Predictive tool to identify sequences causing mutations, genome instability and diseases

May 29 2015, by Vanessa Loh

Scientists from A*STAR's Bioinformatics Institute (BII) have developed an analytical model and computational tool to rapidly and accurately predict the occurrence and locations of R-loop Forming Sequences (RLFSs) in any genome or artificial nucleic acid sequences. R-loops, which are three-stranded RNA and DNA hybrid structures, can be crucial to many normal biological processes and have also been associated with triggering mutations, DNA breaks and diseases. These hybrid structures provide intriguing possibilities for use as novel targets for diagnostics and treatment of diseases including cancer, autoimmune and neurodegenerative conditions.

While R-loops were first described in 1976 and were for many years associated with only a few specific genes, it is only in recent years that understanding of their critical function and prevalence in the genomes has advanced, revolutionising the field.

Scientists from BII's Genome and Gene Expression Data Analysis Division developed the Quantitative Model of R-loop Forming Sequence finder (QmRLFS-finder), making it freely available to accelerate research in this area. Using the QmRLFS-finder, the scientists made a surprising discovery that 75% of well-annotated human genes and/or their vicinities contain RLFS. The tool has also proven to have an accuracy of between 80 to 90% in predicting the location of RLFS in any genome sequence. The high accuracy would significantly accelerate R-loop detecting and dramatically reduce the cost and time taken compared to currently available experimental methods, paving the way

for further improvement and development in the relatively nascent field of R-loop biology.

Its benefit to the global research community and the industry is apparent considering that there is currently only one experimental method developed for genome-wide location of R-loops, which has been applied at the genome level only to a single cell line. This and current non-genome level experimental methods to detect R-loops in a double-helix DNA sequence take a long time, high costs and a high level of expertise found in only a few laboratories worldwide.

The use of the QmRLFS-finder would empower users to conduct research in this field and contribute to growing knowledge of the importance of R-loops to human biology and diseases. The increased understanding of R-loops would also support the development of novel therapies targeting these hybrid structures.

Dr Vladimir Kuznetsov, head of the division and Senior Principal Investigator who led the development of the tool, said "We developed this predictive tool as we foresee that it has a great number of applications that can contribute to the advancement of this field. It is our hope that an increased understanding of R-loop formation and functions will in turn allow us to better predict and treat various diseases."

The QmRLFS-finder has been accessed more than 1200 times by users from more than 20 countries since a paper on its development and use had been recently published in *Nucleic Acids Research*.

More information: "QmRLFS-finder: a model, web server and stand-alone tool for prediction and analysis of R-loop forming sequences." *Nucl. Acids Res*. first published online April 16, 2015 DOI: 10.1093/nar/gkv344

Provided by Agency for Science, Technology and Research (A*STAR), Singapore