

A Genome Commons

In the past week, the complete diploid genomes of James Watson and J. Craig Venter were publicly announced. Astonishingly, most press reports offered no meaningful interpretation of these genomes other than that Watson's genome had many regions not previously seen, and that it masked information about an Alzheimer's gene--because Watson did not want to know about it. I read one report which noted that Venter found that he had a mutation that raises the risk of heart disease, only after reading about the gene in the New York Times. However, Venter said he might have guessed this from his family history.

It is true that we cannot explain most variation in the genome at this time, but certainly we could do better than this. If an individual's genome is so revealing, why was so little revealed?

One problem is that we cannot come to grips with what we know. The effects mutations are scattered amongst hundreds of databases and amidst millions of manuscripts and patent applications. These data are heterogeneous; while some papers discuss the precise effects of a single nucleotide change, many analyses basically offer rules of thumb. All such information could be powerful in personal genome interpretation, if only we could make use of it.

To that end, I propose the creation of a Genome Commons, a public knowledgebase of human variation and its effect. The contents would be culled from the scientific literature and curated by experts who could assess the meaning of the information and make it accessible. I view this repository of our common human inheritance as a necessary resource for research, medicine, and understanding ourselves.

The Genome Commons would be compiled by researchers around the world: anyone with an Internet connection, access to academic journals, and appropriate training could contribute in a systematic manner. I envision a vital role for private enterprise in the genome commons as well, both contributing discoveries to the database and helping to make its contents accessible to clinicians and the public.

Even if all genome knowledge were collected in one place, we presently lack effective public tools to relate these back to a specific individual's genome. Therefore, to be able to offer some interpretation of the variations seen, we also need a Genome Commons navigator. In principle the idea is simple: find the mutations described in the database and see if they are present in an individual's genome. In practice, this is devilishly complicated because our genomes don't come indexed for such analysis and especially because the knowledge we have is so variegated. Furthermore, each of us will have million of differences from any reference genome. Perhaps the greatest challenge is to avoid being deluged with genomic marginalia by somehow ranking these genetic variations according their importance.

In this context, even as we celebrate what we will learn about ourselves from our genomes, we should maintain realistic expectations. Given that most medications don't even take our weight into account, it is hard to believe that many pharmaceutical doses will depend upon minutiae of our genomic complement. Indeed, it remains to be seen whether we will typically learn anything more important from our genomes than that we should use sunscreen, lose weight, exercise more, and stop eating all those salty chips laden with saturated or trans fats.

But in a world of limited time, resources, and personal restraint, a Genome Commons enables productive use of the wealth of genomic information available, helping each of us prioritize healthful activities and therapies to give us the most productive enjoyable lifespans.