

# Korean Genome Project data released for public use

May 28 2020, by Bob Yirka

---



Credit: CC0 Public Domain

A team of researchers with members affiliated with a large number of institutions in Korea, two in the U.S. and one in the U.K. has released data from the initial phase of the Korean Genome Project (Korea1K). In

their paper published in the journal *Science Advances*, the group notes that the data includes information describing 1,094 whole genomes along with 79 quantitative clinical traits.

Korea1K is the largest genome sequencing [project](#) conducted thus far in South Korea—it began back in 2006, and has been supported by a variety of sources since that time. Its goal is to collect, analyze and distribute Korean genome information for use in clinical and ethnographic studies. One of the main uses of the data is expected to be in the area of cancer research.

In the first stage of the project, 1,094 whole genomes were sequenced with an average depth of 31x. The data for each sequenced genome has been paired with 79 associated traits from the person whose [genome](#) was sequenced. In their paper, the team highlights notable details of the data.

Some examples of notable details: Thirty-nine million single nucleotide variants and indels were identified, half of which were found to be doubletons or singletons. The researchers also found that there was better imputation accuracy with Korea1K than there had been with the 1000 Genomes panel, making filtering out cancer samples more effective. They also note that 1,007 of the genomes sequenced as part of the project were newly generated and that information in the data included characterizations of indels, SNVs, transposable element insertions, [copy number variants](#) and human leukocyte antigen types, which have already been compared with [data](#) from other populations.

The researchers also note that over 70 percent of the doubletons or singletons that have been identified thus far have never been reported before—and fewer than 20 percent of them were designated as common. They also found that there were more deletions than insertions in the indels, suggesting possible skewing of variant calling. They conclude their remarks by suggesting that the dataset should provide a strong

reference panel going forward, thereby enhancing personalized medical applications for the people of Korea.

**More information:** Sungwon Jeon et al. Korean Genome Project: 1094 Korean personal genomes with clinical information, *Science Advances* (2020). [DOI: 10.1126/sciadv.aaz7835](https://doi.org/10.1126/sciadv.aaz7835)

© 2020 Science X Network

Citation: Korean Genome Project data released for public use (2020, May 28) retrieved 14 February 2024 from <https://medicalxpress.com/news/2020-05-korean-genome.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.