Together with Palotie’s research team, he has been able to use data analysis to demonstrate that the underlying causes of various neurological diseases consist of numerous genes, instead of a single genetic mutation. For example, there are hundreds of different genes that affect a person’s predisposition to migraine, epilepsy, Parkinson’s disease or Alzheimer’s disease.

Aarno Palotie’s research requires an enormous number of samples. In 1998, while working as a professor at The University of California, Los Angeles (UCLA), he had access to the most extensive research data on migraines, available at the time: the data had been collected from 400 Finnish families suffering from migraines. Over the years, the research data has grown and now covers 1,600 families. Between 2007 and 2013, he carried out research related to migraines, schizophrenia and epilepsy in Cambridge, UK, using extensive research material.

Clinical and research data collected on people is constantly being produced and recorded. The more data there is available for research use, the easier it is to find statistical variables. Extensive amounts of data have the potential of revealing new information if the data is mined and analysed well.

“Studies should be able to utilise sample sizes, that are no longer measured in thousands but in millions,” Aarno Palotie notes. “The large sample sizes have been collected from the diverse donor base of different biobanks. Data from different sources is combined in order to really increase the numbers. This way we can increase the signal and reduce the noise.”

By increasing the signal, Palotie refers to making the data statistically significant. For example, mining rare diseases from the research material requires large amounts of data. It is valuable for researchers if the data has been collected using the same methods. Different laboratories and research facilities may have different practices for collecting, processing and classifying raw data from measuring instruments. The more consistent the data is, the easier it is to analyse.

“However, in real life, absolute harmonisations in disease research is very challenging to achieve. That is why it is important to harmonise the aspects that can be harmonised, in order to make real discoveries and correct interpretations using vast quantities of data.”

Cost efficient genotyping

Genome data has conventionally been collected using sequencing techniques that are used to investigate the base sequence of genes in test tubes. However, sequencing costs are very high if a lot of data is required for the study, such as research into common or chronic diseases. Genotyping has become established as a cost efficient and reliable method. In genotyping, a DNA microarray technique is used to collect genetic data from DNA samples. The samples are studied with a microarray scanner and the collected raw data is then processed. Only the sections of chromosomes known to feature genetic variants related to the

Hundreds of genes could lie behind a single disease

Professor Aarno Palotie from the Institute for Molecular Medicine Finland (FIMM) focuses on genetic analysis of diseases by utilising large quantities of data gathered from subjects.
studied disease are studied in genotyping. After collecting the data, computational methods are put into use. A reference genome (a reference assembly created using DNA sequences from various donors) can be used to predict the variants that were not examined earlier.

The genetic variants that are studied in genome wide association studies (GWAS) are measured from sample sizes varying from hundreds of thousands to millions of samples. The GWAS method is most commonly used when the genetic background of a disease is multifactorial, or polygenic, meaning that hundreds or thousands of genetic variants affect the disease risk. Multifactorial diseases include cardiovascular diseases, allergies, diabetes and mental disorders, for example. An extensive amount of research data is required for a reliable GWAS analysis. The computing power of super computers is needed to analyse the data.

"Sequencing requires even more extensive amounts of data than GWAS techniques do. Sequencing is also expensive compared to the GWAS method. Producing data using the GWAS methods costs a few dozen euros. The method can be used with extensive enough sample materials. Data is standardised, and it preserves well. Data genotyped at different locations can be easily combined."

**The source of migraine**

Migraine is a disorder that causes headaches and is usually considered to stem from a disturbance in the brainstem caused by external factors. One in ten adults suffer from migraines and it is three times more common in females than males. Palotie has studied migraine for a long time. One study utilised samples collected from 375,000 people worldwide. 60,000 of them suffered from migraines. In 2016, his research team identified 30 new hereditary risk factors related to migraine. Many of them are located in genes that regulate vascular function.

In 2018, an article that provided new information on the causes of migraine, written by Palotie and other researchers, was published in the Neuron scientific journal. A significant observation was that migraine is not only affected by certain genes, but a vast number of genes. Palotie talks about gene load.

"For decades, the genetics of diseases have been thought of the way Mendel described them. It is actually far more complex," Palotie says.

Gregor Mendel, who has been called the father of genetics, demonstrated that certain characteristics of a person are inherited by the succeeding generation. Genes can be dominant or recessive. According to Palotie, new research results have shown that it is not that simple. For example, a disease may be affected by a group of genetic mutations, not necessarily only one genetic variant.

"Genetic findings are believed to play a significant role in understanding diseases and forming a basis for developing new treatments."

"The assumption has been that if there is migraine, heart attacks, cancer or another common condition in the family, the genetic variants that cause it are strong and transferred to children from their parents. The migraine study together with other research actually show that the cause of certain diseases is likely to be an accumulation of very common genetic variants. These are some of the same variants that can be found in the entire population. Sometimes however, a person and their spouse simply happen to both have a heavy genetic variant load. When these two gene loads, which contain thousands of genetic variants, come together the risk of disease in their offspring increases."

Palotie and his team also search for similar gene loads related to other neurological diseases. Palotie is currently working on an extensive international study on the genetic background of psychotic disorders. Research material for the study is being collected from over 100,000 people worldwide. Genetic findings are believed to play a significant role in understanding diseases and forming a basis for developing new treatments.

"If there is a sufficient amount of data available on patients it is possible to provide more specific treatment, this is referred to as targeted treatment or more individualised treatment."
Rare genetic variants provide a short cut for biology

Palotie conducts his research from two different perspectives: he looks for both accumulations of common genetic mutations and rare genetic variants. 

“Rare variants may provide a short cut to biology,” he says.

As an example, Palotie mentions a patient with schizophrenia who had a few genes with a strong mutation connected to the disorder. This is very rare with schizophrenia, because predisposition to the disorder is usually the result of a combined effect of thousands of genes.

“However, such a rare occurrence may reveal something about the mechanisms of the disorder. It may be easier to identify a cell’s biological signaling pathway by studying rare mutations in the genes of patients with schizophrenia than it is by studying accumulations of common variants.”

Studying the signaling pathway, or understanding how a cell reacts to communication, plays a key role in understanding disease mechanisms. Cells react to messages they get from their surrounding environment. Often the signal goes all the way to the nucleus and starts to regulate what the gene does. Sometimes cells contain special proteins that are there to stop the signal. For example, cancer cells do not react to many signals intended for them. Instead, cancer cells enhance the signaling pathway, which makes the cell divide and grow.

Finns have a number of genetic variants that are rare in other parts of the world but rich in Finland because of our demographic history. When data collected from Finnish people is combined with data from other populations, we can get more information on signaling pathways. This means that a Japanese patient may benefit from data collected from Finnish patients and vice versa.

“Even if a variant identified in Finland is unknown in Japan the physiology and biology of people is still very similar.”

Even if a variant identified in Finland is unknown in Japan the physiology and biology of people is still very similar. Hopefully, the identified variant helps steer toward the correct cell signaling pathway. When we identify a new cell signaling pathway another genetic variant, that is in-fact connected to the same signaling pathway, may be discovered from another population base. In such case, the finding helps confirm that the signaling pathway is significant in relation to this disease.

The location of the genetic variant matters. Palotie has an example of this. When visiting Iceland, Palotie talked about Leif Groop’s study, as a result of which a rare genetic variant, that protects from type II diabetes, was discovered in the population on the west coast of Finland. Leif Groop’s colleague asked Icelandic researchers whether a similar variant had been identified in the Icelandic population. The Icelandic researchers checked their databases. The same genetic variant had not been found but there was another variant of the same gene.

“This Icelandic discovery confirmed that the gene in question protects against type 2 diabetes. Such a protective genetic variant is obviously very interesting from the perspective of molecular drug design.”

FinnGen and phenotype data

The FinnGen project was launched in autumn 2017. The aim of the project is to record the genomes of 0.5 million Finnish people. The project utilises samples collected by all Finnish biobanks. The data collected on Finnish heritage will be combined with
clinical data from national healthcare registers. The goal is to gain a better understanding of diseases by combining genome and healthcare data. Patient healthcare can only be significantly improved by analysing large quantities of samples.

FinnGen is centred around Finnish phenotype data collected from healthcare registers. In Palotie’s opinion, FinnGen can be considered a prime example of how data from biobanks and healthcare registers can be combined for genome data analysis.

The project has partners from all over the world. The goal is to combine data from Finnish biobank samples with biobank data from other countries and carry out a meta-analysis.

“Combining data from the different sources is a great challenge and meta-analysis is often a more functional solution. FinnGen meta-analysis has been carried out with biobanks located in Great Britain and Japan. The aim is also to get other countries involved.”

According to Palotie, it is crucial to be able to combine the research data collected on Finnish people with data from other countries and populations. New methods for processing and cultivating data are constantly being developed. New methods are useless if researchers do not have access to enough data that can be analysed and inspected.

“Even artificial intelligence has difficulty functioning without enough data.”

Ari Turunen

Institute for Molecular Medicine Finland (FIMM)
The mission of the Institute is to advance new fundamental understanding of the molecular, cellular and etiological basis of human diseases. This understanding will lead to improved means of diagnostics and the treatment and prevention of common health problems. Finnish clinical and epidemiological study materials will be used in the research.

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.

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.

MORE INFORMATION:

Institute for Molecular Medicine Finland (FIMM)
The mission of the Institute is to advance new fundamental understanding of the molecular, cellular and etiological basis of human diseases. This understanding will lead to improved means of diagnostics and the treatment and prevention of common health problems. Finnish clinical and epidemiological study materials will be used in the research.

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.

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.

ELIXIR FINLAND
Tel. +358 9 457 28215 • 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