IMPROVING THE ACCURACY OF GENETIC SEQUENCING

A new genetic sequencing technology developed by BIOPIC researchers has improved ways to analyse DNA.

DNA sequencing is an indispensable laboratory technique under rapid development. It unlocks genetic information carried in DNA segments, the quality of which relies on the level of sensitivity, accuracy and speed of sequencing technology. Eliminating errors in DNA sequencing has proved one of the biggest challenges in the field.

Yanyi Huang's laboratory at PKU's BIOPIC is committed to advancing technologies to facilitate genome sequencing and its applications. The team has developed a new strategy to achieve ultra-accurate sequencing, which could be part of the 'next-generation' of DNA sequencing, according to Huang.

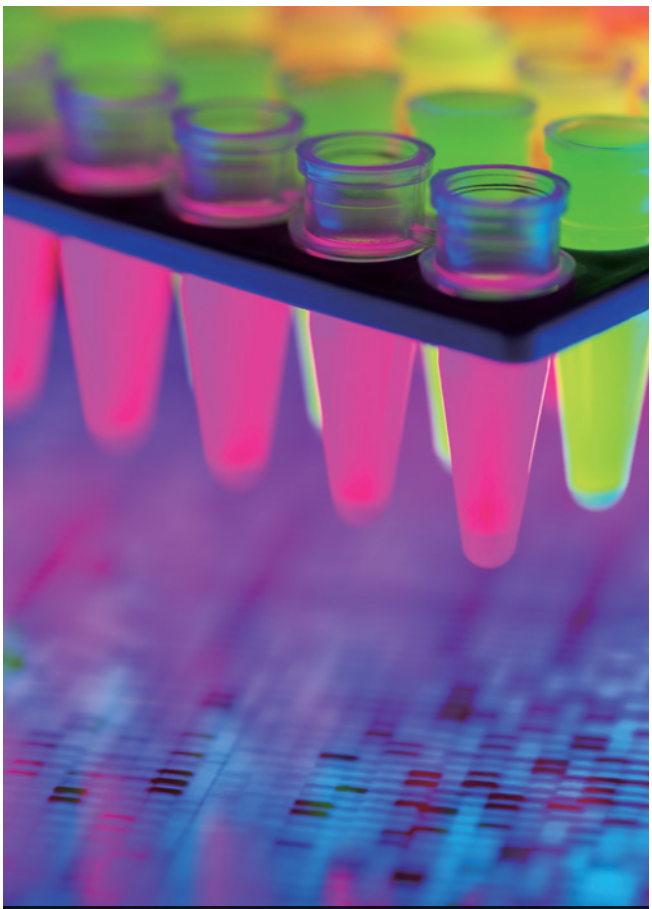
Named error-correction code (ECC) sequencing, Huang's method builds on an emerging technology called fluorogenic sequencing, a sequencing-by-synthesis method that adds, one at a time, nucleotides labelled with a dye into microwells that contain target DNA fragments. When the nucleotide is incorporated into new DNA strands, the dye becomes fluorescent, and the intensity of the signal makes detection possible.

Each sequencing chemistry has its pros and cons, and imperfections lead to errors when deducing sequences from the signals detected during sequencing reactions. At present, all 'next-generation sequencing' methods have higher error rates than the first-generation Sanger sequencing approach.

In ECC sequencing, Huang's team adjusts the way nucleotides are added, resulting in extra information for each DNA fragment sequenced. Drawing analogies with the encoding and decoding strategies used in information communication for error detection and correction, Huang's team has embedded information redundancy in sequencing reads. The key is a specially designed algorithm, which combines the results and acts like a 'course corrector', identifying and fixing any errors.

The team has built a laboratory prototype to perform the ECC experiments and found out that it can generate error-free sequences up to 200 base pairs long, with raw accuracy reaching more than 98%, as reported in their 2017 *Nature Biotechnology* paper.

"We are glad to have developed an elegant way that enables accurate identification of extremely rare genomic variations," Huang says. Huang's group has been investigating the potential of ECC sequencing for nearly a decade, having explored more than 30 fluorophores — chemical compounds that produce fluorescent signals when activated — to find the ones with the brightest and longest lasting signals when combined with nucleotides.



Huang's group has explored more than 30 fluorophores to find the ones with the brightest and longest lasting signals when combined with nucleotides.

To perform ECC sequencing on a larger scale and to make it more user-friendly, Huang is leading his team to develop an automated microfluidic device and prototype.

Early data show comparable results that could place ECC sequencing on equal footing with available commercial sequencing technologies. Huang is confident that ECC sequencing could be a potent tool in precision medicine, including detecting genetic mutations in tumour tissues. ■