Matinées thématiques du groupe « Statistique & Génomique »

du Réseau Interdisciplinaire autour de la Statistique

« L'ANALYSE (MULTI-)OMIQUE »

En webinaire ZOOM

JEUDI 14 OCTOBRE 2021

Inscription à : https://univ-lille-fr.zoom.us/meeting/register/tJ0qc--urDgqE9RIBHhOOFiJHEe50Q0HnPsQ

9h30 – 10h15 : "DECONVPDAC : A SINGLE-CELL BASED QUANTIFIER OF PDAC TUMOR CELLULAR HETEROGENEITY"

Magali Richard, Laboratoire TIMC, Grenoble

The pancreatic ductal adenocarcinoma (PDAC) is a very aggressive and invasive tumoral lesion affecting the pancreas. PDAC incidence increases regularly in Western countries and is expected to become the second leading cause of cancer-related mortality in 2025.
As any solid cancer, PDAC are composed of the 'tumoral mass' which is surrounded by normal epithelial cells, and by a 'stroma', the stromal cells giving support, nutrients and sometimes resistance and metastatic potential to neoplastic cells. The PDAC intra-tumor heterogeneity is a major pathological feature that can confer aggressiveness and chemoresistance. A promising approach to accurately quantify cell type heterogeneity in PDAC relies on the recent emergence of bulk deconvolution algorithms based on single-cell reference profiles. One of the main limitations of these approaches is the accuracy of the single-cell based profiles, which can strongly impair the quantification and the biological interpretation of the inferred tumor composition.
We characterized a pool of 39,000 single cell RNA-Seq from normal and tumoral pancreas from 4 different published studies. Using a curated gene-marker database generated from multiple sources, we were able to assign 17 different cell types, from which we built an integrative set of PDAC cell-type specific gene markers. We then use these markers to systematically assign cell identity of various public scRNA-seq pancreatic dataset (normal and cancerous types) and to revise our current understanding of cell type heterogeneity in PDAC. Finally, we evaluate the performances of bulk transcriptome single-cell based deconvolution algorithms, using our newly generated set of gene markers. The accuracy and robustness of cell type heterogeneity quantification was assessed using in vitro and in vivo dedicated benchmark datasets. From this work, we provide a highly resolutive catalog of cells present in PDACs as well as their markers and the scientific community will be able to benefit from it for downstream analyzes.

10h15-10h30 : Pause

10h30-11h15 : " DROMICS: AN R PACKAGE FOR DOSE-RESPONSE MODELING IN A MULTI-OMICS APPROACH"

Marie Laure Delignette-Muller, Laboratoire de Biométrie et Biologie Evolutive, Villeurbanne

Many tools were developed to analyze omics data in the context of differential analysis, where the purpose is to compare two or more conditions. In ecological risk assessment, the purpose is often different: in order to identify and describe the impact of a contaminant on an organism/population/community, bioassays are performed along a gradient of exposure concentrations and one of the purpose is to derive an effect concentration (e.g. a benchmark dose, BMD) from the concentration-response curve. We started the development of the DRomics tool a few years ago (Larras et al. 2018), as an R package and a Shiny application, to enable dose-response modeling on omics data especially in the context of environmental risk assessment. The workflow proposed in DRomics was built in order to i) identify the responding molecular items (e.g. contigs, metabolites) and ii) optimize the selection of monotonic and biphasic responses from data collected within a typical dose-response design with a great number of tested concentrations and a low number of replicates (or even no replicate). In order to help the understanding of biological functional meaning of results in an AOP context (Adverse Outcome Pathways) we also wanted to provide, in addition to the BMD, a characterization of the dose-response trend (as increasing, decreasing, U-shape or bell-shape). DRomics enables a multi-omics approach, allowing the same workflow to be used with microarray, RNAseq, or other omics data and also continuous anchoring apical data. The R package provides functions to help the interpretation of the workflow results in view of a biological annotation (e.g. from KEGG or GO data bases). We will present you the use of DRomics on an example of multi-omics approach in ecotoxicology (Larras et al., 2020), with a focus on new functions we recently developed to help the biological interpretation of the dose-response modeling results at different omics levels but with a common biological annotation.

****

JEUDI 28 OCTOBRE 2021

Inscription à : https://univ-lille-fr.zoom.us/meeting/register/tJUkduyvrzoiG9WUpE1389NQn_wJklZOJ2FP

9h30 – 10h30 : "MIXOMICS: AN R PACKAGE FOR THE INTEGRATION OF BIOLOGICAL DATA SETS"

Sébastien Déjean, Institut de Mathématiques de Toulouse

It is generally admitted that single 'omics analysis does not provide enough information to give a deep understanding of a biological system, but we can obtain a more holistic view of a system by combining multiple 'omics analyses. In this context, this talk will present the mixOmics R package that proposes multivariate statistical methods to explore and integrate 'omics data sets with a specific focus on variable selection and visualisation.

10h30-10h45 : Pause

10h45 – 11h45 : "KERNEL METHODS IN SYSTEMS BIOLOGY"

Jérôme Mariette, Unité de Mathématiques et Informatique Appliquées de Toulouse

The development of high-throughput sequencing technologies has lead to produce high dimensional heterogeneous datasets at different living scales. To process such data, integrative methods have been shown to be relevant, but still remain challenging.
Here, we propose contributions useful to simultaneously explore heterogeneous multi-omics datasets and to select relevant features in a dataset. To tackle these problems, kernels and kernel methods represent a natural framework because they allow to handle the own nature of each datasets while permitting their combination. In a first part, I will present a multiple kernel framework that allows to integrate multiple datasets of various types into a single exploratory analysis. In a second part, I will introduce feature selection methods that are adapted to the kernel framework and go beyond the well established work in supervised learning by addressing the more difficult tasks of unsupervised learning and kernel output learning.
The proposed methods efficiency is highlighted in the domain of microbial ecology : eight TARA oceans datasets are integrated and analysed.

Nous remercions l'Université de Lille qui a permis l'organisation de ces matinées en visio-conférences grâce à la mise à disposition de Zoom.