Predicting the human genome using evolution

October 29 2015

To gain a clearer picture of health and disease, scientists have now provided an independent reference for all human variation by looking through the evolutionary lens of our nearest relatives. Such a powerful approach has been developed by Temple University professor Sudhir Kumar and colleagues and was detailed in the advanced online publication of *Molecular Biology and Evolution*.

"There are two ways to generate a map of the human genome variation: one is to get genomes of all the humans and build a compilation as the 1,000 Genomes Project and others have undertaken," said Kumar, a Temple University professor and director of the Institute for Genomics and Evolutionary Medicine (iGEM). "The alternative, which is the basis of our approach, is to compile all genome data from other species and predict what the human sequence reference should be."

By observing evolution's "greatest hits" (and misses) and the history of the major themes and patterns of genome conservation (and divergence) across many species, Kumar's approach predicts probable mutations that will be found among people and the fate of human variation.

His research team relied on an evolutionary tree that included 46 vertebrate species spanning over 500 million years of life on Earth to predict the evolutionary probability (EP) of each possibility at each position of our genome. They applied their new method on all protein-coding genes in the human genome (more than 10 million positions). Consistent with the knowledge that most mutations are harmful, they

found very low EPs (lower than 0.05) for a vast majority of potential mutations (94.4 percent).

Next, they produced a complete evolutionary catalog of all human protein variation, or evolutionary variome (eVar), that can be used to better understand human diseases and adaptations. And, it can be directly applied to the genomes of any other species. Their eVar was also compared against available human sequence data from the 1,000 Genomes Project to look at benign and disease mutations, and found that the use of EPs could correctly diagnose them. They also used a cancer benchmark dataset to show that EPs accurately predicted cancer-related mutations.

Lastly, they found a large number (36,691) of variations, that according to the EP data were evolutionarily improbable (EP less than 0.05), but were found 100 percent of the time in the 1,000 Genomes Project data—-which Kumar suggests could be strong candidates for adaptive evolution—-and what may make us uniquely human.

"The fascinating part of the story is that once we know what our ancient evolutionary history predicts our sequence to be, then we can compare this expectation to what we observe in human populations today. When there is a discordance such that an unlikely variant is found in many people, it directly indicates that something has changed about us or the protein," said Kumar.

Provided by Oxford University Press