MINING FOR BIOINFORMATICS JOBS

The growing number of courses and experienced workers in bioinformatics means that the recruitment market has become tougher for new graduates

The publication of the draft sequence of the human genome in 2001 was not only a monumental achievement in science but also signalled the future of bioinformatics. Although bioinformatics had been growing in the mid-1990s, the post-genomic revolution lit the blue touch paper for data-driven biomedical research. With information on gene expression, protein structure and function, and disease susceptibility being applied to the genome sequence, pharmaceutical companies needed the right tools and expertise to make sense of the deluge of data spewing out of laboratories.

Early on, recruited staff usually fell into two groups: biologists who had some computer skills or computer scientists who had some knowledge of biology. In general, companies tended to favour the former. Bioinformaticians need to understand computer languages, but a fundamental and thorough understanding of biology is crucial in making sense of biological data and extracting the meaningful information (BOX 1).

Understandably, there was a dearth of hybrid scientists who were competent in biology and computer science, so those who had skills in both areas had the enviable luxury of being able to choose positions and salaries at will. The emergence of several M.Sc. courses in bioinformatics, chemoinformatics and more IT-focused conversion courses after a scientific undergraduate degree (BOX 2), in response to the lack of an appropriate background in the field, helped to populate this area.

But bioinformatics has become sensitive to market trends. In recent years, companies have come and gone, such as Oakland, California-based Double Twist in 2002. Many leaders in the field, such as Celera and Incyte, have changed their business models from pure informatics-based approaches to applied approaches.

Recruiters at the large pharmaceutical companies say that this is making it a tougher environment for bioinformatics neophytes. More people in general, but also more people with relevant experience, are applying for bioinformatics positions. Companies would prefer to recruit people who have used specific algorithms, rather than those who have only heard of them.

So, neophytes can improve their chances of success if they show they can apply their knowledge to the field. Placements in industry and/or institutes, such as the European Bioinformatics Institute, will look attractive on your CV. Some companies recommend that if you really want a career in bioinformatics, doing a Ph.D. on a relevant project will provide you with experience, and having high-quality publications will strengthen your CV. Another way of gaining experience is to analyse the many data sets that are available on the Internet, such as gene-expression data from microarrays.

Bioinformatics is still a rapidly growing field. Large pharmaceutical companies are relying on informatics to improve efficiency, to allow more rapid decision making around existing projects and to more readily identify new ideas and targets. New areas, such as systems biology, will, if embraced by industry as anticipated, also require the skills of bioinformaticians. Mining for the right background and experience will increase your chances of joining this hottest of fields.

Simon Frantz

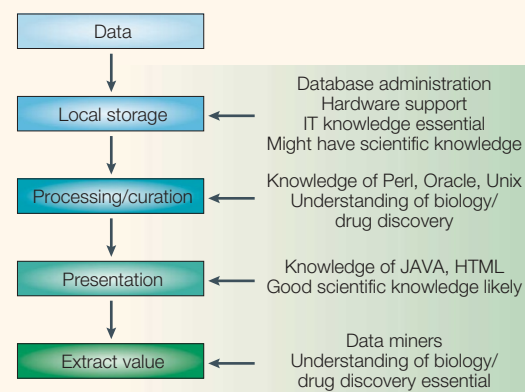**Box 1 | The bioinformatics pathway and skills required**

Database administration. Scaling and rating of raw data and diversity of data types requires dedicated database designers and administrators to manage the flow and storage of information efficiently.

Data processing and curation. Processing and curation of raw data (especially data derived from genomics) is necessary to link efficiently across other disciplines and allow the data to be accessed easily.

Data presentation. Developing mechanisms to view and present the useful data to scientists from a range of disciplines.

Data mining. The ultimate aim of informatics is to capture, curate and present the data in a format that will allow scientists to extract value from the underlying information.

Data → Local storage → Processing/curation → Presentation → Extract value

- Local storage: Database administration / Hardware support / IT knowledge essential / Might have scientific knowledge
- Processing/curation: Knowledge of Perl, Oracle, Unix / Understanding of biology/drug discovery
- Presentation: Knowledge of JAVA, HTML / Good scientific knowledge likely
- Extract value: Data miners / Understanding of biology/drug discovery essential

Adapted from Lanfear, J. Dealing with the data deluge. Nature Rev. Drug Discov. *1, 479 (2002)*

Box 2 | Useful links

Simon Frantz, Associate Editor (News), Nature Reviews Drug Discovery
e-mail: s.frantz@nature.com