Error Correction Code Sequencing and Bit Sequencing

Yanyi HUANG

Department of Advanced Materials and Nanotechnology, Peking University, Mainland China

Email: yanyi@pku.edu.cn

Eliminating errors in next-generation DNA sequencing has proved challenging. I am going to talk about our newly developed error-correction code (ECC) sequencing, a method to greatly improve sequencing accuracy by combining fluorogenic sequencing-by-synthesis (SBS) with an information theory–based error-correction algorithm. ECC embeds redundancy in sequencing reads by creating three orthogonal degenerate sequences, generated by alternate dual-base reactions. This is similar to encoding and decoding strategies that have proved effective in detecting and correcting errors in information communication and storage. In a proof-of-concept experiment, we have shown that, when combined with a fluorogenic SBS chemistry with raw accuracy of 98.1%, ECC sequencing provides single-end, error-free sequences up to 200 bp. I will also cover some of our recent progresses that aim to achieve high throughput (over millions of reads) ECC sequencing using a prototype and a microfluidic-based flow and sealing scheme. ECC approaches should enable accurate identification of extremely rare genomic variations in various applications in biology and medicine.