

Your genetic data

Ben Klemens

6 April 2009

[Or, *The ethical implications of SQL.*]

The figure explains how my work in statistical genetics is all possible. It is what a genetics lab looks like. That's a work bench, like the ones upon which thousands of pipettes have squirted millions of liters of fluid in the past. But you can see that it is now taken up by a big blue box, which hooks up to a PC. Some of these big boxes use a parallel port (like an old printer) and some run via USB (like your ventilator or toothbrush). The researcher puts processed genetic material in on the side facing you in the photo, onto a tray that was clearly a CD-ROM drive in a past life. Then the internal LASER scans the material and outputs about half a million genetic markers to a plain text file on the PC.

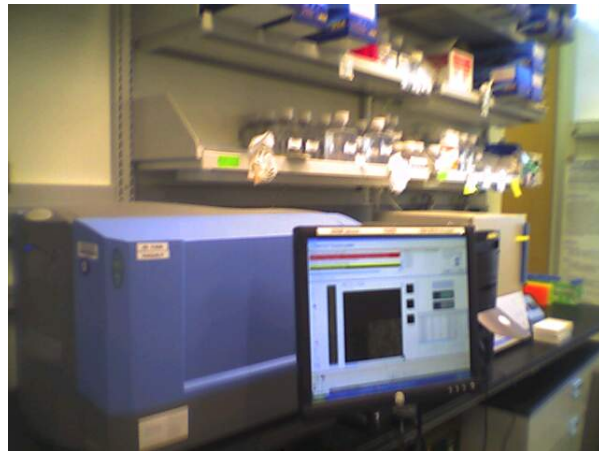


Figure 1: The tools of the data processing field known as Biology

I know I'm not the first to point this out, but the study of human health is increasingly a data processing problem. My complete ignorance regarding all things biological isn't an issue in doing analysis, as long as I know how to read a text file into a database and run statistical tests therefrom.

Implication one: Research methods Historically, the problem has been to find enough data to say something. One guy had to sail to the Galapagos Islands, others used to wait for somebody to die so they could do dissections, and endless clinical researchers today post ads on bulletin boards offering a few bucks if you'll swallow the blue pill.

But now we have exactly the opposite problem: I've got 18 million data points, and the research consists of paring that down to one confident statement. In a decade or so, we went from grasping at straws to having a haystack to sift through.

I've got tools printed in textbooks from the 1970s that will eke out a relationship from a minimum of data, and 5GB of genetic data regarding people with bipolar disorder over on the other screen. Applying one to the other will give me literally thousands of ways of diagnosing bipolar disorder, none of which are in any way trustworthy.

So the analytic technology is not quite there yet. There's a specific protocol for drawing blood that every nurse practitioner knows by heart, and another protocol for breaking that blood down to every little subpart. We have protocols for gathering genetic data, but don't yet have reliable and standardized schemes for extracting information from it.

When we do have such a protocol—and it's plausible that we soon will—that's when the party starts.

Implication two: Pathways If you remember as much high school biology as I do, then you know that a gene is translated in human cells into a set of proteins that then go off and do some specific something (sometimes several specific somethings).

If you know that a certain gene is linked to a certain disorder, then you know that there is an entire pathway linked to that disorder, and you now have several points where you could potentially break the chain. [Or at least, that's how it'd work in theory. Again, there's no set protocol.] There are many ways to discover the mechanism of a disorder, but the genetic root is the big fat hint that can make it all come together right quick. So the scientists still working with squishy biological organisms are also keeping their eye on the database-crunchers for clues about what to poke at.

Then, once there's consensus on a pathway, the drug companies go off and develop a chemical that breaks the destructive chain, and perhaps make a few million per year in the process.

Implication three: Free will versus determinism One person I talked to about the search for genetic causes thought it was all a conspiracy. If there's a genetic cause for mental illness, then that means that it's not the sufferer's fault or responsibility. Instead of striving to improve themselves, they should just take a drug. And so, these genetic studies are elaborate drug-company advertising.

From my casual experience talking to folks about it, I find that this sort of attitude is especially common regarding psychological disorders. See, every organ in the human body is susceptible to misfiring and defects—*except* the brain, which is created in the image of ", and is always perfect.

Annoyed sarcasm aside, psychological disorders are hard to diagnose, and there's a history of truly appalling abuse, such as lobotomies for ill behavior, giving women

hysterectomies to cure their hysteria, the sort of stories that made *One Flew over the Cuckoo's Nest* plausible, &c. Further, there are often people who have no physiological defect in their brains, but still suffer depression or other mood disorders. They get some sun, do some yoga, and everything works out for them.

But none of that means that the brain can not have defects, and that those defects can not be treated.

The problem is that our ability to diagnose is falling behind our ability to cure. We know that certain depressives respond positively to lithium, Prozac, Lexapro, Wellbutrin, Ritalin, Synthroid, and I don't know today's chemical of the month. But we still don't have a system to determine which are the need-of-drugs depressives and which are the get-some-sun depressives.

Or to give a physical example, we don't know which obese individuals have problems because of genetic barriers and which just need to eat less and exercise. It's only harder because, like the brain, the metabolism is an adaptive system that can be conditioned for the better or for the worse, confounding diagnosis. Frequently, it's both behavior and genetics, albeit sometimes 90% behavior and other times 90% genes.

A genetic cause provides genetic tests. If we have a drug based on a genetic pathway, as opposed to a drug like Prozac that just seemed to perk people up, we can look for the presence or absence of that genetic configuration in a given individual. This ain't a silver bullet that will sort people perfectly (if that's possible at all), but having a partial test corresponding to each treatment is already well beyond the DSM checklists we're stuck with now.

From here, there are ethical implications, which I'll save for next time.