

Computer scientists may have what it takes to help cure cancer

The war against cancer is increasingly moving into cyberspace. Computer scientists may have the best skills to fight cancer in the next decade — and they should be signing up in droves.

One reason to enlist: Cancer is so pervasive. In his Pulitzer Prize-winning book, *“The Emperor of All Maladies,”* the oncologist Siddhartha Mukherjee writes that cancer is a disease of frightening fractions: One-fourth of deaths in the United States are caused by cancer; one-third of women will face cancer in their lifetimes; and so will half of men.

As he wrote, “The question is not if we will get this immortal disease, but when.”

Dr. Mukherjee noted that surprisingly recently, researchers discovered that cancer is a genetic disease, caused primarily by mutations in our DNA. As well as providing the molecular drivers of cancer, changes to the DNA also cause the diversity within a cancer tumour that makes it so hard to eradicate completely.

The hope is that by sequencing the genome of a cancer tumour, doctors will soon be able to prescribe a personalised, targeted therapy to stop a cancer's growth or to cure it.

Steve Jobs' cancer

According to Walter Isaacson's new biography *“Steve Jobs,”* a team of medical researchers sequenced the Apple executive's pancreatic cancer tumour and used that information to decide which drug therapies to use. Since Mr. Jobs's cancer had already spread, this effort was even more challenging. Each sequencing cost \$100,000.

Fortunately for the rest of us, the cost of turning pieces of DNA into digital information has improved: The costs dropped a hundredfold in the last three years. The tipping point before widespread use is believed to be \$1,000 per individual genome, which is a reason for the major investment in reducing its cost. Given such dramatic improvement, we could soon afford to sequence the genomes of the millions of cancer patients, which only billionaires could afford a few years ago.

How can computer scientists help?

First, as recently reported in this newspaper [NYT], the cost of millions of short reads of one cell by a gene sequencing machine is dwarfed by the data processing costs to turn them into a single usable three-billion-base-



INSIGHT: Changes to the DNA — the picture shows computer artwork of a double-stranded DNA helix — cause the diversity within a cancer tumour. — Photo: The Hindu Photo Library



pair digital representation of a genome. To make personalised medicine affordable for everyone, we need to drive down the information processing costs.

Second, we need to collect cancer genomes in a repository and make them available to scientists and health professionals. The computer scientist David Haussler of the University of California, Santa Cruz, for example, is creating one. Plans are that this five-petabyte (5,000,000,000,000,000 bytes) store will house more than 20,000 genomes.

Third, finding a personalised, targeted therapy for each tumour among myriad possible combinations of drugs is like finding a very small needle in a very large haystack. Researchers are exploring the engagement of people when traditional hardware and software are not up to the task.

An inspirational example is the Fold it game — developed by the computer scientist Zoran Popovic at the University of Washington — that recently attracted thousands of volunteers to uncover the structure of an enzyme important to H.I.V. research.

Big Data challenge

Cancer tumour genomics is just one example of the Big Data challenge in computer science. Big Data is unstructured, uncurated and inconsistent, and housing it often requires a thousand-fold increase in size over traditional databases. It is not pristine data that can be neatly stored in rows and columns. YouTube alone holds nearly one exabyte of videos, which is one trillion megabytes, or 1,000,000,000,000,000 bytes.

The Big Data research challenge is to develop technology that can obtain timely and cost-effective answers to Big Data questions. A Berkeley team of eight faculty members and 40 Ph.D. students is rising to that challenge via three initiatives: inventing algorithms based on statistical machine learning; harnessing many machines in the cloud; and developing crowd-sourcing techniques to get people to help answer questions that prove too hard for our algorithms and machines.

Algorithms, machines and people gave our new lab its name: the AMP Lab. AMP technology could help the war on cancer. It needs new algorithms to find those needles in haystacks. To process genome data faster and more cheaply, the war needs new infrastructure to use many machines in the cloud simultaneously. And it needs to be able to engage the wisdom of the crowd when the problems of cancer genome discovery and diagnosis are beyond our algorithms and machines.

It may have been true once that expertise in computer science was needed only by computer scientists. But Big Data has shown us that's no longer the case. It is entirely possible that we have the skill sets needed now to fight cancer and to advance sciences in myriad other ways.

The night after we made that argument, I awoke in the middle of the night with this question etched into my mind: Given that millions of people do have and will get cancer, if there is a chance that computer scientists may have the best skill set to fight cancer today, as moral people aren't we obligated to try? (*David Patterson is a professor in computer science at the University of California, Berkeley.*) — **New York Times News Service**