

Orateur : LABORY Justine

Affiliation : Maison de la Modélisation, de la Simulation et des Interactions (MSI), Medical Data Laboratory (MDLab), Institute for Research on Cancer and Aging, Nice (IRCAN)

Titre de l'exposé : ABEILLE: a novel method for ABerrant Ex-pression Identification empLoying machine Learning from RNA-sequencing data

Résumé :

Current advances in omics technologies are paving the diagnosis of rare diseases proposing as a complementary assay to identify the responsible gene. The use of transcriptomic data to identify aberrant gene expression (AGE) have demonstrated to yield potential pathogenic events. However popular approaches for AGE identification are limited by the use of statistical tests that imply the choice of arbitrary cut-off for significance assessment and the availability of several replicates not always possible in clinical contexts. Machine learning methods via neural networks including autoencoders (AEs) or variational autoencoders (VAEs) have shown promising performances in medical fields.

Here, we describe ABEILLE, (ABerrant Expression Identification empLoying machine LEarning from sequencing data), a novel method for the identification of AGE from RNA-seq data without the need of replicates or a control group, using a flexible model obtained after testing several parameters. ABEILLE combines the use of a VAE, able to model any data without specific assumptions on their distribution, and a decision tree to classify genes as AGE or non-AGE. An anomaly score is associated to each AGE in order to stratify them by severity of aberration.

We compare ABEILLE performances to the state-of-the-art alternatives by using semi-synthetic data and a real dataset demonstrating the importance of the flexibility of the VAE configuration to identify potential pathogenic candidates.

Justine Labory, Gwendal Le Bideau, David Pratella, Jean-Elisée Yao, Samira Ait-El-Mkadem Saadi, Sylvie Bannwarth, Loubna El-Hami, Véronique Paquis-Fluckinger, Silvia Bottini, ABEILLE: a novel method for ABerrant Expression Identification empLoying machine Learning from RNA-sequencing data, *Bioinformatics*, 2022;, btac603, <https://doi.org/10.1093/bioinformatics/btac603>
ABEILLE source code is freely available at : <https://github.com/UCA-MSI/ABEILLE>.