mixMint: A multivariate integrative approach to identify a reproducible biomarker signature across multiple experiments and platforms

Florian Rohart[*1]

[1]The University of Queensland Diamantina Institute (UQDI) – Translational Research Institute, QLD 4102, Australia, Australie

Résumé

The reproducibility of biomarker identification across transcriptomics independent studies is often limited by small sample size experiments. One solution is to increase statistical power by combining those studies in an integrative approach. In addition, the advantage is to enable data sharing across research groups and benchmark studies. However, such analysis is not straightforward due to the unwanted systematic variation arising from the use of different commercial platforms in different laboratories with different protocols.

We propose a novel multivariate integration method, MINT that accommodates for unwanted systematic variation, builds an accurate multivariate linear classifier based on a small subset of key discriminative biomarkers. We illustrate the benefits of combining transcriptomics data sets (microarray and RNA-sequencing) with MINT on two case studies and show that the gene signatures obtained are highly predictive, as validated on external studies, and are therefore highly reproducible. MINT compares favourably to two-steps batch effect removal and classification procedures, and provides insightful study-specific outputs to quality control each study to be integrated in the analysis.

The MINT algorithm is implemented as part of the mixOmics R package available on CRAN (http://cran.r-project.org/web/packages/mixOmics/, http://www.mixOmics.org/).

[*]Intervenant