Comparing Japanese genome data with distant relatives

October 15 2018

The dataset from Tohoku Medical Megabank Organization represents the entire Japanese population. Credit: Tohoku University Tohoku Medical Megabank Organization

A genome bank for the Japanese population can better identify rare genetic variants and disease susceptibilities by adding samples from distant areas of the country.

Incorporating samples from different areas far apart in Japan makes it easier to find rare genetic variants on a nationwide scale, according to a new study published in *BMC Genomics*. It is important to identify these

markers as they might indicate which genetic variants lead to certain diseases in a population.

Genetic variations or mutations are common, and pinpointing which variations are responsible for specific traits or diseases is a challenge. One way geneticists try to do this is with genome studies that cover an entire population and look for rare sequences that have not been identified before. They use genotype imputation, a statistical technique to identify unknown genotypes in order to search through large datasets and pick out a group of genes inherited from a parent called haplotypes, which contain a cluster of variations called single nucleotide polymorphisms (SNPs).

Previously, a dataset of 1,070 genomes containing haplotype information was produced as part of the Tohoku Medical Megabank project. Launched in the aftermath of the 2011 earthquake, the project aims to develop tailor-made medical diagnosis and treatment using people's genetic information in Miyagi and Iwate prefectures. The data, known as the '1KJPN' reference panel, was created based on samples from Miyagi.

In a present study, a research team from the Tohoku Medical Megabank Organization at Tohoku University investigated how accurately Miyagi samples represent the diversity of genome variations of the entire Japanese mainland population. They compared the 1KJPN panel with 144, 39 and 35 genome samples from the areas of Iwate, Nagahama and Aki, respectively.

The results showed that while the Miyagi data was a sufficient representative of the entire population, combining the 1KJPN dataset with genome samples from Iwate, Nagahama and Aki improved the efficiency of the genotype imputation, particularly in identifying rare variants or SNPs on a nationwide scale. The combined data was also more accurate than Japanese samples in the 1000 Genome Project, the

largest human genotype database created by the European Bioinformatics Institute.

The comparison of the genome samples from the four areas indicates why the combined data is stronger. The researchers found genetic differentiations increased with distance from Miyagi. For example, populations in the neighbouring regions of Miyagi and Iwate were most similar to each other than populations in Nagahama and Aki, which are more than 700 km (435 miles) and 1,000 km (620 miles) south, respectively.

A deeper analysis of individual genomes identified rare SNPs that are present in Iwate, Nagahama and Aki, but not in Miyagi. More variants were observed in the areas furthest from Miyagi, showing the importance of collecting genomic data from disperse areas to capture a wide range of rare variations.

Interestingly, the team found that Aki samples formed a distinct genome cluster, indicating the Aki population on Shikoku Island is genetically different from populations on neighbouring Honshu Island. This is contrary to the existing notion that genetic differentiation is minimal among populations on Japan's main four islands, which are close together and connected by bridges or tunnels. However, the genetic differences between Aki and the other areas are much smaller than the genetic differences found between Japanese populations and mainland China populations.

In the next step, the researchers hope to collaborate with other groups to produce genetic data from other parts of the country. A combined dataset could be used to identify specific genes associated with common diseases.

 More information: Jun Yasuda et al. Regional genetic differences

among Japanese populations and performance of genotype imputation using whole-genome reference panel of the Tohoku Medical Megabank Project, *BMC Genomics* (2018).

Provided by Tohoku University