New training platform for big data analysis

July 20 2018

Studies indicate that more than 95 percent of researchers in the life sciences are already working with big data or plan to do so—but more than 65 percent state that they only have minimal knowledge of bioinformatics and statistics. "Many are currently lacking training and instruction in this area—that's the real bottleneck that is making broader use of data science methods difficult," says bioinformatician Dr. Björn Grüning of the University of Freiburg. He is coordinating the Galaxy Europe project, which aims to remedy this with interactive, community-based and freely-available online tutorials on data analysis in the life sciences. The participating researchers have presented their project in the journal *Cell Systems*.

Galaxy, an open source platform enabling the analysis of big data for life sciences, can be accessed by scientists via an Internet browser. It requires no programming skills: all settings can be made using a graphic interface. In order to make it easier for other researchers to get started, the Galaxy team has set up a series of tutorials: using a "research story", based on real datasets, users learn step-by-step how they can apply the analytical tools of Galaxy to their work. In addition, all tutorials provide basic knowledge of bioinformatics and statistics. "We want to avoid researchers just using the software like a "Black Box" without understanding what methods these tools are applying and what they can do," explains Grüning. Another goal is to involve scientists around the world in a collaboration to constantly expand the range of tutorials on offer—a combination of the community approach and the Galaxy platform make this possible: more than 60 researchers from all round the world have already contributed to the training material.

After being initiated at Penn State University in the USA, Galaxy was further developed at the University of Freiburg in the Medical Epigenetics special research area and as part of the Deutsches Netzwerk für Bioinformatik-Infrastruktur (German Network for Bioinformatics Infrastructure, de.NBI). The project's European server, which is run by Björn Grüning's team from the work group of Prof. Dr. Rolf Backofen at the Institute of Computer Science, is situated in the IT Services Center at the University of Freiburg.

 More information: Bérénice Batut et al, Community-Driven Data Analysis Training for Biology, *Cell Systems* (2018). DOI: 10.1016/j.cels.2018.05.012

Provided by University of Freiburg