

Oscar Brück ja Kimmo Porkka

Datamassojen hallinta syöpäpotilaiden hoidossa

Syöpäkasvainten molekyyli-tason profiloinnin tuottaman klinisen tiedon määrä lisääntyy nopeasti. Tällaisen käsityksen muodostaminen potilaan taudintilasta ja sen sovittaminen monimutkaisiin hoito- ja seurantaohjelmiin edellyttää suurten tietomassojen sujuvaa integraatiota ja koneellista päätöksenteon tukea. Eräänä ratkaisuna voidaan hyödyntää hajautettuun laskentaan perustuvia edistyneitä algoritmeja, mikä edellyttää yhteisiä klinisiä tietomalleja ja ratkaisukeskeistä päätöksentekoa lupaprosessien osalta. Uudet teknologiat, kuten konenäköön perustuva kuva-analyysi ja tautiprosessin laaja-alainen ja jatkuva simulaatio tuovat sekä lisähaasteita että mahdollisuuksia yksilöllistetyn syövänhoidon toteutukseen.

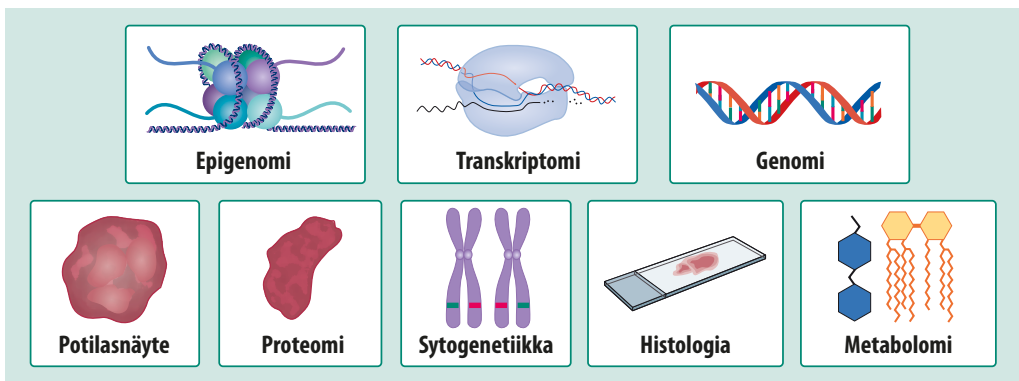
Syöpätautien hoitoprotokollat perustuvat pääasiassa ennusteellisiin tilastomalleihin. Yli 70 vuotta julkaisunsa jälkeen TNM-levinneisyysluokitus on edelleen keskeisin apuväline räätälöimään kiinteiden syöpien hoito-ohjelman intensiteettiä. Syöpäsolujen geneettisiä poikkeavuuksia tutkitaan yhä kattavammin, ja esimerkiksi akuuteissa leukemioissa riskiluokitus pohjautuu nykyään tunnettuihin geneettisiin muutoksiin. Toisaalta suurissa aineistoissa kehitetyt tilastolliset mallit soveltuvat vaihtelevalla menestyksellä yksilötasolle. Osaammeko tulkita riittävästi kertynyttä lääketieteellistä profiloititietoa, keskiarvoistavatko mallit liikaa tuloksia vai onko tieto yhä liian hajanaista edistyneempien ennustemallien kehittämiseksi?

Syöpäsolujen monet kasvot paljastuvat, mutta identiteetti pysyy piilossa

Hankitut tai periytyvät poikkeamat solujen geneettisessä tiedossa ovat yksi syöpätautien klassisista tunnusmerkeistä (1). Yhdysvalloissa käynnistettiin 2000-luvulla kaksi rinnakkaista hanketta, jotka ovat kenties eniten edistäneet ymmärrystämme syövän geneettisestä heterogeenisuudesta. Yhdysvaltojen Broad Institute

ja lääkeyhtiö Novartis karakterisoivat yhdessä noin 1 000 syöpäsolulinjaa 36:sta eri syöpätyypistä, jotka muodostavat The Cancer Cell Line Encyclopedia (CCLE) -tietokannan (2). Yhdysvaltalainen The National Cancer Institute perusti lisäksi The Cancer Genome Atlas -yhteisön (TCGA) kerätäkseen ja analysoidakseen yli 14 000 kliinistä potilasnäytettä useasta sairaalasta. Näytteet kattoivat 33 eri syöpätyypin (3). Miltei yhtä merkittävää kuin aineiston keruu, oli tietojen avoin julkaisu tutkimusyhteisön käyttöön, mikä on poikanut useita monitieteellisiä läpimurtoja muun muassa syöpätautien, genomiikan ja koneoppimisen aloilla.

Tietokantojen perusteellinen analysointi selvensi, kuinka genomisen DNA:n laajat (sytogeneettiset uudelleenjärjestymät) tai pienet (mutaatiot, deleetiot, insertiot) rakenteelliset muutokset sekä epigeneettiset muutokset koodaavien geenien säätelevillä alueilla vaikuttavat RNA-molekyylien ja niistä tuotettavien proteiinien rakenteeseen, määrään ja toimintaan, sekä viime kädessä syöpäkuoleman metaboliaan, histologiaan, tulehdusvasteeseen ja kliniseen taudinkulkuun (KUVA 1). Multiomiikka (kuten genomiikka, transkriptomiikka, proteomiikka, metabolomiikka) on edistänyt lääketutkimusta monialaisesti ja teknologian kehityttyä genomilajuisen molekulaarisen informaation lou-



KUVA 1. Potilasnäytteestä voidaan analysoida soluja, niiden osia ja niiden välisiä yhteyksiä erilaisilla multiomian menetelmillä.

hinta on siirtynyt kudostasolta yksisolutasolle (4).

Geneettisen profiloinnin merkityksen korostuessa on kiinnitetty enemmän huomiota toistuvien invasiivisesti otettujen biopsioiden aiheuttamaan kuormitukseen syöpäpotilaille. Kudoshomeostaasissa osa soluista kuolee ja toiset uusiutuvat ja vapauttavat DNA:ta soluvälitilaan, mikä jakaantuu elimistön muihin nestetiloihin kuten verenkiertoon, aivo-selkäydinnesteeseen, virtsaan, askitekseen, pleuranesteseen tai sylkeen (5). Kasvainsoluista vapautuu tihentyneen solukuoleman ja nekroosin seurauksena suuria määriä nukleiinihappoja. Nestebiopsioiden ottaminen on vaivattomampaa sekä näytteenottajalle, potilaalle että näytteen käsittelijälle. Pieniäkin määriä (jopa tasolla 1 per 2 000 000 nukleiinijuostetta) niin sanottua kiertävää kasvain-DNA:ta (circulating tumor DNA; ctDNA) ja näissä esiintyviä muutoksia kuten yhden nukleotidin variantteja (SNV), translokaatioita, indeleitä ja kopiolumuutoksia (CNV) on mahdollista havaita moderneilla sekvensointimenetelmillä (6). Kiertävän kasvain-DNA:n mittaaminen nestebiopsioista onkin herättänyt kiinnostusta muun muassa taudin varhaisen toteamisen, jäännöstaadin mittaamisen, hoitovasteen ennakoimisen sekä taudinseurannan vuoksi (7).

Digipatologia, kuva-analyysi ja konenäkö ovat myös edenneet laajoin harppauksin skannatuista kudosleikkeistä koostuvien sähköisten kuva-arkistojen tultua julkisiksi (8). Tietokoneiden grafiikkaprosessoreita (Graphics Processing Unit; GPU) hyödyntävät ja pitkälti

konvoluutioneuroverkkoihin perustuvat algoritmit kykenevät tunnistamaan tehokkaasti miljardeista pikseleistä koostuvista mikroskooppikuvista soluja, näiden muodostamia kudosalueita sekä solujen välisiä spatiaalisia yhteyksiä (9). Kuvista voi automaattisesti tunnistaa syövän, sen histopatologisen alatyypin ja jopa elossaloennusteen (10,11). Eräs mielenkiintoinen sovellus on geneettisten muutosten, kuten DNA:n korjausmekanismien häiriöiden seurauksena kehittyvän niin sanotun mikrosatelliitti-instabiliteetin, tunnistaminen patologian leikkeistä konenäön avulla, jolloin tarkentavia sekvensointitutkimuksia ja immunoterapiaa voidaan kohdentaa niistä eniten hyötyville potilaille (12,13).

Molekyyli-tason louhinnasta kliiniseen vasteeseen

BCR-ABL1-fuusiogeneiä estävät tyrosiini-kinasiin estäjät (TKE) mullistivat kroonista myelooista leukemiaa sairastavien potilaiden hoidon 2000-luvun taitteessa (14). Sittemmin TKE-hoitoa on käytetty menestyksekkäästi muun muassa *KIT*-mutatoituneen gastrointestinaalisen stroomatumorin, sekä *FLT3*-muutuneen akuutin myelooisen leukemian hoidossa (15,16). Täsmälääketieteessä vähintään yhtä merkittävä läpimurto on tapahtunut immuno-onkologiassa. Kroonista myelooista leukemiaa sairastavat potilaat pystyvät todennäköisemmin keskeyttämään päivittäisen TKE-lääkityksen, jos heidän verenkierrössään on suurempi osuus kypsiä NK-soluja (17). Lisäksi puolustusjärjes-

telmän liikasääntelyn, kuten PD-L1-molekyylin liikatuotannon, osoitus immunohistokemiallisesti on yhteydessä suotuisaan hoitovasteeseen immuno-onkologiselle anti-PD1-vasta-aineelle useassa eri kiinteässä kasvaintyyppissä (18). Syöpäsolujen poikkeavuuksien ohella on käynyt ilmeiseksi tarve profiloida entistä kattavammin syövän mikroympäristöä.

HYKS:ssä ja Helsingin yliopistossa selvitetiin hiljattain, kuinka syöpäsoluista kertyvää informaatiotulvaa voisi integroida yksittäisten akuuttia myelooista leukemiaa sairastavien potilaiden hoitoon tapauksissa, joissa muut hoitovaihtoehdot olivat ehtyneet (19). Tutkijoista ja klinikoista muodostuva kasvainryhmä määritteli multiomiikasta ja funktionaalisista solumaljoilla määritettävistä lääkevasteista hoitoehtoja 37 potilaalle täsmälääketieteen periaatteita noudattaen. Vaikka useilla tutkittavista edeltävät hoitovasteet standardihoitoon olivat jääneet riittämättömiksi tai tauti oli uusiutunut, kolmasosa potilaista sai täydellisen vasteen, mikä mahdollisti muun jatkohoidon. Vastavanlaisia tuloksia syntyi EXALT-lääketutkimuksessa, jossa 143 edennyttä hematologista syöpää sairastavan potilaan näytteitä käsiteltiin solumaljalla ja lääkevasteita analysoitiin laskennallisella kuva-analyysillä (20). Yksilöityyn lääketieteeseen perustuvia tehokkaita hoitoyhdistelmiä on tunnistettu vain rajallisesti, ja niiden tutkimustulokset ovat hankalasti tulkittavissa (21,22). Toisaalta potentiaalistaan huolimatta omiikkaan perustuvat hoidot soveltuvat toistaiseksi vain harvoille potilaille satunnaistettujen kontrolloitujen kolmannen vaiheen tutkimusten, kustannusten ja ennen kaikkea sopivien lääkkeiden puuttumisen vuoksi (23).

Terveysdatan rajallisuudet

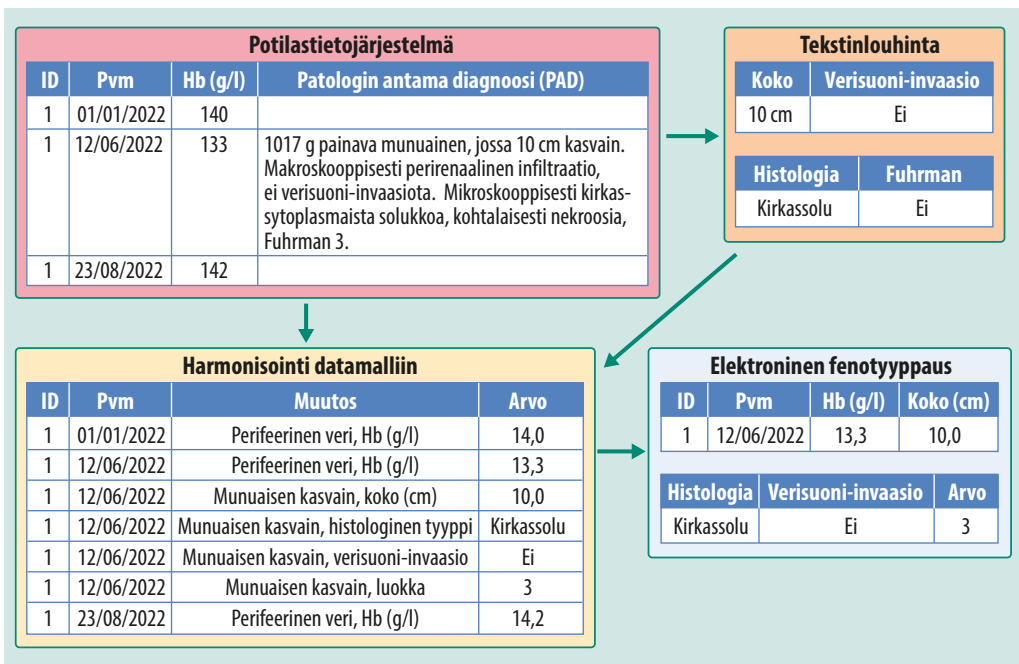
Jotta kalliin lääkekehityksen riskejä saataisiin minimoitua, tutkimuspotilaiden näytteitä kerätään standardoidusti osana lääketutkimuksia. Näytteitä analysoimalla voidaan selvittää tarkemmin lääkeaineen vaikutusmekanismeja sekä tunnistaa tarkasti potilasaloryhmit, jotka hyötyisivät uudesta lääkeaineesta. Akateemiset tutkijat ovat myös aktivoituneet kasvainmerkkiaineiden etsinnässä biopankkien ja yli-

opistojen tutkimuspalvelujen muodostamien infrastruktuurien myötä. Esimerkiksi takautuvassa aineistossa tunnistettu leukemiasolujen ex vivo -vaste kalliille mutta erittäin lupaavalle atsasitidiinin ja venetoklaksin yhdistelmähoidolle on validoitunut myös in vivo -vasteen biomarkkeriksi etenevässä toisen vaiheen tutkimuksessa (24). Vastaavien läpimurtojen myötä hoidon yksilöllistyminen ilmennee myös kansainvälisissä ja kansallisissa hoitosuosituksissa, ja vaikuttavia lääkeaineita voidaan kohdentaa nykyistä biomarkkerivetoisemmin.

Toisaalta kuinka tarkkoja biomarkkereita voimme tunnistaa? Analogiana sateen ennustaminen on mahdollista kohtalaisella todennäköisyydellä muutaman päivän päähän. Toisaalta kaaosteorian mukaisesti sään ennustaminen kahden viikon päähän on hyvin epävarmaa (25). Terveysdataan saatetaan liittää liiankin suuria odotuksia, vaikka useimmat pitävät kasvaimen sisäistä heterogeenisuutta ja ihmisen elimistöä ilmasto haastavampina systeeminä. Onko siis realistista, että voimme ennustaa potilastarkkuudella lääkkeen tarkkaa hoitovastetta tai jäljellä olevaa elinaikaa kuukausitarkkuudella? Osaamme tulkita kliinisesti vain murto-osaa sekvenssidatata. Lisäksi näyttö eri tasoilla ilmenevien molekyylibiologisten muutosten (epigenetiikka, genomiikka, transkriptomiikka, yms.) hyödyntämisestä puoltamaan yksittäistä hoitoa on vielä rajallista. Keskeisiä datalähtöisiä läpimurtoja on vielä todennäköisesti tunnistamatta, mutta nämä vaativat panostuksia korkealaatuisen datan keräämiseen ja tämän jalostamiseen hyödyttämään kliinistä päätöksentekoa.

Laskennallinen fenotyyppi – portti syöpäpotilaan ja tutkijan välillä

Sähköisten potilastietojärjestelmien tulisi myös taipua moniulotteisen datan hallintaan. Sairaaloiden tietotekniseen ympäristöön on investoitu merkittävästi, ja kertyvää tietovarantoa siirretään jatkuvasti tietoaletaksi kutsuttuihin keskitettyihin tietokantapalvelimiin. Terveysdatan määrä on valtava, mutta merkittävä osa siitä ei ole rakenteisessa muodossa ja vaatii muun muassa tekstinlouhinta-algoritmien käyttöä. Esimerkiksi syöpätutkimuksessa olen-



KUVA 2. Potilastiedon jalostaminen tekstinlouhinnalla, harmonisoimalla kansainväliseen datamalliin ja elektronisella fenotyyppauksella.

naisimpiin tietolähteisiin kuuluvat potilaan histopatologiset ja radiologiset lausunnot sekä potilaan yleiskuntoa kuvaavat mittarit kirjataan vain rajallisesti rakenteisessa muodossa. Lisäksi nykyiset potilastietojärjestelmät eivät juurikaan kannusta rakenteisen tiedon kirjaamisen kliinissä työssä esimerkiksi hyödyntämällä tietoa selkeiden ja informatiivisten potilaskohtaisten näkymien muodostamiseksi. Ohjelmistot voisivat myös muodostaa automaattisesti rakenteista tietoa potilaskertomuksesta tai muuttaa vapaata tekstiä rakenteiseksi kliinikoiden käytänteiden mukaan.

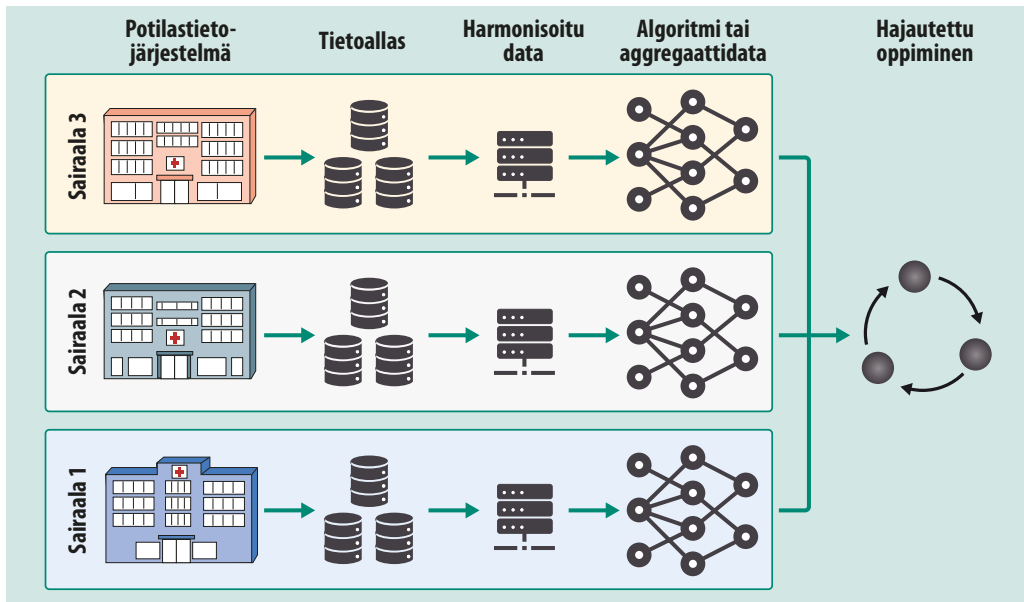
Rakenteinen ja harmonisoitu terveystieto sallisi diagnostiikan, syöpäsairauden etenemisen ja hoitovasteiden määrittämisen automaattisesti hyödyntäen kansainvälisiä lääketieteellisiä suosituksia, kuten WHO:n tautiluokituksia. Laskennallinen fenotyyppin määrittely (digital phenotyping, electronic phenotyping) on perusta usealle potilaskeskeiselle kliiniselle sovellukselle ja hoidon vaikuttavuuden mittaamiselle (26). Tavallisesti tutkimus on nojannut käsin kerättävään terveystietoon ja siitä määritettäviin muuttujiin kuten tautikohtaisiin riskiluokkiin (KUVA 2). Rakenteinen ja harmonisoi-

tu terveystieto sallisi automaattisen potilas- ja tautitaulujen ohjelmoinnin ja tehostaisi merkittävästi kliinistä hoitoa ja tutkimusta. Sairaalatoimintoja voisi seurata lähes reaaliaikaisesti määrittämällä diagnoosikohtaiset suoritteet ja niiden kustannukset ja tutkia kliinisten yksiköiden tietojohdantamista ja resurssien kohdentamista.

HUS:n Syöpäkeskuksessa yhteistyössä TietoEVERY Oy:n kanssa kehitteillä oleva ohjelmisto poimii moniammatillisissa kasvainkouruksissa käsiteltävien potilaiden tiedot interaktiivisiin näkymiin. Automaattinen ja tarkka potilaan terveystilan määrittäminen säästäisi sekä klinikon työaikaä että tehostaisi hoidonohjausta (27).

Yhteistyöstä voimaa

Suomessa on tehty vuosikymmeniä investointeja korkealaatuisten lääketieteellisten rekistereiden rakentamiseksi. FinnGen-hanke ja sen myötä tehdyt tieteelliset havainnot ovat herättäneet globaalia huomiota (28). Sosiaali- ja terveysministeriön kirjoittama laki sosiaali- ja terveystietojen toisiokäytöstä, toisiolakki, on



KUVA 3. Potilastietojärjestelmässä rekisteröityvää henkilötietoa voidaan käsitellä tietoturvallisissa tietoaalassa, harmonisoida yhteiseen datamalliin ja analysoida hajautetulla oppimisella.

kuitenkin hidastanut ja estänyt lääketieteellisen tiedon toissijaista käyttöä (29). Samanaikaisesti riittämättömän julkisen tutkimusrahoituksen myötä kliinisen tutkimuksen tulevaisuus on uhattuna (30).

Sairaaloitten tietoaaltaiden kehittyminen on lähentänyt klinikoita ja tietohallintoa innovoimaan uusia tietoturvallisia ja vaikuttavia ratkaisuja tutkimus- ja kehitystyön ylläpitämiseksi. Terveystietoja on enenevästi harmonisoitu kansainvälisiin, standardoituuihin tietomalleihin (common data model). Suomessa yliopistosairaalat ovat käynnistäneet FinOMOP-hankkeen yhdessä THL:n ja FinnGen-hankkeen kanssa kääntääkseen terveystietoja OMOP (Observational Medical Outcomes Partnership) -tietomalliin (31). OMOP on avoimen kehittäjäyhteisön ylläpitämä standardoitu tietomalli, jonka mukaan kliinisiä tietoja kuten lääkityksiä, verikokeiden tuloksia ja kliinisiä käyntitietoja voidaan tallentaa ja hyödyntää tilastollisissa analyyseissa. Harmonisointi perustuu kansallisiin yhtenäisiin koodistoihin (esim. SNOMED, LOINC, RxNORM) (32). Alkuvuoteen 2023 mennessä lähes 200 terveystiedon rekisterinpitäjää on ilmoittautunut virallisiksi dataharmonisointiin osallistuviksi partnereiksi pelkästään EU-alueella (33), ja on ilmeistä, että OMOP-

standardi tulee olemaan terveystietojen sujuvan jakamisen mahdollistaja.

Hajauta ja hallitse

Kun terveystiedot on harmonisoitu yhteiseen tietomalliin, voi tutkija suunnitella ja toteuttaa tilastolliset analyysinsä usealle rekisterille ilman aineistojen siirtämistä, jakamista tai yhdistämistä (hajautettu tai federoitu laskenta). Esimerkkinä havainnoivan tutkimusasetelman menestyksestä on 4,9 miljoonan verenpaine-potilaan OMOP-aineistossa tehty massiivinen propensiteettikaltaistukseen (propensity score) perustuva tutkimushanke, jossa havaittiin tiatsididiureettien suojaavan kardiovaskulaarisilta päätapahtumilta angiotensiinikonvertaasin estäjiä tehokkaammin (34).

Jopa koneoppimismenetelmien hyödyntäminen on mahdollista hajauttamalla algoritmin kehitys rekisterikohtaisesti ja jakamalla vain mallin parametrit iteraatioiden välillä (KUVA 3). Hajautettu laskenta on byrokraattisesti kevyempää koska vältetään raskaat rekisterinomistajien väliset luovutussopimukset. Laskentatapa on sovitettavissa erilaisiin aineistotyyppiin kuten taulukkomuotoiseen terveystietoon tai matriisimuotoiseen kuvadataan. Hajautettu

oppiminen on toistetusti tuottanut yhtä tarkkoja malleja kuin keskittäessä usean rekisterin aineistot samaan tietokantaan analysoitavaksi (35,36). Säilyttämällä tietojoukot alkuperäisissä tietokannoissaan on myös tietomurron riski vähäisempi.

Kliiniset päätöksentekojärjestelmät

Kliiniset päätöksentekojärjestelmät ovat tietokoneohjelmistoja, jotka tukevat yksilöllisesti hoitosuunnitelmien laatimista (37). Yksinkertaisimmillaan järjestelmät ovat ehdollisia, kynnysarvoihin perustuvia varoituksia. Osa nykyisistä potilastietojärjestelmistä muistuttaa jo nyt varmistamaan lääkeannoksen plasman suuren kreatiniinipitoisuuden vuoksi. Terveysdatan ja ennen kaikkea koneoppimiseen perustuvien ennustemallien kehittymisen ja validoinnin myötä kasvaa tarve myös suorittaa vaativampia laskutoimituksia. Esimerkiksi HUS Neurokirurgia ja Tietohallinto ovat yhdessä luoneet ja ulkoisesti validoineet algoritmin traumaattisen aivovamman liittyvän kuolleisuuden ennustamiseksi hyödyntäen tehohoidon monitoreista kertyvää dynaamista dataa (38).

Odotamme kehitystä tapahtuvan diagnostiikan automatisoimisessa. Erityisesti patologian ja radiologian aloilla on lisääntyvästi panostettu kudosnäytteiden taltioimiseen kuviksi, laajojen kuva-aineistojen tallennukseen sekä kuvien tehokkaaseen analytiikkaan. Nykyiset algoritmit eivät kykene ihmisen tavoin vertaamaan röntgenkuvaleikettä potilaasta aiemmin otettuihin kuviin ja muodostamaan diagnoosiehdotusta esitietoja käyttäen. Sen sijaan algoritmit voivat näyttää radiologille alueet kuvaleikkeistä, joissa on viitteitä patologisista muutoksista ja jopa muodostaa löydöksistä automaattisesti radiologisia lausuntoja (39). Hiljattain OpenAI:n kehittämät vapaata tekstiä prosessoivat laajat kielimallit (large language models) kuten GPT-sovellukset ovat saaneet paljon huomiota, koska näitä ei ole kehitetty yhtä rajallista tehtävää varten. Nauhoitetun ja tekstiksi käännetyn potilas-lääkärinäkökulman voisi jonain päivänä kenties tiivistää automaattisesti sairauskertomus tekstiksi ja jopa tehdä lääketieteellisiä päätelmiä tarvittavista jatkotoimenpiteistä (40).

Ydinasiat

- ▶ Syöpätautien fenotyypin selvittäminen multiomiikalla ja edistyneillä teknologioilla etenee vauhdilla, mutta näyttö yksittäisten hoitopäätösten tueksi on vielä rajallista.
- ▶ Yhteiset tietomallit ja hajautettu laskenta ovat avainasemassa ja saattavat osittain helpottaa vaikuttavaa havainnoivaa tutkimustyötä kiristyneistä tietosuojalajeista huolimatta.
- ▶ Kliiniset päätöksentekojärjestelmät yleistyvät tulevaisuudessa erilaista terveystietoa ja koneoppimismalleja hyödyntäen.

Lopuksi

Yhteisiin tietomalleihin ja hajautettuun laskentaan liittyviä mahdollisuuksia modernisoida potilastietojärjestelmiä ja tehostaa lääketieteellistä päätöksentekoa tulee entisestään kehittää ja ottaa kliiniseen rutiinikäyttöön. Toisaalta matka kohti modernia terveystiedon arkkitehtuuria ja oppivaa terveydenhuoltojärjestelmää vaatii myös kehitystä tukevaa lainsäädäntöä sekä kannustavaa infrastruktuuria innovaatioiden kliiniselle validoinnille. Vaikka potilastyön arjessa multiomiikka ja koneoppivat päätöksentekojärjestelmät saattavat vielä tuntua vierailta, ovat nämä pian keskeisiä työkaluja kohdentamaan viisaammin kalliita hoitoja ja räätälöimään taudinseurantaa syöpäpotilaille. Analogisesti paperiteollisuudelle laadukkaan ja rikkaan systeemilääketieteellisen arvion (”sellun”) tuottaminen nykyaikaisilla menetelmillä onnistuu, mutta tiedon jalostaminen potilasta hyödyttäväksi hoitopäätökseksi (”korkean teknologian tuote”) on vielä lapsenkengissä. ■

OSCAR BRÜCK, tutkimusryhmän johtaja, dosentti
HUS, Syöpäkeskus

KIMMO PORKKA, ylilääkäri, professori
HUS, Syöpäkeskus ja Helsingin yliopisto

TEEMAN TOIMITTAJAT
Sampsu Hautaniemi ja Tuomas Mirtti

KIRJALLISUUTTA

1. Hanahan D. Hallmarks of cancer: new dimensions. *Cancer Discov* 2022;12:31–46.
2. Barretina J, Caponigro G, Stransky N, ym. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 2012;483:603–7.
3. TCGA Research Network. Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* 2013;499:43–9.
4. Stuart T, Satija R. Integrative single-cell analysis. *Nat Rev Genet* 2019;20:257–72.
5. Jahr S, Hentze H, Englisch S, ym. DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells. *Cancer Res* 2001;61:1659–65.
6. Lauer EM, Mutter J, Scherer F. Circulating tumor DNA in B-cell lymphoma: technical advances, clinical applications, and perspectives for translational research. *Leukemia* 2022;36:2151–64.
7. Isomursu A, Kononen J, Kuopio T. Verenkierron solunulkoinen DNA syövän merkkiaineena. *Duodecim* 2015;131:424–32.
8. Väyrynen J, Väyrynen S, Ollila-Raj H, ym. Immunohistokemialliset monivärjäykset kudostutkimuksessa: diagnostiikan arkipäivää ja tulevaisuutta. *Duodecim* 2023;139:531–40.
9. Schalper KA, Rimm DL. Artificial intelligence in digital pathology — new tools for diagnosis and precision oncology. *Nat Rev Clin Oncol* 2019;16:703–15.
10. Ström P, Kartasalo K, Olsson H, ym. Artificial intelligence for diagnosis and grading of prostate cancer in biopsies: a population-based, diagnostic study. *Lancet Oncol* 2020;21:222–32.
11. Chen RJ, Lu MY, Williamson DFK, ym. Pan-cancer integrative histology-genomic analysis via multimodal deep learning. *Cancer Cell* 2022;40:865–78.e6.
12. Brück O, Lallukka-Brück S, Hohtari H, ym. Machine learning of bone marrow histopathology identifies genetic and clinical determinants in MDS patients. *Blood Cancer Discov* 2021;2:238–49.
13. Kather JN, Pearson AT, Halama N, ym. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. *Nat Med* 2019;25:1054–6.
14. Druker BJ, Talpaz M, Resta DJ, ym. Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia. *N Engl J Med* 2001;344:1031–7.
15. Demetri GD, von Mehren M, Blanke CD, ym. Efficacy and safety of Imatinib mesylate in advanced gastrointestinal stromal tumors. *N Engl J Med* 2002;347:472–80.
16. Stone RM, Mandrekar SJ, Sanford BL, ym. Midostaurin plus chemotherapy for acute myeloid leukemia with a FLT3 mutation. *N Engl J Med* 2017;377:454–64.
17. Iländer M, Olsson-Strömberg U, Schlums H, ym. Increased proportion of mature NK cells is associated with successful imatinib discontinuation in chronic myeloid leukemia. *Leukemia* 2017;31:1108–16.
18. Doroshow DB, Bhalla S, Beasley MB, ym. PD-L1 as a biomarker of response to immune-checkpoint inhibitors. *Nat Rev Clin Oncol* 2021;18:345–62.
19. Malani D, Kumar A, Brück O, ym. Implementing a functional precision medicine tumor board for acute myeloid leukemia. *Cancer Discov* 2022;12:388–401.
20. Kornauth C, Pemovska T, Vladimer GI, ym. Functional precision medicine provides clinical benefit in advanced aggressive hematologic cancers and identifies exceptional responders. *Cancer Discov* 2022;12:372–87.
21. Sicklick JK, Kato S, Okamura R, ym. Molecular profiling of cancer patients enables personalized combination therapy: the I-PREDICT study. *Nat Med* 2019;25:744–50.
22. Rodon J, Soria JC, Berger R, ym. Genomic and transcriptomic profiling expands precision cancer medicine: the WINTHER trial. *Nat Med* 2019;25:751–8.
23. Middleton G, Fletcher P, Popat S, ym. The National Lung Matrix Trial of personalized therapy in lung cancer. *Nature* 2020;583:807–12.
24. Kuusanmäki H, Kytölä S, Vääntinen I, ym. Ex vivo venetoclax sensitivity testing predicts treatment response in acute myeloid leukemia. *Haematologica* 2023;108:1768–81.
25. Lorenz EN. Deterministic nonperiodic flow. *J Atmos Sci* 1963;20:130–41.
26. Pathak J, Kho AN, Denny JC. Electronic health records-driven phenotyping: challenges, recent advances, and perspectives. *J Am Med Informatics Assoc* 2013;20:e206–11.
27. Peterzens E. Cost-effectiveness potential of a clinical trial matching application in cancer care. *Väitöskirja*. Helsinki: Aalto University 2021.
28. Kurki MI, Karjalainen J, Palta P, ym. FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature* 2023;613:508–18.
29. Sipilä J. Rekisteritutkijan karu todellisuus. *Lääkärelehti* 12.1.2023.
30. Reito A, Sanmark E, Tuovinen T, ym. Toisiolaki – lääketieteellisen tutkimuksen mahdollistaja vai tukahduttaja? *Suom Lääkäril* 2022;78:e30589
31. Laitinen T, Virkki A, Porkka K. FinOMOP: terveystietojen kansainvälinen harmonisointi. *Duodecim* 2022;138:1761–3.
32. Tiedonhallinta sosiaali- ja terveysalalla. Koodistopalvelu. Helsinki: Terveyden ja hyvinvoinnin laitos 3.12.2020.
33. EHDEN. The European health data & evidence network portal. <https://portal.ehden.eu/>.
34. Suchard MA, Schuemie MJ, Krumholz HM, ym. Comprehensive comparative effectiveness and safety of first-line anti-hypertensive drug classes: a systematic, multinational, large-scale analysis. *Lancet* 2019;394:1816–26.
35. Warnat-Herresthal S, Schultze H, Shastry KL, ym. Swarm learning for decentralized and confidential clinical machine learning. *Nature* 2021;594:265–70.
36. Ogier du Terrail J, Leopold A, Joly C, ym. Federated learning for predicting histological response to neoadjuvant chemotherapy in triple-negative breast cancer. *Nat Med* 2023;29:135–46.
37. Sutton RT, Pincock D, Baumgart DC, ym. An overview of clinical decision support systems: benefits, risks, and strategies for success. *NPJ Digit Med* 2020;3:17.
38. Raj R, Wennervirta JM, Tjerkaski J, ym. Dynamic prediction of mortality after traumatic brain injury using a machine learning algorithm. *NJP Digit Med* 2022;5:96.
39. Tiu E, Talus E, Patel P, ym. Expert-level detection of pathologies from unannotated chest X-ray images via self-supervised learning. *Nat Biomed Eng* 2022;6:1399–406.
40. Lee P, Bubeck S, Petro J. Benefits, Limits, and risks of GPT-4 as an AI chatbot for medicine. *N Engl J Med* 2023;388:1233–9.

SIDONNAISUDET

Oscar Brück: Apuraha (Gilead Sciences, Pfizer), luentopalkkio/asiantuntijapalkkio (Amgen, Novartis, Sanofi, Astellas Pharma, GSK)

Kimmo Porkka: Apuraha (BMS, Celgene, Incyte, Novartis, Pfizer), luentopalkkio/asiantuntijapalkkio (BMS, Celgene, Novartis, Pfizer), luottamustoimet (Euroopan hematologiyhdistys (EHA), hematologien biopankki FHRB)