

nature biotechnology

So long to the silos

***Nature Biotechnology* will now publish all computational biology papers in its research section, a decision that reflects an ongoing need to break down silos surrounding mathematical, computational and basic biology.**

Since this journal launched its computational biology section in 2004, life science has changed. It is now over 15 years since the first human genome sequence was ‘completed’, and technological advances across all areas of biology have seen the breadth and complexity of data increase to unprecedented scales. Large, integrative data initiatives increasingly dominate not only the research sector but also biomedicine, with the worlds of bench and clinic converging to an ever greater extent. For these big data approaches to thrive, however, historical barriers dividing research groups and institutions need to be broken down and a new era of open, collaborative and data-driven science ushered in. One part of that process is to cease thinking of computational biology as a subdiscipline of biology.

Research silos continue to hamper biological research. Historical academic departmental divisions mean that mingling with colleagues from other fields is usually accidental and all too rare. Competition for funding and academic publications encourages an inwardly focused scientific culture, where research questions rarely extend beyond the expertise of the group, and data are often sequestered from the public domain. Those holding research data often regard them as proprietary and may even stigmatize investigators seeking to reanalyze data as “research parasites.”

It is clear that this way of doing research is insufficient and ill-fit for twenty-first century biology.

Support for collaborative, big data projects extends far beyond the research sector. This year, President Barack Obama’s precision medicine initiative is injecting \$215 million to generate genetic, metabolic, environmental and lifestyle data from ~1 million people. The UK Biobank, set up by several medical charities and UK government organizations, is collecting genetic data, phenotypic information from activity monitors and questionnaires, biochemistry profiles, magnetic resonance imaging (MRI) scans and DEXA bone measurements in up to half a million people. Commercial companies, such as 23andMe, are currently receiving 2 million new data points a week from their customers, which is being shared with the wider research community.

These collaborative projects reflect biology’s move to big data generation. But they don’t answer the larger question of how the research community outside these proscribed, well-funded collaborations can increase their involvement in big data research. There are as yet too few groups focused on developing efficient and robust statistical algorithms and dedicated high-performance computing systems, and too few tools available to researchers interested in tackling big data. The scale of the technical challenges of storing, accessing and analyzing the ever-increasing amounts of available data is such that researchers and their institutions need to find new ways of working together; indeed, academic groups and institutions that refuse to find a way to share their ideas and resources risk falling behind.

That is not to say the need has gone unnoticed. In 2012, the US National Institutes of Health launched the Big Data to Knowledge initiative, which

tasked 13 US centers with developing new analysis approaches, methods and software tools to do exactly what the name suggests—translate data into knowledge. The US Defense Advanced Research Projects Agency has also launched its ‘big mechanism’ initiative (<http://www.darpa.mil/program/big-mechanism>). Three years ago, the UK’s Medical Research Council, together with other research councils, charities and health departments, invested ~£90 (\$128) million into improving the capability, capacity and infrastructure of medical informatics over a five-year period. This money has been put towards the establishment of the Farr Institute, whose focus is on developing methods for sharing, integrating and analyzing diverse datasets with currently unachievable speed and scale.

Nature Biotechnology created the Computational Biology section in the early noughties to advertise the journal’s interest in this area and to encourage an exchange of ideas between experimental biologists and computational scientists. In those days, analytical papers often encountered resistance from biologists during peer review, who insisted that computational data be supplemented with wet lab experiments—experiments often beyond the research capabilities of most computational groups. This was one reason we adopted the Analysis format, which does not require extensive experimental validation. We also used the section to showcase Primer articles intended as educational tools for readers describing key approaches to computational analysis, each comprising an introduction together with a worked example in an area of biology.

Going forward, we will wholeheartedly champion the work of those seeking to develop computational analysis approaches. As technologies improve and science becomes more collaborative and focused on high-level data integration, the analytical challenges are mounting. New methods and standardized protocols will be needed to bring together data generated at different times, by different technologies. Effective data sharing will go hand in hand with the development of massive data repositories that enable data to be added and downloaded from anywhere in the world. And developing accurate computationally efficient algorithms capable of modeling biological complexity will remain a top priority.

The time has come to break down the silos within our own journal. Criteria for establishing what is, and what isn’t, a computational biology paper are increasingly difficult to define, and our readers have embraced the importance of computation in nearly all corners of biological research. As a result, we will no longer publish computational analysis approaches in a dedicated section; instead, they will be integrated into our research section.

To mark the occasion, next month we will make a selection of our computational biology content openly available to readers. As we look forward, we envisage a time when the majority of the papers we publish will incorporate computation. Computational biology will become biology.