



HAL
open science

Data integration of transcriptomic and metabolomic data in a cohort of patients with pancreatic neuroendocrine tumors.

Benjamin Marsac, Lucie Coppin, Arnaud Jannin, Jimmy Vandel

► **To cite this version:**

Benjamin Marsac, Lucie Coppin, Arnaud Jannin, Jimmy Vandel. Data integration of transcriptomic and metabolomic data in a cohort of patients with pancreatic neuroendocrine tumors.. JOBIM - Journées ouvertes en biologie, informatique, et mathématiques 2024, Jun 2024, Toulouse, France. hal-04664185

HAL Id: hal-04664185

<https://hal.science/hal-04664185v1>

Submitted on 29 Jul 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Benjamin MARSAC^{1,2}, Lucie COPPIN^{3,4}, Arnaud JANNIN^{3,5} and Jimmy VANDEL¹

¹ - University of Lille, CNRS, Inserm, CHU Lille, Institut Pasteur de Lille, US 41 - UAR 2014 - PLBS, 59000 Lille, France

² - University of Rouen Normandie, Normandie Univ, Master Bioinformatique, 76000 Rouen, France

³ - University of Lille, CNRS, Inserm, CHU Lille, UMR9020-U1277—CANTHER—Cancer—Heterogeneity Plasticity and Resistance to Therapies, 59000 Lille, France

⁴ - CHU Lille, Service de Biochimie et Biologie moléculaire « Hormonologie, Métabolisme-Nutrition, Oncologie », 59000 Lille, France

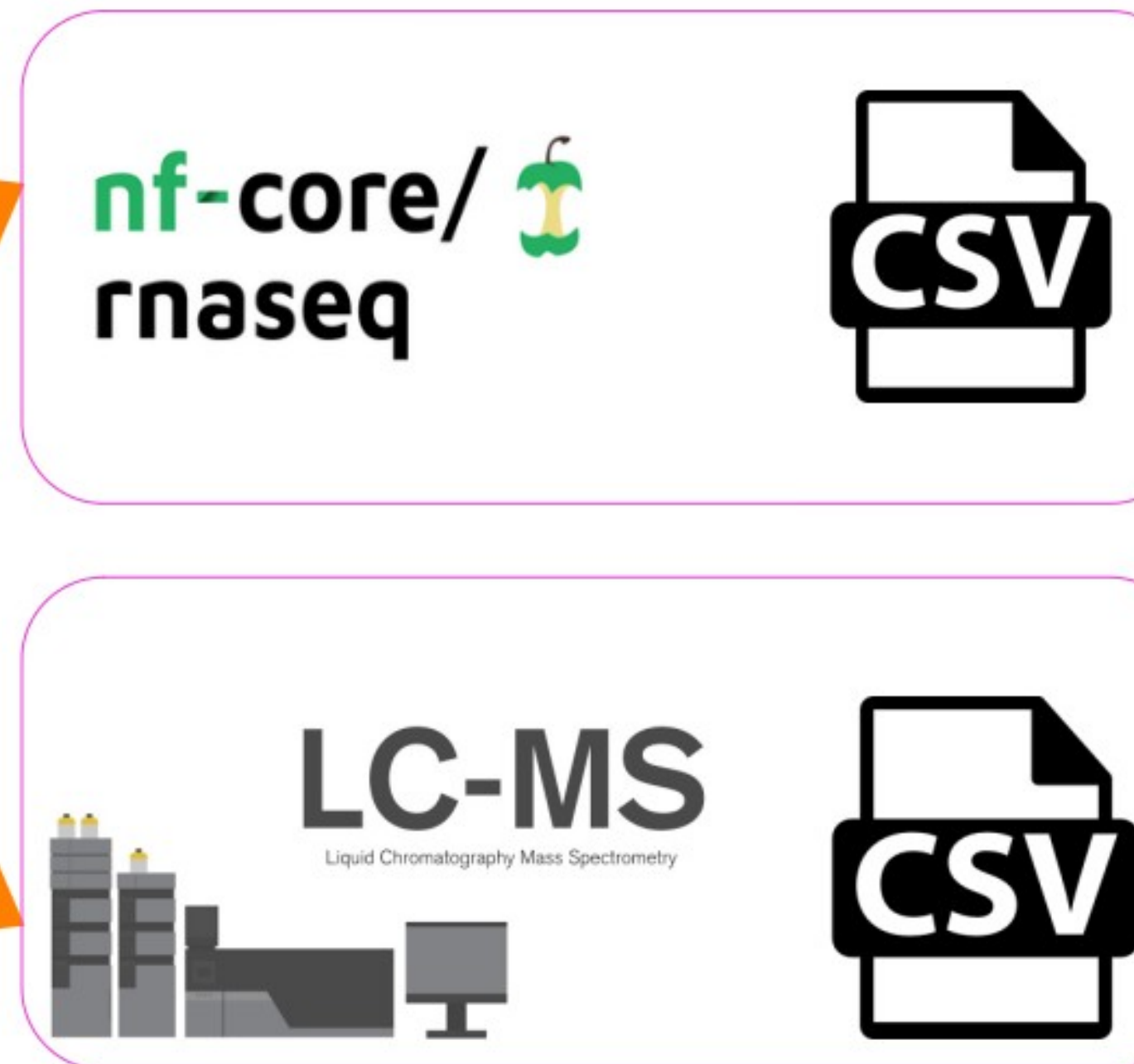
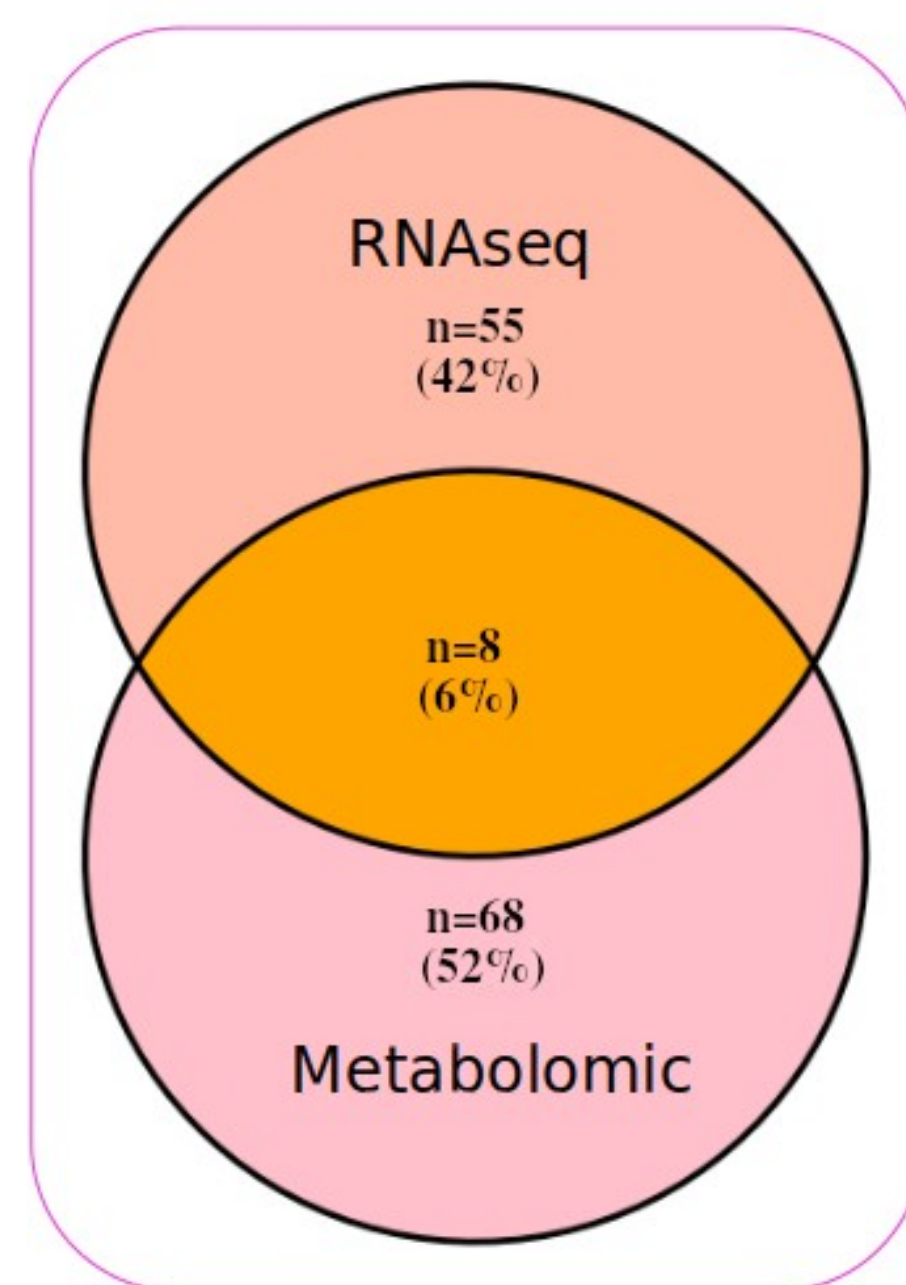
⁵ - CHU Lille, Department of Endocrinology, Diabetology and Metabolism, 59000 Lille, France

Abstract

Pancreatic neuroendocrine tumors (pNETs) form a group of rare tumors, the incidence of which is increasing (0.27 to 1 case per 100,000 individuals from 2000 to 2016), and they have the particularity of being able to produce hormones. Despite the fact that these tumors frequently develop metastases, their relatively slow growth results in a prolonged survival rate (50% at 5 years). The NEMETABO study (led by Dr. Arnaud Jannin, CHU Lille) gathers multi-omics data from two cohorts of patients with pNET: a first cohort with clinical and transcriptomic data (n=63) and then a second cohort with clinical and metabolomic data (n=76), 8 patients of which are common between these two cohorts. Using the data from these two cohorts, we performed analyses integrating transcriptomic and metabolomic data using the RGCCA algorithm. These analyses were initially performed in an unsupervised manner, without any prior assumptions about potential groups, and then in a supervised manner based on certain factors of interest.

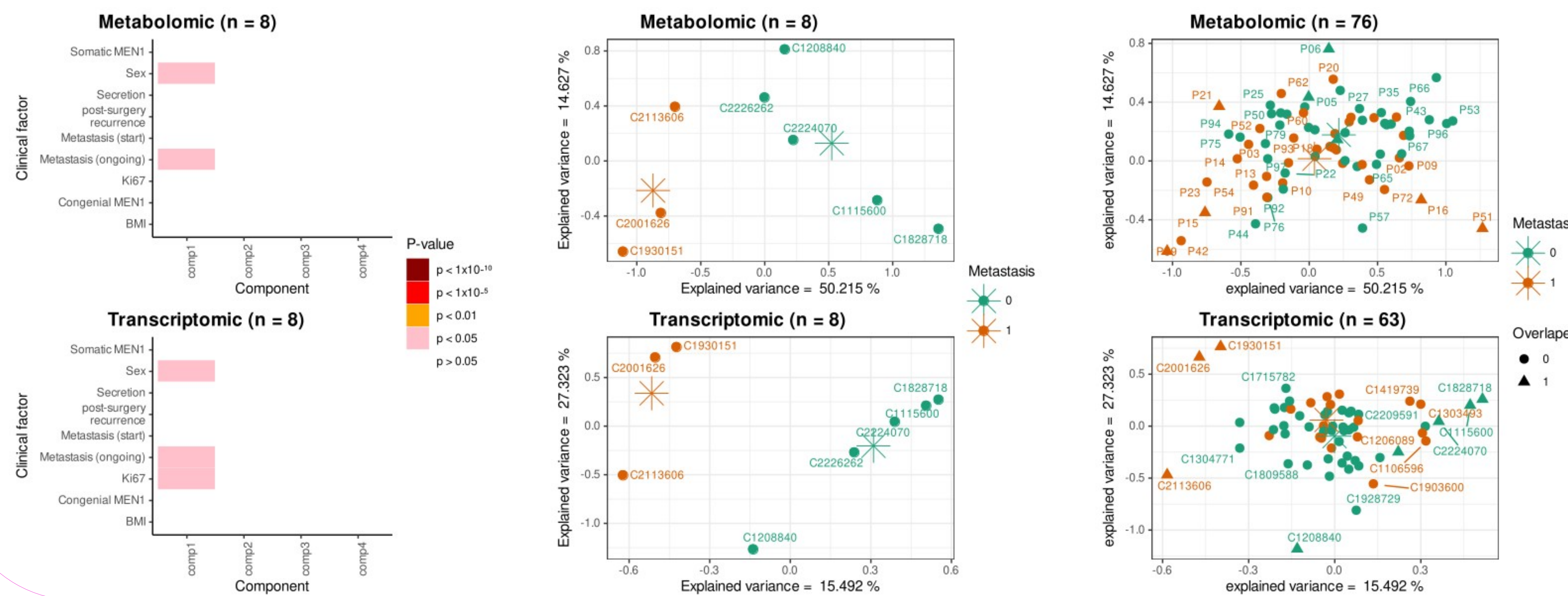
1-Data and pre-processing

RNAseq and metabolomic data were obtained from tumor biopsies and plasma samples of pNET patients. Each omic was first analyzed independently using standard methods. Then the two omics blocks were analysed with RGCCA to ensure correlation between gene and metabolite information. The main difficulty lay in the small number of patients (n=8) at the intersection of these two blocks.



2-Unsupervised RGCCA analysis

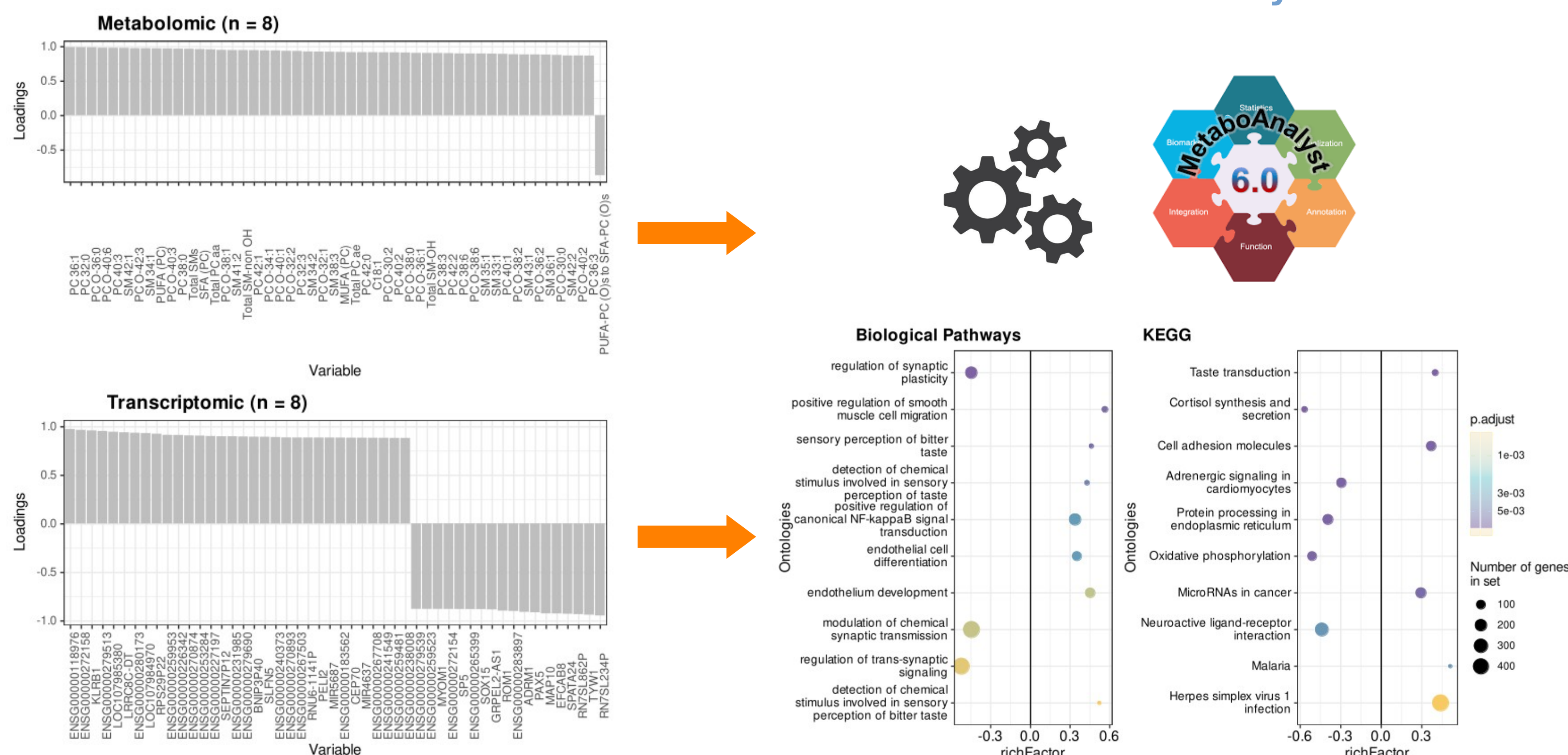
a. Components associations b. Projection of common patients Metastasis c. Projection of other patients Metastasis



The association between the components identified by RGCCA ($\tau=0.5$) and the clinical variables was calculated using a Kruskal-Wallis test (a). Only the first component showed statistically significant associations (p-value < 0.05), notably with sex and the presence of metastases. The projection of the 8 patients onto the first 2 components of RGCCA (b) shows a separation of the patients according to the presence of metastases. Using the component weights, we projected the remaining patients according to the available omics (c). Unfortunately, this projection does not show the same separation.

3-Biological interpretation

a. Loading vectors b. Functional analysis



We have calculated the loading vectors (correlation between omics variables and components) and performed an enrichment analysis. The metabolomic loadings analysis is in progress. Gene Set Enrichment Analysis was performed on transcriptomic loadings using GO and KEGG pathways.

4-Conclusion & perspectives

This integrative analysis is the first to be carried out on patients with pNETs. Despite the size of each cohort, the number of patients with both omic information is very limited (n=8). However, unsupervised RGCCA analysis identified a component statistically linked to the presence of metastases. For now, the projection of other patients on this component has not validated this association. However functional analysis of loadings showed enrichments linked to chemical/sensory and immune transmission pathways. We are currently finalizing enrichment analyses on metabolite loadings, made difficult by the variability in metabolite nomenclature. This will enable us to validate the consistency between functional transcriptomic and metabolomic analyses. We will also use supervised RGCCA for several clinical parameters of interest, such as the MEN1 predisposition gene.