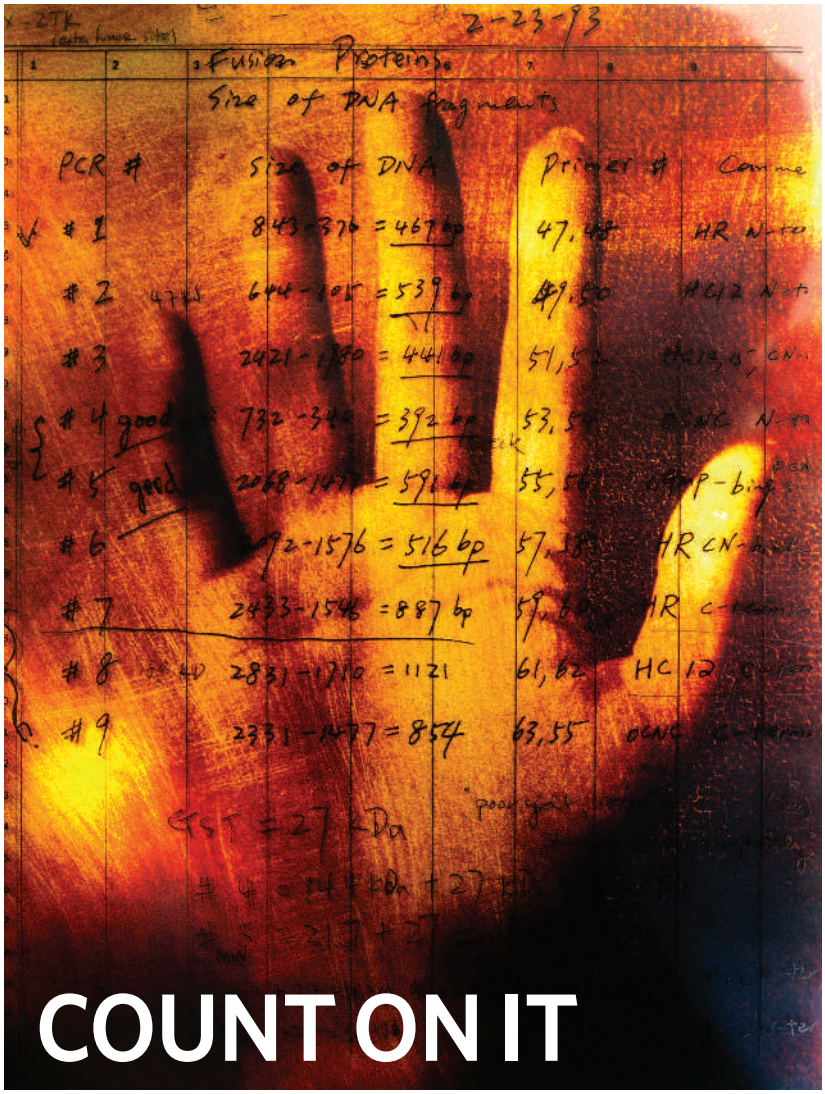
CORBIS



COUNT ON IT

It's a tough brief to fill, but geneticists with mathematical and computing skills find jobs easy to get. **Virginia Gewin** reports.

Andrew Skol is in an enviable position. Unlike many biology postdocs facing a possible lifetime of temporary jobs, Skol went straight to a faculty-level instructor position in the University of Chicago's Department of Medicine after completing his PhD. While others generate the gene-expression profiles, Skol mines the morass of data for biological meaning — a skill that's increasingly in high demand.

As the catalogue of genetic variations grows, so does the ability to identify genes that influence traits of medical interest. DNA microarrays enable biologists to explore the most subtle of human genetic variations, single-nucleotide polymorphisms (SNPs) — differences of only one base pair between the same DNA sequence in two humans. The decreasing cost and increased performance of microarrays and similar technology, which are now able to screen for up to 500,000 SNPs at once, mean that statisticians well versed in genetics are now in demand in both industry and academia.

The 2005 publication of the International HapMap Project — a library of human genetic variation — heralded the dawn of studies that aim to link disease susceptibility to complex genetic patterns. "The demand for people able to relate genetic variation to disease is exploding," says Lisa Brooks, director of the genetic variation programme at the National Human Genome Research Institute in Bethesda, Maryland.

Data provided by projects such as HapMap mean that scientists interested in linking disease to genotype will soon be able to dissect out the multiple causes of susceptibility to complex diseases such as heart disease and cancer. "The analysis focus of large-scale association studies will soon move beyond simple phenotypes, such as simply the presence or absence of cancer, to more complex phenotypes," says Christoph Lange, assistant professor of biostatistics at the Harvard School of Public Health in Boston, Massachusetts.

The increasing complexity of genetic-variation data is driving the need for more sophisticated statistical skills, says Bruce Weir, chair of the biostatistics department at the University of Washington in Seattle. It's not only gene-expression data that have to be taken account of. It is now possible, and desirable, to incorporate proteins and metabolites into the data mix.

Career opportunities in statistical genetics are plentiful, but training is not. Even for a good mathematician, gaining experience in the special blend of statistics, genetics and computer programming required might prove the biggest hurdle to landing a job in this field.

Training opportunities

The top statistical genetics departments in the United States focus primarily on training PhDs and postdocs, but some offer MS or certificate programmes emphasizing statistical genetics. The University of Alabama at Birmingham, for example, offers a certificate in statistical genetics only to those completing a PhD, while Washington University in St Louis, Missouri, offers a master's-level genetic epidemiology certificate with an MS degree. Whether you have an MS or a PhD could make a difference to the sort of job you can expect (see 'Look before you leap').

Despite few formal statistical genetics teaching programmes, there are signs that demand is beginning to fuel new training opportunities. Genetics

LOOK BEFORE YOU LEAP

Even when you get offered a job, there can be pitfalls ahead. Would you be happy simply serving the statistical needs of other researchers? "If that's not what you want and you are being hired as the only statistician in the place, you are in trouble," says David Allison, head of the statistical genetics section at the University of Alabama's biostatistics department.

Companies hire statistical geneticists trained to either MS or PhD level, but MS-level entrants can end up doing routine data analysis for methods that are being created by the PhDs in the development department. "The people who can design experiments to analyse data and interpret results and apply the findings are essential," says Meg Ehm, director of genetic epidemiology and analysis for pharmacogenetics at GlaxoSmithKline, based in London.

She says the company has struggled to find candidates with the appropriate statistics and genetics skills — sometimes taking up to 12 months to fill positions that require computer-programming skills as well. **V.G.**

departments are recruiting specialist instructors, anticipating that student interest won't be far behind. "We're scrambling to hire faculty to educate students about these opportunities," Weir says.

Even where opportunities exist, however, students don't seem to have caught on to the demand for statistical skills. "Often, people find us by chance, and only then get excited by the field," says Wolfgang Huber, statistics group leader at the European Bioinformatics Institute in Cambridge, UK. He is constantly on the lookout for people with a robust mathematical and biological background and the flexibility to learn about unfamiliar fields. He worries that the limited training in current methods that many people get will not serve their long-term career goals. "The field is moving so fast that narrow training might be obsolete five years from now."

Networking pays

In the small community of statistical genetics, the informal network is key. Skol was one of its beneficiaries. He did his PhD with biostatistician Michael Boehnke at the Center for Statistical Genetics at Michigan State University. As with all his students, Boehnke sent an e-mail to colleagues to let them know he had a student ready to leave. "There are many more positions than are even advertised in journals," says Boehnke.

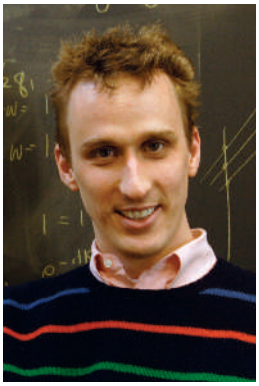
Informal networks are important in Europe as well. "Europe has yet to put in place a much-needed statistics epicentre," says Leena Peltonen, molecular geneticist at the University of Helsinki and the National Public Health Institute in Helsinki, Finland. A few groups offer training in statistical genetics. Françoise Clerget-Darpoux at the National Institute for Health and Medical Research in Paris, France, trains between three and eight master's-level students each year, and she also helps to teach short courses organized by the European Genetics Foundation (see 'Keeping ahead of the game').

In Asia, where demand and student interest are increasing, Japan is the predominant training ground. Institutes in Singapore, for example, rely on statistical geneticists trained in North America or Europe. This is because educational infrastructure has not embraced the team science approach needed for statistical genetics, says Edison Liu, executive-director of the Genome Institute of Singapore.

Genetics statisticians are also urgently needed in industry. Pharmaceutical companies are increasingly turning to pharmacogenomics — the linking of individual genotypes to disease susceptibility or responses to drugs — to salvage the struggling business of drug development (see *Nature* **436,** 746–747; 2005). To do this they need to sift through the roughly 10 million SNPs in the human population to find the few that can explain an individual's susceptibility to disease or good or bad responses to drugs. Some companies are investing heavily in pharmacogenomic approaches and often partner with biotechnology companies developing the complex statistical analyses to identify potential drug targets.

"All the easy things, the low-hanging fruit, have been found in drug development," says John Hooper, chief executive of Genizon Biosciences, a Quebec-based company specializing in gene mapping and genetic analysis. Genizon provides genomic analysis services to pharmaceutical giant Pfizer, and is looking to expand its team of around 30 statisticians and computer programmers who mine genomic data for drug targets. "Statistics is the engine of our business," says Hooper.

Companies are looking in particular for statistical geneticists with computer-programming skills. Entry-level bioinformaticians sometimes think they can analyse complex genetic data well enough, but it's not so simple, says Eric Schadt, senior scientific director for genetics at Rosetta Inpharmatics, in Seattle, Washington state, a subsidiary of Merck. "When you are working with 500,000 SNPs, intersecting with 40,000 traits in 2,000 patients — that leads to terabytes of result data," says Schadt, who is actively recruiting statistical geneticists for Rosetta. "Moving through this scale of data requires top-notch programming skills."



Christoph Lange (top) and Bruce Weir both see a bright future for mathematically minded geneticists.

Industry or academia?

Rosetta Inpharmatics evolved from a bioinformatics company into drug discovery, with an emphasis on methods-driven statistical development. Its high-profile publication record helps recruit scientists who may be unsure about leaving academia. For despite the roughly 30% pay hike that the pharmaceutical industry offers, as well as access to real-world data, many biostatisticians, like Skol, want to stay in academia.

Until pharmacogenomics proves its value, that may be a smart move. But, no matter how you slice the data, biology is becoming more analytical. "The days of molecular biologists as empirical scientists are past," says Hooper. "You can't eyeball 1.5 terabytes of data."

Skol agrees. He's got job security for the next few years, and he knows he is positioned for a future that will demand statistical skills. "As the lines between population genetics, bioinformatics, computational biology and statistics continue to blur, they will all merge into a broader field," he says. ■

Virginia Gewin is a freelance writer in Portland, Oregon.

KEEPING AHEAD OF THE GAME

For those already in the field, short courses are a good way to keep your skills up to date. The European Genetics Foundation holds courses on statistical topics and the UK-based Wellcome Trust runs week-long statistics courses combining lectures with hands-on data-analysis practicals.

The University of Washington in Seattle hosts two summer institute training programmes in statistical genetics — one in Seattle and one in Liège, Belgium. The genetic analysis workshop held every two years by the Southwest Foundation for Biomedical Research in San Antonio, Texas, is another way to gain access to the latest analysis techniques, and a good opportunity to network.

Of the 345 participants last year, 43% were graduate students and postdocs. "Anybody can come, but only if they've done something with the data sets distributed beforehand," says Jean MacCluer (pictured), statistical geneticist and co-organizer of the workshop.

Genetic Analysis Workshop
www.gaworkshop.org
Summer Institute in Statistical Genetics at the University of Washington
www.biostat.washington.edu/sisg07
European counterpart in Liège, Belgium
www.functionalgenomics.org.uk/sections/activitites/2007/Georges/info.htm
Wellcome Trust short course
www.wellcome.ac.uk/doc_wtx026851.html
European Genetics Foundation
www.eurogene.org/index.php