

All breast cancer risk factors evaluated with AI

Breast cancer is the most common type of cancer in women. One quarter of all cancers in women are breast cancers. Until now, genetic risk factors for breast cancer have usually been studied as single factors. Professor Arto Mannermaa intends to explore the big picture and look for more factors which significantly increase the risk of illness when they interact. This is where artificial intelligence comes in.



Mannermaa's team is developing algorithms that can learn on the basis of genomic and clinical data, and identify and predict risk factors. Learning algorithms are also used in the interpretation of mammography images. Genomic and clinical data are integrated to an AI model that not only helps to determine the risk of illness, but also in drawing up individual treatment plans.

"I'm a biologist and geneticist by training. I engaged in the close study of human genetics and specialised as a hospital geneticist. I have been working on clinical studies in genetics laboratories, but have been interested in cancer throughout my researcher career," says Arto Mannermaa.

Mannermaa is a Professor of personalised medicine and biobanking at the University of Eastern Finland. In his research, he has focused on the genetics of breast and ovarian cancer. Since its inception, Manner-

maa's team has been involved in the work of the world's largest genetic epidemiology consortium, the Breast Cancer Association Consortium (BCAC). The consortium has the world's largest centralised collection of breast cancer tissue samples, collected from over 200,000 patients and controls. The collection includes well-annotated data on factors and clinical results related to breast cancer.

"My research team has been engaged in long-term work to determine the genetic risk variants of breast cancer. We can now identify some 200 variants within normal genomic variation which increase the risk of getting breast cancer. Together with the BCAC, we have also learned about the genetic mutations that are major contributing factors to breast cancer. These include among others mutations of the BRCA and PALB-2 gene, which we were involved in

finding to be a contributing factor in breast cancer."

If a woman has a BRCA1 or BRCA2 gene mutation, she has a 60–80% risk of developing breast cancer. The risk impact of PALB-2 is almost the same.

Huge amounts of data

It is estimated that genome accounts for about 30 per cent of susceptibility to breast cancer, while 70 per cent is determined by environmental factors. Risk factors for breast cancer include the total amount of oestrogen during a lifetime, which depends on the number of pregnancies and children, and weight. Other factors include smoking, alcohol consumption and exercise. According to Mannermaa, risk factors tend to have been studied one at a time. Now the BCAC has been able to study the prevalence of breast cancer in the close relatives of pa-

tients. This includes plenty of international material that can be used for making comparisons.

“For example, studies have been done on whether there are more common factors in Finnish cancer data than in international data.”

Research material has been obtained from the Biobank of Eastern Finland and Kuopio University Hospital.

“The research team is grateful to all volunteers who participated in the study. Without their consent, this kind of work would not be possible.”

Finnish data has been compared with data obtained through BCAC, collected from more than 100 research teams around the world.

“Finns form a uniquely isolated population. In Finland there has identified many diseases linked to the Finnish disease genome, and genetic changes related to these. But the situation with multifactorial diseases, such as breast cancer, is surely the same, meaning that some risk factors are related to our population. Some of the genetic risk factors for multifactorial diseases only occur in Finns.”

Although data has been obtained from around the world, for Mannermaa the challenge lies in the fact that the material was collected for different purposes, and is not always in the same format.

“In order to use the material, the data collected from different sources must be unified, which often takes up a large amount of the total time spent on research.”

Aiming for the big picture in breast cancer risk factors

The question Mannermaa wants to answer is which factors have contributed to the onset of breast cancer in a patient. Mannermaa’s team has created an AI model for breast cancer risk factors that is being tested with Finnish and international material.

“We also have material obtained from the Biobank. We are comparing the data of breast cancer patients and healthy individuals and trying to find the interactive combination of all variables that has the greatest influence on the onset of breast cancer.”

One of the study’s targets concerns normal genomic variation, or SNPs. The

rapid development of DNA sequencing techniques has made it possible to determine single nucleotide polymorphisms (SNPs), providing a very accurate estimate of the differences between individuals. SNP is the difference in the DNA chain caused by a mutation within a population. According to some estimates, the human genome has 4–5 million SNPs, located in the DNA chain, either in the inter-gene or gene region. They can act as biomarkers, helping researchers to locate genes related to diseases. Certain SNPs can affect the operation of the gene and thus directly affect the onset of the disease.

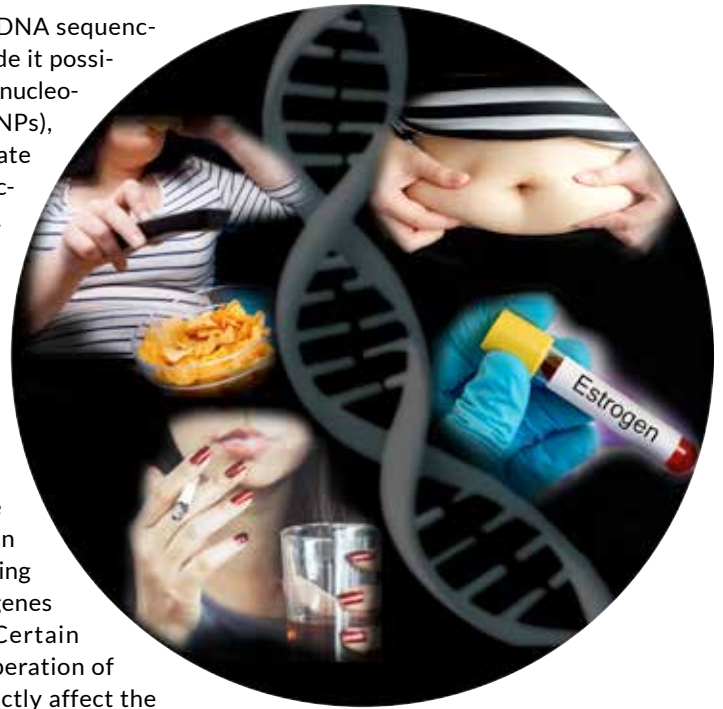
“In practice, we focus on the differences between cancer patients and control groups. We want to learn how many SNPs are in common with these groups, and what the common SNP network is like among cancer patients compared to healthy individuals.”

Mannermaa’s team is working to identify SNPs related to breast cancer, by means of AI and learning algorithms.

“We teach our algorithm to detect SNP networks. With the help of artificial intelligence, we can identify the interactive group of SNPs with the greatest impact on disease risk.”

The results have been promising. The algorithm helped to identify genes close to SNPs, and these SNPs are probably affecting the operation of the genes. We found a gene network related to oestrogen metabolism.

“Oestrogen metabolism is a key component in the development of breast cancer, while another group that we found was related to apoptosis, or programmed cell death. Apoptosis is crucial in cancer development, because cancer cells must be able to prevent programmed cell death. That’s why we believe that the AI models helped us find the correct breast cancer factors.”



Risk factors for breast cancer include the total amount of oestrogen during a lifetime and weight. Other risk factors are smoking, alcohol consumption and exercise.

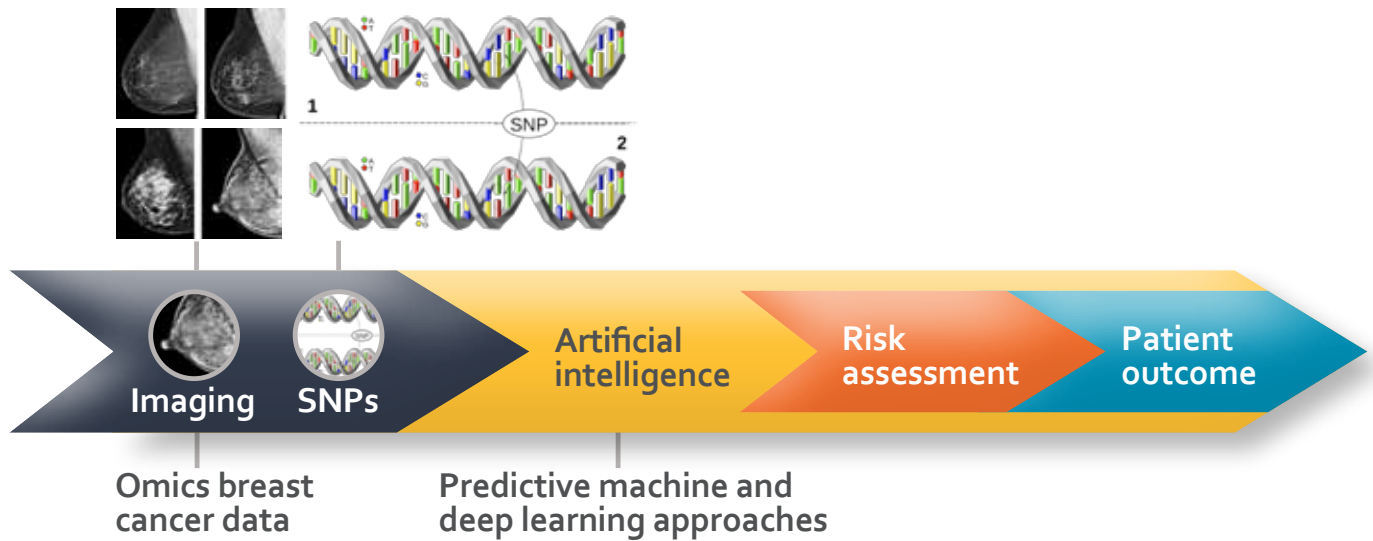
Supercomputing required

The amount of data in Mannermaa’s team’s study is so huge that CSC’s (ELIXIR Finnish centre) supercomputing capacity is required.

“About 200,000 SNPs can be identified from one laboratory sample. Each SNP is compared with all the others. In addition, we simulate genetic variation, in other words what SNPs they have in common but remain unidentified. This means that up to another 10 million SNPs can be added to the equation. Add to this variables from imaging and the biobank, and computing capacity is definitely called for.”

The basic model of the Mannermaa team’s AI is based on genetic data. Clinical variables, i.e. breast cancer risk factors, have now been added to this model. Mannermaa believes that the models will significantly improve diagnostics.

“Artificial intelligence enhances screening and diagnostics. In the future, we can avoid overdiagnosis and use the data to differentiate those who need more accurate screening from those who don’t. This means that certain women do not need



frequent mammography, due to their low risk of developing breast cancer.

Once genetic data is combined with not only known risk factors but also breast cancer diagnosis and treatment, the predictability of the disease will improve and personalised treatment plans can be drawn up.

Biobanks play a crucial role in research of this type. It is essential that all data is available.

“If the person giving a sample has consented to their data being used for biobank purposes, this data is combined with other data. The Biobank Act is the basis for secure data storage, and enables people who have given their consent to cancel it if they wish. Biobank consent is general consent based on the law. Through biobanks, everyone has

the opportunity to participate in research aimed at developing health care.”

Multidisciplinary requirement for effective care

Mannermaa leads the SOTE AI Hub project, funded by the Regional Council of Pohjois-Savo. The project is seeking to improve the use of various data sources and AI in aid of decision-making. The project involves utilising and developing health data in a data lake. The Pohjois-Savo data lake consists of social and health data from the Biobank of Eastern Finland, Kuopio University Hospital and the City of Kuopio.

North Savo AI-Hub



According to Mannermaa, the health data can be used to evaluate the impact of the research results. In addition to receiving plenty of data on the actual patient, we can see the impact of cancer patients’ treatment alternatives and solutions based on new research.

“The model and its prediction can help determine how it affects a patient’s life, and how resources should be allocated. This can make treatment more effective in the future. Patient-specific profiling and individualised treatments help to provide the right treatments for the right patients, and thereby make health care more efficient. This requires a multidisciplinary network.”

Ari Turunen

MORE INFORMATION:

School of Medicine, University of Eastern Finland
<https://www.uef.fi/fi/web/laake>
 Institute of Clinical Medicine,
 University of Eastern Finland
<https://www.uef.fi/fi/web/kliinisenlaaketieteenyksikko>
 Cancer Center of Eastern Finland, CCEF
<http://www.uef.fi/fi/web/ccef>
 The Breast Cancer Association Consortium
<http://bcac.ccge.medschl.cam.ac.uk>

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>