Logical Reasoning in Human Genetics

Human genetics is one of the most active fields in all of modern science, with very large research investment being made worldwide. A proliferation of journals and unprecedented flow of new papers reporting findings from across the spectrum of traits might suggest that this field is succeeding at a rate greater than almost any other in history.

Yet, this has turned out to be a complex area of science. There has been dramatic success in identifying genes associated with classical 'mendelian' diseases, in which it is clear in advance that a gene is responsible. There have probably been hundreds of such successes. In addition, genes that contribute to or even cause some cases of complex chronic diseases have been found. This course will present the strategies that have been used to find these genes, their rationale, and aspects of successful study designs. At the same time, this success has been tempered by the etiological complexity even of the simplest traits, those dominated by effects of a single gene. There is a big difference between identifying a gene and using that knowledge successfully, which has proven to be much more difficult. Even for truly genetic disorders understood for decades, such as sickle cell anemia or cystic fibrosis, successful genetically based interventions are rare, though in some cases genetic counseling has been feasible and effective as a public health measure.

Success with simple diseases—at least in the sense of identifying their causal gene—has led to enthusiastic promises of major relief from disease that will come from the genetic pursuit of complex diseases. But the case has turned out to be rather different. For example, many genes have been claimed to affect the risk of these diseases, yet those effects that are confirmed are usually rare, and/or of much weaker effect than originally claimed. Effects of environmental exposures and genotypes of other loci probably play a major role in the difficulties being faced, and indeed modify the risk associated even with the most clear-cut etiological effects. But even such statements, though widely accepted, have not led to a full understanding of any of these disorders. Despite extensive effort and refinement of current approaches, there is still little empirical evidence or theoretical basis on which to base predictions of imminent success. Much remains to be learned and probably will require new thinking about the problem. Progress can be made by using an understanding of biology, evolution, and genetics, to understand why the attempt to understand them has been frustrating so far.

For example, multiple competing or complementary risk factors are poorly accounted for in genetic and epidemiological studies alike, for reasons that can be explored and understood. Similar claims about the environmental component of these traits, to those often made about their genetics, have motivated many epidemiologists to try to consider genetic factors as nuisance parameters in their studies of environmental risk, while geneticists do roughly the opposite in what is typically perfunctory inclusion of measures of environmental factors in their studies.

This course is designed to look at the overall issues in concept, data, and method, to consider the present state of affairs in terms of the origin and nature of human genetic variation, the nature of genotype-phenotype relationships, and the way that studies are being designed to understand them. We will use computer software to evaluate the power of different study designs to identify genetic or environmental risk factors under a variety of complex sets of assumptions about these genotype $\rightarrow$ phenotype relationships to try and make some conclusions about power and efficiency of studies of various sorts, and the robustness of study designs to the underlying sets of assumptions.

Beginning from evolutionary first principles about the origin and nature of human phenogenetic variation will give a conceptual basis for any investigator's choice about what kinds of biologically sound models might be reasonable for the trait s/he is interested in. This can help motivate decision-making about study design and power to make reliable inferences in studies, based on something more scientific than the usual calculations and models based on assumptions that are often used as the basis for a given research project, and lead, often predictably, to unsatisfying results.

It will be shown that methodological details are often not the critical issue that should be used to design genetic studies with optimal chances of success. Often, the biological problem is not adequately identified, around with to design family or population based ascertainment schema that can most powerfully lead to identifying and characterizing the important genetic (and environmental) factors. Sometimes, it may be better not to attempt such studies. The issue of importance is how (and when or whether) to design a study to find genes, not what statistical methods should be used to analyze the data. Logic reasoning in advance, and in the interpretation of results, cannot provide miracle answers, but an improved basic perspective on these issues might be helpful in planning future studies, and to know better what the limitations of genetic and epidemiological approaches to complex traits and public health may be.