


How the Databases in Our Cells Can Save Lives

Lily Hildebeitel

Imagine a patient in the future with early stage pancreatic cancer. Rather than receive chemotherapy, which kills dividing cells in all types of body tissues, they first have their DNA sequenced. From there, a personalized set of treatments can be given, catered to the specific biomarkers of that person's pancreatic tumor cells. These treatments give the patient a better clinical outlook due to their preciseness based on the patient's specific genome. This field, called "precision medicine", is not far off at all. It is the result of a widely growing discipline of medicine that is focused on using computational methods to analyze genetic data and then make predictions based on said findings.

In the current age of medicine, we treat patients in a very mechanistic way. We treat the symptoms that we observe. When we treat every patient with a certain symptom the same way, the treatment effectiveness varies wildly. It's time for medicine to change its methods.

Let's start with understanding our genome, which is the complete set of our DNA. Each  one of our 37.2 trillion cells carries our entire genome in its nucleus. Our genome is made up of a combination of four different chemical bases: adenine, thymine, cytosine, and guanine. The order of these four bases determines every characteristic about what makes us human. They also code for minuscule proteins in our body, which work hard to regulate countless functions that we take for granted. We refer to regions of DNA that code for proteins as genes.

While a cell in our brain and a cell in our heart may contain the exact same genome, they function differently and use different proteins based on what is needed in that part of the body. This is because not every region of our DNA is "turned on" in every cell. When a part of our DNA is turned on, this is called gene expression.

Considering how sophisticated humans are compared to other species, it is easy to assume that genes make up most of our DNA. Shockingly, only 1-2% of our genome is a gene that codes for a protein. An even smaller percentage of those coding genes are expressed in each cell. This reveals a large mystery that bioinformaticists are working to uncover: what is the purpose of our 98-99% (supposedly) noncoding DNA? Does it do anything at all, or is it just an artifact of genetic evolution? Could it reveal answers about why we get diseases such as cancer or Alzheimer's?

Let's revisit the cancer example. While chemotherapy can be effective in some patients at killing cancer cells, it isn't a one size fits all. Cancer drugs with specificity have higher success

rates in treatment, and development has already started—for example, a lung cancer drug Gefitinib. Gefitinib is only effective at targeting tumors that have a specific mutation, called “EGFR”, in its genetic sequence. Patients with these EGFR tumors who received gefitinib have a high recovery rate compared to general cancer treatments such as chemotherapy (1). If we can start applying these sequencing methods for all diseases of incoming patients, we may start saving more lives.

There is huge promise in using precision medicine to read a person’s genome and genetically cater a medical treatment to them. However, the only thing better than treating is preventing. Success in disease prevention will only come with genetic data analysis, which has massive potential to tell us why diseases arise in the first place.

The link between our genome and diseases can be vague because not all diseases or syndromes directly arise from our DNA. If a person gets the flu, or if a person develops lung cancer after a lifetime of smoking, we can confidently attribute these to external causes. On the other hand, conditions such as Turner syndrome or Down syndrome are directly caused by specific mutations in our genes. And in the middle, we have the gray area: hundreds of conditions that are merely suspected to be linked to our DNA. While we scientists like to think we know the cause of most health conditions, I’ll let you in on a secret: there is way more that we don’t know than what we do know. This is where evolutionary genomics can come into play.

Evolutionary genomics may contain those missing pieces of the puzzle for solving the mystery of why diseases develop in the first place. There is only so much we can do with just our human genome. Data by itself doesn’t mean much unless we can compare it to other sets of data and figure out what is significant. So, in order to maximize what we learn from DNA, we must compare our genes to those of other species.

The main appeal of applying evolutionary genomics to medicine is that it gives us insight into whether or not a disease is unique to humans or other primates as well. If it is, what’s causing it? If we don’t factor in evolutionary genomics, we don’t get the full picture of the disease. Knowing how our genomes evolved allows us to understand more holistically and put us on the road towards disease prevention.

Currently, there is a large gap between the skills needed to analyze all of this data and the skills that incoming scientists are equipped with. If we want to both further understand the origins of human diseases and learn to develop precise treatments, then we need people with the skills, such as coding or navigating bioinformatics softwares, and the passion. It’s exciting to think that many answers to health-related crises lie in the huge databases within each of our

cells. For the first time in human history, we have access to all of the data, but we need incoming scientists to learn how to decode it. Doing so could save millions of lives.