

# DESPERATELY SEEKING BIostatisticians

**New career opportunities exist at the interface between genetics and statistics, but the lack of training that bridges the two subjects makes biostatistician positions difficult to fill.**

At first it seemed straightforward: with the sequence of the human genome in hand, the secrets of biological processes and how genes are linked with disease would begin to reveal themselves. But the genome proved to be more sophisticated than many researchers thought, deriving most of its complexity not through increases in gene number but through more subtle touches, creating diversity through other methods such as alternate splicing.

All of which made the already challenging pursuit of hunting for genes linked with common diseases even more demanding. Most common diseases are complex in nature, and result from the interaction of several genes, each of which makes a small contribution, in conjunction with environmental influences, to the phenotype. As the ability to collect genomic data increased, so too did the importance of developing computational methods that could enable the generation of hypotheses, and stimulate the development of experimental approaches to test them.

Several areas of computational biology are crucial to this pursuit, and all have the same goal: to study the relationship between molecular signposts (such as single-nucleotide polymorphisms (SNPs)), environmental influences and phenotype. Faster and cheaper ways of screening individuals for SNPs have been developed and are no longer the bottleneck. High-density, whole-genome SNP mapping analyses are routinely performed after the systematic collection of DNA from patients. These analyses, including SNP mapping, ultra-high-throughput SNP genotyping, population data and clinical phenotype data create large-scale genetic and pharmacogenetic information. The challenge now is to make sense of it all.

To manage and interpret the deluge of data requires computational methods that are based on statistics and

applied mathematics. This is different from statistical analysis of clinical trials, which incorporates genetic epidemiology, statistical methods and clinical biomarkers. Biostatistical analysis of the genetics of complex traits integrates knowledge in statistics, genetics, epidemiology, computational biology and applied mathematics (FIG. 1). This combination of skills allows the development of sophisticated combinatorial statistical methodologies, such as linkage analysis and linkage disequilibrium mapping.

But, despite the growing use of biostatistical tools in the industry we face the current frustration of the difficulty of filling available positions in this area. Experts in one field are currently trained in-house; for example, we can propose postdoctoral positions in which the candidate has the possibility of broadening their expertise. And we often adapt the position to the skills of the candidate, when it should be the other way round. The candidate would rather be recruited for a 'Methodology' position if he or she is skilled in statistics and mathematics, whereas a candidate experienced in statistical genetics would prefer an 'Analysis & Interpretation' position.

The problem is the lack of training programmes that produce graduates who are either able to, or prepared to, bridge the gap between the several disciplines involved in this field. Training programmes should try to make boundaries between disciplines less rigid, to give students the curiosity to apply their expertise in another field.

Reluctance stems, in part, from the often unappealing manner in which students are introduced to statistics. Genetic programmes tend to focus on biology, but do not address the complexities of statistics and analysis, and too frequently defer them to 'advanced' courses. Terms such as 'chi-square' are taught in a rote-learning manner, without providing any understanding of the meaning of such a value.

Yet the intellectual challenges created by combining both areas of genetics and biostatistics should convince some that this path is worth considering. Pure geneticists willing to embrace statistics have the tools to accurately and effectively analyse and interpret their results, and to build biological hypotheses. Pure statisticians and applied mathematicians will find that knowledge of genetics provides unspoiled areas to explore, and new approaches for solving a new series of problems. The field of drug discovery, like academic science, is increasingly benefiting from interdisciplinary skills, and so young scientists need to be encouraged to learn skills other than their core science subjects, and should be taught these other subjects in a more interesting and applied manner.

## Online links

### FURTHER READING

American Society of Human Genetics — Educational Resources:  
<http://genetics.faseb.org/genetics/ashg/menu-educ.shtml>  
Access to this interactive links box is free online.

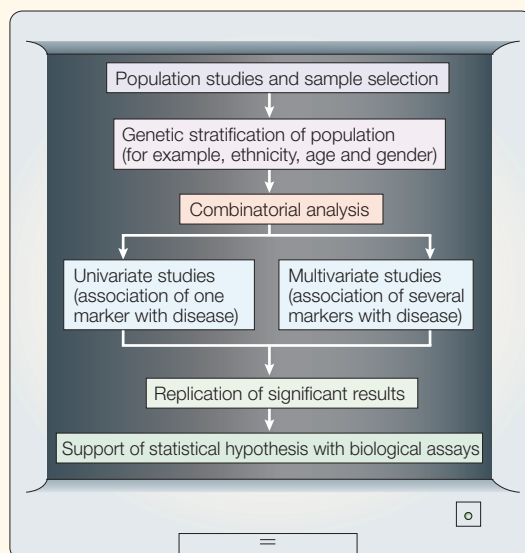


Figure 1 | **Biostatistical analyses applied to genetics.**

Anne Gimalac,  
Communication Manager,  
Sero Genetics Institute,  
Route Nationale 7,  
91030 Evry Cedex, France.  
e-mail:  
[anne.gimalac@serono.com](mailto:anne.gimalac@serono.com)