2009-03-22 OSC NEWS: Can computers find a cure for cancer? (TechRadar.com)

Can computers find a cure for cancer?

Fast-developing computing technology gives researchers hope

TechRadar.com/Computing News

March 22, 2009

Cancer is one of the most dreaded diseases in the developed world. Its forms are many and its symptoms are diverse, but all variants cause pain and suffering.

Finding a cure for the disease is perhaps the most enduring medical dream of all, and increasingly the hope of medical researchers lies with fast-developing computing technology.

Can computers cure cancer? We don't know the answer to that question yet. But what we do know is that no other field shows us more vividly what computers can do. At Ohio State University Medical Center – home of one of the fastest supercomputers in the world – scientists are weighing proteins in order to find and measure the microscopic differences between healthy and abnormal cells.

At the Swedish Medical Center in Seattle, gene-sequencing techniques are providing rich information about brain cancer – considered the most challenging disease to research – and its potential treatment. And at the School of Informatics at Indiana University, researchers have used colossal computers to create a huge database of cell structures, hoping to understand exactly how they work and – most importantly – how they interact with each other.

Though the study of cancer also shows what computers can't do, it's by focusing on and resolving these problems that a cure may be discovered and the power of the technology advanced further.

Finding the magic bullet

One of the main goals in cancer research is to find a 'magic bullet' that can enter the human body, find mutated cells, target specific proteins in order to switch off the cancer's self-replications and destroy the mutated cells. Part of the obstacle to achieving this cure is knowing enough about the cancer cells and molecules.

Jake Chen, an assistant professor at Indiana University, says that this process – called 'finding drug targets' – requires a massive database of biomedical information. His team has developed one of the largest Oracle-powered relational databases, holding about one half of a terabyte. Chen and his team – who have focused their efforts on breast cancer research, one of the most common forms of cancer – are currently analysing tens of GBs of raw mass spectrometry data from 80 blood samples, with more coming soon.

These samples should help to further research into our understanding of the relationships between cancer and normal cells down at the molecular level, which is a particular difficulty at the moment. To help with

this, Chen's team created complex algorithms not widely used in the biomedical field. The algorithms analyse not just the characteristics of individual molecules but also how each one affects others. This is what makes cancer research so complex – the interrelationships that exist and the data analysis required.

Chen says that the closest analogy to this relational study is the Internet itself. For example, the servers for AOL are widely known on the web, and it's easy to see the links between one AOL server and another. Yet there are many servers on the outer edges of the web which only link to a few others. These are the 'molecules' that are harder to understand. When one of them crashes, it can effect that part of the Internet in adverse ways – causing server outages, for example.

Data visualisation software can help researchers understand these 'fringe' areas of systems biology. Correlating the data requires complex algorithms which are still evolving. It might mean culling data from 100 other researchers around the world who have all found a likely protein target, analysing 25,000 genes and a few hundred thousand protein fragments, archiving the data for later retrieval and finally processing the algorithms using the Big Red cluster at Indiana University. It's a highly collaborative effort.

"The answer is in the data set, but the computer is not intelligent enough yet," says Chen. "We need to make the computer smarter. Today's computers are used primarily for managing information; we need to make them smart about interpreting the data."

Chen had an interesting analogy for how this works. When you look at a painting of a face, you can see what it is immediately. A computer can analyse the colours and chemicals of the painting, but it's not clever enough to see the face. Similarly, Chen is trying to produce an algorithm that can see through the noise of intricate molecular interaction networks in cancer and find the critical proteins where drug interventions may occur.

Part of the computational challenge is transferring what we already know about curing cancer in mice to humans. The drugs used to cure cancer in mice could be used for humans, but they might provoke a different set of side effects. The informatics question is how to find a cure that works on 100 per cent of cancer patients. "This exciting conquest will likely go on in the next one to two decades, and will rely on systems biology informatics techniques," says Chen.

Computing the weight of a cell

Creating algorithms to understand the relationships between cells in the human body is one computational challenge. Another research endeavour, underway at the Ohio State University Medical Center, is to weigh the proteins that cause cancer.

A recent finding in cancer research is that protein weight changes are an indicator of cancer. This is a field called proteomics, and it's the study of how proteins work, in the same way that genomics is the study of how DNA works. Cancerous cells cause a change in a protein's amino acid sequence. Weight changes in proteins are also caused by 'post translational' modifications, where a naturally occurring change signalling a cellular event isn't handled properly by cancerous cells.

Dr Michael A Freitas, a research assistant professor at the Ohio State University Medical Center, says that each proteomic study can produce hundreds of thousands of mass measurements. The process involves creating a theoretical space of all database forms in the samples, resulting in a theoretical spectrum greater than 10 to the eighth power. Scientists then compare this theoretical space with an experimental spectrum.

The computational power already exists to analyse the spectra, but there are currently no algorithms that can identify every possible protein form. "We require algorithms with a better understanding of biology," says Dr Freitas. "For example, if we want to consider somatic gene mutations that give rise to changes in a

protein sequence, we have to tell the algorithm to include all known mutations. However, we still don't know all the somatic mutations that may exist. We need algorithms that are capable of finding patterns in the data. These unknown protein forms could lead to breakthrough discoveries in biology and lead to a better understanding of cancer."

Besides the need for better algorithms, there's also an infrastructure challenge. The data from samples accumulates at 1GB of data per hour and requires high-capacity file servers, high-speed networks and dedicated terminals to run the required algorithms, plus access to clusters of computers for processing. "Informatics and high-performance computing are playing a greater role in our efforts to improve cancer patient outcome," says Dr Freitas. "As we improve our ability to detect and identify proteins, we improve our understanding of the cellular machinery. We rely on the use of parallel processing to speed up analysis and distribute the memory requirement across machines."

Dr Freitas noted that two advances in computer technology in particular are helping to further his research. First, multicore processors, 64-bit operating systems and high-speed RAM mean that massive data informatics analysis can be handled on the desktop. Second, graphics card technology has advanced to the point where a field-programmable gate array could be used in a cluster for cancer research.

Currently, researchers use the Ohio Supercomputer Center. The supercomputer consists of the IBM 1350 AMD Opteron cluster (which can handle 22 trillion floating point operations per second) and the HP Itanium 2 cluster. The HP cluster has 516 Intel Itanium 2 processors, more than a terabyte of RAM and 11 terabytes of aggregate disk space, an SGI Altix 3700 with 32 processors and 64 gigabytes of memory and three SGI Altix 250 systems, each with 16 processors and 32 to 64 gigabytes of memory.

Understanding brain cancer

Brain cancer is one of the most poorly understood diseases in the medical profession, partly because we know so little about the brain and partly because there are relatively few cases of it: roughly 20,000 cases are reported per year. Even if it were possible to accumulate data on all 20,000 cases, the cancer is heterogeneous: it varies from patient to patient.

At the Swedish Medical Center in Seattle, and specifically at the Center for Advanced Brain Tumor Treatment (CABTT), raw processing power is required to analyse the genes in cancer patients and determine their genetic makeup. At that point, doctors can diagnose a treatment that is geared specifically to that patient. The more data collected, the better the treatment plan.

Using a 'next-generation genome sequencer', the researchers are able to map three million genomic expressions using desktop computers. Data models run as high as nine terabytes. In a sense, the process is applying programming techniques to biological programming: researchers analyse data that has malfunctioned at the molecular level and try to unlock the puzzle pieces that are causing a tumour or cancerous cells to spread in the brain.

Dr Anup Madan and Dr Greg Foltz are the co-founders of CABTT. Their research is one of comparison: seeing how a gene sequence in a normal brain compares to the gene sequence in a brain cancer patient. (They also use a map of a mouse brain that was generated at the Allen Institute for Brain Science.) Foltz says that brain cancer has a one- to two-year survival rate; yet only a handful of researchers are working on a cure for it.

Located in Seattle, the centre has easy access to several other cancer research institutions and is near to Microsoft and other high-tech firms. Microsoft has a Health and Life Sciences division that helps researchers understand .NET frameworks and assists with programming models.

Collaboration is close, and both the Allen Institute and Microsoft have set up social networking systems where researchers can compare findings – sometimes even annotating data at a molecular or genomic level. "We are being moved to this position in science where the traditional approaches do not work. There are targeted drugs in the pipeline but you could not do enough trials," says Foltz. "The traditional approach is to do an MRI, start the chemotherapy, perform surgery and use cancer drugs. The challenge is to use the computer to tune the treatment according to the genetics of the patient."

What's most interesting about the gene sequencing tests is that to the human eye, all tumour data looks the same. Yet the computer knows the exact differences between two tumours, which helps it to aid medical professionals when finding cancer therapies.

Although most of the medical research has moved to computerised analysis, the researchers are also aware of a basic axiom: the starting data used must be sufficient for the desired result. In the end, the answer is in the genes and molecules of the human body.

What's needed is a computerised system that can understand our bodies completely. As work on this science is progressing, we may yet see this goal achieved.

*****

First published in *PC Plus*, issue 278

http://www.techradar.com/news/computing/can-computers-find-a-cure-for-cancer--586234?artc_pg=3