

Tuomo Mantere ja Katri Pylkäs

Uudet long-read-teknologiat – kohti tarkennettua genomitietoa perinnöllisistä sairauksista ja syövästä

Uusien molekyylibiologisten menetelmien kehityksellä ja laajamittaisella hyödyntämisellä on avainasema eri sairauksien taustalla vaikuttavan geneettisen vaihtelun tutkimisessa. Uuden sukupolven sekvensointimenetelmät ovat mahdollistaneet lukuisien tautigeenien tunnistamisen, ja kustannustehokkuutensa ansiosta ne ovat tulleet tärkeäksi osaksi myös kliinistä rutiinidiagnostiikkaa. On silti tärkeää tiedostaa näiden menetelmien puutteet erityisesti genomien rakenteellisten muutosten havaitsemisessa, jossa hyödynnetään edelleen pääasiassa tavanomaisia sytogeneettisiä menetelmiä. Viimeaikaisen teknologisen kehityksen myötä uudet long-read-menetelmät voivat mahdollistaa entistä tarkemman ja kattavamman genomitiedon muodostamisen sekä korvata nykyisiä tutkimusmenetelmiä tulevaisuudessa.

Kolmesta miljardista emäsparista koostuva ihmisgenomi on rakenteeltaan hyvin monimutkainen. Suurimmaksi osaksi eri ihmisten genomi on identtinen, mutta siinä esiintyy myös vaihtelua, joka vaikuttaa yksilön ilmiäsuun. Osa tästä vaihtelusta voi aiheuttaa sairauden tai vaikuttaa yksilön sairastumisriskiin (1).

Yksittäisten emäsparimuutosten tai emästen pienten lukumäärämuutosten lisäksi genomi on täynnä pituusiltaan vaihtelevia DNA-toistojaksoalueita sekä rakenteellista muuttelua, johon lukeutuvat kaikki kooltaan yli 50 emäsparia kattavat muutokset. Tällaisia ovat esimerkiksi tietyn genomialueen kopiokopio-
muutokset (copy number variation, CNV, deleetiot tai duplikaatiot), insertiot, inversiot, translokaatiot ja muut monimutkaiset genomiset uudelleenjärjestäytymiset. Jos tarkastellaan näissä mukana olevien emäsparien lukumäärää, rakenteelliset muutokset selittävät suuremman osuuden eroista kahden satunnaisen ihmisgenomin välillä kuin pistemutaatiot ja pienet emästen lukumäärän muutokset yhteenlasketuna (2). Pistemutaatioiden tavoin genomien rakenteelliset muutokset voivat liittyä lukuisiin

eri sairauksiin, esimerkiksi syöpään ja kehityshäiriöihin (3).

Genetiikan menetelmät kehittyvät

Geneettisen vaihtelun yhdistäminen eri sairauksiin on aina ollut vahvasti riippuvaista molekyylibiologisten menetelmien kehityksestä (1). Varhaisimmat genetiikan menetelmät, kuten kromosomien raitavärjäys, genomien kytkentäanalyysit ja paikkakloonauus sekä myöhemmin yksittäisten kandidaattigeenien Sanger-sekvensointi ja genomilaajuiset assosiaatiotutkimukset, ovat osaltaan lisänneet ymmärrystä eri sairauksien geneettisestä taustasta.

Viimeksi kuluneen reilun vuosikymmenen aikana uuden sukupolven sekvensointimenetelmät (next-generation sequencing, NGS eli rinnakkaissekvensointi) ovat hallinneet geneettistä tutkimusta ja mahdollistaneet lukuisien uusien tautigeenien tunnistamisen sekä tulleet tärkeäksi osaksi kliinistä rutiinidiagnostiikkaa (4,5). Näiden menetelmien ansiosta geneettinen tutkimus ja geenitestit ovat siirtyneet yksittäisten geenien tarkastelusta laajempien geenipaneelien ja koko genomien tai sen proteiini-

neja koodaavan alueen, eksomin, sekvensointiin. Tämä on mahdollistanut myös merkittäviä edistysaskeleita syövän molekulaaristen mekanismien ymmärtämisessä (6). Nykyisten NGS-menettelmien tärkeimmät vahvuudet ovat kustannustehokkuus, nopeus ja luotettavuus DNA:ssa esiintyvien pienten emäsmuutosten havaitsemisessa (TAULUKKO) (4).

Vaikka NGS-teknologiat ovat kiistatta mul- listaneet genetiikan alan tutkimusta ja diagnos- tiikkaa ennennäkemättömällä tavalla, ihmisge- nomille tyypillinen runsas toistojaksoalueiden esiintyminen, rakenteellinen muuntelu sekä homologia-alueet (kuten toiminnallinen geeni ja sen toimimaton geenikopio eli valegeeni) ai- heuttavat perustavanlaatuisia rajoitteita näille menetelmille. Nykymuotoinen NGS perustuu lyhyisiin (alle 300 emäsparia) sekvenssifrag- mentteihin, niin sanottuihin lukuihin (read) ja niiden linjaamiseen verrokkigenomia vasten (7).

Ihmisen runsaisten toistojaksoalueiden vuoksi yksittäinen lyhyt luku ei aina sisällä tar- peeksi informaatiota sen ankkuroimiseksi var- muudella oikeaan kohtaan genomia. Tämä voi johtaa sekvensointituloksien tulkinnanvaraisuuteen ja virheisiin. Lukujen lyhyiden vuoksi ra- kenteellisten muutosten tunnistaminen NGS:n avulla perustuu useimmiten suoran havainnoi- nin sijasta epäsuoraan päättelyyn (esimerkiksi lukusyvyyden poikkeavuudet) (7).

Useimmat NGS-menettelmät hyödyntävät myös PCR-monistamista, mikä voi johtaa epä- tasaisuuteen sekvensoitavien kohdealueiden sekvenssilukujen määrässä ja heikentää analy- sin luotettavuutta. Erityisesti genomiset alueet, joilla esiintyy runsaasti guaniini (G)- ja sytosii- ni (C) -emäksiä, ovat entsyymaattisesti vaikeasti monistettavia, minkä vuoksi alueelle saadaan vähemmän sekvenssilukuja. Tämä vaikeuttaa GC-rikkaiden alueiden luotettavaa analyysia (8).

Edellä mainittujen puutteiden vuoksi NGS- menettelmät eivät toistaiseksi ole kyenneet kor- vaamaan klassisia sytogenetiikan menetelmiä, joita käytetään kliinisessä diagnosoinnissa laa- jempien kromosomaalisten muutosten havait- semiseen. Uusimmat long-read-teknologiat, mukaan lukien long-read-sekvensointimenetel-

mät ja optinen genomikartointi, ovat kuitenkin nousemassa tutkimusmenetelmiksi, joilla voi- daan täydentää tai osittain jopa korvata vallit- sevia sekvensointimenetelmiä sekä sytogeneti- kan tutkimuksia (2,9).

Long-read-teknologiat

Long-read-sekvensointimenetelmistä (LRS) käytetään myös nimitystä kolmannen sukupol- ven sekvensointimenettelmät. Niiden tärkein etu vallitseviin NGS-menettelmiin verrattuna on nimensä mukaisesti sekvenssilukujen pi- tuus, joka on keskimäärin jopa kymmeniä tu- hansia emäspareja (TAULUKKO).

PCR-monistamisen sijaan LRS-teknologiat perustuvat yksittäisten DNA-molekyylien sek-vensointiin, jolloin monistamiseen liittyviä epätasaisuuksia ei synny ja DNA:ssa mahdol- lisesti esiintyvät modifikaatiot säilyvät. Tämä mahdollistaa esimerkiksi DNA-metylaation ja muiden epigeneettisten muutosten havainnoi- misen suoraan sekvensoimalla (8). Nykyään hallitsevat LRS-menettelmät ovat single mole- cule real-time (SMRT) -sekvensointi (PacBio) sekä nanopore-sekvensointi (Oxford Nanopore Technologies).

Ensimmäisen kaupallisen pitkiin lukuihin perustuvan sekvensointilaitteen lanseerasi Pac- Bio vuonna 2011. Varhaisen vaiheen PacBion SMRT-sekvensoinnin tarkkuus oli hyvin heik- ko, mutta nykyisin ongelma voidaan kiertää sekvensoimalla sama DNA-molekyyli useaan kertaan, ja emätason tarkkuus lähestyykin jo NGS-menettelmiä. Hyvä sekvensointitarkkuus yhdistettynä yli kymmenentuhannen emäspa- rin mittaisiin lukupituuksiin tarjoaa mahdolli- suuden erittäin kattavaan koko genomien ana- lyysiin (9). SMRT-sekvensoinnin merkittävim- mät nykyiset rajoitteet ovat moninkertainen hinta ja hitaus verrattuna NGS:ään sekä me- netelmään vaadittava korkealaatuisen DNA:n määrä (TAULUKKO).

Oxford Nanopore Technologies julkaisi pit- kiin lukuihin perustuvan nanopore-sekvensoi- ntilaitteensa vuonna 2015. Nanopore-sekven- soinnin etuna ovat keskimäärin jopa pidemmät luvut kuin SMRT-sekvensoinnissa sekä edul- lisempi hinta. Nanopore-sekvensaattorista on

TAULUKKO. Yhteenveto NGS- eli rinnakkaissekvensointimenetelmistä, long-read-teknologioista ja tavanomaisista sytogeneettisistä menetelmistä.

Menetelmä	Lukupituus (emäsparia)	Pistemutaatioiden havaitseminen	Rakenteellisten muutosten havaitseminen	Muuta
NGS-menetelmät	75–300	Hyvä tarkkuus ja luotettavuus	Epäsuora havainnointi, lyhyiden lukupituuksien vuoksi puutteellinen	Pitkään käytetty menetelmä, jolle vakiintuneet analyysikäytännöt Voidaan käyttää myös huonolaatuisen DNA:n analysoinnissa
Long-read-sekvensointi	10 000–20 000	Emästason tarkkuus ei vielä NGS-menetelmiä vastaava	Kyllä, usein suora havainnointi	Voidaan käyttää vaativien genomisten alueiden tutkimisessa Menetelmät edelleen kalliita, ja niihin tarvitaan runsaasti hyvälaatuista DNA:ta
Optinen genomikartoitus	> 250 000	Ei, emästason erottelukyky puuttuu	Kyllä, erottelukyky noin 500 emäsparia Havaitsee pienet klonaaliset muutokset ja suuret kromosomaaliset muutokset	Analyysin helppous ja edullinen hinta verrattuna long-read-sekvensointiin Suuret DNA:n laatuvaatimukset, tarvitaan oma DNA:n eristysmenetelmä
Karyotyypitys (G-raitavärjäys)	Analyysi kromosomeittain	Ei, emästason erottelukyky puuttuu	Kyllä, erottelukyky noin 5–10 miljoonaa emäsparia	Kokonaisten kromosomien analyysi, myös Robertsonin translokaatiot Analyysi vaatii jakautuvia soluja
Molekyylikaryotyypitys	–	Ei, emästason erottelukyky puuttuu	Kyllä, erottelukyky noin 1 000–2 000 emäsparia	Ei havaitse balansoituneita muutoksia
FISH-koettimet	–	Ei, emästason erottelukyky puuttuu	Kyllä, vain ennalta määritetyt muutokset	Solutason analyysi

saatavilla myös kannettava versio, jota voidaan hyödyntää erilaisissa kenttäolosuhteissa, esimerkiksi virusepidemioiden valvonnassa (10).

Nykyisen nanopore-teknologian merkittävä heikkous on sekvensointivirheiden suuri määrä. Virheprofiili on myös osittain systemaattinen, jolloin pelkästään sekvenssilukujen määrää lisäämällä sekvensointivirheistä ei päästä eroon (8). LRS-menetelmien data-analyysi on myös merkittävästi varhaisemmassa kehitysvaiheessa verrattaessa pitkään käytössä olleisiin NGS-menetelmiin (11).

Edellä mainittujen uusien sekvensointimenetelmien lisäksi long-read-teknologioihin voidaan lukea myös Bionano Genomicsin kehittämä optinen genomikartoitus. Menetelmä perustuu suuren erottelukyvyn mikroskopiaan sekä automatisoituun kuva-analyysiin, joka mahdollistaa ultrapitkien (yli 250 000 emäsparia) entsymaattisesti leimattujen DNA-molekyylien kuvantamisen (12). Genomi koostetaan näiden kuvien perusteella, ja jokaisen kromosomin leimakuviota verrataan verrokki-geenomiin rakenteellisten muutosten tunnistamiseksi. Optinen genomikartoitus ei perustu

DNA:n sekvensointiin, joten se ei havaitse pistemutaatioita (**TAULUKKO**).

Kohti tarkentuvaa genomitietoa

Harvinaisten geneettisten sairauksien diagnoosi on parantunut huomattavasti NGS:n myötä. Rutiinimaisesti kliinisessä käytössä olevat menetelmät kuten eksomisekvensointi ja mikrosiruihin perustuva molekyylikaryotyypitys eivät silti tarjoa kokonaisvaltaista kuvaa geneettisestä vaihtelusta. Tautiryhmän mukaan eksomi- tai genomisekvensoinnin avulla pystytään diagnosoimaan arviolta noin 30–70 % tutkittavista potilaista (13,14). Uusien tautigeenien ja mutaatioiden tunnistamiseksi onkin tärkeää selvittää uusien teknologioiden tarjoamia mahdollisuuksia.

Rakenteelliset geneettiset muutokset. Osa puuttuvista molekyyylitason diagnooseista voi selittyä havaitsematta jääneillä genomien rakenteellisilla muutoksilla. Eri teknologioiden väliset vertailut ovat osoittaneet, että lyhyisiin sekvenssilukuihin perustuvissa NGS-analyysissä jopa yli 70 % genomien rakenteellisista muu-

TIETOLAATIKKO. Termejä ja käsitteitä.

Genomilla tarkoitetaan ihmisen koko perimää. Se sisältää kaikki proteiineja koodaavat alueet eli geenit sekä ne perimän alueet, jotka eivät koodaa proteiineja.

GC-rikas alue. Genomin alue, jossa esiintyy runsaasti guaniini (G)- ja sytosiini (C) -emäksiä. Alueet ovat entsyymaattisesti vaikeasti monistettavia.

DNA:n toistojaksoalueet. Useimmat geenisekvenssit esiintyvät genomissa vain kerran. Sen sijaan DNA:n toistojaksoalueet esiintyvät useina, jopa tuhansina kopioina. Toistot voivat esiintyä peräkkäin (tandem) tai sattumanvaraisesti (interspersed) ympäri genomia, kuten liikkuvat LINE- ja SINE-retrotransposonielementit.

SMRT-sekvensointi (single molecule real-time) -termiä käytetään PacBio-yrityksen long-read-sekvensointitekniologiaista.

Sytogenetiikka on tieteenala, jossa selvitetään periytyksen ja sairauksien suhdetta kromosomaalisiin muutoksiin. Sytogeneettisen tutkimuksen tärkeimmät menetelmät ovat G-raitavärjäys (karyotyypitys), FISH-koettimet (fluorescence in situ hybridization) ja mikrosiruanalyysiin perustuva molekyylikaryotyypitys.

Genomin rakenteellisella muutoksella viitataan kaikkiin yli 50 emäsparia kattaviin perimän muutoksiin. Tällaisia ovat deleetiot (perimäaineksen häviämät), duplikaatiot (perimäaineksen monistumat), insertiot (uuden perimäaineksen liittyminen DNA-ketjuun), translokaatiot (perimäaineksen järjestäytyminen tavanomaisesta poikkeavalla tavalla yleensä kahden kromosomin välillä), inversiot (kromosomin tietyn osan perimäaineksen järjestyksen muuttuminen käänteiseksi) ja aneuploidiat (kokonaisten kromosomien lukumäärämuutokset).

Uuden sukupolven sekvensointimenetelmät (next-generation sequencing, NGS) eli rinnakkaissekvensointi tarkoittaa lukuisien DNA-ketjujen emäsjär-

jestyksen yhtäaikaista selvittämistä eli sekvensointia. Lyhyisiin lukupituuksiin (alle 300 emäsparia) perustuva menetelmä soveltuu erityisesti yksittäisten pistemutaatioiden sekä pienten insertioiden ja deleetioiden havaitsemiseen. Näihin menetelmiin lukeutuvat koko genomin ja eksomin sekvensointi sekä geenipaneelit.

Verrokkigenomi on kansainvälisen ihmisgenomiprojektin tuottama emäsjärjestys ihmisen genomista, jota on sovitettu käytettäväksi vertailusekvenssinä genomianalyseissa.

Valegeeni on evoluution kuluessa tapahtuneen toiminnallisen geenin osittaisen tai täydellisen kopioitumisen myötä syntynyt toimimaton geenikopio.

Long-read-sekvensoinnilla (long-read sequencing, LRS) viitataan uusiin sekvensointimenetelmiin, joiden avulla voidaan tuottaa kymmenien tuhansien tai jopa satojen tuhansien emäsparien mittaisia DNA-sekvenssilukuja. Näistä teknologioista käytetään myös nimitystä kolmannen sukupolven sekvensointimenetelmät. NGS-menetelmien tapaan näiden teknologioiden avulla voidaan sekvensoida koko genomi tai käyttää kohdennettuja analyysitapoja. Pitkät luvut mahdollistavat ihmisgenomin vaativimpien alueiden analysoinnin sekä rakenteellisten muutosten paremman havaitsemisen.

Nanopore-sekvensointi on Oxford Nanopore Technologies -yrityksen kehittämä LRS-menetelmä, joka perustuu DNA-molekyylin kulkeutumiseen kalvolla sijaitsevan kanavaproteiinin lävitse.

Optinen genomikartoitus perustuu useiden satojen tuhansien emäsparien mitaisten, entsyymaattisesti leimattujen DNA-molekyylien kuvantamiseen. Koko genomi koostetaan näiden leimakuvioiden avulla. Näytteen leimakuvioiden vertaaminen verrokkigenomista saatuihin leimakuvioihin mahdollistaa rakenteellisten muutosten havaitsemisen.

toksista voi jäädä tunnistamatta (2). NGS:stä poiketen LRS-menetelmien pitkät luvut voivat usein kattaa koko rakenteellisen muutoksen emästason tarkkuudella mukaan lukien mahdolliset DNA:n toistojaksot, jotka ovat usein myös osallisina genomin rakenteellisten muutosten muodostumisessa (15).

Long-read-teknologioita on viime aikoina hyödynnetty erityisesti yksigeenisten sairauksien tutkimuksissa. Esimerkiksi aggressiivisen keskushermostosyövän (epätyypillinen teratoidi-rhabdoidikasvain, ATRT) geneettistä syytä

etsittiin perheessä, jonka kaksi lasta oli menettänyt tämän hyvin harvinaisen sairauden vuoksi (16). Yleensä ATRT-kasvaimissa havaitaan *SMARCB1*-kasvunrajoitegeenin inaktivaatio molemmissa alleeleissa, ja noin 30 % näistä syövästä syntyy ituradan uuden heterotsygoottisen *SMARCB1*-mutaation seurauksena.

Tutkitun perheen molempien potilaiden kasvaimissa todettiin proteiinitasolla *SMARCB1*-inaktivaatio, mutta diagnostiset tutkimukset ituradan muutoksen havaitsemiseksi, mukaan lukien eksomi- ja koko genomin sekvensointi

sekä molekyylikaryotyypitys, jäivät tuloksiltaan negatiivisiksi.

Jatkoselvityksissä hyödynnettiin long-read-teknologioita, joista ensin toteutettu optinen genomikartoitus tunnisti noin 2 800 emäsparin kokoisen insertin *SMARCB1*-geenissä. SMRT-sekvensoinnilla toteutettu emästason karakterisointi tunnisti muutoksen genomini liikkuvaksi DNA-elementiksi, retrotransposoniksi, joka oli insertoitunut *SMARCB1*:n introniselle alueelle ja aiheuttanut silmukointivirheen. Sama muutos todettiin myös äidin verinäytteestä lievänä mosaikismina, mikä varmisti lopullisen diagnoosin molekyyllitasolla sekä samalla mahdollisti oikeaan tietoon perustuvan perinnöllisyysneuvonnan ja perhesuunnittelun (16). Vastaavanlaisia transposoni-insertioita on tunnistettu LRS-menetelmien avulla myös perinnölliseen rinta- ja munasarjasyöpäalttiuteen liittyvässä *BRCA1*-kasvunrajoitegeenissä (17).

Toisessa esimerkkitapauksessa tunnistettiin X-kromosomaalisesti periytyvän korioidieremian aiheuttava geenivirhe *CHM*-geenissä, joka on toistaiseksi ainoa sairauteen yhdistetty tautigeeni (18). Perheessä oli todettu mRNA-tasolla *CHM:n* silmukointivirhe jo kaksi vuosikymmentä aikaisemmin, mutta genomitason muutosta ei yrityksistä huolimatta ollut löydetty. Optisen genomikartoituksen ja SMRT-sekvensoinnin avulla taudin aiheuttavaksi geenivirheeksi tunnistettiin lopulta yhden eksonin kattava kääntynyt duplikaatio *CHM*-geenissä (18).

Lukuisten yksittäisten tapausten lisäksi long-read-menetelmiä on onnistuneesti hyödynnetty useiden eri potilasryhmien rakenteellisten muutosten tunnistamisessa ja uusien tautigeenien tunnistamisessa (19,20).

Vaikeasti sekvensoitavat genomiset alueet. Genomin sekvensoinnin osalta on osoitettu, että LRS-menetelmien avulla voidaan sekvensoida jopa noin 35 miljoonaa emäsparia, mikä käsittää yli sata geeniä sellaisilta genomisilta alueilta, jotka eivät ole luotettavasti analysoitavissa nykyisten NGS-menetelmien avulla (21). Näihin sisältyy muun muassa genomini toistojakoalueita ja GC-rikkaita alueita (22). Nähtäväksi kuitenkin jää, löytyykö näiltä alueil-

Ydinasiat

- ▶ Rinnakkaissekvensointi (NGS) -menetelmillä havaitaan erityisen tehokkaasti pienten emästason mutaatioita mutta puutteellisesti genomini rakenteellisia muutoksia.
- ▶ Uusien long-read-teknologioiden avulla voidaan tuottaa entistä tarkempaa tietoa ihmisen genomista ja sen rakenteesta.
- ▶ Optinen genomikartoitus voi tulevaisuudessa korvata nykyisiä sytogeneettisiä tutkimusmenetelmiä.

ta tulevaisuudessa sairauksien kannalta uusia merkittäviä geenejä ja mutaatioita (23).

Peräkkäiset toistojaksolaajentumat. Genomissa esiintyvät lyhyet peräkkäiset eli tandemtoistojaksot muodostuvat yleensä 1–6 emäsparin mittaisista DNA-jaksoista, jotka ovat erityisen alttiita DNA:n kahdentumisen aikana tapahtuville mutaatioille. Toistaiseksi on tunnistettu ainakin 50 yksigeenistä sairautta, jotka ovat seurausta DNA-toistojaksojen laajenemisesta (24).

Toistojaksolaajentumien tutkiminen PCR:n, kloonauksen ja DNA-sekvensoinnin keinoin vaatii erityisiä kohdennettuja analyysitapoja. NGS:n käyttäminen toistojaksojen koon määrittämisessä on ongelmallista, koska valtaosa sairauden aiheuttavista toistojaksolaajentumista on pidempiä kuin NGS:n lukupituudet (11).

LRS-menetelmillä voidaan sen sijaan sekvensoida kokonaisia pitkiä ja GC-rikkaita toistojaksolaajentumia, esimerkiksi särö-X-oireyhtymän (fragile X syndrome) tapauksessa (25). Erityisen lupaavaa näiden teknologioiden osalta on myös genomilaajuinen toistojaksolaajentumien paikantaminen ja sekvensoiminen, joka on mahdollistanut uusien patogeenisien toistojaksolaajentumien tunnistamisen (26,27).

Valegeenit ovat toimimattomia geenejä, jotka ovat syntyneet geeniduplikaatioiden seurauksena ja muistuttavat siksi sekvenssiltään alkuperäistä toiminnallista geeniä. Useilla kliinisesti merkittävillä geeneillä esiintyy genomis-

samme valegenejä, jotka voivat hankaloittaa geneettisen vaihtelun luotettavaa tunnistamista näistä geeneistä ja johtaa jopa vääriin diagnooseihin (28,29).

Käytettäessä LRS-menetelmiä kymmenien tuhansien emäsparien mittaiset luvut sisältävät tarpeeksi muuntelevia kohtia, jotta toiminnallisesta geenistä ja valegeenistä peräisin olevat sekvenssit voidaan luotettavasti erottaa toisistaan. Long-read-sekvensoinnilla on saatu hyviä tuloksia useiden vaativien geenialueiden, kuten *PKD1:n* (autosomissa vallitsevasti periytyvä munuaisten monirakkulatauti) ja farmakogeneettisesti tärkeän *CYP2D6:n* luotettavasta sekvenssianalyysistä (30,31).

Optinen genomikartointi – seuraavan sukupolven sytogeneettinen analyysi?

NGS ei ole toistaiseksi pystynyt korvaamaan kromosomaalisia muutoksia tunnistavia tavanomaisia sytogeneettisiä analyyseja. Näihin menetelmiin lukeutuvat kromosomien G-raita-värjäys (karyotyypitys), FISH-koettimet (fluorescence in situ hybridization) ja mikrosiruihin perustuva molekyylikaryotyypitys, joilla on tärkeä osa diagnoosin määrittämisessä ja riskiarvioinnissa erityisesti hematologisissa syövässä sekä kehityshäiriöiden ja infertiliteetin syiden selvittämisessä.

Kyseisillä menetelmillä on kuitenkin myös hyvin tunnetut heikkoutensa: karyotyypityksen erottelukyky on huono (5–10 miljoonaa emäsparia) ja FISH-koettimet toimivat vain kohdennetusti ennalta määritetyn muutoksen havaitsemisessa. Molekyylikaryotyypityksen erottelukyky on parhaimmillaan noin 1 000–2 000 emäsparia, mutta menetelmä ei havaitse balansoituneita muutoksia (**TAULUKKO**) (32). Tällaisia ovat muun muassa inversiot ja balansoituneet translokaatiot, kuten *BCR-ABL*-geenifuusioon johtava translokaatio kroonisessa myeloidisessa leukemiassa (33). Yksittäisten sytogenetiikan testien puutteiden vuoksi on yleistä, että diagnoosivaiheessa joudutaan käyttämään useiden eri testien yhdistelmää. Tämä on kallista ja vie aikaa.

Long-read-tekniikoihin lukeutuvan opti-

sen genomikartointitekniikan käytettävyyttä kattavana sytogeneettisenä testinä on selvitetty uusissa tutkimuksissa, ja tulokset ovat olleet lupaavia (34,35). Hematologista syöpää sairastavien 52 potilaan näytteille, joista rutiinidiagnostiset sytogeneettiset testit olivat tunnistaneet kromosomaalisia muutoksia, tehtiin optinen genomikartointi (35). Tulokset olivat erittäin yhtenevät aikaisempien sytogeneettisten testien kanssa, ja lisäksi menetelmällä kyettiin tunnistamaan uusia muutoksia. Optinen genomikartointi tarjoaa näin mahdollisen kliinisen käytettävyyden lisäksi helppokäyttöisen keinon löytää uusia syöpään mahdollisesti liittyviä muutoksia (35).

Toisessa monikeskustutkimuksessa analysoitiin kromosomaalisten muutosten osalta 85 näytettä, joista oli aikaisemmin diagnostisesti raportoitu yhteensä 99 muutosta karyotyypityksen, FISH-analyysien sekä molekyylikaryotyypityksen yhdistelmän perusteella (34). Kaikki samat muutokset tunnistettiin käyttämällä yksittäistä optiseen genomikartointuun perustuvaa testiä. Näiden tutkimustulosten perusteella on mahdollista, että optinen genomikartointi voi tulevaisuudessa toimia seuraavan sukupolven sytogeneettisenä testinä.

Optisen genomikartoinnin avulla ei vielä voida tunnistaa kromosomien sentromeerialueiden katkoskohtia, sillä alueilla ei sijaitse genomia näytekohtaisessa koostamisessa käytettäviä leimakohtia. Menetelmän teknistä toteuttamista varten tarvitaan lisäksi aina hyvälaatuista alkumateriaalia, jotta ultrapitkien DNA-molekyylien eristäminen on mahdollista. Edullinen hinta ja data-analyysin helppous ilman kattavaa bioinformatiikkaa voivat kuitenkin mahdollistaa optisen genomikartoinnin käytön diagnostisena sytogeneettisenä testinä lähitulevaisuudessa sekä tutkimuskäytön valitsevilla NGS-menetelmillä saatavan genomitiedon täydentämiseksi. Lisäksi menetelmän herkkyys mahdollistaa pienten klonaalisten muutosten havaitsemisen syöpänäytteistä, esimerkiksi muutosten, jotka esiintyvät alle 10 %:ssa tutkitun näytteen soluista (35). Sekvensointitekniikoilla, mukaan lukien LRS:llä ja NGS:llä, tämä ei toistaiseksi ole vielä kustannustehokasta koko genomia osalta.

Lopuksi

Uudet long-read-tekniologiat ovat viime vuosina kehittyneet nopeasti ja voivat tulevaisuudessa mahdollistaa entistä kattavamman geneettisen analyysin. Erityisesti genomilaajuinen rakenteellisten muutosten tarkempi tunnistaminen sekä toistojaksoalueiden ja -laajentumien tutkiminen voivat mahdollistaa tärkeitä löydöksiä eri sairauksien taustalta. Näiden tekniologioiden yleistymisen ja laajemman

käyttöönoton myötä on oletettavissa, että pääsemme kohti yhä tarkempaa sekä perinnöllisiä sairauksia että syöpää koskevaa genomitietoa. Tekniologioita voidaan hyödyntää myös kansansairauksien ja yksilöllisten ominaisuuksien geneettisten taustatekijöiden tutkimuksissa sekä entistä tarkemman ihmisen verrokkitieteen tuottamisessa sekvensoimalla kaikki kromosomit päästä päähän (36,37). ■

TUOMO MANTERE, FT, Suomen Akatemian tutkijatohtori

Syöpägenetiikan ja tuumoribiologian laboratorio, Syövän ja translationaalisen lääketieteen tutkimusyksikkö, Biocenter Oulu, Oulun yliopisto
Department of Human Genetics, Radboud University Medical Center, Nijmegen, the Netherlands
Twitter: @MantereTuomo

KATRI PYLKÄS, FT, syöpägenetiikan dosentti, akatemiaturkija

Syöpägenetiikan ja tuumoribiologian laboratorio, syövän ja translationaalisen lääketieteen tutkimusyksikkö, Biocenter Oulu, Oulun yliopisto, Nordlab Oulu
Twitter: @Pylkaskatri

VASTUUTOIMITTAJA

Tuomas Mirtti

SIDONNAISUUDET

Kirjoittajilla ei ole sidonnaisuuksia.

KIRJALLISUUTTA

1. Claussnitzer M, Cho JH, Collins R, ym. A brief history of human disease genetics. *Nature* 2020;577:179–89.
2. Chaisson MJP, Sanders AD, Zhao X, ym. Multi-platform discovery of haplotype-resolved structural variation in human genomes. *Nat Commun*, julkaistu verkossa 16.4.2019. DOI: 10.1038/s41467-018-08148-z.
3. Collins RL, Brand H, Karczewski KJ, ym. A structural variation reference for medical and population genetics. *Nature* 2020;581:444–51.
4. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet* 2016;17:333–51.
5. Gilissen C, Hoischen A, Brunner HG, ym. Unlocking Mendelian disease using exome sequencing. *Genome Biol* 2011, julkaistu verkossa 14.9.2011. DOI: 10.1186/gb-2011-12-9-228.
6. Alexandrov LB, Nik-Zainal S, Wedge DC, ym. Signatures of mutational processes in human cancer. *Nature* 2013;500:415–21.
7. Treangen TJ, Salzberg SL. Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat Rev Genet* 2012;13:36–46.
8. van Dijk EL, Jaszczyszyn Y, Naquin D, ym. The third revolution in sequencing technology. *Trends Genet* 2018;34:666–81.
9. Logsdon GA, Vollger MR, Eichler EE. Long-read human genome sequencing and its applications. *Nat Rev Genet* 2020;21:597–614.
10. Quick J, Loman NJ, Duraffour S, ym. Real-time, portable genome sequencing for Ebola surveillance. *Nature* 2016;530:228–32.
11. Sedlazeck FJ, Lee H, Darby CA, ym. Piercing the dark matter: bioinformatics of long-range sequencing and mapping. *Nat Rev Genet* 2018;19:329–46.
12. Cao H, Hastie AR, Cao D, ym. Rapid detection of structural variation in a human genome using nanochannel-based genome mapping technology. *Gigascience*, julkaistu verkossa 30.12.2014. DOI: 10.1186/2047-217X-3-34.
13. Boycott KM, Rath A, Chong JX, ym. International cooperation to enable the diagnosis of all rare genetic diseases. *Am J Hum Genet* 2017;100:695–705.
14. Shendure J, Balasubramanian S, Church GM, ym. DNA sequencing at 40: past, present and future. *Nature* 2017;550:345–53.
15. Zhao X, Collins RL, Lee W-P, ym. Expectations and blind spots for structural variation detection from long-read assemblies and short-read genome sequencing technologies. *Am J Hum Genet* 2021;108:919–28.
16. Sabatella M, Mantere T, Waanders E, ym. Optical genome mapping identifies a germline retrotransposon insertion in <sc> SMARCB1 </sc> in two siblings with atypical teratoid rhabdoid tumors. *J Pathol* 2021;255:202–11.
17. Walsh T, Casadei S, Munson KM, ym. CRISPR-Cas9/long-read sequencing approach to identify cryptic mutations in BRCA1 and other tumour suppressor genes. *J Med Genet* 2021;58:850–2.
18. Fadaie Z, Neveling K, Mantere T, ym. Long-read technologies identify a hidden inverted duplication in a family with choroideremia. *HGG Adv*, julkaistu verkossa 20.7.2021. DOI: 10.1016/j.xhgg.2021.100046.
19. Mantere T, Kersten S, Hoischen A. Long-read sequencing emerging in medical genetics. *Front Genet*, julkaistu verkossa 7.5.2019. DOI: 10.3389/fgene.2019.00426.
20. Miller DE, Sulovari A, Wang T, ym. Targeted long-read sequencing identifies missing disease-causing variation. *Am J Hum Genet* 2021;108:1436–49.
21. Pauper M, Kucuk E, Wenger AM, ym. Long-read trio sequencing of individuals with unsolved intellectual disability. *Eur J Hum Genet* 2021;29:637–48.
22. An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012;491:56–65.
23. Ebbert MTW, Jensen TD, Jansen-West K, ym. Systematic analysis of dark and camouflaged genes reveals disease-relevant genes hiding in plain sight. *Genome Biol*, julkaistu verkossa 20.5.2019. DOI: 10.1186/s13059-019-1707-2.
24. Depienne C, Mandel J-L. 30 years of repeat expansion disorders: What have we learned and what are the remaining challenges? *Am J Hum Genet*. 2021;108:764–85.
25. Loomis EW, Eid JS, Peluso P, ym. Sequencing the unsequenceable: expanded CGG-repeat alleles of the fragile X gene. *Genome Res* 2013;23:121–8.

26. Deng J, Yu J, Li P, ym. Expansion of GGC repeat in GIPC1 is associated with oculopharyngodistal myopathy. *Am J Hum Genet* 2020;106:793–804.
27. Sone J, Mitsuhashi S, Fujita A, ym. Long-read sequencing identifies GGC repeat expansions in NOTCH2NLC associated with neuronal intranuclear inclusion disease. *Nat Genet* 2019;51:1215–21.
28. Mandelker D, Schmidt RJ, Ankala A, ym. Navigating highly homologous genes in a molecular diagnostic setting: a resource for clinical next-generation sequencing. *Genet Med* 2016;18:1282–9.
29. Claes KBM, Rosseel T, de Leeneer K. Dealing with pseudogenes in the next generation era. *Kirjassa: Polisen L, toim. Pseudogenes*. New York: Humana Press 202, s.363–81.
30. Borràs DM, Vossen RHAM, Liem M, ym. Detecting PKD1 variants in polycystic kidney disease patients by single-molecule long-read sequencing. *Hum Mutat* 2017;38:870–9.
31. Qiao W, Yang Y, Sebra R, ym. Long-read single molecule real-time full gene sequencing of cytochrome P450-2D6. *Hum Mutat* 2016;37:315–23.
32. Alkan C, Coe BP, Eichler EE. Genome structural variation discovery and genotyping. *Nat Rev Genet* 2011;12:363–76.
33. Quintás-Cardama A, Cortes J. Molecular biology of bcr-abl1–positive chronic myeloid leukemia. *Blood* 2009;113:1619–30.
34. Mantere T, Neveling K, Pebrel-Richard C, ym. Optical genome mapping enables constitutional chromosomal aberration detection. *Am J Hum Genet* 2021;108:1409–22.
35. Neveling K, Mantere T, Vermeulen S, ym. Next-generation cytogenetics: comprehensive assessment of 52 hematological malignancy genomes by optical genome mapping. *Am J Hum Genet* 2021;108:1423–35.
36. Beyter D, Ingimundardottir H, Oddsson A, ym. Long-read sequencing of 3,622 Icelanders provides insight into the role of structural variants in human diseases and other traits. *Nat Genet* 2021;53:779–86.
37. Miga KH, Koren S, Rhie A, ym. Telomere-to-telomere assembly of a complete human X chromosome. *Nature* 2020;585:79–84.