Alyssa Imbert[1,2], Magali Rompais[3], Julia Novion Ducassou[4], Emmanuelle Mouton-Barbosa[5], Mohammed Selloum[6], Florence Castelli[1], Charlotte Joly[7], Emeline Chu-Van[1], Aurélie Hirschler[3], Marie Courçon[4], Yohann Couté[4], Christophe Lechaplais[8], Natacha Lenuzza[1], Arthur Tenenhaus[9], Laurent Vasseur[6], Marion Brandolini-Bunlon[7], Ludovic Cottret[11], Yves Vandenbrouck[4], Olivier Sand[2], Christophe Bruley[4], Franck Giacomoni[7], Fabien Jourdan[10], Odile Schiltz[5], Pierre Le Ber[8], Claudine Médigue[2,8], Jérôme Garin[4], Yann Herault[6], Dominique Rolin[12], Christophe Junot[1], Marcel Salanoubat[8], Virginie Brun[4], Alain Perret[8], Myriam Ferro[4], Tania Sorg[6], Estelle Pujos-Guillot[7], Anne Gonzalez-de-Peredo[5], François Fenaille[1], Christine Carapito[3], David Vallenet[2,8], and Etienne Thévenot[1] (the ProMetIS consortium)

MetaboHUB National Infrastructure in Metabolomics and Fluxomics: [1]Saclay (MetabolomeIDF, CEA/INRA), [7]Clermont (PFEM, INRA/Clermont Auvergne University), [10]Toulouse (MetaToul, INRA/INSA/INSERM), [12]Bordeaux (PMB, INRA/CNRS/Bordeaux University); IFB Institute of Bioinformatics: [2]Evry (Core UMS3601 and Genoscope, CNRS/CEA/INRA/INRIA/INSERM); ProFI National Infrastructure in Proteomics: [4]Grenoble (EDyP U1038, CEA/INSERM/Grenoble Alpes University) [3]Strasbourg (LSMBO, IPHC UMR7178, CNRS/Strasbourg University), [5]Toulouse (IPBS, Toulouse University/CNRS/UPS), France; PHENOMIN National Infrastructure for Mouse Phenogenomics: [6]Strasbourg (ICS, CNRS/INSERM); France Génomique National Infrastructure: [8]Evry (LABGEM, CEA/CNRS); [9]L2S (CentraleSupelec/CNRS); [11]LIPM (INRA/CNRS).

Proteomics and metabolomics data integration for the study of knock-out mice: from univariate to multi-block analysis

The ProMetIS consortium, gathering 10 laboratories from the 5 national infrastructures for mouse phenogenomics (PHENOMIN), genomics (France Génomique), proteomics (ProFI), metabolomics and fluxomics (MetaboHUB), and bioinformatics (IFB), has designed a case study focusing on the high-throughput phenotyping of mouse models. Liver from homozygous mutant mice for the "Linker For Activation Of T Cells" gene (LAT), together with their wild-type littermates (WT), were analyzed by clinical phenotyping, proteomics, and metabolomics complementary LC-MS technologies. LAT is located on chromosome 16 in humans, in a region that is deleted in case of intellectual disabilities. Moreover, this gene is associated with immunodeficiency [1]. Our objective is to use complementary information from proteomics and metabolomics data to better understand the biological processes involved in the suppression of this gene LAT.

We present a framework to manage and analyze proteomics and metabolomics datasets acquired on the same samples to carry out an integrative analysis. We proceed step by step, by starting from basic univariate statistics and leading to multi-block integration analysis. At first, we analyze each dataset separately with some univariate and multivariate statistical methods (PCA, PLS-DA etc.). We look for differentially expressed proteins and metabolites and search the impacted biological pathways. We present here the first results.

However, it is generally admitted that studying a single kind of omics data is not sufficient to understand all the biological mechanisms involved. The integration of several datasets helps to enhance biological interpretation and provide more robust results. For that, we focus, finally, on a multi-block approach: the regularized generalized canonical correlation analysis method (rGCCA) [2], which enables a broad of exploration of the relationships between the individual datasets, in an unsupervised or supervised context, as well as feature selection.

References

[1] B. Keller et al., Early onset combined immunodeficiency and autoimmunity in patients with loss-of-function mutation in *LAT,* The Journal of Experimental Medicine, 213(7) (2016) 1185

[2] A. Tenenhaus et al., Regularized generalized canonical correlation analysis, Psychometrika 76(2) (2011) 257-284