

Massiiviparalleelisekvensointi (NGS) mikrobiologian näkökulmasta – ja miten se liittyy Leo Tolstoihin



Eeva AUVINEN

on virologian dosentti ja laboraattori HUSLABin Virologian ja immunologian osastolla.



Petri AUVINEN

on virologian dosentti ja tutkimusjohtaja Helsingin yliopiston Biotekniikan instituutissa.

Uudet sekvensointimenetelmät tuovat uusia mahdollisuuksia mikrobiologisen diagnostiikan kehittämiseen. Bakteerien, virusten ja muiden mikrobien samanaikainen tunnistaminen tulee mahdolliseksi. Haasteina on jatkuvasti uudistuva teknologia, valtavan tietomäärän hallinta ja tulosten järjevä hyödyntäminen. Suuri potentiaali tuo mukanaan myös eettisiä kysymyksiä: miten suhtautua kliinisesti merkittäviin yllätyslöydöksiin

ja toisaalta satunnaislöydöksiin, joiden merkitystä ei (vielä) ymmärretä; miten käsitellä mahdollisesti kertyvää humaanisekvenssitietoa. Myös klinisen laboratorio-osaamisen vaatimukset laajenevat: bioinformatiikasta tulee välttämätön ja keskeinen osa diagnostiikan asiantuntemusta.

Mitä NGS tai massiiviparalleelisekvensointi tarkoittaa

Sekvensoinnissa hyödynnetään kullekin eliölle ominaista DNA- tai (joidenkin virusten) RNA-genomia. Genomien sekvensointi alkoi jo vuonna 1977, jolloin ensimmäinen kokonainen bakteriofagin genomi sekvensoitiin silloin upouudella Sangerin sekvensointitekniikalla. Myöhemmin Sanger-sekvensoinnista kehittyi automatisoinnin myötä hyvin tehokas menetelmä, jolla selvitettiin ensimmäisten bakteerien genomeja ja myös ihmisen genomi 2000-luvun alussa. Menetelmä perustuu joko pieniin kloonattuihin DNA-fragmentteihin tai PCR-monistuksella tuotettuihin fragmentteihin. Sanger-sekvensointi oli vallitseva menetelmä aina vuoteen 2005 asti, jolloin niin sanottu *next generation sequencing* (NGS) -menetelmät muuttivat tilannetta dramaattisesti.

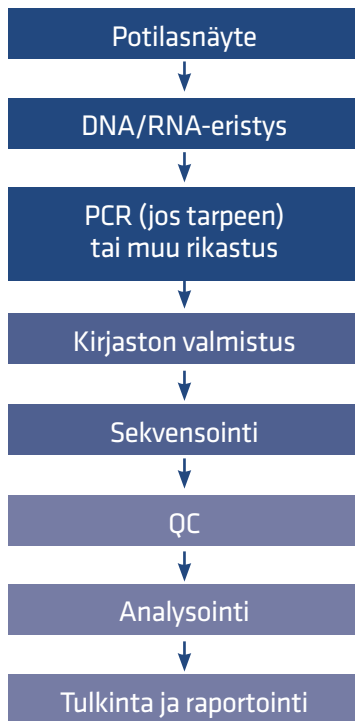
Uusien sekvensointimenetelmien suosittava yhteisnimitys on *massiiviparalleeli-*

sekvensointi. Termi kuvaa hyvin menetelmien yhteistä periaatetta: satoja miljoonia tai jopa miljardi sekvensointireaktiota voidaan tehdä yhdestä näytteestä yhdellä kertaa. Suositeltava lyhenne on kuitenkin NGS. Termi *syväsekvensointi* viittaa etupäässä sekvensoinnin syvyyteen (kattavuus, peitto, coverage), eli kuinka moneen kertaan haluttu alue sekvensoidaan. NGS pitää kuitenkin sisällään paitsi syvää, myös leveää sekvensointia, eli mahdollisimman kattavan sekvenssitiedon hankkimista yhdestä näytteestä yhdellä kertaa. Ensimmäinen kaupallinen NGS-laite oli vuonna 2005 julkaistu Rochen 454 GS20, jollainen hankittiin Helsingin Yliopiston Biotekniikan instituuttiin jo vuonna 2006. Suomen ensimmäisissä NGS-analyyseissä sekvensoitiin bakteerien genomeita.

Menetelmien keskeiset erot

NGS-menetelmissä on joitakin keskeisiä eroja Sanger-sekvensointiin verrattuna. Kuvassa NGS:n kulku mikrobiologisessa diagnostiikassa esitetään pääpiirteissään. Kliinisissä näytteissä olevia mikrobeja rikastetaan mahdollisesti viljelemällä tai PCR-monistuksella ja näytteestä eristetään nukleiinihapot. Sekvensointikirjaston valmistuksen yhteydessä haluttuja kohteita yleensä monistetaan edelleen ja sekvensoinnin vaatimia alukesek-

Jatkuu seuraavalla sivulla



Potilasnäytteen kulku NGS-analytiikassa. QC tarkoittaa sekvensointilaitteesta tulevaa raakadataa, joka on joukko saatua sekvenssiä ja emäskohtaisia laatuaroja edustavia kirjaimia ja numeroita. Datasta poistetaan väärin pituiset ja laadultaan huonot sekvenssit, ja vain testikohtaiset laadukkaat sekvenssit analysoidaan edelleen.

venssejä lisätään. Sekvensointi tapahtuu täysin automaattisesti lasin pinnalla ilman koeputkia ja käsityötä vaativia vaiheita. Hyvin ki- nuokoista näytteistä pystytään analysoidaan lähes kaikki mikrobit, jos vain budjet- ti sen sallii.

Sanger-sekvensoinnilla analysoidaan aina suuri, joskus heterogeeninen joukko molekyyliä, eikä sillä voida tunnistaa vähemmän edustettuja sekvenssivariantteja, vaan pääasiassa vallitsevaa sekvenssiä (Taulukko 1). NGS-analytiikassa taas monistetaan aluksi yksittäisiä molekyyliä, joita kutakin sekvensoidaan useaan kertaan. NGS:llä voidaan analysoida yksittäisestä näytteestä tuhansia tai miljoonia yksittäisiä molekyyliä, ja riippuen halutusta syvyydestä saadaan tarkka kuva näytteen sisältämistä sekvensseistä. Toistuva saman alueen syväsekvensointi mahdollistaa vähäistenkin varianttien löytämisen ja toisaalta tarjoaa mahdollisuuden sekvensoinnin laadunvarmistukseen. Sanger-sekvensoinnilla voidaan yleensä erottaa sekvenssimuunnokset, joiden osuus sekvensoitavasta kohteesta on vähintään 10 %, kun taas NGS:n erottelukyky on parhaimmillaan 1 % tai parempi (Taulukko 1). Jos kohteesta sekvensoidaan esimerkiksi miljoona molekyyliä, mikä NGS:ssä usein tulee kysee-

NGS JA MIKROBIOMI

Altistumme ympäristömme mikrobeille ensimmäisestä hengenvedostamme viimeiseen. Altistuksen lähteitä voivat olla vesi, ilma, pöly, maaperä, ravinto, ihmiskontaktit ja eläinkontaktit. Vain pieni osa näistä mikrobeista on patogeenisia.

Eräs NGS:n tärkeä sovellus on metagenomiikka eli mikrobiologian alalla mikrobiomin tai sen osien (bakteriomin, viromin) monimuotoisuuden selvittäminen ja karakterisointi. Mikrobiomilla tarkoitetaan jonkin anatomisen alueen tai ekologisen lokeron (niche), esimerkiksi suoliston, kaikkia mikrobeja, joihin luetaan bakteerit, arkit, virukset mukaan lukien bakteriofagit sekä sienet (joskus myös muut yksisoluiset eukaryootit). Bakteriomin tutkimiseen hyödynnetään kaikilla bakteereilla esiintyvää ja osin homologista ribosomaalisen RNA:n 16S-geeniä, joka mahdollistaa myös bakteerien tyyppityksen. Bakteriomi voi sisältää myös patogeenisia bakteereita, jotka aiheuttavat taudin vain silloin, kun koko bakteriomin tasapaino järkkyy. Bakteriomin yhteys mm. suolen terveyteen, allergioihin ja lihavuuteen tunnetaan.

Viromia ja sen merkitystä tunnetaan huonommin – ei vähiten siksi, että viruksilla ei ole vastavaa homologista geenialuetta. Elimistössä persistoi monia patogeenisia viruksia, jotka aiheuttavat taudin esimerkiksi immunologisen tilanteen heikentyessä. Mahdollisuutta käyttää viromin koostumusta esimerkiksi immunologisen tilanteen mittarina elinsiirron jälkeen on tutkittu, mutta toistaiseksi ilman läpimurtoja. On luultavaa, että terveen henkilön viromi jatkuvasti stimuloi immuunijärjestelmää, mikä nostaa elimistön immunologista valmiutta.

Terveen ja sairaan henkilön mikrobiomit ovat erilaiset, mutta vaihtelu sairaiden henkilöiden välillä on suurempi kuin terveiden. **Leo Tolstoin** romaanin kuuluisan aloituslauseen ideaa voidaan soveltaa myös tässä: Anna Karenina -periaate mikrobiologiassa kuvaa sitä, että mikrobikoostumuksessa dysbioottisten yksilöiden (joilla mikrobikoostumus on epätasapainossa eli epäterve) välillä on paljon enemmän vaihtelua kuin eubioottisten yksilöiden (joilla mikrobikoostumus on tasapainossa) välillä.

Harvinainen metagenomiikan sovellus mikrobiagnostiikassa voisi tulevaisuudessa löytyä *C. difficile* -antibioottiripulin hoitoon tehtävistä ulosteensirroista, joissa potilaan suoleen siirrettävän ulosteen laji-koostumus selvitetään terveen mikrobiomin karakterisoimiseksi ja myös patogeenien poissulkemiseksi.

seen, saadaan syvälinen kuva eri sekvenssi-varianttien esiintyvyydestä tutkitussa näytteessä monistumisen ollessa tasainen. Kyse on siis hyvin demokraattisesta lähestymistavasta.

Heti NGS-menetelmien tullessa markkinoille oli ilmeistä, että niitä voidaan soveltaa myös diagnostiikassa. Mahdollisuus tunnistaa mikä tahansa agenssi ilman esitietoa taudinaiheuttajasta avaa uusia näkymiä haastavien näytteiden diagnostiikkaan ja tuntemattomien tai muuntuneiden organismien tunnistamiseen. NGS on geneerinen lähestymistapa, joka mahdollistaa minkä tahansa DNA- tai RNA-genomin omaavan eliön tunnistamisen. NGS:n avulla selvitettiin esimerkiksi Saksan vuoden 2011 ravinnoksi käytetyistä iduista aiheutuneen *E. coli* -epidemian aiheuttaja sekä Haitin vuoden 2011 maanjäristyksen jälkeisen koleraepidemian aiheuttaja. Viimeisimmän, vuosien 2014–2016 suuren Ebola-epidemian aikana kokeil-

tiin menestyksellä pientä MinION-NGS -laitetta, jolla selvitettiin Ebola-kantojen epidemiologiaa kenttäolosuhteissa epidemian edetessä.

NGS-alustoja

Vuonna 2005 Biotekniikan instituuttiin hankittu Roche'n 454-laite on nykyään Helsingin yliopiston museossa, ja useat myöhemmätkin laitesukupolvet ovat jo poistuneet käytöstä. Mikrobiologiseen diagnostiikkaan soveltuvat Illumina Inc:n laitteistot (esim. MiSeq, NextSeq 500) ja Thermo Fisher Scientificin laitteet (mm. Ion Torrent), joista Illuminan laitteet ovat tavallisemmin käytettyjä. Sekvensointi tapahtuu pienen pienessä, yhdestä nukleinihappomolekyylistä monistetussa klustereissa lasin pinnalla. Näillä laitteilla saadaan kymmenistä miljoonista satoihin miljooniin sekvenssijaksoa, eli readiä, analyysiä kohden, mutta saatujen

Sekvenssi	Osuus alkuperäisessä näytteessä	Osuus Sanger-readeistä	Osuus NGS-readeistä
-----AA-----	80 %	Pääosa	80 %
-----AT-----	12 %	Pieni signaali	12 %
-----GC-----	4 %	-	noin 4 %
-----TT-----	2 %	-	noin 2 %
-----CG-----	1 %	-	noin 1 %
-----TA-----	1 %	-	noin 1 %

Taulukko 1. Sekvenssivarianttien erottelukyky Sanger- ja NGS-sekvensoinnilla. Saatuja sekvenssiosuuksia kuvaavat luvut ovat viitteellisiä ja vaihtelevat lähestymistavasta riippuen.

sekvenssien pituudet ovat varsin lyhyitä, noin 150–300 emästä. Luotettavaan diagnostiikkaan vaaditaan tavallisesti suuri määrä sekvenssejä.

Edellä mainituilla laitteilla voidaan tehdä kohdennettua sekvenssointia, kuten 16S rRNA-geeniin perustuvaa bakteeritunnistusta tai virusgenomin tietyn osan sekvenssointia genotyyppitystä varten. Laitteet sopivat myös metagenomiikan shotgun-sovelluksiin, joissa tarvitaan suuria määriä sekvenssejä. Uudemmissa PacBio RSII -laitteella (Pacific Biosciences of California Inc.) ja MinION-laitteella (Oxford Nanopore Technologies) analysoidaan yksittäisiä DNA-molekyylejä ilman klusterointia. Saavutetut lukupituudet ovat kymmenistä tuhansista jopa satoihin tuhansiin emäksiin. Klustereita käyttävissä tekniikoissa emäskohtainen virheen todennäköisyys on $\leq 1\%$, kun taas yksittäisiä molekyylejä sekvensoivissa tekniikoissa raa-kadatan virheprosentti on 13–15 %. Kun esimerkiksi PacBio-laitteella samaa sekvenssijaksoa sekvensoidaan tuhansia kertoja, saatavan konsensussekvenssin tarkkuus on erittäin korkea. Laskennallisesti noin 20 lukukerran jälkeen PacBion emäskohtainen tarkkuus on parempi kuin Illuminan.

Kaikilla NGS-laitteilla on omat käyttöalueensa. On tärkeää valita NGS-menetelmä diagnostisen kysymyksenasettelun ja halutun datan luonteen mukaan. NGS-laitteiden kehitysvauhti on ollut poikkeuksellinen. Viimeisten kymmenen vuoden aikana DNA-sekvensoinnin nopeus on kasvanut enemmän kuin tietokoneiden laskentakapasiteetti on lisääntynyt, mutta kustannukset ovat laskeneet. Uusien NGS-laitteversioiden ja uusien laitteiden julkistamistahti vie alaa eteenpäin, mutta haittapuolena on menetelmien standardisoinnin vaikeus diagnostiikka varten.

Sekvensoinnin lähestymistapoja

Diagnostisissa sekvensoinneissa on kaksi pääasiallista lähestymistapaa. Sekvenssointia voidaan tehdä kohdennetusti, jos tiedetään mitä halutaan etsiä ja sekvenssitietoa on

käytettävissä. Tällöin kohdetta tavallisesti rikastetaan ensin PCR-monistuksella spesifeillä alukkeilla tai viljelemällä. Kohdennetussa sekvensoinnissa sekvenssidata on vähemmän kompleksista ja haluttu sekvenssi on rikastuneena. Sekvenssidatan bioinformaattinen käsittely on helpompaa ja automaattisia tulkinta-algoritmeja on mahdollista kehittää. Esimerkkejä kohdennetusta sekvensoinnista on bakteerien osoittaminen ja tyypittäminen 16S rRNA-geenin avulla tai antibiootiresistenssigeenin määrittäminen.

Virologiassa kohdennettua sekvenssointia käytetään virusten tyyppitykseen, lääkeaine-resistenssien määrittämiseen tai resistenttien kantojen varhaiseen toteamiseen nopeasti muuntuvilla viruksilla. Kohdennettu sekvenssointi voi olla myös hyvin laajaa. Koko viromin analyysissä Briese ym. syntetisoivat kaksi miljoonaa oligonukleotidikoetta, jotka kattoivat kaikki tunnetut selkärankaisten virukset yhteensä 207 taksonomisesta ryhmästä (1). Koettimilla pyydystettiin eristetyistä näytteistä virusten nukleiinihapot, ja ne monistettiin ja sekvensoitiin. Tällä lähestymistavalla voitiin löytää uusia viruksia erittäin kompleksisista näytemateriaaleista ja herkistää tavanomaisen NGS:n herkkyyttä vielä 100–10 000 -kertaiseksi.

Shotgun-sekvenssointi tulee kyseeseen silloin, kun sekvensoinnin kohde ei ole tiedossa, kun halutaan sekvensoida mahdollisimman kattavasti, tai kun sekvenssitietoa ei ole saatavilla. Shotgun-sekvenssointia voidaan käyttää mikrobipopulaatioiden karakterisointiin, uusien patogeeneiden hakemiseen tai etiologisen tekijän tunnistamiseen silloin, kun sitä ei tavanomaisilla menetelmillä löydetä. Harvinaisten vaikeiden tautitapausten etiologian selvittelyssä shotgun-sekvenssointi voi tulla kyseeseen, esimerkiksi henkeä uhkaavissa enkefalopatioissa, joissa taudinaiheuttajaa ei löydetä tavanomaisin diagnostisin menetelmin. Shotgun-sekvensoinnin erityisiä haasteita on sivutuotteena kertyvän, kysymyksenasettelun kannalta tarpeettoman sekvenssitiedon suuri määrä. Näytteestä ja lähestymistavasta riippuen jopa 99 % shotgun-sekvensoinnilla saatavasta sekvenssidata-

tasta voi edustaa humanisekvenssejä. Spesifin sekvenssitiedon suodattaminen ja analysointi vaatii suurta tietokonekapasiteettia ja laajaa bioinformaattista osaamista, ja on vaikeasti automatisoitavissa.

NGS:n sovellusalueita mikrobiologisessa diagnostiikassa

On selvää, että valtaosa uusista mikrobiologisen diagnostiikan sovelluksista keskittyy nukleiinihappojen osoittamiseen. Tämän hetken kehitystyössä voidaan erottaa konventionaalisten PCR-menetelmien ohella kaksi megatrendiä: toisaalta mahdollisimman nopeat ja varsin suppeat vier- ja muut pikatestit, ja toisaalta haluttaessa hyvin laajan mittakaavan NGS, joka vielä paljolti hakee käytännön diagnostisia sovelluksia. Tietojemme mukaan Suomen kliinisissä laboratorioissa ei toistaiseksi käytetä NGS-menetelmiä mikrobiologisessa potilasdiagnoosiikassa. Taulukossa 2 esitetään niitä NGS-sovelluksia, joita on käytössä joissakin Euroopan maissa, joita suunnitellaan käyttöön, tai jotka teknisesti olisivat lähitulevaisuudessa toteutettavissa diagnostiikassa.

Ensimmäisinä käyttöön ovat tulleet tai tulossa sellaiset diagnostiset sovellukset, joissa jo käytetään Sanger-sekvenssointia. 16S rRNA-geeniin perustuva bakteerien tyyppitys tehdään jo NGS:llä useissa laboratorioissa Pohjoismaissa. Kokogenomisekvenssointia käytetään tai kokeillaan uusien MRSA-kantojen, invasiivisten meningokokkien ja gonokokkien tyyppitykseen sekä antibiootiresistenssimäärityksiin. Bakteeripatogeeneja ja useita resistenssigeenejä on onnistuttu sekvensoimaan virtsanäytteistä jopa ilman bakteerien rikastamista viljelyllä (2). NGS:ää käytetään myös silloin, kun halutaan varmistaa harvinainen bakteerilöydös, tai näyte on jäänyt bakteeriviljelyssä negatiiviseksi, mutta on vahva epäily bakteerietiologiasta. Bakteerien patogeeneiden determinanttien, kuten virulenssi- ja toksinigeenien osoittaminen on yksi NGS:n sovellus.

Harvinainen metagenomiikan sovellus mikrobiologiassa voisi tulevaisuudessa löytyä *C. difficile* -antibiootiripulin hoitoon tehtävistä ulosteensiirroista, joissa potilaan suoleen siirrettävän ulosteen lajikoostumus selvitetään terveen mikrobiomin karakterisoinniksi ja myös patogeeneiden poissulkemiseksi.

Sienipatogeeneiden tunnistamisessa Sanger-sekvensoinnin avulla hyödynnetään homologisia 18S/28S sRNA-geenejä. NGS-sekvensoinnissa käytetään lyhyempiä ja tarkkan tunnistamisen mahdollistavia ITS (internal transcribed spacer) -alueita. NGS:n käytöstä sienidiagnostiikassa on vielä varsin vähän kokemusta.

Jatkuu seuraavalla sivulla

Sovellus	Sekvensoinnin kohde	Käyttöönoton vaihe
Bakteerityypitys	16S rRNA -geeni	Käytössä
Bakteerien antibioottiresistenssien tunnistaminen	Kokogenomisekvensointi	Käytössä
Uusien MRSA-kantojen tyyppitys	Kokogenomisekvensointi	Käytössä
Invasiivisten meningokokkien tyyppitys	Kokogenomisekvensointi	Käytössä
Gonokokkien tyyppitys	Kokogenomisekvensointi	Tulossa käyttöön
Viljelynegatiivisten bakteerinäytteiden jatkok tutkimus	16S / Kokogenomisekvensointi	Kokeiltavana
Harvinaisten bakteerilöydösten varmistus	16S / Kokogenomisekvensointi	Kokeiltavana
Hepatiitti B -viruksen genotyyppitys ja lääkeaineresistenssien määrittäminen	Polymeraasi, HBsAg	Kokeiltavana
Hepatiitti C -viruksen genotyyppitys ja lääkeaineresistenssien määrittäminen	NS5B, 5'UTR, core; NS5A, NS5B, NS3	Kokeiltavana / Käytössä
HIV-tyypitys ja lääkeaineresistenssien määrittäminen	Integraasi, RT	Kokeiltavana

Taulukko 1. Sekvenssivarianttien erottelukyky Sanger- ja NGS-sekvensoinnilla. Saatuja sekvenssiosuusia kuvaavat luvut ovat viitteellisiä ja vaihtelevat lähestymistavasta riippuen.

Tavallisimpia jo käytössä tai kokeiltavina olevia virologisia sovelluksia on hepatiitti B- ja C-virusten ja HIV:n tyyppitys sekä lääkeaineresistenssien määrittäminen. NGS:n avulla voidaan löytää harvinaisia mutatoituja HIV-kantoja, jotka ovat kliinisesti merkittäviä mutta niin vähäisiä, että Sanger-sekvensoinnilla ne jäisivät havaitsematta. Myös herpes simplex (HSV)- ja muiden herpesvirusten resistenssimäärityksissä ollaan siirtymässä fenotyyppisistä määrityksistä NGS:llä tehtäviin genotyyppisiin määrityksiin.

Tärkeä virologinen NGS-sovellusalue voi löytyä enkefalopatioista, joissa epäillään virusetiologiaa, mutta ei voida tavanomaisilla menetelmillä osoittaa mitään virusta. **Naccache** ym. ovat julkaisseet tapauksen, jossa likvorista ja muista kliinisistä näytteistä etsittiin kymmeniä patogeeneja spesifeillä PCR-testeillä, mutta mitään tautia selittävää tekijää ei löydetty. Lopulta likvornäytteestä tehtiin shotgun-sekvensointi monistamalla näytettä aluksi random heksameereilla. Saadusta sekvenssijakoista eli readeista 0,0012 % edusti astrovirusvarianttia, jota ei voitu osoittaa suolistoinfektioita varten tarkoitetulla astrovirus-PCR-testillä. Diagnoosiksi varmistui neuroinvasiivinen astrovirusinfektio, johon potilas menehtyi joitakin kuukausia NGS:llä saadun diagnoosin jälkeen (3).

Mikrobin patogeeniset kannat tai resistentit muodot saattavat edustaa hyvin pientä osaa näytteessä olevista mikrobeista, mutta ne voivat siitä huolimatta olla kliinisesti merkittäviä. Jos muuntuneen mikrobin sekvenssiä

ei tunneta, NGS on ainoa menetelmä muuntuneiden mikrobikantojen osoittamiseen.

Epidemiologia ja muita sovelluksia

NGS soveltuu epidemiaseurantaan, taudinaiheuttajan tai variantin nopeaan tunnistamiseen puhkeavissa epidemioissa, epidemioiden ennakoitiin ja tartunnan lähteen selvittelyyn. NGS:ää on käytetty esimerkiksi muuntuneiden influenssa A -kantojen varhaiseen osoittamiseen (4). Jo aiemmin mainitun Ebola-epidemian aikana viruskantojen muuntumista seurattiin NGS:n avulla lähes reaaliaikaisesti. Ebola-epidemian jälkeen ryhdyttiin mm. **Bill** ja **Melinda Gatesin** säätiön rahoituksella maailmanlaajuiseen epidemioiden ennaltaehkäisyyn. Reservoareiksi tunnetuista eläinlajeista etsitään NGS:llä vaarallisia bakteereita ja viruksia. Löydösten perusteella ennakoidaan diagnostiikan, rokotetuotannon ja rokotekehityksen tarpeita sekä mm. alueellisia eristystarpeita. Ääriesimerkki on bioterrorismi (pernarutto) ja biologiset aseet, joiden diagnostiikkaan ja karakterisointiin NGS on omiaan.

Omia sovelluksia

Olemme käyttäneet NGS-menetelmiä bakteeriyhteisöjen karakterisointiin useissa sairauksissa. Osoitimme, että ihon mikrobiomin

suurempi diversiteetti suojaa allergialta (5). Toinen äskettäin julkaistu tutkimuksemme osoitti, että eräät aknen hoidossa tehokkaat lääkkeet muuttivat ihon mikrobiomin koostumusta (6). Parkinsonin tautia sairastavien ja terveiden henkilöiden välillä osoitimme eroja suoliston ja suun mikrobiomeissa (7, 8).

Virologiassa olemme käyttäneet PacBio-sekvensointia JC-polyoomaviruspopulaatioiden karakterisointiin neurologisten potilaiden likvornäytteissä ja munuaisensiirtopotilaiden virtsoissa (9, 10). Onnistuimme sekvensoimaan kokonaisia virusgenomeita ja osoittamaan neurologisten PML-potilaiden likvorista useita erilaisia ns. neurorooppisia JC-viruskantoja. Ainoastaan PacBio-menetelmällä päästään karakterisointiin kokonaisten viruskantojen ominaisuuksia, kun sekvenssiä ei koota lyhyistä fragmenteista. Kaikki omista tutkimuksistamme tuotettu sekvenssidata on lähetetty julkisiin tietokantoihin.

Kaupallisia NGS-analytiikassa käytettäviä testejä

Joitakin kaupallisia reagensseja ja tietokoneohjelmistoja on saatavilla mikrobiologiseen NGS-sekvensointiin ja data-analyysiin. Ensimmäisiä kaupallisia virologisia NGS-sovelluksia ovat hepatiitti B- ja C-virusten genotyyppitys ja lääkeaineresistenssimääritykset. Myös HIV-genotyyppitykseen ja tartunnan jäljitykseen sekä HIV:n ja HSV:n lääkeaineresistenssimäärityksiin on kaupallisia reagensseja. Monissa kaupallisissa testeissä tarjotaan PCR- ja sekvensointialukkeet sekä analyysiohjelmisto sekvenssitetiedon tulkitsemiseen. Itse sekvensointi tapahtuu yleisillä NGS-alustoilla. Analyysiohjelmit perustuvat tulosten vertailuun olemassa olevien tietokantojen kanssa.

Hyötyjä ja haasteita

NGS on herkkä menetelmä, jolla voidaan osoittaa pieniä määriä kohdetta hyvin

Vähäisetkin mikrobisekvenssit, kuten lääkeaineresistentit kannat, saattavat olla kliinisesti merkittäviä.

kompleksisesta näytteestä. NGS:n avulla voidaan löytää sellaisia uusia mikrobeja, joita ei voida viljellä laboratoriossa. Se on myös spesifi menetelmä, mutta lyhyitä nukleinihappofragmentteja sekvensoitaessa on vääran tulkinna vaara, jos läheisiltä sukulaismikrobeilta löydytty homologinen sekvenssi. NGS-menetelmän etuja ovat myös geneerisyys ja mahdollisuus löytää sekä tunnettuja että tuntemattomia kohteita. Käyttäjän haasteena on tunnistaa ne vaikeat diagnos-

tiset ongelmat, joissa NGS:ää voidaan hyödyntää.

Kliinisessä käytössä pitää ratkaista, mikä on lääketieteellisesti merkitsevä löydös, mikä taas merkityksetön, ja mikä voi tulla merkitseväksi myöhemmin. Vähäisetkin mikrobisekvenssit, kuten lääkeaine-resistentit kannat, saattavat olla kliinisesti merkittäviä. Karttuvan tutkimustiedon avulla merkityksetön tieto voi osoittautua merkitykselliseksi myöhemmin, esimerkiksi uuden tautiassosiaation tai uuden hoidon ansiosta, jolloin kaiken sekvenssintiedon tallentamiselle on perusteita. Mikrobin sekvenssoinnissa kertyvän humaanisekvenssitiedon poistaminen paitsi tulkinnallisista, myös eettisistä syistä lienee välttämätöntä. Humaanisekvenssien poistamisen menetelmistä ja laatuksiteereistä tulisi olla yhteen sopimus.

Massiiviparalleelisekvenssoinnin etu ja haaste on sen kaikkien osa-alueiden nopea kehitys: näytteenkäsittely, sekvensoitavien kirjastojen valmistus, sekvensointi ja erityisesti tulosten analysointi bioinformatiikan ja tilastotieteen keinoin kehittyvät valtavaa vauhtia. Pulonkaulana on tulosten tulkinta ja sen hitaus silloin, kun valmiita tulkinta-algoritmeja ei ole olemassa. Diagnostisen laboratorion kannalta uusi vaatimus on bioinformatiikan osaaminen, jota Suomessa on aivan liian vähän. Osaamisen kehittäminen on aikaavievää ja vaatii matemaattista ymmärrystä – ja mikrobiologisista sekvenssoinneista puhuttaessa mikrobiologian alan taustakoulutusta. Uusia diagnostisia menetelmiä käyttöön otettaessa ensimmäinen ilmeinen kysymys on hinta. NGS on kalliimpaa kuin Sanger-sekvenssointi, mutta hinnat ovat laskusuunnassa.

Laadunarvioinnin näkökulmia

Laadunarvioinnin ja laboratoriodien välisen tulosvertailun kannalta NGS:n jatkuvasi uudistuvat laboratoriomenetelmät, sekvenssointialustat, tulkinta-algoritmit ja muut tekniset ratkaisut ovat äärimmäisen haastavia. Menetelmien välillä tulee aina olemaan suuria eroja, jotka kasvavat sitä mukaan, mitä yksityiskohtaisempaa tietoa tarkastellaan. Hyvin kuratoituja tietokantoja mikrobilöydösten vertailuun on vasta vähän saatavilla. Jokaisen sekvenssointituloksen tulkinta perustuu julkisiin tietokantoihin, jotka karttavat ja paranevat ainoastaan sinne lähetettävien sekvenssien avulla. Siksi on äärimmäisen tärkeää, että kaikki mikrobisekvenssit lähetetään julkisiin tietokantoihin kaikkien käytettäväksi.

Diagnostisessa käytössä tärkeintä on pystyä erottamaan lääketieteellisesti merkitsevä löydös ja tulkitsemaan sen merkitys. Lopullisen löydöksen lisäksi NGS-prosessin laatua voidaan tarkastella osina: näyttekäsittelyä, halutun kohteen tai sekvenssin rikastamista, sekvenssointia ja bioinformaattista datan käsittelyä kutakin erikseen. Esimerkiksi sekvenssoinnin tuottamaa dataa voidaan käyttää ulkoisen laadunarvioinnin ”näytteenä”, joka laboratorion tulee kyetä analysoimaan.

Eettiset kysymykset

Sekvenssointitiedon kertyminen tuo mukanaan eettisiä kysymyksiä mikrobiologiassa samaan tapaan kuin genetiikassa. Sekvenssitiedon joukossa voi olla satunnaislöydöksiä, jotka eivät liity potilaan sairauteen ja joita ei ole osattu epäillä: esimerkiksi haetussa olostepatogeeneja metagenomiikan lähestymistavoilla on mahdollista löytää HIV-sekvenssejä potilaalta, jota ei ole tiedetty HIV-positiiviseksi. Satunnainenkin löydös voi olla potilaan terveyden kannalta merkityksellinen.

On tärkeää sopia etukäteen menettelytavoista, jotta satunnaislöydökset eivät vaikeuta tulosten tulkintaa liialti. Bioinformatiikan keinoin epätoivottua sekvenssidataa voidaan suodattaa pois ennen raportin laatimista. Satunnaislöydökset eivät ole yksinomaan NGS:n sivutuote. Esimerkiksi MRI- ja CT-kuvauksissa voidaan todeta syöpä muussa tarkoituksessa tehdyn tutkimuksen sivutuotteena. Genetiikassa ihmisen eksomitai kokogenomisekvenssoinnin yhteydessä voidaan löytää lukuisia loss-of-function -mutaatioita, joista osa voi liittyä tunnetuihin perinnöllisiin tauteihin. Sekä potilaan että hoitavan lääkärin tulee olla tietoinen tällaisesta mahdollisuudesta.

Mikrobisekvenssitiedon tallentamista voidaan perustella sillä, että tällä hetkellä merkityksetön sekvenssitieto saattaa tulla tärkeäksi, jos löydetään uusia tautiassosiaatioita tai kehitetään uusia lääkkeitä tai rokotteita. Erityisesti shotgun-sekvenssoinnissa kertyvän humaanisekvenssitiedon osalta tilanne on toinen. Kysymys humaanitiedon käsitteistä ja tallentamisesta silloin, kun se ei ole ollut tutkimuksen kohteena, edellyttää laboratoriodiagnostiikan, lääketieteen, muiden terveydenalan toimijoiden ja etikan asiantuntijoiden välistä keskustelua.

Lopuksi

Massiiviparalleelisekvenssointi mahdollistaa laajan, oireisiin perustuvan laboratorio-

diagnostiikan jopa yksittäisistä näytteistä. Tärkeitä diagnostisia sovellusalueita ovat esimerkiksi lääkeaineresistenssien toteaminen ja vaikeiden, harvinaisten tautitausten diagnostiikka. Metagenomiikan lähestymistavat helpottavat uusien tautiassosiaatioiden ja kokonaan uusien patogeenien löytämistä. Mikrobiomin merkityksen selvittämiseen NGS on ainutlaatuinen työkalu. Löydöksiä voidaan hyödyntää lääkkeiden, rokotteiden ja mikrobien kasvatusmenetelmien kehittämisessä. NGS:llä saadaan valtavasti tietoa, jonka tulkinta on haasteellista, mutta joka tuo mukanaan myös uusia diagnostisia mahdollisuuksia. ■

KIRJALLISUUSVIITTEET

1. Briese T, Kapoor A, Mishra N, Jain K, Kumar A, Jabado OJ, Lipkin WI. Virome capture sequencing enables sensitive viral diagnosis and comprehensive virome analysis. *MBio*. 2015 Sep 22;6(5):e01491-15. doi: 10.1128/mBio.01491-15.
2. Schmidt K, Mwaigwisya S, Crossman LC, Doumith M, Munroe D, Pires C, Khan AM, Woodford N, Saunders NJ, Wain J, O'Grady J, Livermore DM. Identification of bacterial pathogens and antimicrobial resistance directly from clinical urines by nanopore-based metagenomic sequencing. *J Antimicrob Chemother* 72:104-14, 2017.
3. Naccache SN, Peggs KS, Mattes FM, Phadke R, Garson JA, Grant P, Samayoa E, Federman S, Miller S, Lunn MP, Gant V, Chiu CY. Diagnosis of neuroinvasive astrovirus infection in an immunocompromised adult with encephalitis by unbiased next-generation sequencing. *Clin Infect Dis* 60:919-23, 2015.
4. Zhao J, Liu J, Vemula SV, Lin C, Tan J, Ragupathy V, Wang X, Mbondji-Wonje C, Ye Z, Landry ML, Hewlett I. Sensitive detection and simultaneous discrimination of influenza A and B viruses in nasopharyngeal swabs in a single assay using next-generation sequencing-based diagnostics. *PLoS One*. 2016 Sep 22;11(9):e0163175. doi: 10.1371/journal.pone.0163175. eCollection 2016.
5. Hanski I, von Hertzen L, Fyhrquist N, Koskinen K, Torppa K, Laatikainen T, Karisola P, Auvinen P, Paulin L, Mäkelä MJ, Vartiainen E, Kosunen TU, Alenius H, Haahtela T. Environmental biodiversity, human microbiota and allergy are interrelated. *Proc Natl Acad Sci USA* 109:8334-9, 2012.
6. Kelhämä HL, Aho V, Fyhrquist N, Pereira P, Kubin M, Paulin L, Palatsi R, Auvinen P, Tasanen K, Lauerma A. 2017. Isotretinoin and lymecycline treatments modify the skin microbiota in acne. *Exp Dermatol*. 2017 Jun 21. doi: 10.1111/exd.13397. [Epub ahead of print]
7. Scheperjans F, Aho V, Pereira PAB, Koskinen K, Paulin L, Pekkonen E, Haapaniemi E, Kaakkola S, Eerola-Rautio J, Pohja M, Kinnunen E, Murros K, Auvinen P. Microbiota are related to Parkinson's disease and clinical phenotype. *Mov Disord* 30:350-8, 2015.
8. Pereira PAB, Aho VTE, Paulin L, Pekkonen E, Auvinen P, Scheperjans F. Oral and nasal microbiota in Parkinson's disease. *Parkinsonism Relat Disord* 38:61-7, 2017.
9. Seppälä H, Virtanen E, Saarela M, Laine P, Paulin L, Mannonen L, Auvinen P, Auvinen E. Single-molecule sequencing revealing the presence of distinct JC polyomavirus populations in patients with progressive multifocal leukoencephalopathy. *J Infect Dis* 215:889-95, 2017.
10. Seppälä HM, Helanterä IT, Laine PK, Lautenschlager IT, Paulin LG, Jahnuksainen TJ, Auvinen PO, Auvinen E. Archetype JC polyomavirus prevails in a rare case of JC polyomavirus nephropathy and in stable renal transplant recipients with JC polyomavirus viraemia. *J Infect Dis*. 2017 Aug 22. doi: 10.1093/infdis/jix435.