

Uuden sukupolven sekvensointimenetelmien kehitys  
ja hyödyntäminen perinnöllisen rintasyöpäalttiuden  
tutkimuksessa

Ida Korpi

LuK-tutkielma  
Biologian tutkinto-ohjelma,  
Genetiikka  
Oulun yliopisto  
Maaliskuu 2019

# Sisällys

1. Johdanto .....	2
2. Sekvensointimenetelmien kehitys .....	3
2.1 Ensimmäisen sukupolven sekvensointimenetelmät .....	3
2.2 Toisen eli uuden sukupolven sekvensointimenetelmät .....	5
3. Syövän geneettiset taustatekijät .....	10
4. Rintasyövän molekulaariset alatyypit .....	11
4.1. Estrogeenin ja progesteronin reseptorit, HER2 ja Ki67 .....	12
5. Familiaalisen rintasyövän geneettinen tausta .....	13
6. Tutkimusmenetelmät .....	15
6.1. Koko genomin sekvensointi (Whole Genome Sequencing, WGS).....	15
6.2. Eksomisekvensointi.....	17
6.3. Geenipaneelit.....	18
6.4. RNA-sekvensointi .....	18
6.6. Uuden sukupolven sekvensointimenetelmien rajoitteita ja tulevaisuuden näkymää.....	19

## 1. Johdanto

Sekvensoinnilla tarkoitetaan DNA:n emäsjärjestyksen selvittämistä (Heather & Chain, 2016). DNA:n neljä emästä ovat adeniini, guaniini, tymiini ja sytosiini (Sinden, 1994, s. 3-14). RNA-molekyyleissä tymiinin korvaa urasiili. Adeniini ja guaniini ovat kahdesta rengasrakenteesta koostuvia puriineja ja tymiini, sytosiini sekä urasiili puolestaan yhdestä rengasrakenteesta koostuvia pyrimidiineja. Emäs on sitoutuneena sokeriosaan, joka on sitoutuneena fosfaattiosaan. Nukleiinihappomolekyylillä on siis sokerifosfaattirunko, johon emäkset ovat sitoutuneet sokeriosan kautta. DNA-juosteet muodostavat kaksoisjuosterakenteita emästen välisin vetysidoksin. Adeniini muodostaa kaksi vetysidosta tymiinin kanssa ja guaniini kolme vetysidosta sytosiinin kanssa. Näin muodostuvia emäspareja on ihmisen genomissa kolme miljardia (Sinden, 1994, s. 37).

Erilaiset syövät aiheuttivat arviolta 9.6 miljoonaa kuolemaa vuonna 2018 (who.int, noudettu 11.10.2019). Tästä rintasyövän osuus on noin 627 000 kuolemaa. Vuosittain tehdään noin 1.7 miljoonaa rintasyöpädiagnoosia (Tervasmäki, 2018, s. 25). Suomessa vuotuisella tasolla tehdään diagnooseja noin 5 100, ja rintasyöpä aiheuttaa noin 840 kuolemaa vuosittain. Rintasyöpä on naisten yleisin syöpä maailmanlaajuisesti ja aiheuttaa suurimman osan, noin 14 prosenttia, naisten syövästä johtuvista kuolemista vuosittain (Kiiski ym., 2014). Noin joka kahdeksas nainen sairastuu rintasyöpään elämänsä aikana (Tervasmäki, 2018, s. 19). Suomessa diagnosoitiin 30 miestä vuonna 2015 (Tervasmäki, 2018, s. 25). Noin 15 prosentilla rintasyöpätapauksia on familiaalinen tausta, eli rintasyöpä on kehittynyt perinnöllisestä alttiudesta sairastua (Kiiski ym., 2014).

Rintasyöpä on genetiikaltaan hyvin monimuotoinen sairaus, ja alle puolet familiaalisista rintasyöpätapauksista selittyy tunnetuilla riskialleleilla (alleelit = geenin eri muodot) (Desmedt ym., 2012; Lee ym., 2016). Tutkittaessa geneettisesti monimuotoisia sairauksia on ensiarvoisen tärkeää, että käytössä on menetelmiä, joilla voidaan tutkia useiden geenien sekvenssejä, ilmenemistä eli ekspressiota, ja vaikutuksia tehokkaasti sekä ajankäytön että kulujen näkökulmasta (Desmedt ym., 2012). Uuden sukupolven sekvensointimenetelmien kehityksen myötä tällaisia menetelmiä on saatu käyttöön.

## 2. Sekvensointimenetelmien kehitys

### 2.1 Ensimmäisen sukupolven sekvensointimenetelmät

Sekvensointimenetelmien kehitys alkoi yksijuosteisten RNA-molekyylien sekvenssien selvittämisestä RNA-nukleaasi entsyymien avulla (Heather & Chain, 2016). Sekvenssit olivat mikrobista ribosomi- ja siirtäjä-RNA:ta sekä RNA-bakteriofagien eli bakteereja infektoivien RNA-virusten genomeja. Ensimmäisen nukleinihapposekvenssin saivat selville Robert Holley ja hänen kollegansa vuonna 1965. Kyseinen sekvenssi oli alaniiniaminohapon siirtäjä-RNA:n sekvenssi *Saccharomyces cerevisiae* -lajista (leivinihiiva). Samanaikaisesti Frederick Sanger kollegoineen kehitti nukleinihappojuosteen synteesiin ja nukleotidien radioaktiiviseen leimaukseen perustuvaa menetelmää, jonka avulla saatiin selville useita ribosomi- ja siirtäjä-RNA-molekyylien sekvenssejä. Nukleotidien leimaukseen käytettiin fosforia tai tritiumia. Tätä menetelmää hyödyntäen vuonna 1972 Walter Fiers kollegoineen selvitti ensimmäisen geenisekvenssin, joka oli RNA-bakteriofagi MS2:n erään kapsidiproteiinin sekvenssi. Näihin aikoihin radioaktiiviseen leimaukseen perustuvia menetelmiä alettiin soveltaa DNA-molekyylien emäsjärjestyksen selvittämiseen. Ray Wu ja Dale Kaiser liittivät DNA-polymeraasin avulla bakteriofagi lambda:n DNA:n kaksoisjuosteen yksijuosteiseen 5'-päähän radioaktiivisia nukleotideja, mistä johdettiin spesifisten oligonukleotidien käyttö alukkeina DNA-polymeraasille. Tämän ansiosta voitiin saada selville DNA-sekvenssejä muualtakin kuin juosteiden päistä. Menetelmät olivat kuitenkin vielä hitaita ja luennat lyhyitä.

1970-luvun puolivälin tärkeä edistysaskel oli polyakryyliamidigeelin (PAGE) käyttöönotto, mikä helpotti emästen tunnistamista huomattavasti. PAGE:a hyödynnettiin kahdessa edistyksellisessä mutta monimutkaisessa menetelmässä, Frederick Sangerin ja Alan Coulsonin 'plus ja miinus' -menetelmässä sekä Allan Maxamin ja Walter Gilbertin kemiallisten sidosten hajottamiseen perustuvassa menetelmässä. 'Plus ja miinus' -menetelmässä eristetty ja yksijuosteiseksi denaturoitu DNA jaetaan kahdeksaan putkeen, joista jokaisessa DNA-polymeraasi katalysoi kaksi reaktiota: plusreaktion ja miinusreaktion. Plusreaktiossa neljään putkeen lisätään vain yhtä emästä vastaavaa nukleotidia. Miinusreaktiossa toisiin neljään putkeen lisätään kaikkia kolmea muuta emästä vastaavaa nukleotidia. Miinusreaktioiden putkissa

juosteen synteesi päättyy aina silloin, kun seuraava lisättävä emäs olisi plusreaktiossa lisätty emäs. Saadut fragmentit ajetaan PAGE:lla, minkä jälkeen geeli kuvataan ja sekvenssi luetaan. Tällä menetelmällä Sanger, Coulson ja heidän kollegansa selvittivät ensimmäisen DNA-genomin, bakteriofagi PfiX:n genomin. Maxam ja Gilbert eivät hyödyntäneet menetelmässään DNA-polymeraasia. He käsittelivät DNA:ta kemikaaleilla, jotka katkaisivat DNA-juosteen tietyn emäksen/emästen kohdalta. Pyrimidiineille oli omat putkensa ja puriineille vastinemäksen kanssa yhteiset putket. Reaktioiden jälkeen fragmentit ajettiin PAGE:lla. Menetelmä oli ensimmäinen laajemmin käyttöön otettu DNA:n sekvensointimenetelmä ja sen käyttöönottoa voidaankin pitää ensimmäisen sukupolven sekvensoinnin syntyinä.

Vuonna 1977 Sanger kehitti menetelmän, joka mullisti DNA-sekvensointi teknologian. Menetelmä perustuu radioaktiivisesti leimattuihin dideoksinukleotideihin eli ddNTP-molekyyleihin. Molekyylistä puuttuu 3'-hydroksyyli-ryhmä, joka vaaditaan uuden fosfodiesterisidoksen muodostamiseen DNA-juosteen jatkumiseksi. Kullekin ddNTP:lle on oma putkensa. Lisäksi jokaisessa putkessa on emästen dNTP-nukleotideja. Fragmentit ajetaan geelielektroforeesilla. Menetelmän tarkkuuden ja helppokäyttöisyyden ansiosta siitä tuli yleisimmin käytetty DNA:n sekvensointimenetelmä. Seuraavina vuosina menetelmään tehtiin monia parannuksia, joista merkittävimpänä radioaktiivisten nukleotidien korvaaminen fluoresoivilla nukleotideilla, minkä ansiosta voitiin alkaa käyttää yhtä putkea aiemman neljän putken sijaan. Edistysaskelten myötä kehitettiin ensimmäiset kaupalliset sekvensointilaitteet. Sangerin menetelmä olikin eniten käytetty menetelmä noin 30 vuotta eteenpäin sen ensiaskelista alkaen ja nykyisin sillä voidaan tuottaa noin 500 – 1000 emäsparia pitkiä luentoja alhaisilla virhelukemilla (Besser ym., 2018). Virhelukema on suunnilleen luokkaa 0.1-0.01 prosenttia ja sekvensointi maksaa noin 450 euroa miljoonaa emäsparia kohti, ja tällä hinnalla kokonaisen genomin sekvensointi maksaisi noin 1.4 miljoonaa euroa (*CADTH Rapid Response Reports*, 2014). Yhdellä laitteella voidaan tehdä 96 rinnakkaista reaktiota (Rizzo & Buck, 2012). Menetelmän suorituskyky on noin 115 kiloemästä päivässä.

## 2.2 Toisen eli uuden sukupolven sekvensointimenetelmät

1980- ja 1990-luvulla kehitettiin menetelmä, jossa ei tarvittu leimattuja dNTP-molekyylejä (Heather & James, 2016). Menetelmässä hyödynnettiin DNA-polymeraasin lisäksi tulikärpäsistä ja kiiltomadoista eristettyä lusiferaasientsyymiä. Kiinteään faasiin kiinnitetyn DNA:n polymeroitumisreaktiossa vapautuva pyrofosfaatti  $PP_i$  muutetaan entsyymaattisesti katalysoidussa reaktiossa sellaiseksi molekyyliseksi, joka toimii substraattina lusiferaasille, mistä seuraa valoreaktio, jonka aallonpituuden perusteella tunnistetaan, mikä neljästä emäksestä juosteeseen on liittynyt. Reaktioon lisätään yksi emäs kerrallaan, minkä jälkeen suoritetaan pesuvaihe, jossa kiinnittymättömät dNTP:t huuhtoutuvat pesureagenssin mukana pois putkesta. Tässä menetelmässä käytettävät nukleotidit eivät pysäytä juosteen synteesiä peruuttamattomasti, kuten Sangerin menetelmässä käytettävät nukleotidit. Kiinnittymättömien nukleotidien pesuvaiheen jälkeen synteesin jatkumisen estävä kiinnittynyt nukleotidin osa pestään pois (Bowers ym., 2009).

Menetelmän edelleen kehittyessä pesuvaihe korvattiin entsyymaattisella kiinnittymättömän dNTP:n hajotuksella (Heather & James, 2016). Luonnollisten nukleotidien käytön mahdollistamisen lisäksi merkittävä etu ensimmäisen sukupolven menetelmiin nähden oli se, että polymeroitumisreaktiota voitiin seurata reaaliaikaisesti. Menetelmää kutsutaan pyrosekvensoinniksi ja sen edelläkävijöitä olivat Pål Nyrén ja hänen kollegansa. Myöhemmin pyrosekvensointi lisensoitiin Jonathan Rothburgin perustamalle 454 Life Sciences -yhtiölle, jossa sitä edelleen kehitettiin siten, että useiden juosteiden paralleelinen eli rinnakkainen sekvensointi mahdollistui. Yksijuosteinen templaatti-DNA kiinnitetään kuulaan, jonka pinnalla on komplementaarinen vastinjuoste, minkä jälkeen DNA monistetaan emulsio-PCR:lla. Reaktion jälkeen nämä kuulat siirretään kuoppalevyille, kukin kuula omaan kaivoonsa, jossa pyrosekvensointireaktiot tapahtuvat. Jokaisen kaivon pohjalla on CCD-sensori, jonka avulla tunnistetaan, mikä emäs on milloinkin liittynyt. Menetelmällä pystyttiin saamaan 500 - 800 emästä pitkiä lukemia per kaivo (levyllä on tavallisesti useita kymmeniä kaivoja) (Buermans & den Dunnen, 2014). Pyrosekvensoinnissa haasteena oli kuitenkin homopolymeerien eli saman emäksen toistojen emästen lukumäärän tarkka määrittäminen (Heather & James, 2016). 454:n pyrosekvensointimenetelmä oli ensimmäinen

kaupallinen uuden sukupolven menetelmä. Myöhemmin 454:n pyrosekvensointilaitteet siirtyivät Roche-yhtiön omistukseen.

454:n menestyksestä seurasi useiden paralleelisten sekvensointimenetelmien kehityksen alku, näistä merkittävimpana Solexa. Menetelmän kehitys alkoi 1990-luvun puolivälissä Cambridgen yliopistossa ja jatkui siellä, kunnes vuonna 2000 perustettiin Solexan ensimmäiset erilliset toimintatilat (illumina.com, noudettu 7.3.2019). Vuonna 2005 Solexa sekvensoi bakteriofagi phiX-174-genomin eli saman genomin, jonka Sanger aikanaan selvitti ensimmäisenä. Solexan menetelmä oli huomattavasti tehokkaampi, sillä sitä käyttäen voitiin sekvensoida yli kolme miljoonaa emästä yhdellä ajolla. Vuonna 2006 Solexan ensimmäinen kaupallinen sekvensointilaitte, Genome Analyzer, tuli markkinoille. Tällä laitteella saatiin sekvensoitua jo miljardi emästä yhdellä ajolla. Vuonna 2007 Illumina osti Solexan, mistä eteenpäin menetelmää on edelleen kehitetty.

Tässä menetelmässä yksijuosteinen templaatti-DNA pariutuu 5'-päähensä jatkeeksi kiinnitetyn sekvenssin avulla levyllä kiinnitettyyn komplementaariseen oligonukleotidiin, minkä jälkeen sille syntetisoidaan vastinjuoste (Heather & James, 2016). Seuraavaksi syntynyt kaksoisjuoste denaturoidaan ja alkuperäinen juoste pestään pois. Levyllä jäänyt juoste taipuu siten, että sen 3'-pää pariutuu levyllä olevan toisen oligonukleotidin kanssa ja juosteelle syntetisoidaan vastinjuoste. Syntynyt juoste avataan ja näin levyllä jää kopio templaatti-DNA:sta sekä 5'-3-suunnassa että 3'-5'-suunnassa. Tämä 'silta-amplifikaatio' eli silta-PCR toistuu ja levyllä muodostuu templaatti-DNA-klusteri. Klusterin muodostuksen jälkeen 3'-5'-juosteet pestään pois. Klusterin juosteisiin kiinnitetään alukkeet niiden päihin lisättyihin osiin, joilla ne kiinnittyivät levyn oligonukleotideihin, minkä jälkeen klusterit sekvensoidaan. Sekvensoinnissa käytetään sellaisia fluoresoivia dNTP-molekyylejä, joiden kiinnittymisen jälkeen uusi dNTP ei voi kiinnittyä ennen kuin dNTP:n 3'-päähän sitoutumisen estävä fluoroforiosa on pesty pois. Kun dNTP on kiinnittynyt, se viritetään laservalolla, mikä aiheuttaa emäksen fluoroforiosalle ominaisen fluoresenssireaktion, joka tunnistetaan CCD-laitteella. Tämän jälkeen fluoroforiosa pestään entsymaattisesti pois, jolloin 3'-pää vapautuu ja seuraava dNTP voi kiinnittyä. Ensin sekvensoidaan 5'-3'-juosteet. Sekvensoinnin jälkeen lukemajuoste pestään pois, minkä jälkeen templaatti-DNA taipuu ja kiinnittyy 3'-päästään alustalla olevaan oligonukleotidiin, ja monistuu.

Monistumisen jälkeen 5'-3'-juosteet pestään pois ja 3'-5'-juosteet sekvensoidaan. Näin saadaan sekvenssitietoa tutkittavan DNA-sekvenssin molemmista päistä, mikä helpottaa saadun lukeman vertausta referenssigenomiin.

454- ja Illumina-menetelmien jälkeen käytetyin toisen sukupolven menetelmä on luultavasti SOLiD-menetelmä (sanoista Sequencing by Oligonucleotide Ligation and Detection), jonka kehitti Applied Biosystems (myöhemmin fuusiossa Invitrogenin kanssa Life Technologies). Ensimmäisen sukupolven Maxam-Gilbert-menetelmän tapaan SOLiD ei perustu juosteen synteesiin. Tässä menetelmässä hyödynnetään koettimia, joiden sekvenssi tunnetaan ja jotka liitetään sekvensoitavaan juosteeseen DNA-ligaasi entsyymien avulla. Myös tällä menetelmällä saadaan tietoa juosteen molemmista päistä (Buermans & den Dunnen, 2014). Viimeisin merkittävimmistä toisen sukupolven sekvensointimenetelmistä on Ion Torrent, joka perustuu polymerisaatiossa vapautuvien vetyionien ( $H^+$ ) aiheuttaman pH-muutoksen mittaukseen (Heather & James, 2016). Kaivon pohjalla oleva sensori havaitsee pH-muutoksen, mistä seuraa jännitesignaali, jonka voimakkuus on suoraan verrannollinen liittyneiden emästen lukumäärään (Reuter ym., 2015). Ion Torrent ei vaadi kuvausvaihetta liittyneen nukleotidin/nukleotidien tunnistamiseksi, mikä kasvattaa laitteen suorituskykyä lisäämällä nopeutta. Menetelmän kehitti 454:n perustaja Jonathan Rothburg lähdettyään 454:sta (Heather & James, 2016). Myöhemmin Ion Torrent -menetelmä myytiin Life Technologies -yhtiölle. Thermo Fisher Scientific osti Life Technologies -yhtiön vuonna 2014 (thermofisher.com, noudettu 23.7.2019). Tämän menetelmän, kuten myös pyrosekvensointimenetelmien, ongelmana olivat kuitenkin homopolymeerit (Heather & James, 2016). Virhelukema yhden emäksen muutoksille on uuden sukupolven menetelmillä arviolta 0.1-1 prosenttia ja sekvensoinnin hinta noin 45 senttiä miljoonaa emäsparia kohti (Fox ym., 2014; *CADTH Rapid Response Reports*, 2014).

Ihmisen koko genomi sekvensoitiin osana Human Genome Project (HPG) -hanketta vuosina 1990-2003, ja tämän kokonaisen ihmisgenomin tuottaminen maksoi arviolta puolesta miljardista miljardiin Yhdysvaltain dollaria eli 445-890 miljoonaa euroa (genome.gov, noudettu 14.8.2019). Koko HPG-hankkeen hinta oli arviolta kolme miljardia Yhdysvaltain dollaria. Projektissa käytettiin Sanger-sekvensointia. Sanger-



menetelmän tarkkuuden ansiosta virhelukemaksi saatiin alhainen 0.01 prosenttia. Sangerin menetelmä on tähän päivään saakka tarkkuudeltaan paras sekvensointimenetelmä ja sitä hyödynnetäänkin usein NGS-menetelmin saatujen tulosten oikeellisuuden varmistamiseen. Menetelmän alhaisen suorituskyvyn ja korkean hinnan vuoksi haluttiin alkaa kehittää tehokkaampaa ja edullisempaa tapaa suurten genomien ja genomien alueiden sekvensointiin (Reuter ym., 2015). National Human Genome Research Institute (NHGRI) rahoittikin vuonna 2004 70 miljoonan Yhdysvaltain dollarin eli noin 62 miljoonan euron aloitteen DNA:n sekvensointimenetelmien kehittämiseksi. Aloitteen tavoitteena oli, että kymmenessä vuodessa ihmisen kokonaisen genomien sekvensoinnin hinta laskisi tuhanteen Yhdysvaltain dollariin (890 euroa). Ihmisen koko genomien sekvensoinnin hinta putosi noin tuhanteen Yhdysvaltain dollariin vuoden 2015 puolivälin tienoilla, joten aloite oli menestyksellinen (genome.gov, noudettu 14.8.2019). Suorituskykyä arvioitaessa otetaan huomioon reaktioaika, rinnakkaisten reaktioiden lukumäärä sekä lukeman pituus per reaktio (Rizzo & Buck, 2012). Sekvensointiprojektit, joihin Sanger-sekvensoinnilla kuluisi jopa kymmeniä vuosia, voidaan uuden sukupolven menetelmillä saada valmiiksi muutamissa päivissä 10 000-100 000 kertaa pienemmällä kuluilla (Desmedt ym., 2012). Merkittävin Sangerin menetelmän suorituskykyä alentava tekijä on geelielektroforeesi vaihe, joka tarvitaan sekvenssin lukemiseksi (Rizzo & Buck, 2012).

*Taulukko 1.* Esimerkkejä uuden sukupolven sekvensointimenetelmillä toimivista laitteista ja niiden ominaisuuksia. Tehty Buermans & den Dunnen (2014) ja Besser ym. (2018) pohjalta.

<b>Menetelmä</b>	<b>Yhtiö</b>	<b>Laite</b>	<b>Ajon kesto</b>	<b>Luenta emäspareina</b>	<b>Luentoja/ajo</b>
Pyrosekvensointi	Roche	GS FLX Titanium XL +	23 t	700	1 x 10 <sup>6</sup>
Illumina	Illumina	HiSeq2500	12 vrk	2 x 100	3 x 10 <sup>12</sup>
SOLiD	Thermo Fisher Scientific (Life Technologies)	Abi/solid	10 vrk	75 + 35	2.7 x 10 <sup>12</sup>
Ion Torrent	Thermo Fisher Scientific (Life Technologies)	Ion Torrent	4 t	200-400	4 x 10 <sup>6</sup>

*Taulukko 2.* Uuden sukupolven sekvensointimenetelmät sen suhteen, saadaanko sekvenssitietoa samalla ajolla juosteen molemmista päistä. Tehty Hodkinson & Grice (2015) pohjalta.

<b>Menetelmä</b>	<b>Sekvenssitieto juosteen molemmista päistä.</b>
Pyrosekvensointi	Kyllä
Illumina	Kyllä
SOLiD	Kyllä
Ion Torrent	Ei

### 3. Syövän geneettiset taustatekijät

Proto-onkogeenit eli esisyöpägeenit säätelevät solujen kasvua, mitoosia eli tumanjakautumista sekä solunjakautumista (Tervasmäki, 2018, s. 23-24). Mutaatio esisyöpägeenissä voi johtaa sen muuttumiseen syöpägeeniksi. Syöpää aiheuttavat esisyöpägeenien mutaatiot ovat usein sellaisia, että pelkästään toisen alleelin muutos riittää edesauttamaan solun muuttumista syöpäsoluksi. Useimmiten esisyöpägeenien syöpää aiheuttavat mutaatiot ovat sellaisia mutaatioita, joiden seurauksena geeni saa uudenlaisen funktion. Yleensä nämä mutaatiot eivät periydy, koska ne ovat yleisempiä somaattisissa eli muissa kuin sukusoluissa. Yleisesti ottaen erilaisille syöville altistavat periytyvät mutaatiot ovat harvinaisia (Klug ym., 2012, s. 473). Vain noin yhdessä prosentissa kaikista syövästä jokin perinnöllinen tekijä on ollut altistavana tekijänä syövän kehitykselle.

Tuumorisupressorigeenit eli kasvurajoitegeenit nimensä mukaisesti rajoittavat solujen kasvua (Tervasmäki, 2018, s. 23-24). Jos solun DNA:n kaksoisjuosterakenne vaurioituu, nämä geenit estävät solunjakautumisen vaurion korjaamisen ajaksi tai aloittavat apoptoosin eli ohjelmoidun solukuoleman. Jotta kasvurajoitegeenin mutaatio edistäisi syövän syntyä, tulee syöpää aiheuttavan mutaation yleensä ilmetä molemmissa kasvurajoitegeenin alleeleissa. Tavallisesti näissä geeneissä syöpää aiheuttavat mutaatiot johtavat geenin funktion menetykseen. Useiden kasvurajoitegeenien mutaatioiden on havaittu olevan yhteydessä perinnöllisiin syöpäalttiuksiin.

Syöpäsoluilla on mutaatioiden seurauksena monia yleisiä ominaisuuksia, jotka erottavat ne elimistön terveistä soluista (Hanahan & Weinberg, 2000). Ne jakaantuvat itsenäisesti riippumattomina jakautumista säätelevistä signaaleista, eivät vastaa kasvua rajoittaviin signaaleihin eivätkä ajaudu apoptoosiin, kykenevät muodostamaan omia verisuonia, voivat kasvaa ryppäinä muodostaen näin kasvainkudoksen sekä pystyvät levittäytymään eri puolille elimistöä. Normaalissa kudoksessa mutaatioiden lukumääräksi yhdessä solunjakautumisessa on arvioitu yksi mutaatio miljardia nukleotidia kohden (Tervasmäki, 2018, s. 21). Yhdessä syöpäkasvaimessa voidaan havaita jopa 100 000 mutaatiota. Yleensä mutaatioiden lukumäärä syöpäkasvaimessa on noin 10 000 tai vähemmän.

Kasvaimen genomissa ei välttämättä aina havaita selkeää syövän synnyn ja kehityksen aloittanutta mutaatiota tai mutaatioita, vaan jokin muu mekanismi voi olla syövän aiheuttaja (Desmedt ym., 2012). Esimerkiksi DNA:n metylaatio on tällainen tekijä. Metylaatioissa esimerkiksi jokin tärkeä kasvurajoitegeeni voi hiljentyä. Syöpäsoluissa voi myös ilmetä

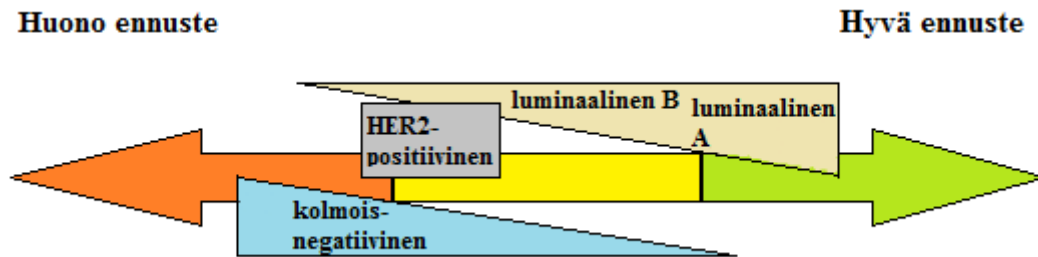
uusia mutaatioita hoitojen seurauksena. Tutkittaessa geneettisesti monimuotoisia ja monimutkaisia sairauksia, kuten rintasyöpää, on ensiarvoisen tärkeää, että käytössä on menetelmä, jolla voidaan tutkia useiden geenien vaikutusta tehokkaasti sekä ajankäytön että kulujen näkökulmasta. Tämä on mahdollistunut uuden sukupolven sekvensointimenetelmien kehityksen myötä.

#### 4. Rintasyövän molekulaariset alatyypit

Rintasyöpäkasvaimet jaetaan viiteen molekulaariseen alatyyppiin perustuen siihen, mitkä geenit niissä ovat aktiivisia (Dai ym., 2015). Alatyypit ovat luminaalinen A-tyyppi, luminaalinen B-tyyppi, HER2-positiivinen, kolmoisnegatiivinen sekä normaalityyppi. Tärkeimmät luokittelun selittäjät ovat erot estrogeeni- ja progesteronireseptoreiden (ER ja PR) ja HER2-proteiinin ekspressioissa. Suurin osa rintakasvaimista on luminaalisia. Luminaalisten kasvainten solut muistuttavat maitotiehyen luminaalisia eli sisimmän kerroksen epiteelisoluja. Kolmoisnegatiivisista kasvaimista suurin osa on sellaisia, joissa solut muistuttavat elimistön tyvikalvojen soluja sekä maitotiehyiden luminaalisia soluja ympäröiviä myoepiteelisiä soluja. Normaalityypin kasvaimet muistuttavat molekulaarisilta ominaisuuksiltaan luminaalisen A-tyypin kasvaimia, mutta solut muistuttavat enemmän normaaleja rintarauhasen soluja ja ennuste on huonompi.

*Taulukko 2.:* Rintasyövän molekulaaristen alatyyppeiden ominaisuuksia. Tehty Dai ym. (2015) pohjalta.

Alatyyppi	ER	PR	HER2	Ki67	tyvikalvo markkeri	Prosentuaalinen osuus tapauksista
luminaalinen A-tyyppi	+	+	-	-		~ 55
luminaalinen B-tyyppi	+	+	+ tai -	+		~ 15
HER2-positiivinen	-	-	+			~ 10
kolmoisnegatiivinen	-	-	-		+	~ 12
normaalityyppi	+	+	-	-		~ 8



Kuva 1. Eri alatyypin kasvainten ennusteet. Ennuste perustuu kasvaimen koon, muodon, imusolmukkeisiin levinneisyyden asteen sekä kauas levinneiden etäpesäkkeiden määrän ohella siihen, mitä molekulaarista alatyypistä kasvain edustaa (Tervasmäki, 2018, s. 26; Dai ym., 2015). Tehty Dai ym. (2015) pohjalta.

#### 4.1. Estrogeenin ja progesteronin reseptorit, HER2 ja Ki67

Suurin osa rintakasvaimista syntyy maitorauhasten tiehyiden epiteelisoluista (Anderson, 2002). Rintarauhasessa maitorauhaset vastaavat maidontuotannosta. Näitä rauhasia yhdistävät maitotiehyet, joita pitkin maito pääsee kulkemaan ja erittymään. Sisimpänä maitotiehyessä on luminaalisten epiteelisolujen kerros. Tätä ympäröi myoepiteelisolujen kerros, jota ympäröi vielä tiehyen rinnan sidekudokseen kiinnittävä tyvikalvo. Estrogeeni hormoni stimuloi maitotiehyiden kasvua ja progesteroni hormoni maitorauhasten kehitystä. Estrogeenin  $\alpha$ -muodon reseptorin ja progesteronireseptorin yliekspressiolla on rintasyöpäkasvaimen kehitystä edistävä vaikutus ainakin kehityksen alkuvaiheessa. HER2 on kasvutekijä proteiini ja se tunnetaan myös nimellä ERBB2 (Banerji ym., 2012). HER2:n yliekspressio lisää kasvaimen kasvunopeutta (Dai ym., 2015). Ki67 on antigeeni proteiini, joka liittyy vahvasti solujen lisääntymiseen (Dowsett ym., 2011). Se ilmenee kaikissa muissa solusyklin vaiheissa paitsi G0-vaiheessa, jossa solut eivät jakaudu. Sen ekspressiotasojen mittausta voidaan hyödyntää rintasyövän diagnostiikassa.

## 5. Familiaalisen rintasyövän geneettinen tausta

Niin kutsutuissa rintasyöpäperheissä sairauden esiintyvyys on suurempi verrattuna muuhun populaatioon (Tervasmäki, 2018, s. 28-29). Naisella, jolla on ainakin yksi rintasyöpään sairastunut ensimmäisen asteen sukulainen, on kaksinkertainen riski sairastua rintasyöpään verrattuna sellaiseen naiseen, jolla tällaista sukulaista ei ole. Rintasyöpäperheissä havaitaan useampia tapauksia ennen 45. ikävuotta, bilateraalisena eli molemmissa rinnoissa sekä useampia tapauksia miehillä. Perinnöllistä rintasyöpäalttiutta lisäävät alleelit voidaan jakaa kolmeen luokkaan: harvinaiset korkean riskin alleelit, harvinaiset kohtuullisen riskin alleelit ja yleiset alhaisen riskin alleelit. Harvinaiset korkean riskin alleelit kasvattavat kantajansa sairastumisen todennäköisyyden noin 10-20 kertaiseksi ja harvinaiset kohtuullisen riskin alleelit noin 2-6 kertaiseksi verrattuna tällaista alleelia kantamattoman henkilön riskiin. Yleisillä alhaisen riskin alleeleilla sairastumisriski on alle kaksinkertainen. Kaksostutkimuksilla on osoitettu, että suurin osa näistä familiaalisista rintasyöpätapauksista selittyy geneettisillä tekijöillä. Noin kymmenessä prosentissa rintasyöpätapauksista sairastumisen taustalla on ituradan eli sukusolulinjan mutaatio/mutaatioita (Tervasmäki, 2018, s. 19). Näistä arviolta 15-20 prosenttia selittyy BRCA1- ja BRCA2-geenien mutaatioilla, noin 5 prosenttia harvinaisten kohtuullisen riskin alleelien mutaatioilla ja noin 14 prosenttia yleisillä alhaisen riskin alleeleilla, eli alle puolet familiaalisista rintasyöpätapauksista selittyy tunnetuilla riskialleeleilla (Lee ym., 2016).

Merkittävin harvinaisista korkean riskin alleeleista, BRCA1, löydettiin 1990-luvun alussa kytkentäanalyysillä ja paikkakloonauksella (Tervasmäki, 2018, s. 30-31; Kiiski ym., 2014). Kytkentäanalyysissä tutkitaan niin kutsuttujen geenimerkkien, kuten esimerkiksi toistojaksojen rakenteiden, mahdollisia yhteyksiä sairauksiin (de la Chapelle ym., 1994). BRCA1-geenin koodaama proteiini aktivoi solusyklin tarkistuspisteiden toimintaa sekä osallistuu DNA:n kaksoisjuosteaurioiden korjaamiseen (Tervasmäki, 2018, s. 31). Tästä geenistä on löydetty 858 rintasyöpäriskiä merkittävästi nostavaa mutaatiota (Clark ym., 2012). BRCA1-geenin mutaatiota kantavalla naisella on 70-80 prosentin riski sairastua rintasyöpään elämänsä aikana. Kolmoisnegatiivisen alatyypin kasvaimet ovat yleisempiä BRCA1-mutaation kantajilla verrattuna muiden altistavien mutaatioiden kantajiin (Dai ym., 2015). Toinen merkittävä geeni tässä luokassa on BRCA2, joka myöskin toimii DNA:n kaksoisjuosteaurioiden korjauksessa (Tervasmäki, 2018, s. 31). Geenin yhteys rintasyöpään havaittiin 1990-luvun alkupuoliskolla. BRCA2-geenin mutaatiota kantavalla

naisella on 50-60 prosentin riski sairastua rintasyöpään elämänsä aikana (Clark ym., 2012). 3-5 prosentissa kaikista rintasyöpätapauksista esiintyy mutaatio joko BRCA1- tai BRCA2-geenissä. 15 prosentissa kolmoisnegatiivisen tyyppin rintasyöpiä on havaittu mutaatio joko BRCA1- tai BRCA2-geenissä (Kiiski ym., 2014). Somaattiset eli muissa kuin ituradan soluissa esiintyvät mutaatiot näissä geneeissä ovat harvinaisia (esiintyvyys noin 0.1 prosenttia) lukuunottamatta joitain populaatioita (esimerkiksi Ashkenazi juutalaiset), joissa somaattisten mutaatioiden esiintyvyys on suurempi johtuen perustajanvaikutuksesta (Tervasmäki, 2018, s. 31). Lisäksi heterotsygootteina eli esiintyessään vain jommassakummassa alleelissa mutaatiot altistavat rintasyövälle, mutta homotsygootteina muille sairauksille.

Harvinaisten kohtuullisen riskin alleelien esiintyvyys populaatioissa on yleensä alle yhden prosentin ja niissä esiintyvien erilaisten rintasyövälle altistavien mutaatioiden määrä on pienempi kuin harvinaisilla korkean riskin alleeleilla (Tervasmäki, 2018, s. 32-33). Näiden alleelien löytämiseksi on hyödynnetty pääasiassa kandidaattigeeni lähestymistapaa, missä syöpään yhteydessä olevia geenireittejä tutkitaan siten, että niihin liittyvien mutaatioiden esiintyvyyttä rintasyöpää sairastavilla verrataan mutaatioiden esiintyvyyteen terveillä kontroleilla. Myös korkean riskin alleeli BRCA2 löydettiin kandidaattigeeni lähestymistapaa hyödyntäen (Kiiski ym., 2014). Merkittävimmät geenit tässä luokassa ovat ATM, CHEK2, PALB2 ja NBS1 (Tervasmäki, 2018, s. 32-33). Osa näiden geenien alleeleista on vaikutuksiltaan verrattavissa korkean riskin harvinaisiin alleeleihin. Esimerkiksi eräs ATM-geenin mutatoitunut muoto kasvattaa rintasyöpäriskin jopa 15-kertaiseksi, ja Australiassa esiintyy PALB2-geenin mutatoitunut muoto, jonka kantajalla on 91 prosentin riski sairastua rintasyöpään ikävuoteen 70 mennessä. ATM-geenin tuotteella on tärkeä tehtävä DNA:n kaksoisjuosteaurioiden korjaamisessa, sillä se aktivoi fosforyloimalla proteiineja kuten BRCA1, p53 ja CHEK2. Myös CHEK2 on muita proteiineja fosforyloiva proteiini. Se säätelee BRCA1- ja p53-proteiinien aktiivisuutta sekä solusyklin tarkistuspisteiden ja apoptoosin signaalireittejä. PALB2-geeni on mahdollisesti merkittävin epäedullisesti mutatoituneena rintasyövälle altistavista geneistä BRCA1- ja BRCA2-geenien jälkeen. PALB2 säätelee BRCA2-proteiinin vakautta sekä sijoittumista tumassa. Riski sairastua rintasyöpään ikävuoteen 80 mennessä on ATM-geenin mutaation kantajalla 28, CHEK2-mutaation kantajalla 30 ja PALB2-mutaation kantajalla 50 prosenttia (Lee ym., 2016). Ainakin ATM-, PALB2- ja NBS1-mutaatiot aiheuttavat

heterotsygoottisina kohonneen rintasyöpäriskin, mutta homotsygootteina altistavat muille sairauksille (Tervasmäki, 2018, s. 33-34).

Yleisten alhaisen riskin alleelien esiintyvyys vaihtelee viidestä yli viiteenkymmeneen prosenttiin (Tervasmäki, 2018, s. 35-36). Käytetyin menetelmä näiden riskialleelien löytämiseksi on ollut genomin laajuinen assosiaatiotutkimus (Genome-Wide Association Studies, GWAS), missä useimmiten tutkitaan yleisten yhden nukleotidin varianttien (Single Nucleotide Variant, SNV) yhteyksiä sairauksiin (käytetään myös nimeä Single Nucleotide Polymorphism, SNP). GWA-tutkimuksissa hyödynnetään mikrosiruja, joilla on koettimia tietyille varianteille (Koboldt ym., 2013). Tähän luokkaan kuuluvat esimerkiksi FGFR2 ja TOX3/TNRC9, joiden lisäksi on tunnistettu yli 150 muuta alhaisen riskin alleelia (Tervasmäki, 2018, s. 35-36). Arviolta 30 epäsuotuisaa mutaatiota alhaisen riskin alleeleissa vaadittaisiin kolminkertaisen rintasyöpäriskin syntymiseksi. Alleeleilla on keskenään monimutkaisia vuorovaikutuksia, mikä hankaloittaa yksittäisen alleelin aiheuttaman riskin suuruuden arvioimista.

## 6. Tutkimusmenetelmät

Uusien rintasyövälle altistavien perinnöllisten riskialleelien löytämiseksi on tutkimuksessa hyödynnetty uuden sukupolven menetelmien kehityksen myötä kytkeäanalyysin, kandidaattigeeni lähestymistavan ja GWA-tutkimusten ohella ja sijaan koko genomin sekvensointia, (koko) eksomin sekvensointia sekä erilaisia geenipaneeleita (Tervasmäki, 2018, s. 20).

### 6.1. Koko genomin sekvensointi (Whole Genome Sequencing, WGS)

Uuden sukupolven sekvensointimenetelmin tehdyissä assosiaatiotutkimuksissa saadut tulokset korostavat niillä tehtävän WGS-tutkimuksen tärkeyttä eri sairauksien tutkimuksessa, sillä yli 80 prosenttia sairauksiin yhteydessä olevista varianteista sijoittuu eksonien eli genomin proteiineja koodaavien alueiden ulkopuolelle (Rizzo & Buck, 2012). Uuden sukupolven menetelmillä kokonainen ihmisen genomi voidaan sekvensoida alle vuorokaudessa yhdellä laitteella kun taas yhdellä Sanger-



sekvensointilaitteella tähän kuluisi arviolta jopa 60 vuotta (Reuter ym., 2015; Rizzo & Buck, 2012).

Erilaisissa sekvensointiprojekteissa on havaittu, että tyypillisesti jokaisella on genomissaan kolmesta ja puolesta neljään miljoonaan yhden nukleotidin varianttia (SNV) sekä useita satoja tuhansia pieniä indeleitä (Reuter ym., 2015). Indeleillä tarkoitetaan pieniä insertioita ja deleetioita eli nukleotidisekvenssiä tietyssä genomin kohdassa pidentäviä (insertio) tai lyhentäviä (deleetio) mutaatioita (Desmedt ym., 2012). Näiden yhden nukleotidin varianttien sekä indeleiden joukkoon kuuluu useita geenin funktion menetykseen johtavia mutaatioita (Reuter ym., 2015). Tällaisen mutaation ilmeneminen kasvurajoitegeenissä edistää syövän kehitystä ja monien kasvurajoitegeenien mutaatioiden on havaittu olevan yhteydessä periytyvään alttiuteen sairastua rintasyöpään (Tervasmäki, 2018, s. 23-24; Tervasmäki, 2018, s. 21). Indelit jäävät useimmiten havaitsematta Sanger-sekvensointia käytettäessä (Koboldt ym., 2013). WGS-tutkimuksissa on sekvensoitu useita syöpägenomeja ja saatu arvokasta tietoa sairauden etenemistä edistävistä mutaatioista sekä geenien ekspressiota säätelevistä reiteistä (Rizzo & Buck, 2012).

Genomissa on myös suurempia (yli 100 emäsparia) rakenteellisia variantteja (Structural Variants, SV) (Tervasmäki, 2018, s. 37-38). Yleisimpiä näistä ovat kopiomäärä variantit (Copy Number Variants, CNV) eli yli yhden kiloemäksen deleetiot, insertiot tai muut DNA:n uudelleenjärjestäytymiset, kuten translokaatiot eli genomin osan siirtymät uuteen paikkaan. Familiaaliseen rintasyöpään liittyvissä geeneissä, kuten BRCA1 ja BRCA2, on havaittu CN-variantteja. Se, että osalla NGS-menetelmiä saadaan saman ajan aikana sekvenssitietoa juosteen molemmista päistä, helpottaa sekä indeleiden että CN-varianttien havaitsemista (Reuter ym., 2015). Sama sekvenssi tulee kuitenkin sekvensoida useampaan kertaan (vähintään noin 20-30 kertaa) luotettavan tuloksen saamiseksi.

GWA-tutkimuksissa havaitaan useimmiten yleisiä yhden nukleotidin variantteja, joiden vaikutukset ovat pieniä (Koboldt ym., 2013). GWA-tutkimuksilla onkin löydetty suuri määrä familiaaliseen rintasyöpään liittyviä yleisiä alhaisen riskin alleeleja. Harvinaisilla varianteilla, joiden vaikutukset ovat suurempia, arvellaan olevan merkittävä vaikutus vielä selittämättömiin tekijöihin geneettisesti monimutkaisten sairauksien periytyvyysasteessa. Lisäksi johtuen viimeaikaisesta koko ihmispopulaation kasvusta

harvinaisten geenivarianttien määrä on suuri. Tämä kasvattaa koko genomien laajuisten tutkimusten tarvetta.

## 6.2. Eksomisekvensointi

Eksoneiden sekvensoinnista käytetään nimitystä eksomisekvensointi. Koko eksomin sekvensointia (Whole Exome Sequencing, WES) hyödyntäen on viime vuosina löydetty useita sellaisia DNA:n kaksoisjuosteaurioiden korjauksessa toimivia geenejä, jotka voivat olla yhteydessä perinnölliseen rintasyöpäalttiuteen (Tervasmäki, 2018, s. 47). Koska eksonit kattavat vain noin yhden prosentin koko genomista, voidaan useita eksomeita sekvensoida rinnakkain samassa ajassa ja samaan hintaan kuin yksi kokonainen genomi (Desmedt ym., 2012). Eksomisekvensointi tarjoaa mahdollisuuksia sellaisten varianttien löytämiseksi, joiden toiminnalliset vaikutukset ovat suuria (Koboldt ym., 2013). Tutkimalla rintasyöpäperheitä voidaan havaita harvinaisia variantteja pienemmällä otannalla, koska saman perheen jäsenistä useamman voidaan olettaa kantavan harvinaista varianttia. Yhdistämällä rintasyöpäperheiden eksomi- ja WG-sekvensointia (WGS) tutkijat voivat tehdä homotsygotiakoitosta selvittääkseen, mitkä alleelit vaikuttavat heterotsygoottisina perinnöllistä rintasyöpäalttiutta lisäävästi.

Suomessa esiintyy useita perustajamutaatioita perinnöllisissä rintasyövän riskigeneissä, kuten FA-geeneihin kuuluvissa BRCA2-, PALB2- ja RAD51C-geeneissä, minkä vuoksi suomalaiset rintasyöpäperheet ovat olleet useiden tutkimusten kohteena (Kiiski ym., 2014). PALB2 tunnetaan myös nimellä FANCN (Antoniou ym., 2014). FA-geenit ovat Fanconin anemia -geenejä, jotka toimivat DNA:n kaksoisjuosteaurion korjauksessa ja joissa esiintyvillä tietyillä mutaatioilla on heterotsygoottisina havaittu olevan rintasyöpäriskiä kasvattava vaikutus (Kiiski ym., 2014). Suomalaisten rintasyöpäperheiden eksomisekvensointi tutkimuksessa on havaittu esimerkiksi myöskin FA-geeneihin kuuluvan FANCM-geenin yhteys perinnölliseen rintasyöpäalttiuteen. Lisäksi tämän geenin mutaatioiden havaittiin altistavan erityisesti kolmoisnegatiiviselle rintasyövän alatyypille.

### 6.3. Geenipaneelit

Geenitestejä mutaatioille korkean riskin harvinaisissa BRCA1- ja BRCA2-alleleissa on ollut käytössä jo yli kaksi vuosikymmentä (O’Leary ym., 2017). Vaikka hoitomenetelmiä onkin kehitetty monien muidenkin geenien mutaatioista johtuville kasvaimille, niin kriteereitä testauksen tarpeellisuudelle näiden geenien suhteen ei ole saatu määriteltyä. Tähän pyritään löytämään ratkaisuja esimerkiksi geenipaneelien avulla. Koska mutaatiot yksittäisissä kohtuullisen riskin alleleissa ovat harvinaisia, on tehokkaampaa testata useita genejä NGS-menetelmillä samanaikaisesti samalla geenipaneelilla. Näin saadaan enemmän tietoa nopeammin suuremmallakin otannalla. Geenipaneelien toiminta perustuu koettimien käyttöön.

Yhtenä geenipaneelien suurena etuna pidetään sitä, että niiden avulla voidaan testata samanaikaisesti monia sellaisia genejä, joiden vaikutus kasvaimen fenotyyppiin on samanlainen (Hall ym., 2014). Kun vaikutus kasvaimen fenotyyppiin on sama eri geneillä, ei voida pelkästään kasvaimen ulkomuodon perusteella ennustaa, mikä mutaatio tai mitkä mutaatiot ovat johtaneet kasvaimen syntyyn ja edistäneet sen kehitystä. On siis olemassa sellaisia geenipaneeleita, joihin on kiinnitetty fenotyyppiin samalla tavoin vaikuttavien geenien koettimia. Havaittaessa useampia mutaatioita samalla geenipaneelilla, voidaan myös saada arvokasta tietoa siitä, esiintyvätkö jotkin tietyt mutaatiot usein samassa kasvaimessa sekä siitä, millaisia vuorovaikutuksia eri geneillä on. Etenkin yleisillä alhaisen riskin alleleilla on keskenään monimutkaisia vuorovaikutuksia, mikä hankaloittaa yksittäisen alleelin aiheuttaman riskin suuruuden arvioimista (Tervasmäki, 2018, s. 36). Geenipaneelien ongelmana on toistaiseksi VUS-variantit (Variants with Unknown Significance) eli sellaiset variantit, joiden yhteys perinnölliseen rintasyöpään ei ole ennestään selvä (O’Leary ym., 2017).

### 6.4. RNA-sekvensointi

RNA-sekvensointia voidaan hyödyntää geeniekspression mittaamiseen, ekspressoituvien mutaatioiden, yhdistyneiden geenien sekä erilaisten RNA-molekyylien muokkauksen tutkimiseen (Desmedt ym., 2012). Sekvensoimalla RNA-molekyyliä voidaan saada tietoa mutaatioista, jotka eivät esiinny DNA:ssa vaan

RNA:ssa. Ennen sekvensointia lähetti-RNA tulee kääntää cDNA-molekyyliksi käänteiskopioijaentsyymien avulla. cDNA-sekvenssi käännetään valmiista lähetti-RNA-molekyylistä, joten se ei sisällä introneita.

RNA-molekyylien muokkaus esimerkiksi vaihtoehtoisella silmukoinnilla on tärkeä rintasyövän geneettistä monimutkaisuutta kasvattava tekijä. Silmukoinnissa lähetti-RNA-molekyylistä poistetaan intronit eli geenin ei-koodaavat alueet. Myös eksoneita voidaan poistaa. Vaihtoehtoisella silmukoinnilla voidaan tuottaa samasta geenistä erilaisia proteiineja valikoimalla poistettavat osat. Koska uuden sukupolven sekvensointimenetelmillä saadaan sekvenssitietoa juosteen molemmista päistä samanaikaisesti, voidaan tunnistaa yhdistyneitä geenejä ja silmukointivariantteja.

RNA-sekvensoinnilla on havaittu esimerkiksi, että aggressiivisesti metastasoivilla eli etäpesäkkeitä muodostavilla rintasyöpäsoluilla esiintyy enemmän EMT-tapahtumiin (sanoista Epithelial to Mesenchymal Transition) liittyvää vaihtoehtoista silmukointia. EMT-prosessissa ilmenevän solujen fenotyypin muutoksen arvellaan lisäävän niiden kykyä levitä eri puolille elimistöä. Koko transkriptomin sekvensoinnilla (Whole Transcriptome Sequencing, WTS) saadaan tutkimustietoa koko eksomin kattavalta alueelta ja voidaan vertailla geeniekspressiota terveillä ja rintasyöpää sairastavilla genomien laajuudesta.

## 6.6. Uuden sukupolven sekvensointimenetelmien rajoitteita ja tulevaisuuden näkymää

Uuden sukupolven sekvensointimenetelmät vaativat sekvensoitavan genomien alueen monistuksen, jotta saadaan riittävän voimakas signaali kiinnittyneen nukleotidin tunnistamiseksi (Buermans & den Dunnen, 2014). Monistuksessa voi tapahtua virheitä, koska DNA-polymeraasi ei koskaan toimi sadan prosentin tarkkuudella, ja nämä virheet voidaan sitten tulkita alkuperäisessä templaattisekvenssissä oleviksi varianteiksi (Rizzo & Buck, 2012). Illumina-tekniikan heikkoutena on se, että pieni määrä klusteria muodostavista juosteista pitenee eri tahdissa kuin muut juosteet, joko jäljessä tai edellä (Buermans & den Dunnen, 2014). Etenevien monistussykliä aikana tällaisia virheitä voi kertyä, mikä häiritsee oikean signaalin saamista klusterista ja voi johtaa virheellisiin sekvenssin tulkintoihin.

Suurimmalla osalla NGS-menetelmiä saadaan saman ajon aikana sekvenssitietoa juosteen molemmista päistä, helpottaa sekä indeleiden että CN-varianttien havaitsemista (Reuter ym., 2015). NGS-menetelmien luennat ovat keskimäärin lyhyempiä kuin Sanger-sekvensoinnilla saatavat luennat eivätkä menetelmien tuottamat luennat välttämättä edusta tasaisesti kaikkia genomin alueita, vaan saadut lukemat voivat ikään kuin keskittyä sellaisille genomin alueille, jotka sisältävät paljon GC-toistojaksoja (eng. GC bias) (Rizzo & Buck, 2012). Tämä hankaloittaa esimerkiksi CN-varianttien tunnistamista. Sama sekvenssi tuleekin sekvensoida useampaan kertaan (vähintään noin 20-30 kertaa) luotettavan tuloksen saamiseksi. Lisäksi tällainen luentojen epätasainen jakautuminen voi johtaa siihen, että osa genomin alueista ei tule sekvensoiduksi riittävän montaa kertaa tarpeeksi tarkasti, jolloin mutaatioita jää huomaamatta. Standardeja siihen, kuinka monta kertaa sama sekvenssi tulisi lukea, kehitetään vastaamaan kunkin menetelmän virhelukemia ja sitä, miten tarkkaa tietoa missäkin tutkimuksessa vaaditaan. Suositukset vaihtelevat välillä 30-100 kertaa, mikä hidastaa WG-sekvensoinnin hinnan laskua. NGS-menetelmillä tuotetun tiedon soveltamiseksi sairauksien tutkimukseen ja klinisiin ratkaisuihin tarvittaisiin helpotusta saadun datan analysointiin. Tähän voisi toimia ratkaisuna eri tietokantojen (WGS, WES, RNA-sekvensointi) dataa yhteen keräävä tietokanta.

Ion Torrent -menetelmän virhelukema yhden nukleotidin varianteille on 0.1 prosenttia, ja Illumina-menetelmän virhelukema yhden nukleotidin varianteille on hyvin lähellä tätä. Ion Torrent -menetelmän merkittävin virheitä aiheuttava tekijä on ongelmat homopolymeerien tunnistamisessa. Mikäli sekvensoitavassa juosteessa esiintyy kaksi A-emästä peräkkäin, on pH-arvon muutos kaksinkertainen verrattuna yhden A-emäksen aikaansaamaan pH-arvon muutokseen. Kuitenkin verrattaessa kuuden ja viiden A-emäksen toistoa, on pH-arvon muutos enää 1.2-kertainen, eli suhteellinen muutos pH-arvossa alenee. Tämän vuoksi todennäköisyys tunnistaa emästen toistojen lukumäärä oikein pienenee. Menetelmään tehdyistä parannuksista huolimatta yli kuuden emäksen homopolymeerit johtavat kohonneisiin virhelukemiin (Reuter ym., 2015). Lisäksi Ion Torrent antaa sekvenssitietoa vain juosteen yhdestä päästä saman ajon aikana (Hodkinson & Grice, 2015). Yleisin virhe Illuminalle on yhden nukleotidin muutokset ja Ion Torrentille indelit (Reuter ym., 2015). SOLiD-menetelmän luennat puolestaan ovat huomattavasti lyhyempiä kuin Illumina- ja Ion Torrent -menetelmien luennat (Buermans & den Dunnen, 2014). Toisaalta SOLiDin virhelukema yhden nukleotidin

varianteille on alhaisempi ja voi olla jopa Sanger-sekvensoinnin virhelukeman tasoa. Suuri luennan pituus oli pitkään Roche-yhtiön pyrosekvensointilaitteiden vahvuus, mutta vuonna 2013 niiden kilpailukyky muiden sekvensointimenetelmien rinnalla todettiin riittämättömäksi, ja 454-teknologiaa hyödyntävät laitteet poistettiin tuotannosta/käytöstä vuonna 2016.

Templaatin monistusta vaativiin menetelmiin on viitattu toisen eli uuden sukupolven menetelminä ja sellaisiin menetelmiin, jotka eivät vaadi templaatin monistusta, kolmannen sukupolven menetelminä (Reuter ym., 2015). Niillä voidaan siis sekvensoida yksittäisiä DNA-molekyylejä. Kolmannen sukupolven menetelmiä ovat esimerkiksi Pacific Biosciences -yhtiön SMRT-menetelmä (sanoista Single Molecule Real-Time) ja Oxford Nanopore -yhtiön Nanopore-menetelmä. Kolmannen sukupolven menetelmät ovat vähemmän herkkiä GC-toistojaksoille, koska monistusvaihetta ei vaadita (GC-toistojen korostuminen monistetussa templaattissa lisää toisen sukupolven menetelmien epätarkkuutta). Luentojen tasaisemman jakautumisen lisäksi kolmannen sukupolven menetelmillä saadaan hyvin pitkiä lukemia (Rizzo & Buck, 2012). Luennat voivat olla jopa 60 kiloemästä pitkiä (Reuter ym., 2015). Luentojen tasaisempi jakautuminen ja suuri pituus helpottavat esimerkiksi CN-varianttien tunnistamista (Rizzo & Buck, 2012). Virhelukemat erilaisille varianteille vaihtelevat kuitenkin toistaiseksi SMRT-menetelmässä välillä 11-15 prosenttia (Besser ym., 2018). Nanopore-menetelmälle on raportoitu seuraavanlaisia virhelukemia: insertiot 4.9, deletiot 7.8 ja substituutiot 5.1 prosenttia (Reuter ym., 2015). Oxford Nanopore on kehitellyt kannettavan muistitikun kaltaisen MinION-laitteen, joka voidaan kiinnittää tietokoneen USB-porttiin, mutta tämän laitteen virhelukema on kuitenkin 12-38 prosenttia (Besser ym., 2018). Vaikka kolmannen sukupolven menetelmät ovat lupaavia tulevaisuuden sekvensointiteknologian kannalta, on niiden suhteen tehtävä vielä hyvin suuri määrä tutkimus- ja kehitystyötä. Parhaalla mahdollisella sekvensointimenetelmällä saataisiin tuotettua tarkasti, nopeasti ja edullisesti pitkiä luentoja templaattina yksittäinen molekyyli, eli siinä yhdistyisivät toisen ja kolmannen sukupolven menetelmien parhaat puolet (Rizzo & Buck, 2012).

## Lähteet

1. Anderson, E. (2002). Progesterone receptors - animal models and cell signaling in breast cancer: The role of oestrogen and progesterone receptors in human mammary development and tumorigenesis. *Breast Cancer Research*, 4(5), 197. <https://doi.org/10.1186/bcr452>
2. Antoniou, A. C., Casadei, S., Heikkinen, T., Barrowdale, D., Pylkäs, K., Roberts, J., ... Tischkowitz, M. (2014). Breast-Cancer Risk in Families with Mutations in *PALB2*. *New England Journal of Medicine*, 371(6), 497–506. <https://doi.org/10.1056/NEJMoa1400382>
3. Banerji, S., Cibulskis, K., Rangel-Escareno, C., Brown, K. K., Carter, S. L., Frederick, A. M., ... Meyerson, M. (2012). Sequence analysis of mutations and translocations across breast cancer subtypes. *Nature*, 486(7403), 405–409. <https://doi.org/10.1038/nature11154>
4. Besser, J., Carleton, H. A., Gerner-Smidt, P., Lindsey, R. L., & Trees, E. (2018). Next-generation sequencing technologies and their application to the study and control of bacterial infections. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases*, 24(4), 335–341. <https://doi.org/10.1016/j.cmi.2017.10.013>
5. Bowers, J., Mitchell, J., Beer, E., Buzby, P. R., Causey, M., Efcavitch, J. W., ... Thompson, J. F. (2009). Virtual terminator nucleotides for next-generation DNA sequencing. *Nature methods*, 6(8), 593–595. <https://doi.org/10.1038/nmeth.1354>
6. Buermans, H. P. J. & den Dunnen, J. T. (2014). Next generation sequencing technology: Advances and applications. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, 1842(10), 1932–1941. <https://doi.org/10.1016/J.BBADIS.2014.06.015>
7. *CADTH Rapid Response Reports*. (2014). Noudettu 23.7.2019 osoitteesta <https://www.ncbi.nlm.nih.gov/books/NBK274079/>
8. Clark, S. L., Rodriguez, A. M., Snyder, R. R., Hankins, G. D. V., & Boehning, D. (2012). STRUCTURE-FUNCTION OF THE TUMOR SUPPRESSOR BRCA1. *Computational and Structural Biotechnology Journal*, 1(1), e201204005. <https://doi.org/10.5936/CSBJ.201204005>
9. Dai, X., Li, T., Bai, Z., Yang, Y., Liu, X., Zhan, J., & Shi, B. (2015). Breast cancer intrinsic subtype classification, clinical use and future trends. *American journal of cancer research*, 5(10), 2929–2943. Noudettu 29.5.2019 osoitteesta <http://www.ncbi.nlm.nih.gov/pubmed/26693050>
10. de la Chapelle, A., Hästbacka, J., Lehesjoki, A.-E., Sulisalo, T., Kere, J., Tahvanainen, E. & Sistonen, P. (1994). Noudettu 28.8.2019 osoitteesta <https://www.duodecimlehti.fi/lehti/1994/7/duo40142>
11. Desmedt, C., Voet, T., Sotiriou, C., & Campbell, P. J. (2012). Next-generation sequencing in

- breast cancer: first take home messages. *Current opinion in oncology*, 24(6), 597–604.  
<https://doi.org/10.1097/CCO.0b013e328359554e>
12. Dowsett, M., Nielsen, T. O., A'Hern, R., Bartlett, J., Coombes, R. C., Cuzick, J., ... Hayes, D. F. (2011). Assessment of Ki67 in Breast Cancer: Recommendations from the International Ki67 in Breast Cancer Working Group. *JNCI Journal of the National Cancer Institute*, 103(22), 1656–1664. <https://doi.org/10.1093/jnci/djr393>
  13. Fox, E. J., Reid-Bayliss, K. S., Emond, M. J., & Loeb, L. A. (2014). Accuracy of Next Generation Sequencing Platforms. *Next generation, sequencing & applications*, 1. <https://doi.org/10.4172/jngsa.1000106>
  14. Hall, M. J., Forman, A. D., Pilarski, R., Wiesner, G., & Giri, V. N. (2014). Gene Panel Testing for Inherited Cancer Risk. *Journal of the National Comprehensive Cancer Network*, 12(9), 1339–1346. <https://doi.org/10.6004/jnccn.2014.0128>
  15. Hanahan, D., & Weinberg, R. A. (2000). *The Hallmarks of Cancer Review evolve progressively from normalcy via a series of pre*. *Cell* (Vsk. 100). Noudettu 29.5.2019 osoitteesta <https://www.cell.com/action/showPdf?pii=S0092-8674%2800%2981683-9>
  16. Heather, J. M., & Chain, B. (2016). The sequence of sequencers: The history of sequencing DNA. *Genomics*, 107(1), 1–8. <https://doi.org/10.1016/J.YGENO.2015.11.003>
  17. History of Illumina Sequencing and Solexa Technology. (ei pvm.). Noudettu 7.3.2019 osoitteesta <https://www.illumina.com/science/technology/next-generation-sequencing/illumina-sequencing-history.html>
  18. Hodkinson, B. P., & Grice, E. A. (2015). Next-Generation Sequencing: A Review of Technologies and Tools for Wound Microbiome Research. *Advances in wound care*, 4(1), 50–58. <https://doi.org/10.1089/wound.2014.0542>
  19. Kiiski, J. I., Peltari, L. M., Khan, S., Freysteinsdottir, E. S., Reynisdottir, I., Hart, S. N., ... Nevanlinna, H. (2014). Exome sequencing identifies FANCM as a susceptibility gene for triple-negative breast cancer. *Proceedings of the National Academy of Sciences of the United States of America*, 111(42), 15172–15177. <https://doi.org/10.1073/pnas.1407909111>
  20. Klug, W. S., Cummings, M. R., Spencer, C. A. & Palladino, M. A. (2012). *Concepts of Genetics*. Pearson Education.
  21. Koboldt, D. C., Steinberg, K. M., Larson, D. E., Wilson, R. K., & Mardis, E. R. (2013). The Next-Generation Sequencing Revolution and Its Impact on Genomics. *Cell*, 155(1), 27–38. <https://doi.org/10.1016/j.cell.2013.09.006>
  22. Lee, A. J., Cunningham, A. P., Tischkowitz, M., Simard, J., Pharoah, P. D., Easton, D. F., & Antoniou, A. C. (2016). Incorporating truncating variants in PALB2, CHEK2, and ATM into the BOADICEA breast cancer risk model. *Genetics in medicine : official journal of the American College of Medical Genetics*, 18(12), 1190–1198. <https://doi.org/10.1038/gim.2016.31>



23. Life Technologies. (ei pvm.). Noudettu 23.7.2019 osoitteesta  
<https://www.thermofisher.com/fi/en/home/brands/life-technologies.html>
24. O’Leary, E., Iacoboni, D., Holle, J., Michalski, S. T., Esplin, E. D., Yang, S., & Ouyang, K. (2017). Expanded Gene Panel Use for Women With Breast Cancer: Identification and Intervention Beyond Breast Cancer Risk. *Annals of Surgical Oncology*, 24(10), 3060–3066.  
<https://doi.org/10.1245/s10434-017-5963-7>
25. Reuter, J. A., Spacek, D. V., & Snyder, M. P. (2015). High-Throughput Sequencing Technologies. *Molecular Cell*, 58(4), 586–597.  
<https://doi.org/10.1016/J.MOLCEL.2015.05.004>
26. Rizzo, J. M., & Buck, M. J. (2012). Key Principles and Clinical Applications of “Next-Generation” DNA Sequencing. *Cancer Prevention Research*, 5(7), 887–900.  
<https://doi.org/10.1158/1940-6207.CAPR-11-0432>
27. Sinden, R. R. (1994). *DNA structure and function*. Academic Press.
28. Tervasmäki, A. (2018). *HEREDITARY PREDISPOSITION TO BREAST CANCER-EVALUATING THE ROLE OF RARE COPY NUMBER VARIANT, PROTEIN-TRUNCATING AND MISSENSE CANDIDATE ALLELES*. Oulu: University of Oulu Graduate School; University of Oulu, Faculty of Medicine; University of Oulu, Biocenter Oulu; Oulu University Hospital. <http://jultika.oulu.fi/files/isbn9789526220826.pdf>