₇) were associated with genes enriched in synaptic transmission, neurogenesis, cell development, cell differentiation, neuron development, neuron projection development and neuron projection morphogenesis (P-Value < 10e-5). This analysis also reveal associations between Pb (Metal) and DDT (Organochlorine) with multiple RNA regulation and metabolic processes, respectively. **Cross-omic analysis** found a significant correlation between genes RP11, RNU6 and HTR3E with particle matter.

AVAILABILITY

R package **MultiDataSet** was released with Bioconductor 3.3. Package **rexposome** is currently being used within **HELIX** project and its release is planed with Bioconductor 3.4. **CTDq** package is still in development, and it is planed to be released with **rexposome** in Bioconductor 3.4. Current version of **rexposome** and **CTDq** are available upon request.

FUNDING

This work has been supported by the **Spanish Ministry of Economy and Competitiveness** (MTM2015-68140-R) and **HELIX** Project supported by **European Comission** FP7 (GA#308333).

