Statistical tests for bivariate spatial association in multi-omics data

Stijn Hawinkel*[1,2], Wangjun Hu[3], Britta Velten[3,4], and Steven Maere[1,2]

[1]Department of Plant Biotechnology and Bioinformatics, Ghent University – Belgium
[2]VIB center for Plant Systems Biology – Belgium
[3]Centre for Organismal Studies – Germany
[4]Interdisciplinary Center for Scientific Computing – Germany

Abstract

Spatial biology has entered a new era of multimodal profiling, with multiple, high-dimensional omics types being measured on consecutive tissue slices, or co-profiled on the same slice. Interest then lies in statistical testing for spatial association between the features of the different modalities, to gain insight in the corresponding biological processes. One major challenge thereby is the multitude of bivariate combinations, leading to high computational demands. Another difficulty is the difference in image shape and spatial resolution between the modalities, which implies no one-to-one matching between the measurement spots of both modalities, even after alignment. As a result, common statistical measures such as joint distributions and correlations are not defined, and tests need to rely on spatial vicinity only. We argue that the existing bivariate association tests Moran's I, Geary's C and Lee's L test an inappropriate null hypothesis, which implies absence of spatial autocorrelation in any of the features, and demonstrate this point using real data reshuffling. Moreover, sample matching combined with off-the-shelf statistical tests, such as Pearson correlation, can be inefficient and yield inflated type I errors. As a solution, we adapt existing tests for the detection of spatially variable genes, Moran's I, the modified t-test, Gaussian processes and generalized additive models (splines), to the bivariate case with non-overlapping coordinate sets. We develop inference methods for single sections as well as for replicated experiments with multiple sections, and compare their performance in parametric and non-parametric simulations. Finally, we apply the newly developed methods to the integration of a spatial transcriptomics and metabolomics mouse dataset obtained through co-profiling. The full suite of tests is available as the R-package sbivar from GitHub.

Keywords: spatial, hypothesis test, bivariate, multiomics

*Speaker