ClusterDE: a post-clustering differential expression method

Carson Zhang

LMU Munich

carson.zhang@campus.lmu.de

Abstract   In typical differential expression analysis, a clustering algorithm is applied to scRNA-seq data, and then a differential expression test is conducted in order to identify genes that are differentially expressed between the clusters. However, this procedure constitutes "double dipping", as it first clusters the data to identify cell types, and then uses those same clusters to identify cell-type marker genes. This leads to an inflated FDR for DE genes. Song et al. (2023) propose ClusterDE, a post-clustering DE method that controls the FDR of DE genes. ClusterDE generates a synthetic null dataset that preserves the structure of the real data, computes differences between this null dataset and the real data, then performs FDR control on the results. Simulations and real data analysis demonstrate that ClusterDE controls the FDR and identifies cell-type marker genes as top DE genes, successfully distinguishing them from housekeeping genes.

Table of contents

Introduction

Scientific writing is a crucial part of the research process, allowing researchers to share their findings with the wider scientific community. However, the process of typesetting scientific documents can often be a frustrating and time-consuming affair, particularly when using outdated tools such as LaTeX. Despite being over 30 years old, it remains a popular choice for scientific writing due to its power and flexibility. However, it also comes with a steep learning curve, complex syntax, and long compile times, leading to frustration and despair for many researchers. [1]

Paper overview

In this paper we introduce Typst, a new typesetting system designed to streamline the scientific writing process and provide researchers with a fast, efficient, and easy-to-use alternative to existing systems. Our goal is to shake up the status quo and offer researchers a better way to approach scientific writing.

By leveraging advanced algorithms and a user-friendly interface, Typst offers several advantages over existing typesetting systems, including faster document creation, simplified syntax, and increased ease-of-use.

To demonstrate the potential of Typst, we conducted a series of experiments comparing it to other popular typesetting systems, including LaTeX. Our findings suggest that Typst offers several benefits for scientific writing, particularly for novice users who may struggle with the complexities of LaTeX. Additionally, we demonstrate that Typst offers advanced features for experienced users, allowing for greater customization and flexibility in document creation.

Overall, we believe that Typst represents a significant step forward in the field of scientific writing and typesetting, providing researchers with a valuable tool to streamline their workflow and focus on what really matters: their research. In the following sections, we will introduce Typst in more detail and provide evidence for its superiority over other typesetting systems in a variety of scenarios.

Overview of differential expression methods

ClusterDE

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Vestibulum dapibus dictum leo, vel molestie felis hendrerit at. In diam nisl, rhoncus dapibus facilisis at, laoreet ac nunc. Sed vehicula metus id faucibus tempus. Etiam at laoreet orci, mattis efficitur sem. Aenean eu malesuada nisl. Mauris tempor pulvinar magna, eu tempus felis aliquet nec. In nec iaculis nisi. Fusce at arcu enim. In ipsum massa, pretium vitae pharetra in, vulputate vel felis.

$a + b = gamma$

Mauris egestas laoreet aliquam. Integer non convallis risus. Ut eu felis ut augue ultricies tempus. Integer posuere lacinia ex, ac sollicitudin tortor tempus id. Morbi dictum elit at orci accumsan, vitae volutpat dui scelerisque. Pellentesque vitae lectus id nisl posuere placerat pharetra nec ligula. Donec laoreet ligula hendrerit velit tempus porttitor. Vivamus eleifend imperdiet magna quis pellentesque. Pellentesque blandit congue nisi ac ultricies. Sed velit ante, pretium nec facilisis ac, ultricies vitae orci. In hac habitasse platea dictumst. Phasellus massa mauris, tincidunt sed facilisis eu, venenatis a leo.

Fusce id risus mauris. Morbi tincidunt interdum odio, a laoreet sem pulvinar condimentum. Integer suscipit lobortis quam vitae hendrerit. Duis nec convallis mi. Etiam cursus felis eget sem tincidunt, at vulputate urna mattis. Aenean nec libero condimentum, posuere tellus eget, ultrices dui. In quis magna ac metus facilisis porta. Integer et massa dignissim, tempus massa quis, feugiat ante. Aliquam a tortor ligula. Praesent aliquam, lorem id ullamcorper rutrum, nunc orci posuere

nulla, id fermentum mi erat a justo. Sed ut sapien vulputate, elementum ante ac, maximus lorem. Phasellus metus est, ultricies et ipsum ac, volutpat lobortis turpis. Vestibulum accumsan elementum vehicula.

Other differential expression methods

Data analysis

Simulation study

Bibliography

[1] R. Astley and L. Morris, "At-scale impact of the Net Wok: A culinarically holistic investigation of distributed dumplings", *Armenian Journal of Proceedings*, vol. 61, pp. 192–219, 2020.