“Data in the biobank can be used for practically any health research,” says research manager Kaisa Silander, who has contributed to the descriptions and classification of THL Biobank’s population cohorts.

“We have plenty of data from various population research studies and when these are put together, researchers have an impressively large material at their disposal.”

Silander considers the material to be significant also by international standards. “I think it is as valuable as UK Biobank or Estonia Biobank. When biobank data is combined with health registers, we have health information about certain people over a period of 40 years. We know exactly, for example, which diseases a certain person has had.”

Silander has been working with cohort data for a long time. Projects funded by e.g. EU Health programmes built the infrastructure. After that she helped combine the metadata for research cohorts managed by the Finnish Institute for Health and Welfare (THL) into a searchable database. “We built a common infrastructure for data storage, because all these cohorts contain similar data. First we created a catalogue in which these cohorts were described in a uniform manner. We saved the variable metadata into the same database, in order to make them easy to search and to find. The description of the variables was done in Finnish and English. The same protocol will also be used for future cohort description,” says Silander.

THL Biobank was established in 2014. The population cohorts collected by THL had two traditional lines of research. Between 1965 and 1980 population health examination surveys were carried out by driving around the country on buses combining biobank data with data from health registers enables research towards personalised treatment

THL Biobank contains a large amount of data on the health and lifestyles of the Finnish population, the collection of which began already in the 1960s. When this is combined with genetic data stored in the biobank and with national health registers, illness risk factors can be effectively identified and predicted.
modified into mobile clinics. The mobile clinic examined more than 50,000 Finns around the country. THL's population cohorts are composed of information provided by the participants of the health surveys, of a baseline clinical examination, and of a sample bank. Most of the samples have been turned into data. For some participants up to 40 years or more of health register follow-up can be obtained. The mobile clinic was followed up in 2000–2001 by the nationwide Health 2000 study and a further follow-up of the latter in 2011–2012. The mobile clinic research included pulmonary diseases, heart conditions, anaemia and iron deficiency, diabetes, kidney and urinary tract conditions, thyroid conditions, calcium metabolism diseases and coronary artery disease. Later the range of diseases has been extended, and today biobank studies can be conducted on any disease.

The FINRISK study has collected information mainly about risk factors for cardiovascular diseases and diabetes, originally in eastern and central Finland. Later the study was expanded to cover several other areas in Finland, and the range of diseases studied was also increased. In 2017, these two lines of research were combined in the FinHealth study.

"The cohorts contain plenty of lifestyle data obtained with survey questionnaires," says Silander.

The questionnaires include questions about smoking, alcohol use and diet, sleeping and exercise habits. The health checks have included height, weight, blood pressure and other measurable matters. Blood, stool and urine samples have also been taken.

"The samples are used to determine biomarker levels, such as lipids and inflammation (C-reactive protein, CRP). CRP is a protein produced by the liver, and its concentration in the body increases rapidly during inflammation."

Kaisa Silander hopes that in the future we would be able to obtain much more biomarker data describing changes in a person's body that may give an indication of an illness.

"There are good methods for high throughput biomarker analysis. Currently we have at the biobank information about more than two hundred biomarkers, which were analyzed by NMR spectroscopy. However, there are laboratories that can produce thousands of biomarkers from a single serum sample. This type of information, combined with the FINRISK material, for example, would be valuable. There are still suitable serum samples for many of the FINRISK participants."

"The cohorts contain plenty of lifestyle data obtained with survey questionnaires," says Silander.

THL Biobank’s sample collections contain genotype data of densely mapped single variants of DNA, and also data of whole genome sequence, and exome sequence data of the protein coding regions of DNA.

All exome and whole genome sequence data related to the FINRISK and Health 2000 were produced at Washington University and the Broad Institute / Massachusetts Institute of Technology, which are among the leading laboratories for genomic studies. The FINRISK cohort includes data of 10,000 exome sequences and 4000 whole genome sequences.

"Now it will be possible to create personalized treatment methods. When lifestyle data is combined with genetic data, better drug treatments can be developed," says Marjonen.

According to Marjonen, DNA samples can also be used to produce epigenetic data.

"If you can combine health register, questionnaire and genome data, it certainly enables extensive research areas," says postdoctoral researcher Heidi Marjonen. She works as a genome expert at THL Biobank, processing genomic data of all THL Biobank cohorts.

Genomic data of cohorts

Thanks to the FinnGen research project, THL Biobank may offer genetic data for other biobank studies.

The goal of the FinnGen project, started in autumn 2017, is to collect the genome data of half a million Finns. The project utilises samples collected by all Finnish biobanks. Genome data is combined with data available in national healthcare registers. This gives a better understanding on how diseases develop, and identify new treatments. Phenotypes used in FinnGen are age, gender, height, weight and smoking. Genome data created in the FinnGen project is returned twice a year to the biobanks.

"If you can combine health register, questionnaire and genome data, it certainly enables extensive research areas," says postdoctoral researcher Heidi Marjonen. She works as a genome expert at THL Biobank, processing genomic data of all THL Biobank cohorts.

"Now it will be possible to create personalized treatment methods. When lifestyle data is combined with genetic data, better drug treatments can be developed," says Marjonen.

According to Marjonen, DNA samples can also be used to produce epigenetic data.

"If you can combine health register, questionnaire and genome data, it certainly enables extensive research areas," says postdoctoral researcher Heidi Marjonen. She works as a genome expert at THL Biobank, processing genomic data of all THL Biobank cohorts.

"Now it will be possible to create personalized treatment methods. When lifestyle data is combined with genetic data, better drug treatments can be developed," says Marjonen.

"If you can combine health register, questionnaire and genome data, it certainly enables extensive research areas," says postdoctoral researcher Heidi Marjonen. She works as a genome expert at THL Biobank, processing genomic data of all THL Biobank cohorts.

"Now it will be possible to create personalized treatment methods. When lifestyle data is combined with genetic data, better drug treatments can be developed," says Marjonen.

"If you can combine health register, questionnaire and genome data, it certainly enables extensive research areas," says postdoctoral researcher Heidi Marjonen. She works as a genome expert at THL Biobank, processing genomic data of all THL Biobank cohorts.

"Now it will be possible to create personalized treatment methods. When lifestyle data is combined with genetic data, better drug treatments can be developed," says Marjonen.

"If you can combine health register, questionnaire and genome data, it certainly enables extensive research areas," says postdoctoral researcher Heidi Marjonen. She works as a genome expert at THL Biobank, processing genomic data of all THL Biobank cohorts.

"Now it will be possible to create personalized treatment methods. When lifestyle data is combined with genetic data, better drug treatments can be developed," says Marjonen.

"If you can combine health register, questionnaire and genome data, it certainly enables extensive research areas," says postdoctoral researcher Heidi Marjonen. She works as a genome expert at THL Biobank, processing genomic data of all THL Biobank cohorts.
Personal polygenic risk score

Heidi Marjonen took part in a study in which more than 3,000 persons were given information of their 10-year disease risk for the most common diseases in Finland. When genetic data is combined with clinical data, one can predict individual’s disease susceptibility. The overall 10-year risk evaluation was based on the genetic data and other traditional risk factors, such as gender, age, body mass index, blood pressure and cholesterol levels. The genetic risk was calculated as a personal polygenic risk score, taking into account millions genetic variations.

This data was stored on the ePouta platform for sensitive data in Finland’s ELIXIR node CSC, which enables a secure transfer between the portal’s user interface and the database.

The polygenic risk score, says Heidi Marjonen, is a major research trend. The risk score is a single value that reveals the genetic burden of a disease.

“Researchers receive information about the genome in a convenient way, which allows to study the effect of genome on a disease or other traits in an individual.”

Ari Turunen

Researchers can buy a service from THL Biobank in which a personal polygenic risk score for a specific disease for a person in cohort under study is calculated. This means that the researcher does not have to request the ‘raw’ genetic data and learn to do the necessary calculations.