Error correction coding paves an avenue to ultra-high accuracy of DNA sequencing

With the support from the National Natural Science Foundation of China, Prof. Huang Yanyi (**黄岩谊**) led a team at Peking University to demonstrate a novel approach, which combined fluorogenic sequencing-by-synthesis (SBS) chemistry with an information theory-based error-correction coding scheme to greatly reduce the sequencing errors. This work was published online on November 6, 2017 in *Nature Biotechnology* (doi: 10.1038/nbt.3982).

In the paper the authors present a strategy for DNA sequencing, ECC sequencing, that can greatly improve sequencing accuracy and read length using a dual-base flowgram combined with fluorogenic SBS chemistry. The ECC sequencing approach allows a mixture of two types of nucleotide substrates to be introduced into each reaction cycle. The synthetic strands expose free $3'-OH$ groups that can be continuously extended until no nucleotides in the mixture can be further incorporated. Although each of such reactions provides only one degenerate sequence with partially defined base composition, one DNA template can be sequenced three times with three orthogonal combinations of dual-base mixes to provide three degenerate sequences, from which an unambiguous sequence can be accurately deduced. Sequencing errors in any degenerate sequence can be further identified and corrected in this approach. The ECC sequencing approach does not require more sequencing reaction time than SNA, but provides higher confidence of the sequence accuracy through the extra information received in the orthogonal flowgrams. In principle, ECC sequencing strategy can be applied to any SNA-based sequencing chemistries. The research team has built a laboratory prototype DNA sequencer to demonstrate the complete ECC process using fluorogenic SBS chemistry, and obtained single-end reads up to 250 bp with the first 200 bp free of error. ECC approaches should enable accurate identification of extremely rare genomic variations in various applications in biology and medicine.

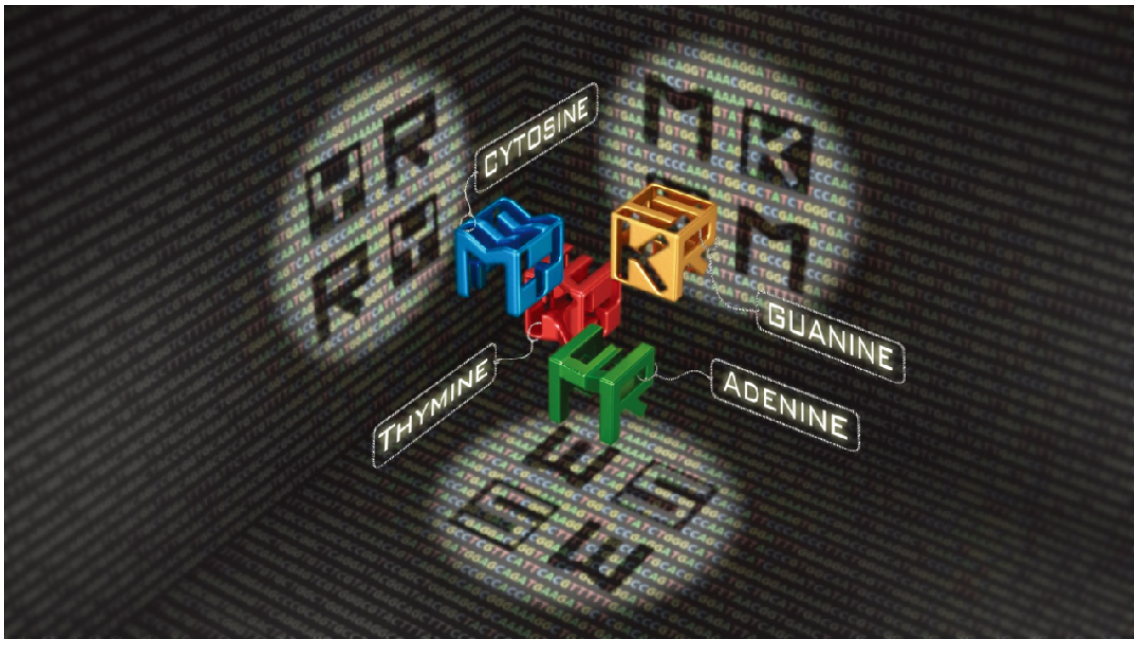

Figure The degenerate bases can be sequenced three times, similar to having three orthogonal projections.