

The boom in bioinformatics

Washington. Employment opportunities in the field of bioinformatics as it relates to genome research have never looked better. A lack of innovative new products in the traditional drug pipeline has sparked considerable interest by many pharmaceutical companies in the new tier of gene-hunting companies, all of whom have made a significant investment in this area. The acute shortage of qualified people largely results from the newness of this field and from a scarcity of formal training programmes that integrate the disciplines of biology and computer science. At least for the foreseeable future, therefore, this is one field where the demand for people with these dual skills will continue to outstrip the supply.

This shortage of qualified people can leave prospective employers in a quandary as to whether to employ a computer scientist who seems fascinated by the biology or a molecular biologist who has been bitten by the computer bug. Mistakes, if they are made, can prove costly. "If the computer system is not done right, at some point that becomes a bottleneck to the overall research," says Tom Slezak, a principal investigator in the informatics core at the Lawrence Livermore National Laboratory Human Genome Center, Livermore, California. "People have to be able to shoot from the hip and hit more often than they miss," he says.

So what are the right set of skills? According to Slezak: "this is a UNIX/C/X-Windows/database type of world."

Training programmes

There is still a "glaring need for training programmes and training opportunities where people at the masters and PhD level would have had formal training in both areas", says Kenneth H. Fasman, informatics director of the Genome Data Base at Johns Hopkins University, Baltimore, Maryland. GDB, as it is called, is a centralized repository for human gene mapping data. "Most of these academic programmes, where they exist at all, are still steaming fresh out of the oven," he says.

One such training programme, run by the University of Pennsylvania, is attempting to bridge that gap and provide researchers with a solid grounding in both the computational side and the molecular biology. The Penn Computational Biology Training Program is only about six months old and offers interdisciplinary training at both the graduate and post-doctoral level with funding from the US National Science Foundation. Funding is available for four graduate students and two post-docs at steady state, says Chris Overton, research associate professor in depart-

To take full advantage of the vast wealth of genetic data emanating from the US government-funded Human Genome Project, and other independent efforts, new software tools and databases must be developed for the collection, ordering and interpretation of the data, so that it can then be used for product development.

ments of genetics and of computer and information sciences, and one of 21 faculty taking part in the programme. In addition, both Merck and SmithKline Beecham recognize the need for qualified people with these dual skills and have agreed to provide support for the training of one post-doc each.

Graduates enter the programme through the biology, genetics or computer and information science departments, but are then given additional training in the complementary discipline. Particular areas of interest include physical and genetic mapping, biological databases, multiple sequence alignment, phylogeny construction, sequence search and analysis, statistical methods, and discrete algorithms and combinatorial optimization in biology. Training will also be offered at the post-doctoral level for researchers with interdisciplinary research interests. (For more information contact: <http://www.cbil.upenn.edu/~dsearls/NSF.html>)

Opportunity knocks

As a discipline, bioinformatics is critical to the success of the smaller gene-hunting companies like InCyte Pharmaceuticals (Palo Alto, California) and Human Genome Sciences (HGS) (Gaithersburg, Maryland), the commercial arm of a partnership that also includes The Institute for Genomic Research. Both of these companies operate high-throughput cDNA sequencing efforts. Companies like Myriad Genetics (Salt Lake City, Utah), who last year isolated the elusive *BRCA1* breast and ovarian cancer gene, Mercator Genetics (Menlo Park, California), Millennium Pharmaceuticals (Cambridge, Massachusetts) and Sequana Therapeutics (La Jolla, California) are taking a different approach. They, instead, are hoping to identify human genes associated with a specific disease by looking at large numbers of families in which that disease is particularly prevalent.

Carlos Zamudio, director of bioinformatics and engineering at Sequana, which is focusing on some of the more complex diseases like asthma, diabetes, obesity and osteoporosis, says Sequana has made a heavy investment in bioinformatics. This area now makes up roughly

one-third of its research effort but Zamudio says this could grow to 50 per cent as the company moves from finding genes to characterizing genes. Ultimately that will mean training people in-house, he says.

Although slower off the mark, many of the larger pharmaceutical companies are now seeing real value in gene mapping and sequence data. They are either showering money around among the smaller gene-hunting companies, or are setting up bioinformatics advisory boards and attempting to lure experts away from academia. One of the first to jump on the human genome bandwagon was SmithKline Beecham who, in 1993, made a \$125-million investment in HGS. Others have subsequently stepped up to the plate, including Merck (which is funding a gene sequencing effort at Washington University), Pfizer (which has linked up with InCyte) and Hoffmann-La Roche (which has done a deal with Millennium).

Salaries

The job opportunities for someone with the right skills may look rosy, but the money isn't bad either. The reason: simple supply-and-demand economics. Someone with a masters in computer science can typically command more than an equivalent-level molecular biologist, and probably more than someone at the PhD level. As Slezak puts it: "Good computer people cost a lot more than good biologists, bad computer people cost more than good biologists." Zamudio says he has employed some people where the supply is so low that basically they can command their own salary. "Mostly, they're not experienced enough to know how far they can take it," he quips.

Networking

Several upcoming meetings might be of interest: (1) The 4th International Conference on Bioinformatics & Genome Research, San Francisco, California, USA, 5-7 June '95, sponsored by Cambridge Healthtech Institute (tel. +1 617-487-7989 or <http://id.wing.net/~chi/homepg.html> via the WWW); and (2) two back-to-back meetings in Cambridge, UK — Intelligent Systems in Molecular Biology, 16-19 July '95 (contact Dominic Clark at ICRF, London, dac@bison.lif.icrf.uk or <ftp://ftp.icrf.net.uk/icrf-public/ismb/ismb95.html> via WWW) and the Second Meeting on the Interconnection of Molecular Biology Databases, 20-22 July '95 (contact Peter Karp at the Stanford Research Institute, California, pkarp@ai.sri.com or <http://www.ai.sri.com/people/pkarp/mimbd.html> via WWW).

Diane Gershon