COMMENT

ILLUSTRATION BY PAUL BLOW

Reward bioinformaticians

Biological data will continue to pile up unless those who analyse them are recognized as creative collaborators in need of career paths, says **Jeffrey Chang**.

The US Precision Medicine Initiative, announced in January, relies on bioinformatics. The US$215-million project calls for collecting medical, physiological and genomic data from more than one million people in the United States, and aims to find patterns across individuals to improve health care. It does not address a worsening deficiency in the scientific community: biological data are accumulating faster than people's capacity to analyse them. For example, Sendurai Mani, a pioneer in cancer metastasis and genomics at the University of Texas MD Anderson Cancer Center, laments that he is constantly searching for bioinformaticians — those who analyse such data.
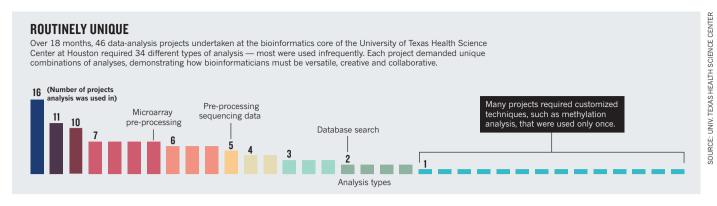
One explanation for the shortage seems obvious: there are not enough bioinformaticians. If so, the solution seems straightforward: train and hire more. As someone tasked with addressing this need at my institute, the University of Texas Health Science Center at Houston, I can vouch that the situation is not so simple.

The scientific community has failed to craft attractive career paths for those who do the analyses it increasingly requires. Institutions and funding bodies must carve out a viable place for bioinformaticians who focus on collaborations, and reward them for their abilities to navigate the myriad demands of multidisciplinary projects.

Biologists are increasingly finding that questions that are initially based on a single protein or gene quickly expand to require large-scale experiments. To support them,

dozens of institutions have set up centralized bioinformatics facilities. Most were established in the past decade[1]; what many consider the first bioinformatics 'core' was set up about 20 years ago by Fran Lewitter at the Whitehead Institute in Cambridge, Massachusetts.

Until the past few years, my institute had limited bioinformatics expertise. In fact, grant reviewers deemed some applicants' data-analysis plans unlikely to succeed because of this. In part to address this, the university hired more faculty members, including myself, in 2010. We established our bioinformatics service centre in 2012.

To give greater support to researchers, our centre set out to develop a series of standardized services. We documented the ▶

ROUTINELY UNIQUE

Over 18 months, 46 data-analysis projects undertaken at the bioinformatics core of the University of Texas Health Science Center at Houston required 34 different types of analysis — most were used infrequently. Each project demanded unique combinations of analyses, demonstrating how bioinformaticians must be versatile, creative and collaborative.

16 (Number of projects analysis was used in)

Microarray pre-processing

Pre-processing sequencing data

Database search

Many projects required customized techniques, such as methylation analysis, that were used only once.

Analysis types

projects that we took on over 18 months. Forty-six of them required 151 data-analysis tasks. No project was identical, and we were surprised at how common one-off requests were (see 'Routinely unique'). There were a few routine procedures that many people wanted, such as finding genes expressed in a disease. But 79% of techniques applied to fewer than 20% of the projects. In other words, most researchers came to the bioinformatics core seeking customized analysis, not a standardized package.

This experience is not unique to our centre. "There are no 'one-size-fits-all' analyses," says Ian Korf, interim director of the Bioinformatics Core at the University of California, Davis.

Another issue is that projects usually become more complex as they go along. Often, an analysis addresses only part of a question and requires follow-up work. For example, when we unexpectedly found a weak correlation between a protein receptor and a signalling response, getting a more-robust answer required refining our analysis to accommodate distinct signalling pathways orchestrated by the protein. An expansion typically doubled the time spent on a project; in one case it increased the time fivefold.

Service cores such as ours are doing research. Our centre is more a partner than a routine service provider — the input affects the success of the project and shapes broader questions. As Korf puts it, "Our service isn't just doing. It's also part of the thinking."

PURE OR APPLIED

In our service core, surprisingly, only 5% of the time was spent on pure bioinformatics — that is, developing new algorithms that merit their own publications. Although we put together tools and analysis pipelines in creative ways, nearly all of our time was spent on applied bioinformatics.

The skill set required for each is different. Simon Andrews, head of bioinformatics at the Babraham Institute in Cambridge, UK, says that the most important factor for applied projects is interacting with biologists, "to know what goes on in the lab, and know what experiments are easy or hard."

Yet applied bioinformaticians do not lead their own research projects, so they do not slot cleanly into an independent faculty position, despite their PhD-level training. Service-core jobs are not seen to be as prestigious as faculty posts and offer fewer opportunities for advancement and leadership. Cores are not currently well placed to recruit researchers who have the necessary drive and expertise. Michael Rebhan, who leads a bioinformatics group at the Novartis Institutes for Biomedical Research in Basel, Switzerland, recalls his trouble setting up a helpdesk-type service. "It didn't work out and was difficult to attract really good scientists."

At the Bioinformatics Shared Resource at the University of Utah, co-director David Nix has found a solution: a separate career ladder for staff in his core, with titles and salary bands attached. Still, the potential for advancement is limited. The resource relies on people being determined to help others at the expense of professional prestige.

I have experienced these tensions myself. Although I do methodological research in my own lab, I spend increasingly more time on applying and adapting algorithms in the service core. This creates a richer research environment that boosts the institution's prestige. However, I am mainly evaluated on the success of my individual research programme. There is a need at my institute to engage in collaborative work, but this output is not captured by the standard metrics of achievement.

The research system does not recognize bioinformaticians for doing what the scientific community needs most. "People realize the importance, but currently there are no real solutions," says Xiaole Liu, a bioinformatician at the Dana-Farber Cancer Institute in Boston, Massachusetts, and at Tongji University in Shanghai, China. This is why it can take more than six months to fill positions at a core, why many of biology's brightest are leaving science for technology companies, and why conventional biologists wait nine months to get help to dissect their data.

GAP FILLING

Relatively little effort is needed to help train, recognize, reward and retain applied bioinformaticians. Physics, with its long history of massive, highly integrative projects has begun to sort out career paths, authorship norms and due credit for those in purely collaborative roles — most notably in beam-line service provision. The main barriers to doing so in bioinformatics are inertia and a lack of imagination. Organizations such as the International Society for Computational Biology and national funding bodies should take up the cause. Biologists should be vocal about the value their core collaborators bring.

In research institutes, there should be a formal career track optimized for applied bioinformatics. Like in tenure-track positions, promotions should be based on a body of work. A successful bioinformatician is one who contributes to successful collaborations, so evaluation schemes must recognize that applied bioinformaticians may never be corresponding or lead authors, but they are still accomplished scientists. Evaluation criteria should also consider whether a bioinformatician is adaptable, as shown by projects across diverse authors and specialities.

Granting agencies should develop principles to evaluate collaborative staff. When researchers apply for grants to cover our services, they want to name core-centre personnel who have published in the area of analysis. Reviewers should instead judge bioinformatics collaborators by their prior success in working with biologists. Naming these people on grant applications would give them deeper motivation for projects. Finally, there need to be more opportunities for biologists themselves to learn bioinformatics skills[2].

These steps should increase the pool of talented people who see applied bioinformatics as a promising career. Otherwise, research will stall. ■

Jeffrey Chang *is co-director of the Bioinformatics Service Center and assistant professor of integrative biology and pharmacology at the University of Texas Health Science Center at Houston, Texas, USA.*
e-mail: jeffrey.t.chang@uth.tmc.edu

1. Lewitter, F., Rebhan, M., Richter, B. & Sexton, D. *PLoS Comput. Biol.* **5,** e1000372 (2009).
2. Lewitter, F. & Rebhan, M. *PLoS Comput. Biol.* **5,** e1000368 (2009).