Branden Tarlow
May 28, 2002
Biochemistry 118Q
Dr. Doug Brutlag
Final Paper

It's not just *your* genes: using genomics to identify, trace, and treat foreign pathogens

With the announcement that the first copy of the human genome has been completed, many eyes have turned to the genomic field. One bi-product of human genome research is the availability of a vast new set of technological tools. Because of the human genome project, genotype screening techniques have greatly proliferated — they are more widely used and widely available for a growing number of researchers and medical professionals. In the last decade a combination of new genomic strategies, classic molecular fingerprinting techniques of restriction fragment-length polymorphism analysis (RFLP), and epidemiological strategies have helped us better understand disease. Elucidating the genetic sequence of individual pathogen strains can help deliver the most efficacious treatment to patients, trace diseases in populations in order to design better public health policy, and lead to wrongful death convictions. In this paper, I will discuss these applications in addition to ethical and technical concerns relating to public health and the legal arena where applicable.

To begin with, genotypic identification of pathogen strains allows offers the possibility to better understand the mechanisms of bacteria or viruses. Though this technology is not new or novel, it has been used in new ways to identify the properties of un-culturable bacteria. For example, by amplifying the DNA with PCR, scientists were able to identify and characterize the bacterium associated with Whipple's disease, a

systemic illness affecting the nervous and gastrointestinal system.[i] In addition, genotypic screening has allowed doctors to distinguish between separate diseases with similar phenotypes—a growing trend in biology. In the 19th century, a variety disease that we now may call cancer and heart disease (and many others) were clumped under the common title "consumption," which was treated by bloodletting. While this now seems absurd, the progress in medicine is a macrocosm of how molecular epidemiology will refine the treatment of pathogens. A more modern corollary can be found in the discovery of hepatitis. Molecular analysis identified patients afflicted with "hepatitis" to show that hepatitis A, B, C, D, and E are separate diseases.[ii] Identifying variation in the genotypes of diseases with similar phenotypes allows medical professionals to better track the spread of disease. By knowing what strain of a disease a patient has, doctors can prescribe the most efficacious drug to fight that disease. This approach is especially useful in treating drug resistant bacterial infections. In the worst of infections, initially choosing the correct drug (that the bacterium is *not* resistant to) instead of operating by trial and error may make a life and death difference. Doctors, however, must ask whether running such an assay is time-effective or cost-effective. Though the host environment is a great variable in disease characteristics, understanding of the pathogenesis of the disease can greatly aid treatment.

Genotype identification is especially applicable in treating human immunodeficiency virus (HIV). Though it is believed to be an incurable disease, the progression of the illness can be treated with a variety of drug cocktails to enhance and lengthen the life of an afflicted person. A major problem in administering HIV drugs is that the drug targets constantly change: HIV has an extremely high mutation rate due to

the inexact nature of the reverse-transcriptase enzyme; nearly 1 in a 1000 base pairs is incorrectly copied. For many years, HIV has been treated with protease inhibitors, drugs that prevent HIV proteins from being correctly divided and AZT, a drug that inhibits reverse-transcription and therefore slowing retrovirus action. But in its hyper-evolution, HIV quickly develops drug resistance. Consequently, to have the most efficacious treatment of the retrovirus, it is important to know which drugs the HIV strain is resistant to and which one work the best. For example, a doctor might enter the RNA sequence of a patient's HIV into the computer and receive an output that says that AZT is only 20 % effective on that patient while another drug works is 80 % effective. Currently, databases with this information receive thousands of scan per day from all over the world. This technology uses bioinformatics to predict the most likely match based upon alignments in the database.[iii]

Because HIV has a great number of nucleotide polymorphisms, it is difficult to treat. But by the same token, the strains take on a personalized aspect that allows epidemiologist to track its evolution. For example, HIV on the African continent is easily distinguished from that in Europe or Asia. Thus, by simply looking at a strain and comparing it to a database, researchers can roughly identify the source of the virus. Even in a smaller cluster of people, a high degree of genetic relatedness of HIV strains can suggest common infection sources.

Several years before the O.J. Simpson trial, genetics, molecular biology, and epidemiology joined forces to help demonstrate that a dentist infected several of his patients with HIV. The dentist, who was symptomatic with AIDS, continued to practice and five of his patients who had very low risks for HIV infection, were found to have the

virus. Molecular analysis of highly variable regions of the HIV genome were analyzed to show a very significant correlation between the doctor and his patients. In the study, investigators identified eight signature nonconsecutive nucleotides in a region called C2-V3 sequences. The theory is that mutations randomly arise in the genetic code in the retrovirus action of HIV. Thus, when a mutation occurs in one host, it is heretically transferred from one HIV molecule to a daughter HIV molecule. Thus, a newly infected person will receive the same strain of HIV as their host. After the moment of infection, the genomes will diverge from each other in an essentially random manner (this microevolution is influenced by the host's immune system, drugs, and other variable environmental factors). By means of a statistical comparison, it was shown that five dental patients shared 7 or more of the signature nucleotides with the dentist while local controls (HIV-positive people in the same geographic area who didn't have contact with the dentist or the other patients) shared 2 or fewer of these markers.[iv] Statistical methods showed this indicated a high degree of correlation ($p = 8E-6$).

Critics of this study are quick to point out that the case of the Florida dentist was an observational study—not an experiment. Therefore, conclusions only suggest correlation (in this case $p = 8E-6$), but do not say that the dentist *caused* his patients to have HIV—though this is inferred. For the purposes of creating health policy, this epidemiological approach is more than adequate. But is it accurate or ethical to use correlation data in a legal trial? Personally, I believe that science is faithful source of analysis; with great p-values and rigorous peer review, I think that correlation is equally reasonable to other forms of evidence. In addition, a technical criticism of the Florida dentist case is that the researchers used "bootstrapping" methods, or a form of sampling

bias where they hand picked eight "signature" nucleotides that had significant results, to compare to the patients and the local controls. They ignored the other 112 nucleotides in that region that didn't provide as clear results.[v] Though the defense would like you to believe this argument discredits the evidence, statistical experts treat this as a subtle issue; the use of correlation holds water in an otherwise strong legal context.

In addition, molecular epidemiology can be used to show the disease transmission pathway of a person with several sources of exposure. Though this source of tracking might be fascinating, it seems merely academic: it doesn't provide the newly infected individual with much new information. Public health officials who trace the methods of disease spread, however, find transmission data extremely valuable. For example, tuberculosis researchers at Stanford University write, "Molecular epidemiological approaches have provided novel insights into the transmission dynamics of tuberculosis and have helped to refocus and refine these practices."[vi] In this case, molecular epidemiology was used to debunk a myth that only 10% of tuberculosis cases were the result of new infection. Instead, their finding that between 31% and 38% are new infections allowed public health officials to enact measures aimed at reducing the spread of new infections.[vii] Because the genotype of latent virus strains vary with that of new infections, epidemiologists designed clusters describing the variation in the population. And all of this is possible because of genomic technologies: "Automated DNA sequencing allows for the direct comparison of specific genes among large populations of isolates and the determination of the complete genomic sequence of two strains.[viii]

The science of comparing two strains is extremely simple in contrast to the lofty goals of elucidating mechanism and curing diseases. The consideration of

microorganisms influence on human genomics only muddies the water when searching for "disease genes." What diseases are diseases are genetically caused? What diseases might be linked to persistent infections? Though cancer is mostly thought of as a genetic disease, the role of viruses in the mechanism of infection is not clear. Epidemiological data, for example, has shown 100-fold associations for hepatitis B virus and liver cancer and human papillomavirus types 16 or 18 are about 30 times more common in women with cervical cancer.[ix] This data provides a correlation not a causal link, but this data will eventually lead researchers to study the interactions of microorganisms on oncogenes.[x] It was less than a decade ago that *Helobactor pyloris* was linked to ulcers, but still this fact has not penetrated the general public or medical profession. Following this example, might some "genetic disease" actually be the work of latent microorganisms yet to be discovered? Person-virus, person-environment, environment-host, and variable genetic penetrance only further complicate things. For now, it appears that the answers to these questions rest not only in our genes but also in our genetic world.

---

[i] Retman et al. "Identification of the Whipple's Disease Bacillus" *The New England Journal of Medicine*. V. 327 n. 5, 30 July 1992.
[ii] Butler D. "Epidemiology set to get fast-tracked treatment" *Nature.* v. 414, 8 Novermber 2001.
[iii] http://www.ncbi.nlm.nih.gov/retroviruses/subtype/makepage.cgi?page=sub&type=0 . Also, in class, we visited http://hivdb.stanford.edu/ that predicted drug efficacy based upon the genotype of the HIV subtype.
[iv] Ou C et. al. "Molecular Epidemiology of HIV Transmission in a Dental Practice," *Science* V. 256, 22 May 1992, pp. 1165- 1171.
[v] Smith TF and Waterman MS, "The Continuing Case of the Florida Dentist" *Science* V. 256 22 May 1992.
[vi] "Kato-Maedo M, Bifani PJ, Kreiswirth BN, and Small PM. "The nature and consequence of genetic variability within *Mycobacterium tuberculosis," The Journal of Clinical Investigation* V. 107, n. 5, March 2001, pp. 533 - 537.
[vii] Kato-Maeda M, Small PM. "How molecular epidemiology has changed what we know about tuberculosis." *West J Med* V. 172, April 2000, pp. 256-259.
[viii] Kato-Maedo et al., *J Clin Inv.* p. 535.

[ix] Danesh, John, et. al. "A Human Germ Project?" *Nature.* v. 389, 4 September 1997.
[x] Potter JD. "At the interfaces of epidemiology, genetics and genomics," *Nature.* V. 2, February 2001, pp. 142- 147.

Other references:
Small PM et al. " Epidemiology of tuberculosis in San Francisco: A population-based study using conventional molecular methods" *New England Journal of Medicine* V. 330, 16 June 1994, pp. 1703-1709.