As a consequence, scRNA-seq still requires a marriage of bioinformatics and wet-lab expertise. With this in mind, many academic centers set up hybrid workflows that incorporate user-friendly commercial algorithms and open-source software, which can be customized. Ivarsson notes that commercial package Qlucore aspires to offer a software for single-cell platforms that can adapt to diverse workflows. "Our vision is not to lock in users — rather the opposite, to make it easier to interact," he says. This is also true for commercially produced SeqGeq, which is designed to work with externally developed or custom-written R software packages. Dolomite's Fischer says that, in her experience, most researchers don't even think about the software aspect of the workflow until they begin wrestling with data — and at that point, flexibility and availability of support are the critical considerations.

Such interoperability is especially important for users looking to venture beyond routine applications like assessing differential gene expression. These cutting-edge applications remain works in progress and will almost certainly require open-source algorithms. "We're quickly moving past 'we did an scRNA-seq experiment and here are the cell types we see'," says Trapnell, whose group has developed computational tools for mapping the developmental trajectories of millions of individual cells in parallel. New tools are also enabling physical mapping of RNA-seq data in multicellular samples. For example, 10x Genomics' Visium Spatial Gene Expression Solution maps gene activity within tissue specimens, and although the company has developed software called Space Ranger to facilitate analysis, other algorithms will surely be needed to make the most of this still-novel experimental approach.

The ultimate challenge is integration. "The bottleneck for many researchers isn't necessarily the early stages of processing data, but the later, more in-depth analysis to integrate results," says Theis, who is Director of the Institute of Computational Biology at the Helmholtz Zentrum Munich. How to best combine data from many different studies using different samples — and, most likely, different experimental workflows — is still very much an unsolved problem. "Right now, it's sort of like a magic trick," says Kharchenko. "You put all your data in a pot and then say 'integrate!'" This problem becomes even more daunting when one begins to contemplate combining scRNA-seq data with other -omic data layers, such as chromatin accessibility, DNA methylation or protein expression.

The perfect software tool to deliver all answers is unlikely to emerge. "You can provide tools to make sure that you make no mistakes," says Corselli. "But ultimately it's still the responsibility of the scientist to look at the data and be rigorous." ❐

Michael Eisenstein
Philadelphia, PA, USA

Published online: 9 March 2020
https://doi.org/10.1038/s41587-020-0449-8

Acknowledgements
Additional reporting by Chris Lieu, Redwood City, CA, USA.

Around the world in a month

Credit: Map: © iStockphoto; Flags: pop_jop / DigitalVision Vectors / Getty

Published online: 9 March 2020
https://doi.org/10.1038/s41587-020-0451-1