

Distribution-free complex hypothesis testing for single-cell RNA-seq differential expression analysis

Marine Gauthier^{1,2,3}, Denis Agniel^{5,6}, Rodolphe Thiébaud^{1,2,3,4}, Véronique Godot^{3,7},
Boris P. Hejblum^{1,2,3}

¹ Univ. Bordeaux, Inserm, Bordeaux Population Health Research Center, UMR 1219, Bordeaux, France

² INRIA Bordeaux Sud Ouest, Talence, France

³ Vaccine Research Institute, Créteil, France

⁴ CHU, Bordeaux, France

⁵ Rand Corporation, Santa Monica (CA), USA

⁶ Harvard Medical School, Boston (MA), USA

⁷ Inserm, U955, Team 16, Univ. Paris-Est, Créteil, France

E-mail for correspondence: marine.gauthier@u-bordeaux.fr

Abstract: Single-cell RNA-seq (scRNA-seq) quantifies gene expression at the cell resolution. State-of-the-art methods for scRNA-seq Differential Expression Analysis (DEA) often rely on strong distributional assumptions that are difficult to verify in practice. Furthermore, while the increasing complexity of clinical and biological single-cell studies calls for greater tool versatility, the majority of existing methods only tackles the comparison between two conditions. We propose a novel, distribution-free, and flexible approach to DEA for single-cell RNA-seq data. This new method, called `ccdf`, tests the association of each gene expression with one or many variables of interest (that can be either continuous or discrete), while potentially adjusting on additional covariates. To test such complex hypotheses, `ccdf` uses a conditional independence test relying on the conditional cumulative distribution function, estimated through multiple regressions. `ccdf` includes an asymptotic test as well as a permutation test (when the number of observed cell is not sufficiently large). `ccdf` exhibits good statistical performance in various simulation scenarios considering complex experimental designs (*i.e.* beyond the two condition comparison), while retaining competitive performance with the state-of-the-art in a two condition benchmark.

Key words: single-cell; conditional cumulative distribution function; conditional independence test; differential expression analysis; distribution-free test