Basics of Bioinformatics

(250) What is Bioinformatics? - YouTube

https://www.youtube.com/watch?v=W-Ov2cUaYQY

Transcript:

(00:01) 80%. That's how similar our protein coding genes are to those of mice. Humans are larger, smarter and live longer than mice, not to mention the fact that humans don't have fur or a tail. The reason we have been able to calculate our similarity to mice is because of a massive effort to sequence the genomes of humans, an undertaking called the Human Genome Project, as well as from sequencing the genomes of mice and many other organisms.

(00:29) Genome sequencing involves reading through the A's, T's, G's and C's that make up our DNA, and it has given us a lot of information about what our genes are and how they are organized. And it has helped us improve how we diagnose and even treat human disease. But genome sequencing involves generating very large sets of data, so we need powerful tools to decipher all those ATGC's.

(00:55) This is where the rapidly growing field bioinformatics comes in. Extremely powerful computers are being used to store and manipulate all of this data, and the people behind the computers are bioinformaticians, scientists who are often trained both in biology as well as math or computer science. These multidisciplinary researchers develop methods and software tools to program computers to dig through and make sense of all of this data.

(01:24) So how did bioinformatics help us learn that humans and mice have 80% similarity in their protein coding genes? One way bioinformaticians approach this type of question is by looking for small sequences in one genome that match the other genome. This is like doing a word search on your computer. But instead of searching for the word Shakespeare in your English term paper, you scan for ATTGCACGTCTA.

(01:52) Once matching areas are found, researchers design algorithms that can scan past the ends of both sequences to see just how far the matching regions extend. So in this case, do the letters after the CTA continue to match between the mouse and human sequences? If they do not and a difference is present, we can analyze the following sequence and start to figure out which genes are involved in many of the traits that make us humans different from mice, like brain development and longevity.

(02:24) The flip side of this is that we can also find sequences of the genome that are highly similar across different species. In addition to sharing parts of our genome with mice, humans have genes in common with plants, flies, and even microscopic bacteria. Thes regions are called conserved genes, and because their shared across many species, they likely code for proteins that are essential for life on earth.

(02:53) In addition to finding the similarities and differences between the genomes of different organisms, sequencing technology has also allowed us to pick up differences in DNA between different people. This has been particularly important because it helps us better understand

human disease. Let's say you knew a specific DNA base in a gene was different from person to person, and you wanted to see if that difference was important for a disease, like diabetes.

(03:19) You could sequence that base in 50 healthy people and 50 people with diabetes. If 47 of 50 people with diabetes had an A, while only 5 out of 50 without diabetes did, that would be strong evidence of association between the A variant and the disease. But importantly, it does not mean the variant causes the disease.

(03:42) Researchers have developed technologies to look across hundreds of thousands of sites across the genome for these kinds of single base differences and have looked for correlations between certain bases and disease. These experiments are known as genome-wide association studies. Sequencing and bioinformatic analysis are also becoming increasingly important for the diagnosis and treatment of cancer, because cancer cells often have many mutations or changes in the nucleotide code compared to a patient's normal cells.

(04:17) As we learn more and more about the human genome and the cost of sequencing decreases, doctors can order sequencing of parts of their patients' genomes for a relatively low cost. By using algorithms to compare a patient's tumor cells to the normal genome, as well as to the tumors of many other patients, doctors can quickly pinpoint the changes in the DNA that are causing the cancer cells to grow uncontrollably.

(04:44) This helps them choose the best treatment for their patients. As our ability to sequence genomes continues to increase, bioinformatics will need to continue to develop faster and more advanced algorithms to handle these massive data sets. An important part of this field has been the development of large centralized databases of genome sequences that can be accessed by anyone.

(05:09) The challenge for the future will be to continue to grow these databases in a way that helps scientists make important new discoveries while preserving the privacy of patients.