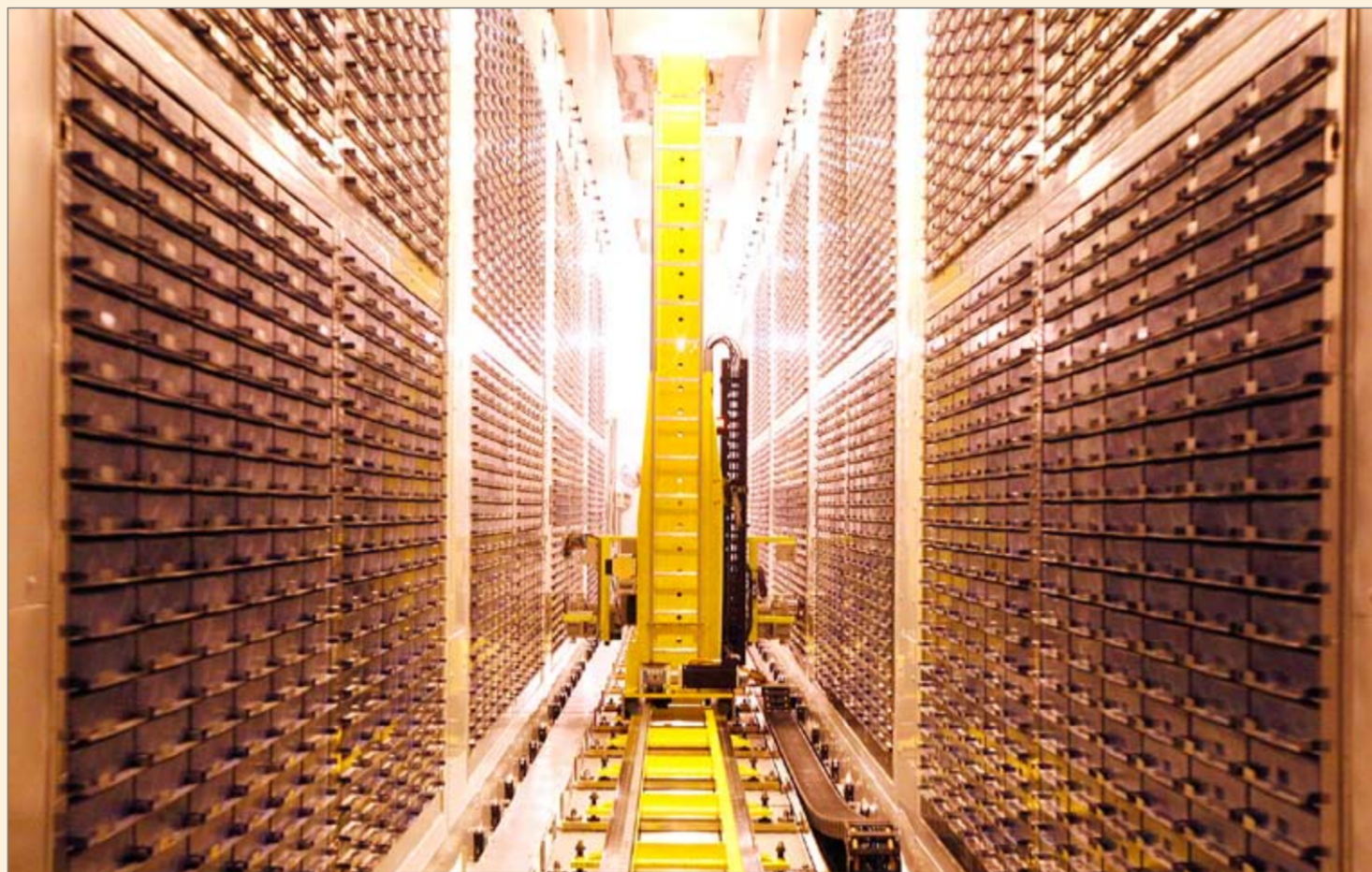


SUNDAY FEATURES

SUNDAY, MAY 30, 2010

13



Left: A storage robot deposits samples in the world's largest blood and urine sample freezer at Biobank, near Manchester, England.
Below: Biotechnicians work in the sequencing laboratory at the J. Craig Venter Institute in Rockville, Maryland.

PHOTOS: REUTERS

The geneticists who are saving lives

Ten years on from the revelation that scientists had cracked the human genome, the phenomenal capacity of modern computers is starting to exploit the potential of that discovery for the fight against disease

BY ROBIN MCKIE
THE OBSERVER, LONDON

A 6m banner hangs down a high, wood-paneled wall in the lofty entrance of the Sanger Institute in Cambridgeshire, eastern England. With its delicate bands of pink, gray and green, it looks like an enormous, abstract watercolor. Closer inspection reveals its true nature, however. Those pale stripes are made up of streams of different colored letters.

More than a million — 1,346,000 to be precise — are printed on the banner. And not just any old letters. Only As, Cs, Gs and Ts, each representing one of the chemical bases of DNA, the stuff of our genes, are inscribed there. This is no giant “watercolor,” but a printout of part of an X-chromosome, one of the packets of genetic material that lie curled within our cells and which direct chemical operations inside our bodies.

However, it is the size, not the content, of the banner that makes it so remarkable — for it turns out those letters represent only a small section of an X-chromosome. A read-out of a whole one would require a further 114 banners, each crammed with more than a million letters. And that is just the start. To display the letters that make up all 23 pairs of chromosomes in the human genome would require a staggering 2,226 banners. The institute would look like a wallpaper factory during a clearance sale.

And that is why the Sanger banner is so important. It vividly demonstrates the human genome's extraordinary complexity and puts into perspective the remarkable effort that went into its unraveling. That task was completed in June 10 years ago when the genome's first rough draft was published. This success was hailed as one of the greatest achievements of modern science, a point that was emphasized this month when US biologist Craig Venter revealed that he had assembled an entire computer-generated synthetic genome — not of a human but of a synthetic “bug” — and inserted it into bacteria which had then begun to replicate. Geneticists were playing at God, claimed newspaper headlines.

This is an exaggeration, but Venter's success does demonstrate how much genetics has changed in the past two decades. This is no longer a lab-bench science that relies on test tubes and pipettes to study living organisms. Instead, it has become utterly dependent on the swelling power of the microprocessor. Look at that banner of genetic data. It could not have been generated without the staggering capacity of modern computers.

The Sanger Institute employs more than 800 scientists and is devoted to the study of biology. Yet its banks of computers now rival those built by Cern in Switzerland to analyze results from the myriad particle collisions produced by its Large Hadron Collider. Each sequencing machine at the

center generates strips of letters from small pieces of DNA and this data is then processed by computers to produce a read-out of a full genome — of a human, or an animal like a dog or a bacterium.

It took almost a decade of processing DNA this way in the 1990s to help to produce the 3 billion letters that made up the first sequenced human genome. By 2008, the institute was processing data at a rate of 100 million letters a day, roughly a genome a month, says Phil Butcher, head of IT at the Sanger. “Today, we can handle so much data that we are producing a couple of genomes every 24 hours,” he says. “And, of course, we are making improvements all the time.”

So vast is the institute's array of computers that it is planning to build its own 2.5-megawatt power station, a combined heat and power unit, which will generate electricity for them. Biology has become as reliant on computer power as hard sciences like particle physics or astronomy. It is an astonishing transformation. The question is: How did it happen? More to the point, what has it achieved so far and what is it likely to achieve in the near future?

Chronic myeloid leukemia is a cancer

of white blood cells that usually occurs in the middle-aged and elderly. It is triggered by a genetic mutation that interferes with chemical messages that help to control cell division, leading to the uncontrolled growth of white blood cells. In the past, prognosis was poor — survival periods of around five months were typical. However, prospects for patients with chronic myeloid leukemia have changed dramatically in recent years, thanks to doctors' new understanding of the human genome, according to Dr David Adams, a Sanger Institute geneticist and cancer expert. “The drug that has changed everything is called Gleevec,” he says, “and it was derived from our new, computer-driven understanding of the genome.”

By studying a key section of the human genome, scientists realized that a mutation there produces a specific protein (called “bcr-abl”) that in turn triggers a cascade of chemical reactions in a patient's body that results in chronic myeloid leukemia. Awareness of the protein's role allowed scientists to develop a drug that could block its activity and so halt the proliferation of white blood cells.

“Patients who have the specific mutation that causes chronic myeloid leukemia will respond to the use of Gleevec and will go into remission quite profoundly,” says Adams. “It was understanding the specific genetics of this disease that led to the realization this drug could help.”

It is an encouraging tale that has since been repeated for several other genome-driven anti-cancer drugs, although it is important to note, says Adams, that the success of these drugs is hit and miss — sometimes they produce no effect. But when they do have an impact, it is invariably profound.

MEDICAL IMPACT

This has important consequences, he argues. In the next 10 years, once computing power has reached the stage when it will become possible to provide full read-outs, easily and cheaply, of everyone's genome, doctors will be able to determine exactly who will benefit from specific cancer drugs and who will not. Similarly, other types of medicine will have their efficacy judged in advance. “If you had the full genome sequence of everyone, you would know exactly who will respond to a drug and who will not. It will be of enormous benefit,” says Adams.

That goal, although distant, does reveal the importance of scientists' current obsession with decoding not just a single genome but of generating thousands of different ones, a task that now absorbs a host of follow-on projects, including the Cancer Genome Project, the 1,000 Genomes Project and others; these require the constant running of the institute's huge rooms of computers. By pinpointing changes in a few base pairs possessed by some individuals and not by others, scientists can discover why the former group might be prone to a particular disease but not the latter. Another example is provided by Crohn's disease, an extremely painful inflammatory disease of the intestines whose origins have, until recently, defied the attention of scientists.

HOW THE SANGER INSTITUTE DELVED INTO THE SECRETS OF LIFE

BY ROBIN MCKIE
THE OBSERVER, LONDON

The Sanger Institute is named after Fred Sanger, the 91-year-old British scientist who is the fourth, and only living person, to have won two Nobel prizes: one for his involvement in sequencing proteins, the other for his work on the sequencing of genes. The institute, originally called the Sanger Center, was opened in the village of Hinxton, near Cambridge, in October 1993.

During the early years, the institute worked on sequencing relatively simple life forms such as the nematode worm and various types of yeast. Later, its operations were ramped up after it received considerable backing from the Wellcome Trust. As a result, it was able to play a major part in the sequencing of the human genome project.

The project was carried out by laboratories across the world with the main work being shared between the US National Institutes of Health and the Sanger Institute. After years of effort, the first draft of the genome, which mapped out the 3 billion base pairs from which our genes are constructed, was published on June 26, 2000. “This is the most important, most wondrous map ever produced by humankind,” US President Bill Clinton announced.

DNA was first isolated by Swiss physician Friedrich Miescher who, in 1869, discovered a microscopic substance in the pus of discarded surgical bandages. But deoxyribonucleic acid's key role in biological inheritance was not fully appreciated until Francis Crick and James Watson demonstrated, in 1953, that it possessed a double-helical structure. During cell division this helix divides into its separate strands, which then migrate to the two new cells. These each grows a second, complimentary strand to form a new double helix, which takes control of that cell's functioning.

WHAT MAKES UP DNA

BY ROBIN MCKIE
THE OBSERVER, LONDON

Coiled inside each of the cells of our bodies are tiny packets of deoxyribonucleic acid, DNA, whose strands are twisted in the shape of a double helix and are made up of long strings of chemicals called bases. A total of four different bases — adenine, cytosine, guanine and thymine, usually represented by their initial letters, A, C, G and T — combine to make up our DNA.

There are around 3 billion of these bases in our DNA, which is broken up into packets of chromosomes. Each of us has 23 pairs of chromosomes, which contain a total of 26,000 genes. These special sections of DNA direct the development of the growing embryo, control our bodies' behavior and determine our physical traits: whether we will be short or tall, have brown or blue eyes or have black or blond hair. Outside our genes, there are other sections of DNA that scientists have studied. As yet, they can find no purpose for these “junk DNA” sections. On the Sanger banner, the bases that make up junk DNA are colored green while those that make up genes are colored red. Hence those pale-colored bands.

