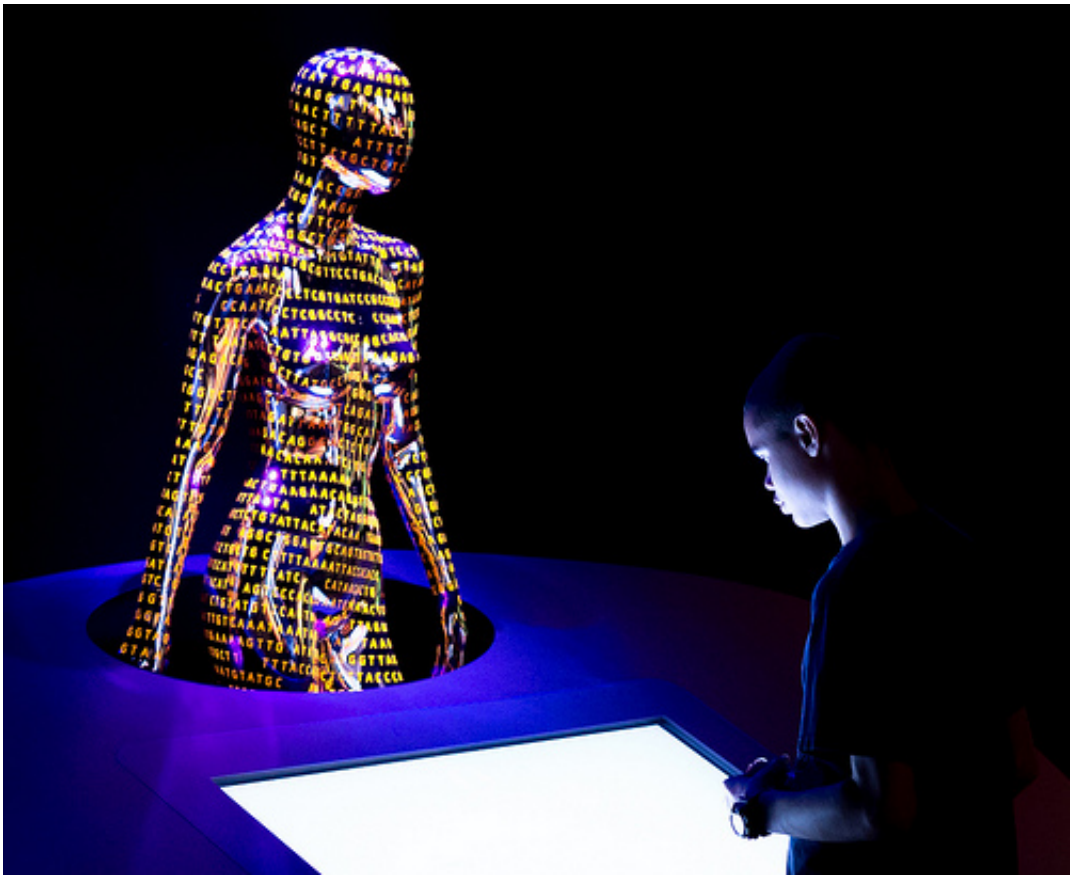
Big data opens up vast frontiers in genetic research

April 20 2015



The epigenome acts like a signpost—controlling how the underlying DNA sequence is used to produce different cells like neurons, heart cells and skin cells. Credit: Victoria Pickering

Professor Ryan Lister from University of Western Australia is using big data to examine DNA in new ways that could greatly increase the

understanding of the human body and improve agricultural techniques.

Prof Lister, who was awarded the 2014 Prime Minister's Prize for Life Science, is studying the epigenome—a layer of code sitting 'on top' of DNA that controls how cells produce different forms and functions.

"All the cells of our body have the same genome sequence—the instructions for all the parts of our bodies," Prof Lister says.

"In a way it's like a parts list for creating a cell or creating an organism."

While there may be hundreds of different cell types in an organism, these are all formed from the same underlying genetic information.

The epigenome acts like a signpost—controlling how the underlying DNA sequence is used to produce different cells like neurons, heart cells and skin cells.

Prof Lister and colleagues have developed a technique that allows them to produce the first complete maps of one major type of epigenetic signpost called DNA methylation.

According to Prof Lister, this has only been possible because of recent advances in DNA sequencers.

These instruments produce vast quantities of data, but not necessarily in terms of raw data size.

Taken directly from a sequencer, the human genome is "only" around 200 gigabytes, the equivalent of four normal BluRay discs.

However, the quantity and complexity of the data sets means this is a big data issue, requiring solutions far beyond those used for traditional data

issues.

To sequence a genome, DNA is taken from cells and 'chopped' into fragments which are 200-300 letters long.

"This is put into the DNA sequencer and it would, in parallel, determine the sequence of these DNA fragments—up to six billion of them," Prof Lister says.

"It's like taking a huge book and shredding it into billions of pieces and trying to stitch it back together.

"The bottlenecks aren't in the experiments but in analysing, interpreting and processing the data."

The techniques developed by Prof Lister and his colleagues can be used to identify the epigenome of any cell in any organism.

The next step in the research will examine how the epigenome in plants changes due to difficult conditions such as drought and how it could be engineered to produce tougher plants.

Prof Lister's research will also study the changes in the epigenome in the human brain during development, which could improve understanding of learning and memory, as well as neurological disorders.

Provided by Science Network WA