



Population geneticists are in short supply, as the need to translate large data sets into disease-susceptibility traits grows, says Eugene Russo

Genetics is in a state of the many relying on the few. The many are the scientists proficient at producing genomics data. The few are the experts capable of interpreting the relevance of those data across populations. As the quantity of genetics data increases, it is becoming apparent to the many that the few are not enough.

A select subset of that few, population geneticists (or, as they are sometimes now more accurately called, molecular population geneticists) — researchers with not only genetics, but also mathematics and statistical know-how — are in increasingly high demand. Their role is becoming more important as large disease-susceptibility gene projects such as the haplotype-map project (see next page) are being proposed.

“There are way more jobs than people to fill them,” says Michael Boehnke, professor of biostatistics and director of the Center for Statistical Genetics at the University of Michigan. Jobs for population and statistical geneticists are so common, says Boehnke, that qualified PhD students routinely skip

postdoctoral study to accept immediate faculty appointments. Meanwhile, molecular biologists often struggle to find appointments after doing several postdocs.

Boehnke — who himself started as a mathematician before parlaying his skills into a ‘biomathematics’ programme at the University of California, Los Angeles — is the sole recipient of the only National Human Genome Research Institute (NHGRI) training grant that specifically emphasizes population genetics. The programme currently funds ten pre- and postdoctoral students.

Boehnke, whose own centre has had problems finding permanent staff in these fields, recalls how he recently advertised the impending availability of two of his students to 15 colleagues at

companies and universities across the country. Within two days, he received eight e-mails expressing interest.

The current abundance of interest contrasts sharply with the field’s early days. Population geneticists were once limited largely to the realms of maize and fruit flies. “Population genetics has been this kind of esoteric field,” said Pennsylvania State University population geneticist Andrew Clark. “It got relegated to the back of journals like *Genetics*.”

Eric Lander, director of the Whitehead Institute in Cambridge, Massachusetts, agrees. But he senses a change. “Now we’re realizing that the human population, because it is so well-studied medically, can be a place for primary discovery.”

THEORETICAL HISTORY

Population-genetics pioneers such as Ronald Fisher, J. B. S. Haldane and Sewell Wright dealt primarily in theoretical terms, then tested their ideas with plant and animal models. But today’s population geneticist has something those experts of the 1920s and 1930s lacked: data. Loads of them.

The Human Genome Project and projects born of it continue to churn out vast amounts of sequence data. That data flood has also produced disease-susceptibility projects such as the search for genetic markers in its wake. The largest effort so far has been the search for single nucleotide polymorphisms (SNPs) — variations in only one base pair.

The public-private SNP Consortium and private companies such as Celera Genomics have produced several million SNPs in the past two years, in the hope of identifying significant markers of disease. But, while SNPs are relatively easy to find, their use in tracking down disease susceptibility may be limited.

A new project under discussion could make these data more useful by mapping sets of SNPs that are collectively inherited in close proximity to one another. But before tens of millions of dollars are devoted to mapping these haplotypes, academic and industrial scientists must first determine the project’s feasibility and utility.

Formal discussions began this summer. But the input of population geneticists may determine whether and how the project will proceed. Before they decide, they must discuss the prevalence of haplotype structures in the genome and the degree to which one

Potential partners

Early discussions have hinted at the possibility of a hap-map project public-private consortium, a model recently adopted by the SNP Consortium. Possible members include:

The SNP Consortium
<http://snp.cshl.org>
 Whitehead Institute
<http://www.wi.mit.edu>
 Sanger Center
<http://www.sanger.ac.uk>

US National Human Genome Research Institute
<http://www.nhgri.nih.gov>

Genome Canada
<http://www.genomecanada.ca>
 University of Tokyo’s Human Genome Center
<http://www.hgc.ims.u-tokyo.ac.jp>

Orchid Biosciences
<http://www.orchid.com>
 Sequenom Industrial Genomics
<http://www.sequenom.com>



population's haplotype map will apply to another. The need for population geneticists already extends beyond any prospective haplotype map project. These scientists are essential to everything from positional cloning of complex traits to tracking species' origins. The NHGRI is doing a few things to encourage more scientists to enter that field, in addition to the training-grant programme that helped Boehnke. Although it is not petitioning the population-genetics community as a whole to submit grants, the institute is encouraging scientists with mathematics, engineering and computer backgrounds who have biology skills to apply for career development grants. According to NHGRI deputy director Elke Jordan, in recent years the NHGRI, and other institutes including the National Institute of Mental Health and the National Institute of General Medical Sciences, have awarded many more research grants to population geneticists and the like. "Their role is changing, as they are now moving much more into mainstream biomedical research," says Jordan. "Molecular-biological and population-biological approaches are merging to complement and enhance each other."

ENCOURAGING CROSS-TALK

Mike Boyce-Jacino of Orchid BioSciences, a biotechnology firm based in Princeton, New Jersey, and a potential partner in the haplotype-map project (see "Potential partners", page 4) sees a growing need for population geneticists in both industry and academia who can determine the best populations to study, the necessary density of markers, and how to interpret results. "There is a shortage of people who can both develop the algorithms for analysing the data and also actually interpret the results," he said. "There aren't too many people that understand the link between the genetics side and statistics." A course or two in statistics will not suffice, says Andrew Clark. Nor will the packaged statistical software programs used by many molecular biologists. One obstacle to better understanding of complex human diseases is that relatively few medical geneticists know about evolution, and relatively few population geneticists understand the intricacies of disease epidemiology. "The problem is there's not enough cross-talk," said Columbia University genome

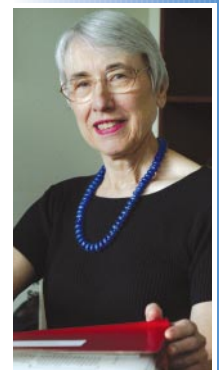
centre medical geneticist Joseph Terwilliger. In order to understand variation in populations, scientists must also understand history, evolution and selection. "And selection is not something that matters in the typical molecular biology lab day to day," says Lander. Eugene Russo is a freelance science writer based in Takoma Park, Maryland.

Obstacles for hap-map project

A haplotype-map project promises to speed up disease-gene discovery, but it faces major hurdles. SNPs are simple genetic variations loosely associated with disease. Haplotypes are collections of these variations that occur in close enough proximity to be inherited together. Rather than genotyping SNPs one at a time, the project would generate a haplotype map of the entire genome. Much of the SNP information within a haplotype is redundant; scientists may therefore require only a few SNPs within a haplotype in order to link it with a disease. Findings reported in *Nature* in May (Reich, D. E. *et al.*, *Nature* 411, 199-204; 2001) and in the October issue of *Nature Genetics* suggest that haplotype structures are consistent across the genome. But researchers must still identify the necessary density of markers and analytical tools and the nature of the population samples to be used. Some critics argue that doing the project across the entire genome is a waste of time and money. The map's usefulness across populations hinges on the 'common disease-common variant' hypothesis, assuming that most variants for disease are in all populations. But, asks Lisa Brooks of the NHGRI, "How different are populations in their haplotype structure?" Mike Boyce-Jacino of Orchid BioSciences voices ethical concerns: "When you go to haplotype level, people are more nervous that you're associating gene function almost on a population-specific basis," he warns. E.R.

Top left: Eric Lander
Above: SNP

Elke Jordan



Mike Boyce-Jacino

