



An Explosion of Bioinformatics Careers

Big data is everywhere, and its influence and practical omnipresence across multiple industries will just continue to grow. For life scientists with expertise and an interest in bioinformatics, computer science, statistics, and related skill sets, the job outlook couldn't be rosier. Big pharma, biotech, and software companies are clamoring to hire professionals with experience in bioinformatics and the identification, compilation, analysis, and visualization of huge amounts of biological and health care information. With the rapid development of new tools to make sense of life science research and outcomes, spurred by innovative research in bioinformatics itself, scientists who are entranced by data can pursue more career options than ever before. **By Alaina G. Levine**

Today's bioinformaticists are in for a real treat. With a seemingly endless stream of biological data being generated across sectors, there is high demand for talented, experienced professionals at the crossroads of biology, statistics, and computer science. Scientists who can analyze large amounts of information and present it in a clear manner to decision makers are finding the sky is the limit in terms of jobs and career pathways, especially in the big pharma and biotech sectors.

"It's a fun place to be and an exciting time to be in big data," remarks **Sriram Mohan**, professor of computer and software

engineering at Rose-Hulman Institute of Technology, who is spending his sabbatical developing bioinformatics software for Avalon Consulting, a data management firm.

And what an immense amount of data it is, due in part to a paradigm shift in the field, from data generation to data analysis, says **W. Jim Zheng**, associate professor in the School of Biomedical Informatics at The University of Texas Health Science Center at Houston. Now, with so much data being produced because of easier and more cost-effective tools, there is an even greater need for specialists who can make sense of the mountains of information in such a way that is meaningful for scientists and clinicians, and ultimately beneficial to customers and patients.

The increase in job opportunities is also being driven by a change in how bioinformatics is perceived in industry and academia. Previously, "scientists and companies used to look at bioinformatics as a tool," says **Wim Van Criekinge**, a professor of bioinformatics at Ghent University in Belgium and chief scientific officer at MDxHealth, a company developing epigenetics-based cancer diagnostics. Bioinformaticists would be called upon to answer a question about data; their role was to run an algorithm on a database that provided that answer. "But the subject has evolved from a service, like histology, to its own research arena.... Bioinformaticists are now the motor of the innovation," he adds. They not only answer the data inquiries, but also, more importantly, determine what questions need to be asked in the first place.

As a result, "there are many opportunities for scientists to pursue a bioinformatics/big data career in the biotech/big pharma industry at the moment," notes **Jared Kaleck**, senior director of computational chemistry/biology and formulation development at executive search firm Klein Hersh International.

Discovering Where the Careers Are

To begin to recognize where the jobs are and how the career tracks are accessed, it helps to understand how bioinformaticist positions are organized within different firms. In pharmaceutical and larger biotech companies, big data scientists may find themselves working in one of three different types of organizational structures. In one, all of the big data scientists and bioinformaticists work out of a central core. This large team could be concentrated in research and development (R&D) or information technology (IT) departments, and the scientists work almost like consultants on projects throughout the company, and are lent out as needed.

In a second model, bioinformaticist positions are decentralized, and located within different therapeutic areas. For example, at Johnson & Johnson (J&J), **Patrick Ryan** leads the epidemiology analytics group. As a clinical informaticist, he develops statistical methods to analyze "observational databases," such as electronic health records, to map disease patterns in order to better understand "the real-world effect of our medicines, and to develop safety protocols *continued*>

Upcoming Features

Postdocs: Broadening Your Skills —Aug. 22 ■ *Faculty: Professionalism 1*—Sept. 12 ■ *Faculty: Professionalism 2*—Oct. 3



“The deeper you understand the biology, the better you do your job in this area.”
—W. Jim Zheng

and mitigate risk for the patients,” he explains. His team is part of an overall epidemiology department which reports to the chief medical officer of J&J. But he notes that the company also has a robust informatics and IT division, whose mission is to “provide technical perspective on how to manage and analyze data.”

The third kind of organizational structure found in big pharma is a hybrid of the other two.

Christian Reich, global head of discovery informatics at AstraZeneca, shares that his company currently follows this model, although he notes that the trend sees enterprises restructuring themselves to follow one of the other paradigms every

few years. His job entails overseeing a principal group of 25 specialists, but other informaticists are sprinkled throughout the company. Similarly at Pfizer, bioinformaticists are embedded in therapeutic units as well as core centers of excellence, says **Susan Stephens**, senior director of research and development business technology at Pfizer.

Genentech follows a similar mixed model, explains senior director of bioinformatics, **Robert Gentleman** (who is also a co-developer of the statistical computing and graphics programming language known as R). Bioinformaticists are organized in a core center, but they “integrate with different functional areas,” he says. “They are in one department, but day-to-day they work directly with disease area specialists.”

Depending upon where they are housed in their company, big data scientists can expect to have varying tasks. In R&D, bioinformaticists conduct research on new approaches to analyzing data and help design and possibly even build the analysis tools utilized by scientists throughout the company, says Reich. Here, the idea is to examine existing open-source algorithms to apply them in novel ways, or to create entirely new algorithms that rely heavily on mathematical and statistical expertise. “The goal is to put together a platform so the data analysis can be done easily, and return high-quality results,” he adds.

Bryn Roberts, global head of operations, which includes informatics for Pharma Research and Early Development at Roche, notes that in his company’s hybrid structure, informaticists and data scientists are involved in a wide range of activities. They develop and support software systems; they procure and make external scientific content available and actionable by scientists throughout the company; they implement and maintain workflow systems, such as e-lab notebooks, for both drug discovery and regulated functions; and they support and perform data, image, and text mining and

analysis to support scientific decision making.

At Genentech, bioinformatics scientists participate in all levels of the investigative process, from helping to design experiments that will find genetic markers for disease, to leveraging their bioinformatics skills to help find biomarkers that will aid in patient selection.

Elsewhere in big pharma, big data scientists may be tasked to investigate trends in diseases and drug development and discovery, which can involve collaboration with the marketing team. They may also provide quantitative support for business decisions, such as in which therapies firms should invest, says **George Telthorst**, director of the Center for the Business of Life Sciences at Indiana University.

Contract Research Organizations (CROs) also offer much for those interested in the big data profession. As **Dimitris Agrafiotis**, vice president of informatics and chief data scientist of Covance, one of the world’s largest CROs, attests, “CROs are becoming the R&D engine of the pharmaceutical industry.” Covance data scientists can expect to be involved in myriad projects across the entire drug development continuum, from biomarker discovery to preclinical development, clinical trials, health economics and outcomes research, and even marketing.

Beyond working in big pharma or biotech, there are also opportunities for data scientists in industry support companies, such as those that produce bioinformatics software and other data analysis tools. Furthermore, bioinformaticists are recruited by health insurance corporations and hospital management organizations.

Even academia has seen an uptick in bioinformatics career opportunities, as the discipline itself is expanding. Zheng recalls a time in the early days of genomics when doing big data research meant scientists had to leave the university lab and head to industry, but the tide has changed. Now, programs like the National Institutes of Health’s Big Data to Knowledge funds academic research in bioinformatics.

Pursuing Big Data Skills

Experts agree that the most successful bioinformaticists (and the ones who land the jobs) are those who have a multitude of skills. But the starting point is always knowledge of life sciences, also referred to as “domain expertise” in the industry. In fact, “the deeper you understand the biology, the better you do your job in this area,” says Zheng. Hiring managers specifically seek out scientists who have doctorates in various areas of life sciences, including molecular and cellular biology, chemistry, genetics, immunology, and epidemiology. At Genentech, Gentleman looks for candidates who possess expertise in the biology of a particular disease.

Additional critical skills are required for big data careers in industry, such as text mining, ontology, data integration, machine learning, and information architecture. A superior “quantitative ability,” as Gentleman calls it, which covers a range of statistical capabilities, is a must, as are overarching computing skills. These include core programming abilities, such as coding in C++ or Java, or scripting **continued>**

in PERL or Python, says Van Criekeing. It is vital to be able to navigate operating systems like UNIX and Linux as well as have knowledge of common tools such as Hadoop and NoSQL databases, adds Mohan. Experience in data visualization and building effective user interfaces, as well as familiarity with hardware, buttresses your marketability.

In addition to scientific problem-solving skills, bioinformaticists must have business proficiency. “Bioinformatics is a team sport,” says **Stephen Ruberg**, distinguished research fellow, advanced analytics of Eli Lilly and Company, and thus project management, teambuilding, and communications experience is a requirement. In fact, “being able to communicate with the other scientists is really the most important skill we look for,” says Gentleman.

Nimbleness and the ability to adapt quickly are also fundamental. “It’s a fast-paced environment,” says Van Criekeing. “You have to have a mindset of constantly using new tools, or you will become obsolete in two years.”

Landing Big Data Jobs

It would be ideal if companies could find candidates who have all of the above skill sets, but sources indicate that that is wishful thinking. More often than not, hiring decisions are made based on the immediate needs of the team, especially given their interdisciplinary nature. “We look for people whose expertise complements the existing group’s skills,” says Roberts. However, just because you lack a specified talent or interest area as noted in a recruitment ad, doesn’t mean you shouldn’t apply anyway. “We share CVs internally all the time,” says Stephens. So even if she can’t bring you in to her group at Pfizer, she may be able to find another team at the company for which you would be a good fit.



Vincent Fusaro

In some cases, companies are growing their own talent, as a result of the lack of large numbers of qualified, multi-skilled candidates. At Roche, “we offer continuous training in various areas and encourage our staff to attend conferences, publish, or pursue higher degrees,” says Roberts. Pfizer data scientists have myriad chances to pursue professional development, and are also granted time to

try out new techniques, says Stephens, something she refers to as “sandbox opportunities.”

Experience plays a major role in gaining access to jobs. Kaleck highly recommends doing an industrial postdoc or internship, but in absence of these, scientists might consider “bridge” programs, like the Insight Data Science Fellows Program. This fully supported, six-week training opportunity offers postdoctoral fellows the chance to work on real-world problems for the likes of Facebook and Microsoft. This appealed to **Vincent Fusaro**, whose Ph.D. is in bioinformatics. As a fellow, he gained expertise in databases, Python, machine learning, and data visualization, which helped him land a position as a self-described “data ninja” for Invitae, a genetic information company. Today he is responsible for software engineering, data analysis, and pipeline and product development, among other tasks.

The Expanding Big Data Universe

Data scientists can expect the field to change and evolve in novel ways in the near future. But the bottom line is that “companies are growing their bioinformatics,” says Kaleck. “There are 100% more job opportunities opening up in bioinformatics than ever before,” much of which is driven by an increase in venture capital investment.

Given that big data “is the hottest field on the planet,” says Agrafiotis, those who have the requisite skills and expertise often have their pick of opportunities. “I have to fight Google, Amazon, LinkedIn, and hedge funds to hire the top people. They are valuable in any industry.”

In particular, the future of big data in big pharma and biotech sectors is bright and exciting. “Bring your expertise to health care,” says Telthorst, “and you’ll know you’re going to make a difference, at the patient level and at the societal level.”

Alaina G. Levine is a science writer based in Tucson, AZ.

DOI: 10.1126/science.opms.r1400143

Featured Participants

AstraZeneca
www.astrazeneca.com

Avalon Consulting, LLC
www.avalonconsult.com

Covance
www.covance.com

Eli Lilly and Company
www.lilly.com

Genentech
www.genentech.com

Ghent University
www.ugent.be/en

Indiana University
www.indiana.edu

**Insight Data Science
Fellows Program**
www.insightdatascience.com

Invitae
www.invitae.com/en

Johnson & Johnson
www.jnj.com

Klein Hersh International
www.kleinhersh.com

MDxHealth
www.mdhealth.com

Pfizer
www.pfizer.com

Roche
www.roche.com

**Rose-Hulman Institute of
Technology**
www.rose-hulman.edu

**University of Texas Health
Sciences Center at Houston**
www.uth.edu