15 human genomes each week

July 2 2008

The Wellcome Trust Sanger Institute has sequenced the equivalent of 300 human genomes in just over six months. The Institute has just reached the staggering total of 1,000,000,000,000 letters of genetic code that will be read by researchers worldwide, helping them to understand the role of genes in health and disease. Scientists will be able to answer questions unthinkable even a few years ago and human medical genetics will be transformed.

The amount of data is remarkable: every two minutes, the Institute produces as much sequence as was deposited in the first five years of the international DNA sequence databases, which started in 1982. It is a global milestone.

"I am delighted that our rapid adoption of next-generation sequencing technologies has been so successful in driving forward our biomedical research," says Dr Harold Swerdlow, Head of Sequencing Technology at the Wellcome Trust Sanger Institute. "Our internal projects, our work with external collaborators and our participation in major international programmes are all benefiting from our success. "

The Institute has major roles in projects such as The 1000 Genomes Project, The International Cancer Genome Consortium and the second round of the Wellcome Trust Case Control Consortium, all of which will depend on DNA sequence to uncover genetics variants that are important for human disease. Next-generation sequencing is also enabling the Institute's own research portfolio.

"The Sanger Institute is positioned to take on challenges and to answer questions that are daunting to most," says Professor Allan Bradley, Director. "We can explore important biomedical questions in a way that few can match, and next-generation sequencing is a vital part of that quest."

The 1000 Genomes Project, launched in January 2008, will produce a map of DNA sequence variants of unparalleled accuracy. Expected to take three years, the Project is currently in a pilot phase. The Sanger Institute is ahead of schedule and has deposited more than 300 billion bases to date, more than half of the global total so far.

"The 1000 Genomes Project is exploring the genome at a resolution nobody has attempted before," says Dr Richard Durbin, who co-heads the Project. "Our goals are ambitious and all of us are still learning, but we can already see that, through the efforts of the Sanger Institute and our partners in the consortium, the results will have a major impact on our understanding of human genetics and disease."

Next-generation sequencing platforms can uncover a wide range of variants in genomes, from single-base changes (called single nucleotide polymorphisms, or SNPs) to larger regions that can be absent from some people or duplicated in others (called copy number variants, or CNVs). Before the Human Genome and HapMap Projects - in which the Sanger Institute played a leading role - the extent of CNVs in human biology was not appreciated. With those tools to hand, scientists could begin to map CNVs across the genome and understand their role in common disease.

It is not only inherited variants that the scientists can tackle using next-generation sequencing platforms. The Sanger Institute's Cancer Genome Project team, co-led by Professor Mike Stratton and Dr Andy Futreal, has searched for genes that are mutated in common cancers for eight

years. Until now, that has meant a piecemeal approach, focussing either on a few samples or only a few hundred regions from the genome. While this is a hugely successful method, next-generation sequencing means that all genes and gene regions in many cancer samples can be looked at simultaneously.

"We have already published results from a study of lung cancer samples that illustrate the complexity and diversity of cancer genomes and have obtained more data in six months than in the previous five years," explains Professor Stratton. "The advent of the next-generation sequencing technologies allows us now to search for all the types of somatic change in cancer genomes and to begin complete resequencing of whole cancer genomes, acquiring full catalogues of somatic changes, ultimately in thousands of cancers as a leading player in the International Cancer Genome Consortium."

The Pathogen Sequencing teams, who used conventional sequencing methods to decode the genomes of MRSA, Cdiff and the parasites that cause diseases such as malaria and sleeping sickness, are gathering a rich harvest of data.

"To tackle pathogens we need to understand how they vary, how they acquire new abilities to cause infection and how they spread through populations," says Professor Julian Parkhill, Head of Sequencing and the Pathogen teams. "Together with colleagues in Vietnam and Kathmandu, we are using this new technology to uncover the fine variation that will enable us to understand the transmission of typhoid fever in South-East Asia, and with colleagues in the UK we will be able to investigate how MRSA and Cdiff spread in our hospitals."

Source: Wellcome Trust Sanger Institute