

Using artificial intelligence for error correction in single cell analyses

January 24 2019

Modern technology makes it possible to sequence individual cells and to identify which genes are currently being expressed in each cell. These methods are sensitive and consequently error prone. Devices, environment and biology itself can be responsible for failures and differences between measurements. Researchers at Helmholtz Zentrum München joined forces with colleagues from the Technical University of Munich (TUM) and the British Wellcome Sanger Institute and have developed algorithms that make it possible to predict and correct such sources of error. The work was published in *Nature Methods* and *Nature Communications*.

A visionary project of enormous scope, the Human Cell Atlas aims to map out all the tissues of the human body at various time points with the goal of creating a reference database for the development of personalized medicine, i.e. the ability to distinguish healthy from diseased [cells](#). This is made possible by a technology known as [single-cell RNA sequencing](#), which helps researchers understand exactly which genes are switched on or off at any given moment in these tiny components of life. "From a methodological point of view, this represents an enormous leap forward. Previously, such data could only be obtained from large groups of cells because the measurements required so much RNA," Maren Büttner explains. "So the results were always only the average of all the cells used. Now we're able to get precise data for every single cell," says the doctoral student at the Institute of Computational Biology (ICB) of the Helmholtz Zentrum München.

The increased sensitivity of the technique, however, also means increased susceptibility to the batch effect. "The batch effect describes fluctuations between measurements that can occur, for example, if the temperature of the device deviates even slightly or the processing time of the cells changes," Maren Büttner explains. Although several models exist for the correction of these deviations, those methods are highly dependent on the actual magnitude of the effect. "We therefore developed a user-friendly, robust and sensitive measure called kBET that quantifies differences between experiments and therefore facilitates the comparison of different correction results," Büttner says.

Besides the batch effect, a phenomenon known as dropout events poses a major challenge in single-cell sequencing. "Let's say we sequence a cell and observe that a particular gene in the cell does not emit any signal at all," explains Dr. Dr. Fabian Theis, ICB Director and professor of Mathematical Modeling of Biological Systems at the TUM. "The underlying cause of this can be biological or technical in nature: either the gene is not being read by the sequencer because it is simply not expressed, or it was not detected for technical reasons," he explains.

To recognize these cases, bioinformaticians Gökçen Eraslan and Lukas Simon from Theis's group used a large number of sequences of many single cells and developed what is known as a [deep learning algorithm](#), i.e. artificial intelligence which simulates learning processes that occur in humans (neural networks).

Drawing on a new probabilistic model and comparing the original and reconstructed data, the algorithm determines whether the absence of a gene signal is due to a biological or technical failure. "This model even allows cell type-specific corrections to be determined without two different cell types becoming artificially similar," Fabian Theis says. "As one of the first deep learning methods in the field of single-cell genomics, the algorithm has the added benefit that it scales up well to

handle data sets containing millions of cells."

But there is one thing the method is not— and this is important to emphasize: "We're not developing software to smooth out results. Our chief goal is to identify and correct errors," Fabian Theis explains. "We're able to share these data, which are as accurate as possible, with our colleagues worldwide and compare our results with theirs"—for example, when the Helmholtz researchers contribute their algorithms and analyses to the Human Cell Atlas, because reliability and comparability of the data are of paramount importance.

More information: Maren Büttner et al. A test metric for assessing single-cell RNA-seq batch correction, *Nature Methods* (2018). [DOI: 10.1038/s41592-018-0254-1](https://doi.org/10.1038/s41592-018-0254-1)

Gökçen Eraslan et al. Single-cell RNA-seq denoising using a deep count autoencoder, *Nature Communications* (2019). [DOI: 10.1038/s41467-018-07931-2](https://doi.org/10.1038/s41467-018-07931-2)

Provided by Helmholtz Zentrum Muenchen

Citation: Using artificial intelligence for error correction in single cell analyses (2019, January 24) retrieved 27 April 2024 from <https://phys.org/news/2019-01-artificial-intelligence-error-cell-analyses.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.