



August 29, 2014

Tohoku University

Tohoku Medical Megabank Organization

**Partial Release of Data from Tohoku Medical Megabank Project's  
Whole Genome Reference Panel**

~ToMMo to enable website search of SNP frequencies with allele frequencies >5%~

Tohoku University Tohoku Medical Megabank Organization (ToMMo) is conducting large-scale cohort studies, along with creating a completed Whole Genome Reference Panel<sup>1</sup> following its completion of whole genome sequencing for 1,000 people<sup>2</sup> in late November 2013. We hereby announce that based on the progress made in our data analysis, ToMMo has decided to publicly make available the data from its whole genome sequencing for 1,000 people, namely, data on single nucleotide polymorphism (SNP)<sup>3</sup> frequencies with allele frequencies<sup>4</sup> of 5% or greater. Our platform will offer capabilities to perform SNP searches from the reference SNP ID number and Gene Symbol and to identify location information on the international human genome reference sequence.

**Background**

Since 2012, Tohoku University and Iwate Medical University have been working on the Tohoku Medical Megabank project designed to support the reconstruction from the Great East Japan Earthquake. The project is being implemented through ToMMo and Iwate Tohoku Medical Megabank Organization, established by Tohoku University and Iwate Medical University, respectively. Both organizations are currently conducting cohort studies<sup>5</sup> to create a biobank<sup>6</sup> of 150,000 people in Miyagi and Iwate Prefectures. As of August 2014, around 50,000 people have registered since May 2013. By November 2013, ToMMo completed whole genome sequencing for 1,000 healthy participants in the cohort studies living in Miyagi Prefecture. This is the world's first homogenous (using independent facilities and equipment and identical protocols) and high-quality (sequencing at 30x coverage per person; 90 billion bases on average) whole genome sequencing of 1,000 people. ToMMo has continued to perform studies aiming to create the Whole Genome Reference Panel based on the whole genome

sequencing data on 1,000 people. We have hereby decided to release the SNP frequency (allele frequency >5%) data obtained from the studies we have completed and make the data available for use by researchers across the world.

## **Method**

ToMMo is conducting data analysis with a view to creating the completed Whole Genome Reference Panel for 1,000 people using a super computer system<sup>\*7</sup> installed in late March 2014. It is expected that the completed reference panel will serve as a foundation for wide-ranging medical and science research and open the way towards personalized preventative health and personalized medicine.

ToMMo has made progress in its data and verification analyses. On this basis, SNP frequencies from the Whole Genome Reference Panel with allele frequencies of 5% or greater will be made publicly available through the National Bioscience Database Center (NBDC)<sup>\*8</sup> and ToMMo's portal site<sup>\*9</sup>, Integrative Japanese Genome Variation Database. The portal site will offer capabilities to perform SNP searches from the reference SNP ID number and Gene Symbol and identify location information on the international human genome reference sequence. (The frequency information being released integrates the data of many people and is not the sequences, etc. of individual people.)

The data being released will be taken from the draft version of the ToMMo Whole Genome Reference Panel which ToMMo aims to create. The information disclosure constitutes one effort to promote broad applications of the outcomes of whole genome sequencing which began in 2013.

## **Explanation of the Terms**

\*1 Whole Genome Reference Panel: Tohoku University ToMMo aims to create the Whole Genome Reference Panel, which will integrate the results of whole genome sequencing of a vast number of people and will compile data on the frequency of DNA sequence polymorphisms, among other data. ToMMo expects that the Panel will offer reference data for future genome research. The whole genome sequencing of 1,000 people will enable the detection of rare SNPs up to a frequency of approximately 0.5%. The Panel will be integrated with data on blood and urine analysis as well as medical records and information on lifestyle habits, and will be created to serve as a database with a wide range of applications for research in Japan targeting next generation medicine.

\*2 completion of whole genome sequencing for 1,000 people: In November 2013, ToMMo completed whole genome sequencing for 1,000 healthy participants in the cohort studies of the Tohoku Medical Megabank project living in Miyagi Prefecture.

Reference: Press Release “Tohoku Medical Megabank Organization High-accuracy whole genome sequencing of 1,000 healthy Japanese people: A base to search for the causes of diseases ~Collection of over 15 million new gene polymorphisms~”

<http://www.megabank.tohoku.ac.jp/english/news/detail.php?id=684&c1=4>

\*3 single nucleotide polymorphism (SNP): Single nucleotide polymorphism refers to a variation in a genome base sequence that occurs only at one base at a frequency of 1% or more in a group.

\*4 allele frequencies: Here, allele frequency refers to the frequency of bases (A, T, G, C) of which few observations were made among the polymorphisms of a group. In this example, allele frequency is the frequency among the selected group of 1,000 Japanese people. It is known that there are SNPs whose allele frequencies vary significantly among different populations.

\*5 cohort studies: A study in which groups of people with certain characteristics are followed over time to determine the relationship between diseases and factors, including environmental factors, such as lifestyle habits, and genetic factors.

\*6 biobank: A system of collecting and storing biological samples and providing them for research purposes. The Tohoku Medical Megabank Project's biobank collects biological samples, such as blood and urine samples, from the participants of the cohort studies (a study in which groups of people with certain characteristics are followed over time to determine the relationship between diseases and factors, including environmental factors, such as lifestyle habits, and genetic factors).

\*7 super computer system: ToMMo has both databank and sequencing functions since July 23, 2014. It has started full-scale operations of Japan's largest supercomputer system in the life sciences field.

\*8 National Bioscience Database Center (NBDC): An institution established at the Japan Science and Technology Agency to promote the integrated use of life science research databases, which are created by research institutes and other organizations in Japan.

\*9 ToMMo's portal site "Integrative Japanese Genome Variation Database"

URL: <http://ijgvd.megabank.tohoku.ac.jp/>

**Contact**

Tohoku University Tohoku Medical Megabank Organization

Group of *in silico* analysis

Prof. Masao Nagasaki

Tel: 022-273-6051

E-mail: [ijgvd@ig-nagalab.megabank.tohoku.ac.jp](mailto:ijgvd@ig-nagalab.megabank.tohoku.ac.jp)

(Press officer)

Tohoku University Tohoku Medical Megabank Organization

Fuji Nagami

Tel: 022-717-7908 Fax: 022-717-7923

E-mail: [f-nagami@med.tohoku.ac.jp](mailto:f-nagami@med.tohoku.ac.jp)