

## Rintasyövän kaikkia riskitekijöitä arvioidaan tekoälyn avulla

Rintasyöpä on naisten yleisin syöpäsairaus. Neljäsosa kaikista naisten syöivistä on rintasyöpää. Rintasyövän geneettisiä riskitekijöitä on tähän mennessä yleensä tutkittu yksittäisinä tekijöinä. Professori Arto Mannermaa haluaa tutkia laajempaa kokonaisuutta ja etsiä useampia tekijöitä, jotka yhdessä vuorovai-  
kutuksessa toistensa kanssa vaikuttavat merkittävästi sairastumisriskiin. Avuksi tähän tulee tekoäly.



Mannermaan ryhmässä kehitetään genomidatan ja kliinisen datan perusteella oppivia algoritmeja, jotka tunnistavat ja ennustavat riskitekijöitä. Oppivia algoritmeja hyödynnetään myös mammografiakuvien tulkitsemisessä. Genomidata ja kliininen data yhdistetään tekoälymalliksi, joka auttaa paitsi sairastumisriskin määrittämisessä, myös yksilöllisten hoitosuunnitelmien tekemisessä.

”Olen koulutukseltani biologi ja perinnöllisyystieteilijä. Perehdyin ihmisgenetiikkaan ja erikoistuin sairaalageneetikoksi. Olen työskennellyt kliinisen tutkimuksen parissa genetiikan laboratorioissa, mutta koko tutkimusurani aikana olen ollut kiinnostunut syövästä”, sanoo Arto Mannermaa.

Mannermaa on Itä-Suomen yliopiston henkilökohtaisen lääketieteen ja biopank-

kitoiminnan professori. Tutkimustyössään hän on tarkastellut erityisesti rinta- ja munasarjasyövän genetiikkaa. Mannermaan ryhmä on ollut alusta asti mukana maailman suurimman rintasyövän geneettisen epidemiologian konsortiumin (Breast Cancer Association Consortium, BCAC) toiminnassa. Konsortiumilla on maailman laajin keskitetty rintasyövän kudoksenäytteiden kokoelma, joka on kerätty yli 200 000 potilaalta sekä verrokilta. Kokoelma sisältää hyvin annotoitua dataa rintasyöpään liittyvistä tekijöistä ja kliinisistä tuloksista.

”Tutkimusryhmässäni on pitkään tehty työtä nimenomaan rintasyövän geneettisten riskivarianttien määrittämiseksi. Nyt tiedetään jo noin 200 genomien normaalivaihteluun kuuluvaa varianttia, jotka kasvattavat rintasyöpään sairastumisen riskejä. Yhteistyössä

BCAC-konsortiumin kanssa on myös selvitetty geenimutaatioita, jotka altistavat vahvasti rintasyöväälle. Näitä ovat mm. BRCA-geenien mutaatiot ja PALB-2-geenin mutaatiot, joita olimme löytämässä rintasyövän taustalta.”

Jos naiselta löytyy BRCA1- tai BRCA2-geenin mutaatio, hänellä on 60–80 prosentin riski sairastua rintasyöpään elämänsä aikana. PALB-2:n riskivaikutus on lähes samaa luokkaa.

### Valtava määrä dataa

Arvioiden mukaan perimä selittää noin 30 prosenttia alttiudesta sairastua rintasyöpään ja 70 prosenttia selittyy ympäristötekijöillä. Rintasyövän riskitekijöitä ovat estrogeenin elämänaikainen kokonaismäärä, johon vaikuttavat mm. raskauksien ja lasten määrät, ja paino. Lisäksi muita tekijöitä ovat tupai-

kointi, alkoholi ja liikunta. Mannermaan mukaan riskitekijöitä on aikaisemmin tutkittu lähinnä yksi kerrallaan. Nyt BCAC-konsortiumin avulla on voitu tutkia esimerkiksi rintasyöpään sairastuneiden lähisukulaisten rintasyöpien yleisyyttä. Mukana on paljon vertailtavaa, kansainvälistä aineistoa.

”On esimerkiksi selvitetty, onko suomalaisessa syöpäaineistossa yhteisiä tekijöitä, jotka ovat yleisimpiä kuin kansainvälisessä aineistossa.”

Tutkimusmateriaalia on saatu Itä-Suomen biopankista ja Kuopion yliopistollisesta sairaalasta.

”Tutkimusryhmä on kiitollinen kaikille tutkimukseen osallistuneille vapaaehtoisille. Ilman heidän suostumustaan ei tällainen työ olisi mahdollista.”

Suomalaista aineistoa on vertailtu aineistoon, joka on saatu BCAC-konsortiumin kautta ja jonka on kerännyt yli 100 tutkimusryhmää kaikkialta maailmasta.

”Suomalaiset muodostavat omalaatuisen isoituneen populaation. Suomessahan on tunnistettu paljon suomalaisen tautiperimään kuuluvia tauteja ja niihin liittyviä geneettisiä muutoksia. Mutta aivan varmasti monitekijäisissä taudeissa, kuten rintasyöpässä, on samanlainen tilanne eli osa riskitekijöistä liittyy meidän populaatioomme. Osa monitekijäisten tautien geneettisistä riskitekijöistä esiintyy vain suomalaisilla.”

Vaikka dataa on saatu ympäri maailmaa, haaste on Mannermaan mukaan se, että materiaalia on kerätty eri tarkoituksiin ja se ei aina ole samanlaisessa muodossa.

”Jotta materiaalia voisi käyttää, tulee eri tahoilta kerätty aineisto yhdenmukaistaa. Se vie usein suuren osan koko tutkimukseen käytettävästä ajasta.”

## Rintasyövän riskitekijöiden kokonaiskuvaan pyritään

Mitkä kaikki tekijät ovat vaikuttaneet siihen että henkilö on saanut rintasyövän, on se kysymys, johon Mannermaa haluaa vastauksen. Mannermaan ryhmä on luonut rintasyövän riskitekijöiden tekoälymallin, jota testataan suomalaisessa ja kansainvälisessä aineistossa.

”Käytössämme on myös biopankista saatua aineistoa. Vertaamme rintasyöpään sairastuneiden ja terveiden henkilöiden dataa

ja yritämme saada vastauksen, mikä on se kaikkien muuttujien vuorovaihteinen yhdistelmä, joka eniten vaikuttaa rintasyövän syntyyn.”

Yksi tutkimuksen kohteista ovat perimän normaali vaihteluun kuuluvat tekijät, snipit. DNA-sekvensointitekniikoiden nopea kehittyminen on mahdollistanut yhden nukleotidin polymorfismien (single nucleotide polymorphism, SNP) selvittämisen, joka antaa hyvin tarkkan arvion yksilöiden välisistä eroista. SNP eli snippi on populaatiossa esiintyvä mutaation aiheuttama ero DNA-ketjussa. Ihmisen genomissa on joidenkin arvioiden mukaan 4-5 miljoonaa snippiä ja ne sijaitsevat DNA-ketjussa joko geenien välisellä alueella tai geenien alueella. Ne voivat toimia biomarkkereina, jolloin ne auttavat tutkijoita paikallistamaan geenit, jotka liittyvät sairauksiin. Jotkut SNP:t voivat vaikuttaa geenin toimintaan ja vaikuttaa siten suoraan sairauden syntymiseen.

”Käytännössä tutkimme syöpään sairastuneiden ja verrokkiryhmien eroja. Haluamme tietää, kuinka paljon on yhteisiä snippejä ryhmien välillä ja minkälainen yhteinen snippiverkosto on syöpään sairastuneilla terveisiin verrattuna.”

Mannermaan ryhmä pyrkii tekoälyn ja oppivien algoritmien avulla tunnistamaan rintasyöpään liittyvät snipit.

”Opetamme algoritmia tunnistamaan snippiverkoston. Tekoälyn avulla pystymme tunnistamaan tautiriskiä eniten vaikuttavan vuorovaihteisen snippien ryhmän.”

Tulokset ovat olleet lupaavia. Algoritmin avulla tunnistettiin geenit snippien lähellä ja nämä snipit todennäköisesti vaikuttavat geenien toimintaan. Löydettiin geeniverkosto, joka liittyy estrogeenin aineenvaihduntaan.

”Estrogeenimetabolia on tärkeä rintasyövän kehittämisessä ja toinen havaitsemamme ryhmittymä liittyi apoptoosiin eli itseohjattuun solukuolemaan. Apoptoosi on absoluuttisen tärkeä syövän kehityksessä, sillä syöpäsolujen pitää pystyä torjumaan itseohjattu solukuolema. Uskomme siksi, että



*Rintasyövän riskitekijöitä ovat estrogeenin elämänaikainen kokonaismäärä ja paino. Lisäksi muita tekijöitä ovat tupakointi, alkoholi ja liikunta.*

löysimme oikeita rintasyövän tekijöitä tekoälymallin avulla.”

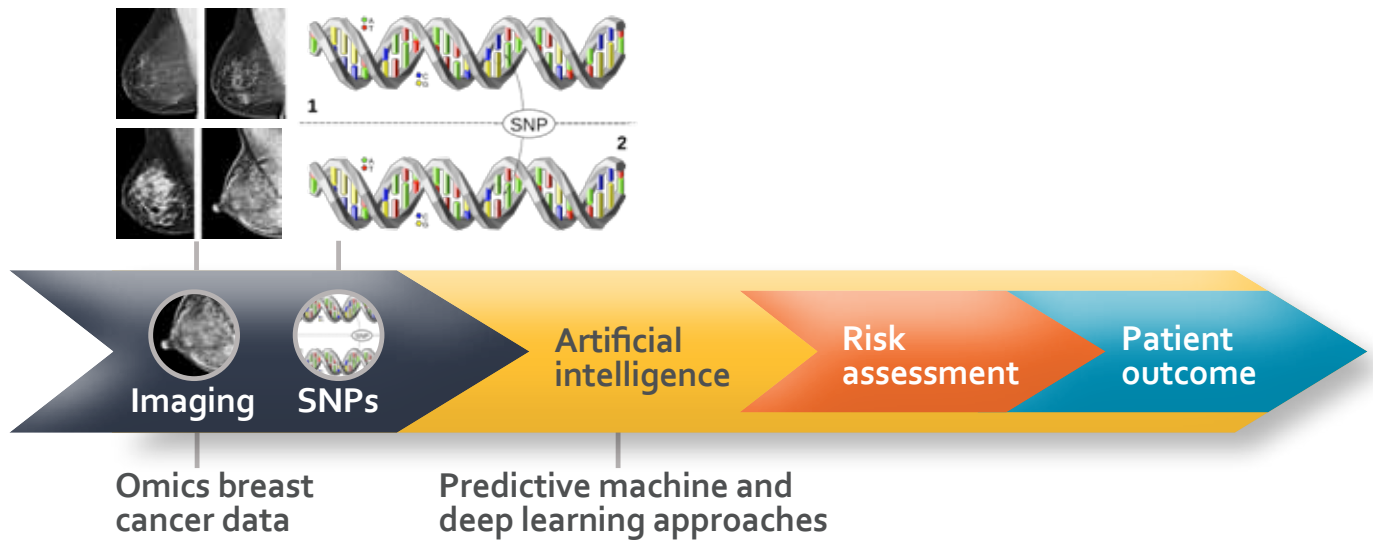
## Superlaskentaa tarvitaan

Mannermaan ryhmän tutkimuksissa käytetyn datan määrä on niin iso, että siihen tarvitaan ELIXIRin Suomen keskuksen CSC:n superlaskentakapasiteettia.

”Yhdestä laboratorionäytteestä saadaan määritettyä n. 200 000 snippiä. Jokaista snippiä verrataan toisiinsa. Lisäksi simuloidaan mitä geneettistä vaihtelua, eli määrittämättä jääneitä snippejä niiden välissä on. Näin laskennallisesti määriteltäviä snippejä saadaan riskiarviointiin lisää jopa 10 miljoonaa. Kun tähän tuodaan vielä kuvantamisesta ja biopankista saadut muuttajat, laskentateholla on tarvetta.”

Mannermaan tutkimusryhmän tekoälyn perusmalli perustuu geneettiseen dataan. Nyt malliin on tuotu kliinisiä muuttujia eli tunnettuja rintasyövän riskitekijöitä. Mannermaa uskoo, että malli tuo merkittäviä parannuksia diagnostiikkaan.

”Tekoälyllä tehostetaan seulontaa ja diagnostiikkaa. Tulevaisuudessa voidaan välttää yli diagnosointia ja arvioida datan perusteella, ketkä tarvitsevat tarkempaa



seulontaa ja ketkä eivät. Osan ei siis tarvitse mennä mammografiaan usein, koska heillä on alhainen riski sairastua rintasyöpään.”

Kun geneettinen data yhdistetään mukaan tunnetuihin riskitekijöihin, ja rintasyöpään diagnoosiin ja hoitoon liittyviin tekijöihin se parantaa rintasyöpään ennustettavuutta ja yksilöllisiä hoitosuunnitelmia.

Biopankit ovat ehdoton edellytys tällaiselle tutkimukselle. Olennaista, että kaikki data on saatavissa.

”Jos näytteen antaja on antanut biopankkisuostumuksen, yhdistetään tämä data muuhun dataan. Biopankkilaki on perustana tietoturvaliselle datan säilytykselle, ja se takaa suostumuksen antaneelle myös vapaaehtoisen tutkimuksista kieltäytymisen mahdollisuuden. Biopankkisuostumus on

lakiin perustuva yleinen suostumus. Biopankkitoiminnan myötä kaikilla on mahdollisuus osallistua tutkimuksiin, joilla pyritään kehittämään terveydenhoitoa.

### Monitieteisyys ehdoton edellytys tehokkaalle hoidolle

Mannermaa johtaa Pohjois-Savon liiton rahoittamaa SOTE AI Hub -hanketta, jonka tarkoitus on tehostaa eri datalähteiden ja tekoälyn käyttöä päätöksenteon tueksi. Hankkeessa hyödynnetään ja kehitetään terveysdataa tietoaaltaassa. Pohjois-Savon tietoaaltaassa on Itä-Suomen biopankin, Pohjois-Savon sairaanhoitopiirin ja

North Savo AI-Hub



Kuopion kaupungin sosiaali- ja terveysdataa.

Terveystiedon avulla saadaan Mannermaan mukaan arvioitua tutkimustulosten vaikuttavuus. Sen lisäksi, että saadaan paljon dataa itse potilaasta, nähdään syöpään sairastuneen hoitovaihtoehtojen ja uusien tutkimukseen pohjautuvien ratkaisujen vaikuttavuus

”Mallin ja sen ennusteen avulla voidaan selvittää, miten se vaikuttaa potilaan elämään ja miten resursseja kannattaisi kohdentaa. Näin voidaan tulevaisuudessa tehostaa hoitoa. Potilaskohtaiset profiloinnit ja yksilölliset hoidot edistävät oikeiden hoitojen antamista oikeille potilaille ja sitä kautta tehostavat terveydenhuoltoa. Tähän tarvitaan monitieteistä verkostoa.”

**Ari Turunen**

#### LISÄTIETOJA:

Lääketieteen laitos, Itä-Suomen yliopisto  
<https://www.uef.fi/fi/web/laake>

Kliinisen lääketieteen yksikkö,  
Itä-Suomen yliopisto  
<https://www.uef.fi/fi/web/kliinisenlaaketieteenyksikko>

Itä-Suomen yliopiston syöpäkeskus  
<http://www.uef.fi/fi/web/ccef>

CSC – Tieteen tietotekniikan keskus Oy on valtion omistama, opetus- ja kulttuuriministeriön hallinnoima, voittoa tavoittelematon osakeyhtiö. CSC ylläpitää ja kehittää valtion omistamaa keskitettyä tietotekniikkainfrastruktuuria.  
<http://www.csc.fi>  
<https://research.csc.fi/cloud-computing>

ELIXIR rakentaa infrastruktuurin bioalan tutkimuksen tueksi. Se yhdistää 21 Euroopan maan ja Euroopan molekyylibiologian laboratorion EMBL:n johtavat organisaatiot yhteiseksi biologisen informaation infrastruktuuriksi. Sen Suomen keskus on CSC – Tieteen tietotekniikan keskus Oy.  
<http://www.elixir-finland.org>  
<http://www.elixir-europe.org>

SUOMEN ELIXIR

Puh. +358 9 457 2821 • e-mail: [servicedesk@csc.fi](mailto:servicedesk@csc.fi)  
[www.elixir-europe.org/about-us/who-we-are/nodes/finland](http://www.elixir-europe.org/about-us/who-we-are/nodes/finland)

[www.elixir-finland.org](http://www.elixir-finland.org)

ELIXIR PÄÄMAJA

EMBL-European Bioinformatics Institute  
[www.elixir-europe.org](http://www.elixir-europe.org)