Finnish research team sequences the genomes of thousands of individuals with diabetes to look for genetic risk factors

Diabetes is a major chronic disease that comes with a number of associated conditions that pose remarkable challenges. Such diseases include diabetic kidney disease, diabetic retinopathy, coronary heart disease and strokes. Now, a Finnish research team sequences the genomes of thousands of individuals with diabetes to look for genetic risk factors.

Individuals with diabetes have a higher risk of cardiac diseases in comparison to the rest of the population. A third of individuals with type 1 diabetes, earlier also called juvenile diabetes, develop kidney disease that has a remarkable impact on mortality and the risk of cardiac disease. On the other hand, diabetic retinopathy is the most significant cause of blindness among working-age population.

Finnish children and young adults have the highest risk of type 1 diabetes in the world. Type 2 diabetes is often considered a disease of the western life style, but the highest patient concentrations are found in middle-income countries with China and India topping the statistics of individual countries.

"Diabetes is associated with remarkable and severe complications.

The associated conditions have a major impact on the quality of life and life expectancy," explains **Niina Sandholm**, Genetic Epidemiologist at Folkhälsan Research Centre. Sandholm is involved in the FinnDiane research project, the objective of which is to identify hereditary and environmental risk factors predisposing to diabetic complications. FinnDiane-study is a collaboration project between University of Helsinki, Helsinki University Hospital (HUS) and Folkhälsan Research Center

According to Sandholm, genetic data could be beneficial to young patients in particular, already at an early stage prior to the emergence of risk factors.

"Currently, the use of genetic data in clinical treatment is mostly associated with rare diseases, but our results and earlier research suggest that extensive genetic data could also be utilised in the early prevention of common diseases."

One of the largest research projects on diabetes

FinnDiane, established by Professor **Per-Henrik Groop in** 1997, is a follow-up study participated in by almost 8,000 individuals with type 1 diabetes. The participants are recruited at 80 hospitals and health centres across Finland. It is one of the most extensive research materials on type 1 diabetes and associated complications in the world. Now, this material is used for sequencing the genome of over 1,800 patients.

Sandholm has experience in research projects where genome-wide association study (GWAS) is applied as research

method. The method is particularly useful when the genetic background of the studied disease is complex. It enables identifying genetic variants that either increase the risk of diabetes or protect from the disease. The GWAS method involves identifying genetic variants from the participants' blood samples. The number of these variants ranges from hundreds of thousands to millions, and the number of patients may vary between thousands and hundreds of thousands.

A GWAS study on 5,600 FinnDiane participants with type 1 diabetes revealed, for example, a new genetic locus related to cardiac diseases close to the DEFB127 gene. It is the most extensive study of its kind to date. Locus means the location of a DNA sequence on a chromosome. The variation of a sequence is called an allele.

The same study that identified the DEFB127 gene also revealed other genetic factors predisposing to cardiac diseases.

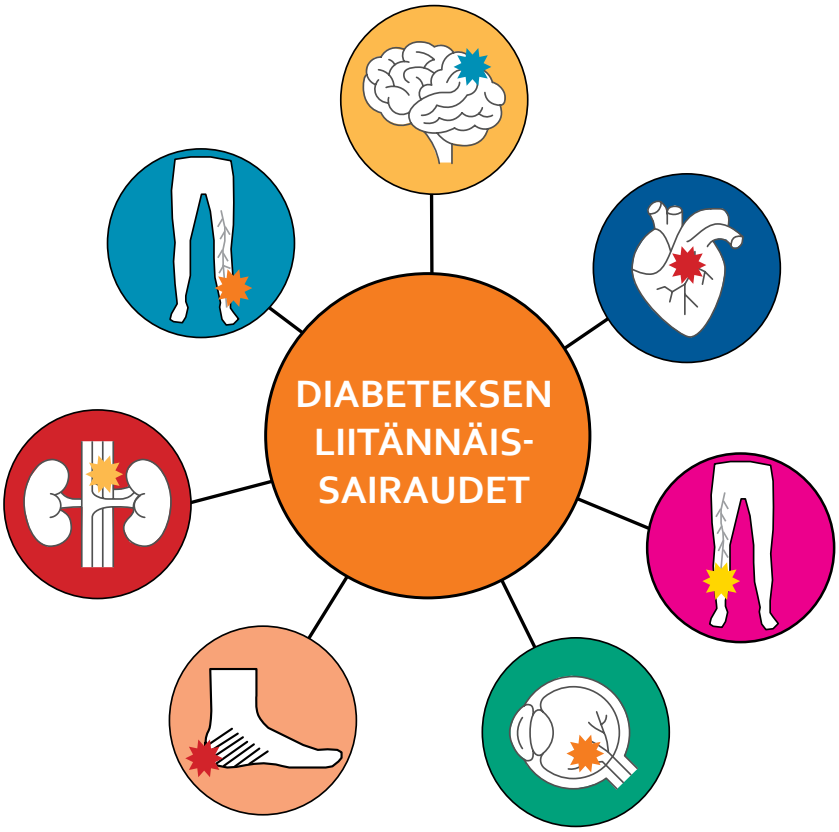
"Many predisposing genetic factors have been identified for cardiac diseases and other common medical conditions. One of the most significant factors lies within the region of genes CDKN2A and CDKN2B. Diabetics have a remarkably higher risk of cardiac disease than the rest of the population, and there is little knowledge of the related genetic factors. Our research indicated that the same genetic area CDKN2A/B affects the risk of cardiac disease also in individuals with type one diabetes."

A third of individuals with type 1 diabetes develop a kidney disease. Some may develop renal failure that may, at worst, lead to the need for dialysis treatment or a kidney transplant.

Another study analysed various data sources to identify connections to kidney disease in 27,000 diabetics. GWAS is a fast and economic method, but it cannot identify all variants. Therefore, the research team have turned to sequencing the entire genomes of patients.

"Variants identified with the GWAS method are common, and individual variants' impact on the risk of developing a condition is quite moderate.

The objective of sequencing is to identify rare variants that may have a significant impact on developing a condition at the



400 million individuals in the world have diabetes. Half of them have diabtetic complications. 35% have a genetic risk for the diabetic kidney disease (nepropathy). Other complications are eye diseases (glaucoma. retinopathy), diabetic foot, neuropathy, stroke, heart attack, and peripheral artery disease

level of individual patients. The worst case scenario is that such a variant prevents the functioning of an entire protein."

New variants identified from an enormous amount of data

According to Sandholm, the research results may help predict the risk of developing a condition or lead the way in developing new medicines.

"The broader goal of genetic research is to identify variants that affect the risk of developing a disease or directly cause a disease. This enables a better understanding of the causes of diabetic complications."

The ultimate goal is to learn to prevent and find cures to diseases associated with diabetes.

"Our aim is to read the entire DNA sequence of all patients. This will result in a huge amount of data," says Sandholm.

"The sequencer produces DNA data in strands of 150 base pairs. To verify the data, our goal is to read each one of the three billion DNA base pairs an average of 30 times. This means that there will be 600,000 strands of 150 base pairs per patient."

To map the entire sequence, the sequenced strands must be placed in the correct order with the help of a reference human genome. This requires enormous computing capacity which is provided by the ELIXIR centre at CSC, the Finnish IT Centre for Science.

"The purpose is to organise the data so that it would enable identifying how each base pair variant affects a given disease at the level of individual patients. Our aim is to identify rare variants that cannot be identified with the GWAS method. Only a few patients in the sample exhibit rare variants."

Single nucleotide polymorphisms (SNP), or variants in the DNA base pairs, are a sort of end result of the data processing.

"The variation in the DNA sequence is expressed as SNPs. Each patient has either no alleles or one or two variants. They are markers that indicate which diseases the variant may cause."

The research group has already sequenced the entire genomes of 600 patients.

"Based on the initial results, we identified individual variants that are clearly associated with the risk of stroke, for example. There are also variants in genes that have previously been associated with congenital kidney diseases. Now it appears that variants in the same genes also affect the development of diabetic kidney disease."

Along with her colleagues, Niina Sandholm studies the protein-coding parts of the gene and gene regulatory regions that may have links to risk factors contributing to diabetes.

"The area between genes – 95% of the genome – contains plenty of regulatory regions that determine which gene appears in which tissue. As such, the DNA sequence is the same in each human cell, but gene regulation causes eyes to develop into eyes and kidneys into kidneys. In this respect, gene regulatory regions and their changes play a key role."

Genome sequencing on an exceptional scale

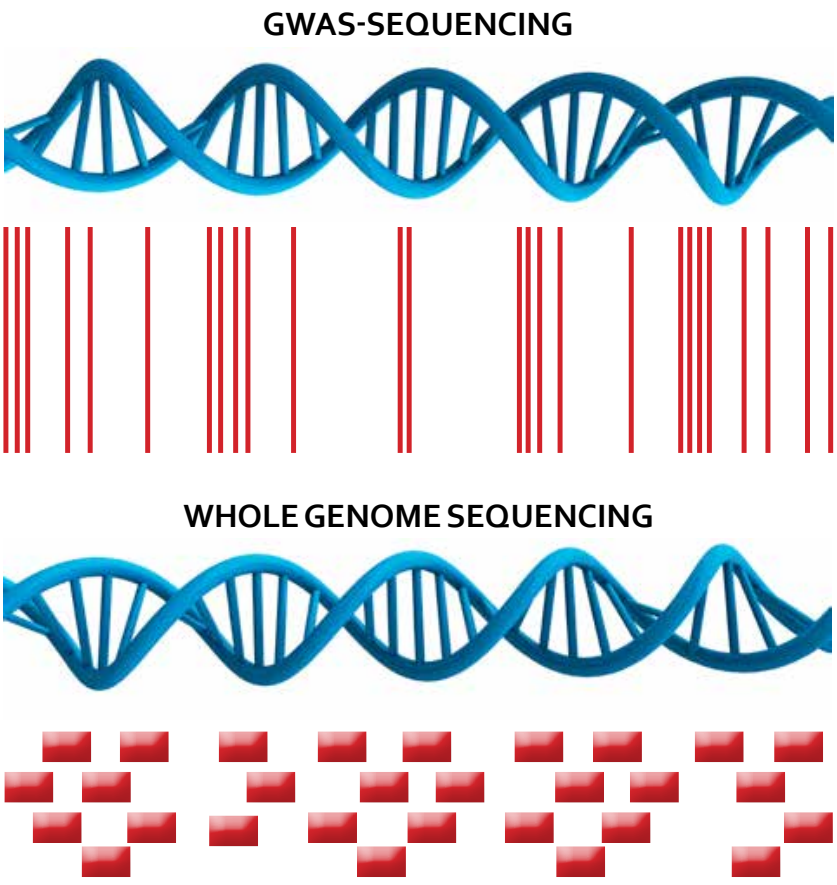
This study is one of the world's first to sequence the entire genome this extensively with regard to a specific disease. For the time being, the sequencing of the entire genome is relatively rare.

"The current trend is to sequence the exome with a focus on the protein-coding parts. However, it is only a matter of time that sequencing the entire genome becomes more common. ELIXIR, for one, invests in the development of full genome sequencing and genomic data processing methods."

CSC provides the ePouta service for processing sensitive data. In the ePouta cloud service, virtual private servers operate on CSC's computing platform under increased data security. The users receive dedicated cloud resources which are separated from CSC's other computing environments. The FinnDiane research group uses the computing cluster of Institute for Molecular Medicine Finland (FIMM) which is connected to CSC's sensitive data computing platform via the ePouta light path. By scaling the computing resources, the light path enables faster processing of the project data.

In addition, the researchers have been allocated a remarkable amount of storage space to store the genomic data.

Ari Turunen



Whole genome sequencing of 1880 individuals with type 1 diabetes. Short sequence reads of 150 DNA bases. Average 30× coverage for each base. Since a human being has 3 billion bases, this means 600 million reads per person. Reading one genome takes one day per person.

MORE INFORMATION:

Folkhälsan
https://folkhalsan.fi/fi/

FinnDiane
http://www.finndiane.fi

CSC – IT Center for Science
is a non-profit, state-owned company administered by the Ministry of Education and Culture. CSC maintains and develops the state-owned, centralised IT infrastructure.
http://www.csc.fi
https://research.csc.fi/cloud-computing

ELIXIR
builds infrastructure in support of the biological sector. It brings together the leading organisations of 21 European countries and the EMBL European Molecular Biology Laboratory to form a common infrastructure for biological information. CSC – IT Center for Science is the Finnish centre within this infrastructure.
http://www.elixir-finland.org
http://www.elixir-europe.org

ELIXIR FINLAND
Tel. +358 9 457 2821s ● e-mail: servicedesk@csc.fi
www.elixir-europe.org/about-us/who-we-are/nodes/finland

www.elixir-finland.org

ELIXIR HEAD OFFICE
EMBL-European Bioinformatics Institute
www.elixir-europe.org